The Epidemiology of Diabetes and Cancer

SDC
http://BendixCarstensen.com/DMCa/EpiDMCa
February 2014
1.0

Compiled Monday 28th April, 2014, 17:39
from:
C:/Bendix/Steno/DM-register/NDR/projects/Cancer/papers/EpiDMCa/Epi-DM-Ca.tex

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

Marit Eika Jørgensen — Steno Diabetes Center, Gentofte, Denmark
maej@steno.dk

Søren Friis — Danish Cancer Society, Copenhagen, Denmark
friis@cancer.dk

Contents

1 Introduction

The link between diabetes and cancer occurrence is well known, and firm population-based studies have demonstrated this already almost 25 years ago [1, 15, 7].

There has been a large increase in the literature on cancer occurrence subsequent to diabetes over the last few years, notably since the publication in 2009 of the papers on the potential effects of insulin Glargine on diabetes occurrence [12, 6, 8, 9].

More recent studies have (as did also Adami *et al.*[1]) demonstrated that there is a strongly elevated excess risk of cancer in newly diagnosed diabetes patients compared to the rather moderate excess seen more than 2–3 years after diagnosis of diabetes [10, 2, 5]

Another group of studies have demonstrated that there is an increased mortality among cancer patients with pre-existing diabetes, compared to cancer patients without [13, 14]

Over the last 5 years there has been an outburt of studies (many quite samll and inconclusive) focussing on putative effects of specific forms of diabetes medication on cancer occurrence. These studies are of course all marred by the prbelems of confounding by indication, which is vividly illustrated by Andersson *et al.* [2], where *all* drugs, be it insulins of varoius forms of OADs are associated with a strongly elevated cancer incidence RR just after initiation of the drug and a subsequent drop of the RR to something very close to 1.

Thus any long-term effects of drugs are likely to be small in terms of modification of cancer occurrence, and we shall therefore not include this in the broader discussion of the relationship between diabetes and cancer occurrence, but rather describe the general population impact of the two diseases.

2 Incidence of cancer

Incidence studies have broadly shown the same *ratios* of rates between diabetes patients and persons without diabetes; figure **??** compares the RR of different types of cancer between diabetes patients and non-diabetics from the major population based studies of diabetes occurrence.

It has been almost absent in the litterature to discuss the observed RRs between diabetes patients and non-diabetics in the light of the actual *size* and *shape* of the age-specific cancer incidence rates. The Danish data that we shall use for illustration of some of these points shows that the average increase by age is 10.2% per year for men and 7.2% per year for women, corresponding to 65 and 42% increse over 5 yera. So broadly speaking, apart from liver and pancreatic cancer, the risk elevation for incident cancer among diabetes patients correspond to the risk increase over less than 5 years of age.

3 Mortality after cancer diagnosis

4 The broader picture

The above mentioned studies are all aimed at describing the *relationship* cancer incidence rates or mortality rates of cancer patients *between* persons with and without diabetes.

This is illustrated in figure 1 where cancer incidence rates are shown in red and mortality rates among cancer patients in black. Studies in diabetes and cancer have traditionally

only focused on pairwise comparison of the thick and thin transition rates in figure 1. This is of course a reflection of a wish to describe variations between these pairs of rates that may give clues to mechanisms that places diabetes patients at different (normally higher) risk than non-diabetics.

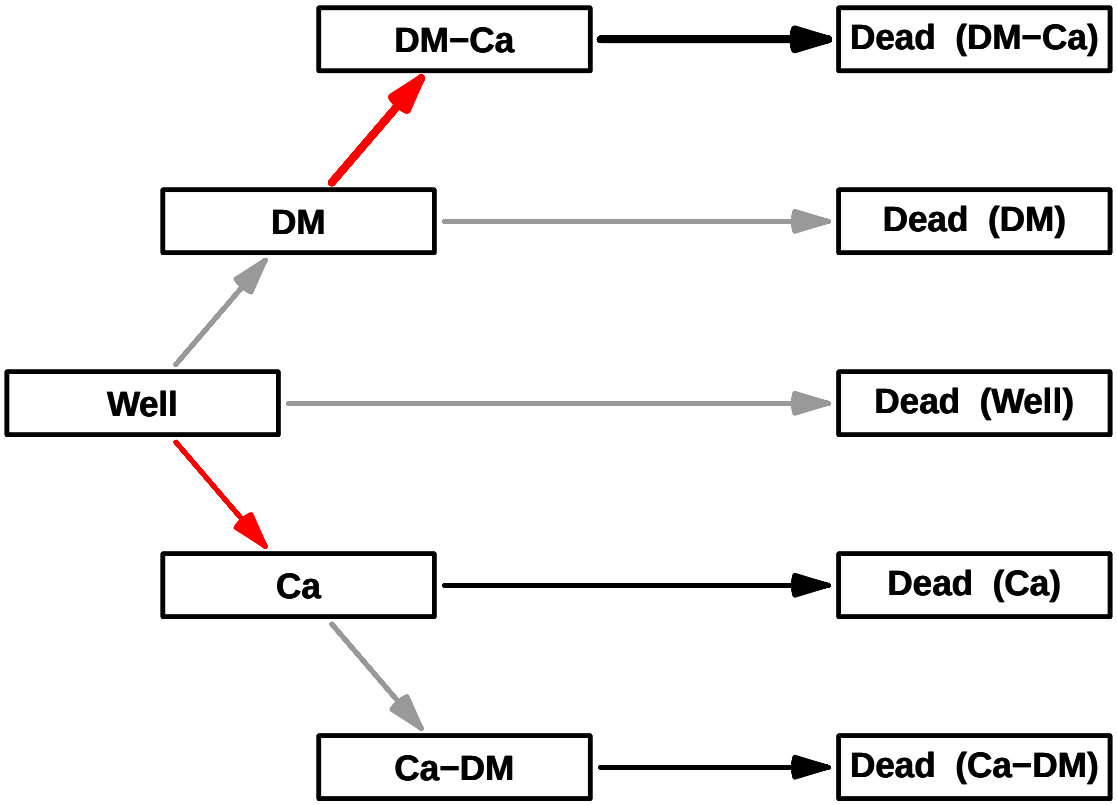

Figure 1: *Transition rates in a population exposed to occurrence of diabetes and cancer. The red transitions represent cancer incidence and the black ones death after a cancer diagnosis.*

For all of the rates in figure 1, the major determinant is age, so by only comparing the *rates* only, the magnitude of the population impact is lost.

As an example of this we shall use Danish data to estimate all 9 rates shown in figure 1 by sex, age and by calendar time.

This will provide the possibility to quantify the fraction of a birth cohort that will end in each of the 5 "death" states. It will also be possible to show what fraction of persons in a given age that will eventually contract cancer, depending on whether they have diabetes at the time.

Finally it will be possible to quantify the effect of increasing the risk of cancer and/or death without cancer on the scenario.

4.1 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, the period is for most types of events quite short, so ignoring the duration effects will have only minor influence on the summary measures.

5 Methods

Ww merged the Danish National Diabetes Register [3, 4] with the Danish Cancer Register, and classified all follow up after 1995 and after any of the diagnoses by sex, age, calendar time and date of birth in 1-year classes. We classified deaths and diagnoses of diabets and cancer similarly.

We also extracted the total population size and number of deaths from the Human Mortality Data Base

6 Discussion

There is increasing risk of cancer among diabetes patients with increasing severity of the diabetic disease. It is not clear (let alone discernible) whether this a result of the disease-processes associated with diabetes or if latent cancers contribute to the deterioration of the diabetic status of patients.

with the exception of liver and pancreatic cancer, it is also clear that the excess risk is not orders of magnitude larger, from most cancers the elevated rates are some 20-50% higher and for oter forms of cancer there is no excess risk.

When taking death as a competing riks to cancer incidence into account, the excess mortality among diabetes patients is a quantitatively much larger concern that the excess of cancers.

[11, 14]

References

[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] C. Andersson, A. Vaag, C. Selmer, M. Schmiegelow, R. Sørensen, J. Lindhardsen, G. H. Gislason, L. Køber, and C. Torp-Pedersen. Risk of cancer in patients using glucose-lowering agents: a nationwide cohort study of 3.6 million people. *BMJ Open*, 2(3), 2012.

[3] 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.

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

[5] B. Carstensen, D. R. Witte, and S. Friis. Cancer occurrence in Danish diabetic patients: duration and insulin effects. *Diabetologia*, 55(4):948–958, Apr 2012.

[6] H. M. Colhoun and the SDRN Epidemiology Group. Use of insulin glargine and cancer incidence in Scotland: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetologia*, 52:1755–1765, Sep 2009.

[7] 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.

[8] C. J. Currie, C. D. Poole, and E. A. Gale. The influence of glucose-lowering therapies on cancer risk in type 2 diabetes. *Diabetologia*, 52:1766–1777, Sep 2009.

[9] L. G. Hemkens, U. Grouven, R. Bender, C. Günster, S. Gutschmidt, G. W. Selke, and P. T. Sawicki. Risk of malignancies in patients with diabetes treated with human insulin or insulin analogues: a cohort study. *Diabetologia*, 52:1732–1744, Sep 2009.

[10] J. A. Johnson, S. L. Bowker, K. Richardson, and C. A. Marra. Time-varying incidence of cancer after the onset of type 2 diabetes: evidence of potential detection bias. *Diabetologia*, 54:2263–2271, Sep 2011.

[11] J. A. Johnson, B. Carstensen, D. Witte, S. L. Bowker, L. Lipscombe, A. G. Renehan, and the Diabetes and Cancer Research Consortium. Diabetes and cancer (1): evaluating the temporal relationship between type 2 diabetes and cancer incidence. *Diabetologia*, 55(6):1607–1618, Jun 2012.

[12] J. M. Jonasson, R. Ljung, M. Talbäck, B. Haglund, S. Gudbjörnsdóttir, and G. Steineck. Insulin glargine use and short-term incidence of malignancies-a population-based follow-up study in Sweden. *Diabetologia*, 52:1745–1754, Sep 2009.

[13] K. Ranc, M. E. Jørgensen, S. Friis, and B. Carstensen. Mortality after cancer among patients with diabetes mellitus: effect of diabetes duration and treatment. *Diabetologia*, 57(5):927–934, May 2014.

[14] A. G. Renehan, H. C. Yeh, J. A. Johnson, S. H. Wild, E. A. Gale, H. Møller, and the Diabetes and Cancer Research Consortium. Diabetes and cancer (2): evaluating the impact of diabetes on mortality in patients with cancer. *Diabetologia*, 55(6):1619–1632, Jun 2012.

[15] 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.

7 Overview of rate computation

7.1 Data

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 either cancer or diabetes (or both). Thus we have follow-up (and deaths) of patients in the Danish population corresponding to all boxes in figure 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.

First we need to attach the relevant packages:

```
library( foreign )
library( Epi )
library( RCurl )
# A function to fish out data from the HMDB
source( "C:/stat/R/BxC/Examples/HMD2R.r" )
# HMD2R
```

7.2 Total population follow-up

Along the same lines we can derive the number of deaths in the class ("Well","DK") by subtracting the number of deaths in all other classes from the total number of deaths in the population. To that end we first retrieve the total number of deaths from the human mortality database:

7.2.1 Mortality data from Human Mortality Database

In order to fetch mortality from the HMD in 1×1 Lexis triangles we need 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`, respctively.

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:

```
DK <- HMD2R( "DNK",
             wanted = "Deaths_lexis",
           username = HMDBusr,
           password = HMDBpwd )$Deaths_lexis[,1:5]
   *** Fetching... Deaths_lexis
```

```
newnames <-  c("P","A","C","F","M")
cbind( names( DK ),   newnames )
```

```
                newnames
 [1,] "Year"   "P"
 [2,] "Age"    "A"
 [3,] "Cohort" "C"
 [4,] "Female" "F"
 [5,] "Male"   "M"
```

```
       names( DK ) <- newnames
DK <- subset( DK, A < 100 & P > 1994 & P < 2010 )
DK$U <- with( DK, P-A-C )
M.dk <- reshape( DK, 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","U","D.tot")]
str( M.dk )
```

```
 'data.frame':        6000 obs. of  5 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 ...
  $ U    : int  0 1 0 1 0 1 0 1 0 1 ...
  $ D.tot: num  179 21 13 8 2 7 4 6 5 8 ...
```

The data frame `M.dk` now have the number of deaths in Lexis triangles between 1995-01-01 and 2008-12-31 in the ages between 0 and 100.

7.2.2 Population data from the Human Mortality Database

Then we get the population size (at 1 Jan each year) in one-year classes in a simular range:

```
DK <- HMD2R( "DNK",
            wanted = "Population",
          username = HMDBusr,
          password = HMDBpwd )$Population[,-5]
```

```
   *** Fetching... Population
```

```
str( DK )
```

```
 'data.frame':        19758 obs. of  4 variables:
  $ Year  : int  1835 1835 1835 1835 1835 1835 1835 1835 1835 1835 ...
  $ Age   : int  0 1 2 3 4 5 6 7 8 9 ...
  $ Female: num  17344 15133 13797 13170 13104 ...
  $ Male  : num  18133 15441 14039 13374 13310 ...
```

```
newnames <-  c("P","A","F","M")
cbind( names( DK ),   newnames )
```

```
               newnames
 [1,] "Year"   "P"
 [2,] "Age"    "A"
 [3,] "Female" "F"
 [4,] "Male"   "M"
```

```
       names( DK ) <- newnames
DK <- subset( DK, A < 101 & P>1994 & P<2011 )
```

— and compute the population exposure (person-years, risk time) in Lexis triangles:

```
Y.dkM <- N2Y( A=A, P=P, N=M, data=DK )
Y.dkF <- N2Y( A=A, P=P, N=F, data=DK )
Y.dk <- rbind( cbind( Y.dkM, sex="M" ),
               cbind( Y.dkF, sex="F" ) )
Y.dk$sex = factor( Y.dk$sex, levels=c("M","F") )
names( Y.dk )[3] <- "Y.tot"
str( Y.dk )
```

```
 'data.frame':        6000 obs. of  4 variables:
  $ A    : num  0.333 1.333 2.333 3.333 4.333 ...
  $ P    : num  1996 1996 1996 1996 1996 ...
  $ Y.tot: num  17961 17938 17445 17631 16769 ...
  $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
```

The data frame `Y.dk` now have the amount of follow-up time in Lexis triangles between 1995-01-01 and 2008-12-31 in the ages between 0 and 100. The function `N2Y` automatically returns the mean age and period in `A` and `P`.

7.2.3 Total population data

We can 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':        6000 obs. of  6 variables:
  $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
  $ P    : num  1996 1996 1997 1997 1998 ...
  $ sex  : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
  $ Y.tot: num  16972 17961 16425 17392 16402 ...
  $ U    : int  0 0 0 0 0 0 0 0 0 0 ...
  $ D.tot: num  137 179 134 189 152 172 132 142 95 156 ...
```

We now have all deaths and follow-up time in the total population after 1995 distributed by Lexis-triangles.

7.3 The follow-up after DM and Ca

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.

7.3.1 Follow-up records

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

```
1                                   "Program: DMCaLex.sas"       17:30 Wednesday, April 23, 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           2.39 seconds
      cpu time            0.43 seconds


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

-----------------------------------------------------------------
C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\EpiDMCa\sas\DMCaLex.sas
-----------------------------------------------------------------
NOTE: Libref HER was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\EpiDMCa\sas
NOTE: Libref DATA was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\papers\EpiDMCa\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 = '31DEC2009'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 = '31DEC2009'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                     D51
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           0.07 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.cancer ;
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 1748815 observations read from the data set DMCA.CANCER.
NOTE: The data set WORK.CANCER has 1286419 observations and 32 variables.
NOTE: DATA statement used (Total process time):
      real time           8.51 seconds
      cpu time            1.70 seconds


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

```
NOTE: There were 1286419 observations read from the data set WORK.CANCER.
NOTE: The data set WORK.CANCER has 1286419 observations and 32 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           14.77 seconds
      cpu time            2.54 seconds


164
165        * Then merge with the diabetes register ;
166        data DMCR;
167          merge cancer
168                DMCA.diabetes ;
169          by id ;
170          keep id sex diag
171                  doBT doDM doCA doDD ;
172        * Demografic dates collected from CRG and NDR ;
173          doBT = min( D_foddto  , D_fdsdato ) ;
174          doDD = min( D_statdato, D_dodsdto ) ;
175        * Event-dates ;
176          doDM = D_inkldto ;
177          doI  = D_ins ;
178          doCA = D_diagnosedato ;
179        * If date of diabetes or cancer is equal to date of death, remove it ;
180          if doDD gt .z then do;
181             if doDM ge doDD then doDM = . ;
182             if doCA ge doDD then doCA = . ;
183             end ;
184        * If date of diabetes and cancer is the same, diabetes first ;
185          if doDM eq doCA then doDM = doCA - 2 ;
186          if doDM > .z or doCA > .z ;
187        * Only persons alive on 1.1.1995 (or born later) ;
188          if doDD gt '31DEC94'd or doDD le .z ;
189        * Only persons with one or the other disease ;
190          if doDM > .z or doCA > .z ;
191        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).
      463041 at 174:10   62702 at 185:36
NOTE: There were 1286419 observations read from the data set WORK.CANCER.
NOTE: There were 437593 observations read from the data set DMCA.DIABETES.
NOTE: The data set WORK.DMCR has 912764 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           2.04 seconds
      cpu time            0.99 seconds


192
193        * The dataset DMCR now has a record for each person who has either a
194        * a diabetes diagnosis or a cancer diagnosis. Persons with more than
195        * one recorded tumour are represented by a record for each tumour ;
196        * We then construct the records of follow-up in different states ;
197
198        data toLex ;
199          set DMCR ;
200          id = _n_ ;
201          keep id sex diag
202               doBT doCa doDM doDD
203               entry exit en_st ex_st ;
204          length en_st ex_st $5 ;
205        *** Only Cancer ;
206          if ( doDM le .z ) then do ;
207             entry = max( doCa, &truncdate. ) ;
208             en_st = "Ca" ;
209             exit  = min( doDD, &censdate ) ;
210             if exit eq doDD then ex_st = "Dead" ; else
211                                  ex_st = en_st ;
212             if entry lt exit then output ;
213             end ;
214        *** Only diabetes ;
215          else
216          if ( doCa le .z ) then do ;
217             entry = max( doDM, &truncdate. ) ;
218             en_st = "DM" ;
219             exit  = min( doDD, &censdate ) ;
220             if exit eq doDD then ex_st = "Dead" ; else
221                                  ex_st = en_st ;
222             if entry lt exit then output ;
223             end ;
224        *** DM before Cancer ;
225          else
226          if ( doCa gt doDM ) then do ;
227             * from DM to Ca ;
228             entry = max( doDM, &truncdate. ) ;
229             en_st = "DM" ;
230             exit  = min( doCa, &censdate ) ;
231             if exit eq doCa then ex_st = "DM-Ca" ; else
232                                  ex_st = en_st ;
```

```
233             if entry lt exit then output ;
234             * from Ca to end ;
235             entry = max( doCa, &truncdate. ) ;
236             en_st = ex_st ;
237             exit  = min( doDD, &censdate ) ;
238             if exit eq doDD then ex_st = "Dead" ; else
239                                  ex_st = en_st ;
240             if entry lt exit then output ;
241             end ;
242        *** Cancer before DM ;
243          else
244          if ( doCa lt doDM ) then do ;
245             * from Ca to DM ;
246             entry = max( doCa, &truncdate. ) ;
247             en_st = "Ca" ;
248             exit  = min( doDM, &censdate ) ;
249             if exit eq doDM then ex_st = "Ca-DM" ; else
250                                  ex_st = en_st ;
251             if entry lt exit then output ;
252             * from DM to end ;
253             entry = max( doDM, &truncdate. ) ;
254             en_st = ex_st ;
255             exit  = min( doDD, &censdate ) ;
256             if exit eq doDD then ex_st = "Dead" ; else
257                                  ex_st = en_st ;
258             if entry lt exit then output ;
259             end ;
260        run ;

NOTE: There were 912764 observations read from the data set WORK.DMCR.
NOTE: The data set WORK.TOLEX has 981476 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           2.06 seconds
      cpu time            0.51 seconds


261
262        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\EpiDMCa\data\allPT.xpt
263        proc copy in = work
264                 out = allPT ;
265          select toLex ;
266        run;

NOTE: Copying WORK.TOLEX to ALLPT.TOLEX (memtype=DATA).
NOTE: There were 981476 observations read from the data set WORK.TOLEX.
NOTE: The data set ALLPT.TOLEX has 981476 observations and 11 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           26.28 seconds
      cpu time            0.85 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           56.30 seconds
      cpu time            7.12 seconds
```

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

```
dc <- read.xport( file="./data/allPT.xpt" )
names( dc ) <- gsub( "_", ".", tolower( names(dc) ) )
str( dc )
```

```
 'data.frame':        981476 obs. of  11 variables:
  $ diag : num  NA NA 101 999 NA NA NA 70 22 70 ...
  $ id   : num  1 2 3 4 5 6 7 8 9 10 ...
  $ sex  : num  2 2 2 2 1 1 2 2 1 2 ...
  $ doca : num  NA NA 12965 15645 NA ...
  $ dobt : num  14610 14610 -21549 15341 15341 ...
  $ dodd : num  NA NA 12989 NA NA ...
  $ dodm : num  16700 17757 NA NA 16713 ...
  $ en.st: Factor w/ 4 levels "Ca","Ca-DM","DM",..: 3 3 1 1 3 3 3 2 1 1 ...
  $ ex.st: Factor w/ 5 levels "Ca","Ca-DM","Dead",..: 4 4 3 1 4 3 3 3 3 3 ...
  $ entry: num  16700 17757 12965 15645 16713 ...
  $ exit : num  18262 18262 12989 18262 18262 ...
```

```
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 )
```

```
       diag              id         sex             doca             dobt
  Min.   : 21.0    Min.   :     1   M:472653   Min.   :1943     Min.   :1860
  1st Qu.: 70.0    1st Qu.:228080   F:508823   1st Qu.:1995     1st Qu.:1925
  Median : 91.0    Median :456421              Median :2001     Median :1935
  Mean   :222.3    Mean   :456412              Mean   :1999     Mean   :1937
  3rd Qu.:241.0    3rd Qu.:684676              3rd Qu.:2006     3rd Qu.:1947
  Max.   :999.0    Max.   :912764              Max.   :2010     Max.   :2010
  NA's   :341300                               NA's   :341621
       dodd             dodm          en.st          ex.st             entry
  Min.   :1995     Min.   :1942     Ca   :522493   Ca   :204878   Min.   :1995
  1st Qu.:1999     1st Qu.:1995     Ca-DM: 33623   Ca-DM: 43094   1st Qu.:1995
  Median :2003     Median :2001     DM   :382477   Dead :430993   Median :2001
  Mean   :2003     Mean   :2001     DM-Ca: 42883   DM   :247428   Mean   :2001
  3rd Qu.:2007     3rd Qu.:2006                    DM-Ca: 55083   3rd Qu.:2006
  Max.   :2011     Max.   :2010                                   Max.   :2010
  NA's   :487935   NA's   :494529
       exit
  Min.   :1995
  1st Qu.:2003
  Median :2010
  Mean   :2006
  3rd Qu.:2010
  Max.   :2010
```

```
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) )
summary( Ldc )
```

```
 Transitions:
      To
 From     Well     DM DM-Ca     Ca Ca-DM   Dead  Records:  Events: Risk time:  Persons:
   DM        0 247428 40856      0     0  94193    382477   135049  2447391.0    382477
   DM-Ca     0      0 14227      0     0  28656     42883    28656    97527.8     42883
   Ca        0      0     0 204878 27964 289651    522493   317615  2470788.6    522493
   Ca-DM     0      0     0      0 15130  18493     33623    18493   138559.2     33623
   Sum       0 247428 55083 204878 43094 430993    981476   499813  5154266.7    912661
```

We can also 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 )
```

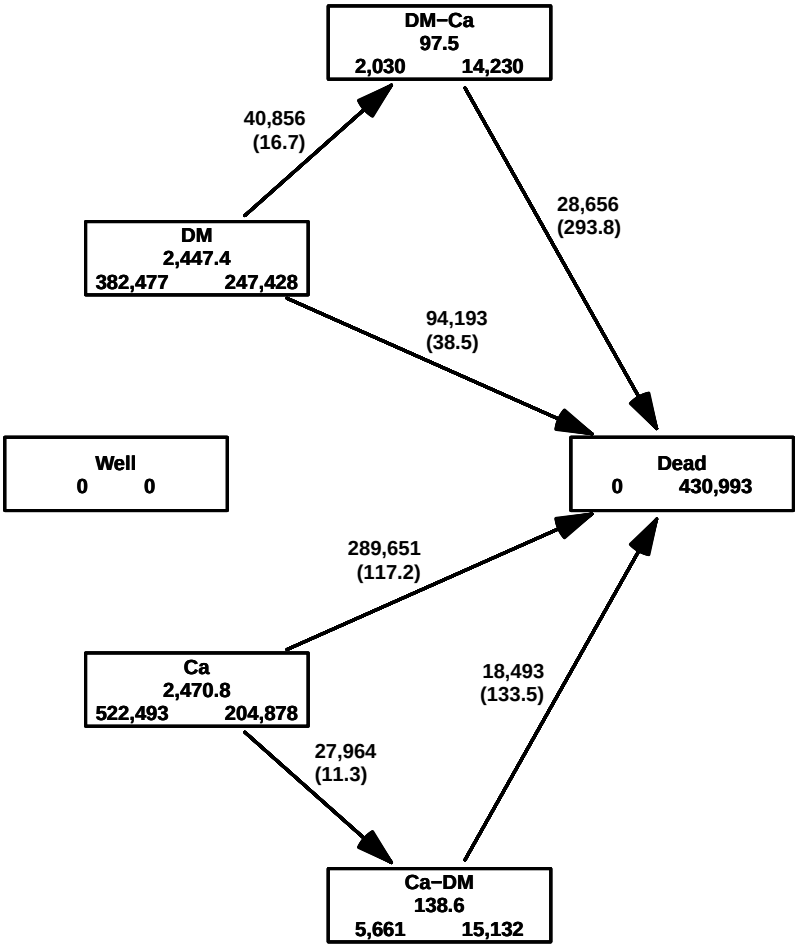

Figure 2: *The follow-up of the patients alone. The central number in each box is the amount of follow-up time (in 1000 PY) and the two number at the bottom are the number of persons that enter resp. exit the study in the state.*

7.4 Setting up 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.

7.4.1 Patient follow-up

In order to get the risk time among patients obtain this 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 aggredated 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 )
n.chunks <- 100
lm <- round( seq(0,nrow(Ldc),,n.chunks+1) )
for( i in 1:n.chunks )
{
whr <- (lm[i]+1):(lm[i+1])
sLx <- splitLexis( Ldc[whr,], 0:100, time.scale="age" )
sLx <- splitLexis( sLx,   1995:2010, 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" )
}
Agg <- transform( Agg, A = A + (1+U)/3,
                       P = P + (2-U)/3 )
Agg <- subset( Agg, A<100 & A>0 )
str( Agg )
save( Agg, file="./data/Agg.Rda" )
```

7.4.2 Non-patient follow-up

Now `Agg` contains all the follow-up among the patients, but we also need the follow-up, deaths, DM and cancer diagnoses, so we must aggregate `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 now merge the patient risk time and deaths with the total population and subtract them to get the risk time and deaths from thw well state:

```
str( All.dk )
 'data.frame':        6000 obs. of  6 variables:
  $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
  $ P    : num  1996 1996 1997 1997 1998 ...
  $ sex  : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
```

```
  $ Y.tot: num  16972 17961 16425 17392 16402 ...
  $ U    : int  0 0 0 0 0 0 0 0 0 0 ...
  $ D.tot: num  137 179 134 189 152 172 132 142 95 156 ...
```

```
str( Ptt.dk )
```

```
 'data.frame':        6000 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':        6000 obs. of  10 variables:
  $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
  $ P    : num  1996 1996 1997 1997 1998 ...
  $ sex  : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
  $ U    : int  0 0 0 0 0 0 0 0 0 0 ...
  $ Y.tot: num  16972 17961 16425 17392 16402 ...
  $ D.tot: num  137 179 134 189 152 172 132 142 95 156 ...
  $ Y.ptt: num  1.136 2.738 2.567 0.936 2.197 ...
  $ D.ptt: num  0 3 2 0 4 0 0 0 0 0 ...
  $ Y    : num  16971 17958 16423 17391 16399 ...
  $ D.dd : num  137 176 132 189 148 172 132 142 95 156 ...
```

7.4.3 Incident cases of DM and Cancer

We must also tabulate the number of newly diagnosed DM and Cancer cases (incidences) — the transition from the "Well" state. This is simply a tabulation of the entry age and date for records with `lex.Cst` equal to either "DM" or "Ca" with an entry date greater than 01Jan1995:

```
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 < 100 & A > 0 )
```

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

```
str( Well )
 'data.frame':        6000 obs. of  10 variables:
  $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
  $ P    : num  1996 1996 1997 1997 1998 ...
  $ sex  : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
  $ U    : int  0 0 0 0 0 0 0 0 0 0 ...
  $ Y.tot: num  16972 17961 16425 17392 16402 ...
  $ D.tot: num  137 179 134 189 152 172 132 142 95 156 ...
  $ Y.ptt: num  1.136 2.738 2.567 0.936 2.197 ...
  $ D.ptt: num  0 3 2 0 4 0 0 0 0 0 ...
  $ Y    : num  16971 17958 16423 17391 16399 ...
  $ D.dd : num  137 176 132 189 148 172 132 142 95 156 ...
```

```
str( Inc )
 'data.frame':        5972 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 ...
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':        6000 obs. of  13 variables:
  $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
  $ P    : num  1996 1996 1997 1997 1998 ...
  $ sex  : Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
  $ U    : int  0 0 0 0 0 0 0 0 0 0 ...
  $ Y.tot: num  17961 16972 17392 16425 17363 ...
  $ D.tot: num  179 137 189 134 172 152 142 132 156 95 ...
  $ Y.ptt: num  2.738 1.136 0.936 2.567 1.125 ...
  $ D.ptt: num  3 0 0 2 0 4 0 0 0 0 ...
  $ Y    : num  17958 16971 17391 16423 17362 ...
  $ D.dd : num  176 137 189 132 172 148 142 132 156 95 ...
  $ D.dm : num  1 0 0 4 1 0 2 0 1 1 ...
  $ D.ca : num  4 3 2 3 1 3 4 6 4 3 ...
  $ state: Factor w/ 6 levels "Well","DM","DM-Ca",..: 1 1 1 1 1 1 1 1 1 1 ...
str( Agg )
 'data.frame':        21011 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 merge the two databases:

```
dcd <- rbind( Well[,names(Agg)], Agg )
str( dcd )
 'data.frame':        27011 obs. of  9 variables:
  $ A    : num  0.333 0.333 0.333 0.333 0.333 ...
  $ P    : num  1996 1996 1997 1997 1998 ...
  $ U    : num  0 0 0 0 0 0 0 0 0 0 ...
  $ sex  : Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
  $ state: Factor w/ 6 levels "Well","DM","DM-Ca",..: 1 1 1 1 1 1 1 1 1 1 ...
  $ Y    : num  17958 16971 17391 16423 17362 ...
  $ D.ca : num  4 3 2 3 1 3 4 6 4 3 ...
  $ D.dm : num  1 0 0 4 1 0 2 0 1 1 ...
  $ D.dd : num  176 137 189 132 172 148 142 132 156 95 ...
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  382959 289438 431103 75450.8
 DM     40854      0  93885  2446.8
 DM-Ca      0      0  28648    97.5
 Ca         0  27958 289131  2468.4
 Ca-DM      0      0  18465   138.5
 Dead       0      0      0     0.0
```

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

8 Modelling of rates

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

```
> library( Epi )
> clear()
> load( file="./data/dcd.Rda" )
> addmargins( xtabs( cbind(D.dm,D.ca,D.dd) ~ state, data=dcd ), 1 )
  state     D.dm   D.ca   D.dd
    Well  289438 382959 431103
    DM         0  40854  93885
    DM-Ca      0      0  28648
    Ca     27958      0 289131
    Ca-DM      0      0  18465
    Dead       0      0      0
    Sum   317396 423813 861232
```

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 im modelling of the age-effects, we make histograms of the age-distribution of the events:

```
> par( mfrow=c(5,3), mar=c(3,1,1,1), mgp=c(3,1,0)/1.6 )
> par( mfg=c(1,1) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.dm), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="DM | Well" ) )
> par( mfg=c(1,2) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.ca), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Ca | Well" ) )
> par( mfg=c(1,3) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | Well" ) )
> par( mfg=c(2,2) ) ; with( subset( dcd, state=="DM" ),
>                           hist( rep(A,D.ca), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Ca | DM" ) )
> par( mfg=c(2,3) ) ; with( subset( dcd, state=="DM" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | DM" ) )
> par( mfg=c(3,3) ) ; with( subset( dcd, state=="DM-Ca" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | DM-Ca" ) )
> par( mfg=c(4,1) ) ; with( subset( dcd, state=="Ca" ),
>                           hist( rep(A,D.dm), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="DM | Ca" ) )
> par( mfg=c(4,3) ) ; with( subset( dcd, state=="Ca" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | Ca" ) )
> par( mfg=c(5,3) ) ; with( subset( dcd, state=="Ca-DM" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | Ca-DM" ) )
```

8.1 Prerequisites for the natural splines

```
> library( Epi )
> library( splines )
> ( a.kn <- c(5,1:9*10) )

   [1]  5 10 20 30 40 50 60 70 80 90

> ( p.kn <- 1995.5 + 0:3*4.5 )

  [1] 1995.5 2000.0 2004.5 2009.0

> ( c.kn <- 1900 + 1:9*10 )

  [1] 1910 1920 1930 1940 1950 1960 1970 1980 1990
```

With this in place we can model the rates from each of the states, we fit either age-cohort or age-period models for the rates. Note that we are entering the person-years in units of 1/10 of a year, because we subsequently will compute cumulative probabilities over intervals of lenth 1/10 year:

```
> # Men
> cm.w2dm <- glm( D.dm ~ Ns(   A, knots=a.kn ) +
+                        Ns( P-A, knots=c.kn ),
+                        offset = log(Y*10),
+                        family = poisson,
+                        data = subset( dcd, state=="Well" & sex=="M" ) )
> cm.w2ca  <- update( cm.w2dm, D.ca ~ . )
> cm.w2dd  <- update( cm.w2dm, D.dd ~ . )
> cm.dm2ca <- update( cm.w2ca, data = subset( dcd, state=="DM" & sex=="M" ) )
> cm.dm2dd <- update( cm.w2dd, data = subset( dcd, state=="DM" & sex=="M" ) )
> cm.ca2dm <- update( cm.w2dm, data = subset( dcd, state=="Ca" & sex=="M" ) )
> cm.ca2dd <- update( cm.w2dd, . ~ . + state,
+                              data = subset( dcd, state %in% c("Ca","Ca-DM","DM-Ca" )
+                                                                  & sex=="M") )
> # Women
> cf.w2dm  <- update( cm.w2dm, data = subset( dcd, state=="Well" & sex=="F" ) )
> cf.w2ca  <- update( cf.w2dm, D.ca ~ . )
> cf.w2dd  <- update( cf.w2dm, D.dd ~ . )
> cf.dm2ca <- update( cf.w2ca, data = subset( dcd, state=="DM" & sex=="F" ) )
> cf.dm2dd <- update( cf.w2dd, data = subset( dcd, state=="DM" & sex=="F" ) )
> cf.ca2dm <- update( cf.w2dm, data = subset( dcd, state=="Ca" & sex=="F" ) )
> cf.ca2dd <- update( cf.w2dd, . ~ . + state,
+                              data = subset( dcd, state %in% c("Ca","Ca-DM","DM-Ca")
+                                                                  & sex=="F" ) )
```

The correponding models for crossectional rates are simply fitted:

```
> # Men
> pm.w2dm  <- update( cm.w2dm , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.w2ca  <- update( cm.w2ca , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.w2dd  <- update( cm.w2dd , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.dm2ca <- update( cm.dm2ca, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.dm2dd <- update( cm.dm2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.ca2dm <- update( cm.ca2dm, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.ca2dd <- update( cm.ca2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> # Women
> pf.w2dm  <- update( cf.w2dm , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.w2ca  <- update( cf.w2ca , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.w2dd  <- update( cf.w2dd , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.dm2ca <- update( cf.dm2ca, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.dm2dd <- update( cf.dm2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.ca2dm <- update( cf.ca2dm, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.ca2dd <- update( cf.ca2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )


> round( ci.exp( pf.w2ca ), 3 )
```

```
                      exp(Est.)    2.5%   97.5%
(Intercept)               0.000   0.000   0.000
Ns(A, knots = a.kn)1      2.095   1.832   2.397
Ns(A, knots = a.kn)2      5.856   5.384   6.370
Ns(A, knots = a.kn)3     14.181  13.128  15.318
Ns(A, knots = a.kn)4     37.612  35.108  40.295
Ns(A, knots = a.kn)5     72.677  67.891  77.801
Ns(A, knots = a.kn)6    121.710 113.823 130.142
Ns(A, knots = a.kn)7    191.942 178.685 206.184
Ns(A, knots = a.kn)8     82.470  72.671  93.590
Ns(A, knots = a.kn)9    170.229 155.709 186.103
Ns(P, knots = p.kn)1      1.047   1.029   1.066
Ns(P, knots = p.kn)2      1.425   1.378   1.474
Ns(P, knots = p.kn)3      1.192   1.178   1.207
```

8.2 Computing the 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.

Since we have restricted ourselves to a scenery where we have only one time scale, namely age we can do the clculations in closed form by setting up the transition probability matrix for small intervals (of length `int` years. For the sake of completeness we also we also set up s asimilar matrix for the RR by period; mainly for showing the estimated RRs by period / cohort:

```
> int <- 0.1
> a.pt <- seq(0,102,int)[-1] - int/2
> p.pt <- seq(1995,2010,,100)
> c.pt <- seq(1900,2000,,100)
> ( 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 <-
+ anam <- list( from = states,
+                  to = states,
+                 age = a.pt,
+               scene = c("Cross","Long"),
+                 sex = c("M","F"),
+                what = c("Est","lo","hi") )
> pnam[3] <- list( p.pt ) ; names( pnam )[3] <- "per"
> cnam[3] <- list( c.pt ) ; names( cnam )[3] <- "coh"
> TR <- NArray( anam ) ; TR[is.na(TR)] <- 0
> PR <- TR[1,,,,,1]
> names( dimnames( PR ) )[1] <- "State"
> pRR <- NArray( pnam ) ; pRR[is.na(pRR)] <- 0
> cRR <- NArray( cnam ) ; cRR[is.na(cRR)] <- 0
```

First we fill in the estimated rates from the models just fitted, but we then must first have the relevant contrast matrices to extrcat the estimated rates:

```
> p.ref = 2005
> c.ref = 1950
> aM <- Ns(                  a.pt   , knots=a.kn )
> pR <- Ns( rep(p.ref,length(a.pt)), knots=p.kn )
> cR <- Ns( rep(c.ref,length(a.pt)), knots=c.kn )
> pA <- cbind( 1, aM, pR )
> cA <- cbind( 1, aM, cR )
> pM <- Ns( p.pt, knots=p.kn ) - Ns( rep(p.ref,length(p.pt)), knots=p.kn )
> cM <- Ns( c.pt, knots=c.kn ) - Ns( rep(c.ref,length(c.pt)), knots=c.kn )
```

First we fill in the age-specific rates that will later be used in the calculations of state occupancy probabilities:

```
> TR["Well" ,"DM"   ,,"Cross","M",] <- ci.exp( pm.w2dm , ctr.mat=pA )
> TR["Well" ,"Ca"   ,,"Cross","M",] <- ci.exp( pm.w2ca , ctr.mat=pA )
> TR["Well" ,"D-W"  ,,"Cross","M",] <- ci.exp( pm.w2dd , ctr.mat=pA )
> TR["DM"   ,"DM-Ca",,"Cross","M",] <- ci.exp( pm.dm2ca, ctr.mat=pA )
> TR["DM"   ,"D-DM" ,,"Cross","M",] <- ci.exp( pm.dm2dd, ctr.mat=pA )
> TR["Ca"   ,"Ca-DM",,"Cross","M",] <- ci.exp( pm.dm2ca, ctr.mat=pA )
> TR["Ca"   ,"D-Ca" ,,"Cross","M",] <- ci.exp( pm.ca2dd, ctr.mat=cbind(pA,1,0) )
> TR["DM-Ca","D-DC" ,,"Cross","M",] <- ci.exp( pm.ca2dd, ctr.mat=cbind(pA,0,0) )
> TR["Ca-DM","D-CD" ,,"Cross","M",] <- ci.exp( pm.ca2dd, ctr.mat=cbind(pA,0,1) )
> TR["Well" ,"DM"   ,,"Cross","F",] <- ci.exp( pf.w2dm , ctr.mat=pA )
> TR["Well" ,"Ca"   ,,"Cross","F",] <- ci.exp( pf.w2ca , ctr.mat=pA )
> TR["Well" ,"D-W"  ,,"Cross","F",] <- ci.exp( pf.w2dd , ctr.mat=pA )
> TR["DM"   ,"DM-Ca",,"Cross","F",] <- ci.exp( pf.dm2ca, ctr.mat=pA )
> TR["DM"   ,"D-DM" ,,"Cross","F",] <- ci.exp( pf.dm2dd, ctr.mat=pA )
> TR["Ca"   ,"Ca-DM",,"Cross","F",] <- ci.exp( pf.dm2ca, ctr.mat=pA )
> TR["Ca"   ,"D-Ca" ,,"Cross","F",] <- ci.exp( pf.ca2dd, ctr.mat=cbind(pA,1,0) )
> TR["DM-Ca","D-DC" ,,"Cross","F",] <- ci.exp( pf.ca2dd, ctr.mat=cbind(pA,0,0) )
> TR["Ca-DM","D-CD" ,,"Cross","F",] <- ci.exp( pf.ca2dd, ctr.mat=cbind(pA,0,1) )
> TR["Well" ,"DM"   ,,"Long" ,"M",] <- ci.exp( cm.w2dm , ctr.mat=cA )
> TR["Well" ,"Ca"   ,,"Long" ,"M",] <- ci.exp( cm.w2ca , ctr.mat=cA )
> TR["Well" ,"D-W"  ,,"Long" ,"M",] <- ci.exp( cm.w2dd , ctr.mat=cA )
> TR["DM"   ,"DM-Ca",,"Long" ,"M",] <- ci.exp( cm.dm2ca, ctr.mat=cA )
> TR["DM"   ,"D-DM" ,,"Long" ,"M",] <- ci.exp( cm.dm2dd, ctr.mat=cA )
> TR["Ca"   ,"Ca-DM",,"Long" ,"M",] <- ci.exp( cm.dm2ca, ctr.mat=cA )
> TR["Ca"   ,"D-Ca" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, ctr.mat=cbind(cA,1,0) )
> TR["DM-Ca","D-DC" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, ctr.mat=cbind(cA,0,0) )
> TR["Ca-DM","D-CD" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, ctr.mat=cbind(cA,0,1) )
> TR["Well" ,"DM"   ,,"Long" ,"F",] <- ci.exp( cf.w2dm , ctr.mat=cA )
> TR["Well" ,"Ca"   ,,"Long" ,"F",] <- ci.exp( cf.w2ca , ctr.mat=cA )
> TR["Well" ,"D-W"  ,,"Long" ,"F",] <- ci.exp( cf.w2dd , ctr.mat=cA )
> TR["DM"   ,"DM-Ca",,"Long" ,"F",] <- ci.exp( cf.dm2ca, ctr.mat=cA )
> TR["DM"   ,"D-DM" ,,"Long" ,"F",] <- ci.exp( cf.dm2dd, ctr.mat=cA )
> TR["Ca"   ,"Ca-DM",,"Long" ,"F",] <- ci.exp( cf.dm2ca, ctr.mat=cA )
> TR["Ca"   ,"D-Ca" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, ctr.mat=cbind(cA,1,0) )
> TR["DM-Ca","D-DC" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, ctr.mat=cbind(cA,0,0) )
> TR["Ca-DM","D-CD" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, ctr.mat=cbind(cA,0,1) )
```

8.3 Average change in cancer incidence rates

We can derive the annual change in cancer incidence rates between ages 40 and 70:

```
> rt <- TR["Well","Ca",c(401,701),"Cross",,1]*10^5
> round( (exp( log( rt[2,]/rt[1,] ) / 30 )-1)*100, 1 )
     M    F
  10.6  7.2
```

8.4 Secular trends

We the use almost the same code to fill in the RRs associated with period:

```
> pRR["Well" ,"DM"   ,,"Cross","M",] <- ci.exp( pm.w2dm , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"Ca"   ,,"Cross","M",] <- ci.exp( pm.w2ca , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"D-W"  ,,"Cross","M",] <- ci.exp( pm.w2dd , subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"DM-Ca",,"Cross","M",] <- ci.exp( pm.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"D-DM" ,,"Cross","M",] <- ci.exp( pm.dm2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"Ca-DM",,"Cross","M",] <- ci.exp( pm.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"D-Ca" ,,"Cross","M",] <- ci.exp( pm.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM-Ca","D-DC" ,,"Cross","M",] <- ci.exp( pm.ca2dd, subset="Ns\\(P", ctr.mat=pM )
```

```
> pRR["Ca-DM","D-CD" ,,"Cross","M",] <- ci.exp( pm.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"DM"   ,,"Cross","F",] <- ci.exp( pf.w2dm , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"Ca"   ,,"Cross","F",] <- ci.exp( pf.w2ca , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"D-W"  ,,"Cross","F",] <- ci.exp( pf.w2dd , subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"DM-Ca",,"Cross","F",] <- ci.exp( pf.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"D-DM" ,,"Cross","F",] <- ci.exp( pf.dm2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"Ca-DM",,"Cross","F",] <- ci.exp( pf.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"D-Ca" ,,"Cross","F",] <- ci.exp( pf.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM-Ca","D-DC" ,,"Cross","F",] <- ci.exp( pf.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca-DM","D-CD" ,,"Cross","F",] <- ci.exp( pf.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> cRR["Well" ,"DM"   ,,"Long" ,"M",] <- ci.exp( cm.w2dm , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"Ca"   ,,"Long" ,"M",] <- ci.exp( cm.w2ca , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"D-W"  ,,"Long" ,"M",] <- ci.exp( cm.w2dd , subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"DM-Ca",,"Long" ,"M",] <- ci.exp( cm.dm2ca, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"D-DM" ,,"Long" ,"M",] <- ci.exp( cm.dm2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca"   ,"Ca-DM",,"Long" ,"M",] <- ci.exp( cm.dm2ca, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca"   ,"D-Ca" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM-Ca","D-DC" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca-DM","D-CD" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"DM"   ,,"Long" ,"F",] <- ci.exp( cf.w2dm , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"Ca"   ,,"Long" ,"F",] <- ci.exp( cf.w2ca , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"D-W"  ,,"Long" ,"F",] <- ci.exp( cf.w2dd , subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"DM-Ca",,"Long" ,"F",] <- ci.exp( cf.dm2ca, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"D-DM" ,,"Long" ,"F",] <- ci.exp( cf.dm2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca"   ,"Ca-DM",,"Long" ,"F",] <- ci.exp( cf.dm2ca, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca"   ,"D-Ca" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM-Ca","D-DC" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca-DM","D-CD" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, subset="Ns\\(P", ctr.mat=cM )
```

8.5 Transition rates

We have now estimates all 9 transition rates between states both using age-perido and age-cohort models. Of particular interest is the relationships between the cancer incidence rates between persons with and without diabetes, and the relationship between mortality rates between persons with and without DM and Cancer.

We would also like to see the ratio of the rates, so we need a small function that computes the ratio of two sets of estimates with the correct resluting ci:

```
> rates2RR <-
+ function( num, den )
+ {
+ # num and den are assumed to be 3-column-matrices with columns
+ # rate, lower, upper.
+ seln <- log( num[,3]/num[,1] )/1.96
+ seld <- log( den[,3]/den[,1] )/1.96
+ lrr  <- log( num[,1]/den[,1] )
+ exp( cbind( lrr, sqrt(seln^2+seld^2) ) %*% ci.mat() )
+ }
```

For the cross-sectional rates we make a plot of the age-specific cancer incidence rates along the RR by calendar time:

```
> clr <- c("limegreen","blue","red","black") # well, dm, ca, RR
> ylm <- c(5,5000)
> rrr <- 100
> par( mfrow=c(2,2), mar=c(3,2,1,3), oma=c(1,2,1,0),
+      mgp=c(3,1,0)/1.6, bty="n", las=1 )
> # Cancer incidence among men
> ciw <-  TR["Well","Ca"   ,,"Cross","M",] * 10^5
> cid <-  TR["DM"  ,"DM-Ca",,"Cross","M",] * 10^5
> rrw <- pRR["Well","Ca"   ,,"Cross","M",]
```

```
> rrd <- pRR["DM"  ,"DM-Ca",,"Cross","M",]
> plci <- function() {
+ matplot( a.pt, cbind( ciw, cid ),
+       lwd=rep(c(3,1,1),2), lty=1, col=rep(clr[1:2],each=3),type="l",
+       xlim=c(20,120), ylim=ylm/5, log="y", xaxt="n", xaxs="i",
+       ylab="", xlab="" )
+ matlines( a.pt, rates2RR( cid, ciw )*rrr,
+           lwd=c(3,1,1), lty=1, col=clr[4],type="l" )
+ abline( h=rrr )
+ matlines( p.pt-1890, cbind( rrw, rrd )*rrr,
+       lwd=rep(c(3,1,1),2), lty=1, col=clr[1:2],type="l" )
+ axis( side=4, at=5:20/10 * rrr, labels=FALSE, tcl=-0.3 )
+ axis( side=4, at=c(0.5,1,2) * rrr, labels=c(0.5,1,2) )
+ axis( side=4, at=3 * rrr, labels="RR", tcl=0 )
+ axis( side=1, at=seq( 20,100, 5), labels=F )
+ axis( side=1, at=seq( 20,100,20) )
+ axis( side=1, at=seq(105,120,5), labels=F )
+ axis( side=1, at=seq(110,120,10), labels=c(2000,2010) )
+ mtext( c("Age","Date"), at=c(60,112.5), side=1, line=3/1.6 )
+ }
> plci()
> # Cancer incidence among women
> ciw <-  TR["Well","Ca"   ,,"Cross","F",] * 10^5
> cid <-  TR["DM"  ,"DM-Ca",,"Cross","F",] * 10^5
> rrw <- pRR["Well","Ca"   ,,"Cross","F",]
> rrd <- pRR["DM"  ,"DM-Ca",,"Cross","F",]
> plci()
> # Mortality among men
> mtw  <-  TR["Well" ,"D-W" ,,"Cross","M",] * 10^5
> mtd  <-  TR["DM"   ,"D-DM",,"Cross","M",] * 10^5
> mtc  <-  TR["Ca"   ,"D-Ca",,"Cross","M",] * 10^5
> mtcd <-  TR["Ca-DM","D-CD",,"Cross","M",] * 10^5
> mtdc <-  TR["DM-Ca","D-DC",,"Cross","M",] * 10^5
> rrw  <- pRR["Well" ,"D-W" ,,"Cross","M",]
> rrd  <- pRR["DM"   ,"D-DM",,"Cross","M",]
> rrc  <- pRR["Ca"   ,"D-Ca",,"Cross","M",]
> rrcd <- pRR["Ca-DM","D-CD",,"Cross","M",]
> rrdc <- pRR["DM-Ca","D-DC",,"Cross","M",]
> plmt <- function() {
+ matplot( a.pt, cbind( mtw, mtd, mtc, mtcd, mtdc ),
+       lwd=rep(c(3,1,1),2), lty=rep(c(1:3),c(9,3,3)),
+       col=rep(clr[c(1:3,3,3)],each=3),type="l",
+       xlim=c(20,120), ylim=ylm, log="y", xaxt="n", xaxs="i",
+       ylab="", xlab="" )
+ # matlines( a.pt, rates2RR( mtd, mtw )*100,
+ #           lwd=c(3,1,1), lty=1, col=clr[4],type="l" )
+ lines( c(105,120), c(100,100) )
+ matlines( p.pt-1890, cbind( rrw, rrd, rrc, rrcd, rrdc )*rrr,
+       lwd=rep(c(3,1,1),2), lty=rep(c(1:3),c(9,3,3)),
+       col=rep(clr[c(1:3,3,3)],each=3),type="l" )
+ axis( side=4, at=5:20/10 * rrr, labels=FALSE, tcl=-0.3 )
+ axis( side=4, at=c(0.5,1,2) * rrr, labels=c(0.5,1,2) )
+ axis( side=4, at=3 * rrr, labels="RR", tcl=0 )
+ axis( side=1, at=seq( 20,100, 5), labels=F )
+ axis( side=1, at=seq( 20,100,20) )
+ axis( side=1, at=seq(105,120,5), labels=F )
+ axis( side=1, at=seq(110,120,10), labels=c(2000,2010) )
+ mtext( c("Age","Date"), at=c(60,112.5), side=1, line=3/1.6 )
+ }
> plmt()
> text( rep(23,3), ylm[2]*0.7^(3:1),
+       c("Well","DM","Cancer"),  col=clr[1:3], adj=0, font=2 )
> # Mortality among women
> mtw  <-  TR["Well" ,"D-W" ,,"Cross","F",] * 10^5
> mtd  <-  TR["DM"   ,"D-DM",,"Cross","F",] * 10^5
> mtc  <-  TR["Ca"   ,"D-Ca",,"Cross","F",] * 10^5
```

```
> mtcd <-  TR["Ca-DM","D-CD",,"Cross","F",] * 10^5
> mtdc <-  TR["DM-Ca","D-DC",,"Cross","F",] * 10^5
> rrw  <- pRR["Well" ,"D-W" ,,"Cross","F",]
> rrd  <- pRR["DM"   ,"D-DM",,"Cross","F",]
> rrc  <- pRR["Ca"   ,"D-Ca",,"Cross","F",]
> rrcd <- pRR["Ca-DM","D-CD",,"Cross","F",]
> rrdc <- pRR["DM-Ca","D-DC",,"Cross","F",]
> plmt()
> mtext( "Cancer incidence rates per 1000 PY",
+        side=2, line=1/1.6, at=0.75, las=0, outer=TRUE )
> mtext( "Mortality rates per 1000 PY",
+        side=2, line=1/1.6, at=0.25, las=0, outer=TRUE )
> mtext( "Men"  , side=3, line=-1, at=0.25, las=0, outer=TRUE )
> mtext( "Women", side=3, line=-1, at=0.75, las=0, outer=TRUE )
```

8.6 Transition probabilities

Now we have the transition rates corresponding to 1/10 year in the array `TR`, but we need to fill in the diagonals to get a proper transition matrix. To this ens we need a function that does this properly; note that the entries in `TR` are cumulative rates corresponding to a period of length 1/10 year (well, formally `int`). Thus if transition cumulatieve rates *from* a given state are, say, Λ_1, Λ_2 Λ_3, the 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 must be be 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 tehn be converted to 0 before the diagonal is
+ # filled with the survival probabilities
+ res[is.na(res)] <- 0
+ diag( res ) <- exp( -sm )
+ res
+ }
```

We can then convert the matrices of cumulative transition intensities to matrices of transition probabilities. From now on we do not need the cis any more, so we skip them:

```
> TR <- TR[,,,,,1]
> print.table( round(        TR[,,800,1,1]  *10^3 ), zero.print="." )
         to
  from    Well DM DM-Ca Ca Ca-DM D-W D-DM D-Ca D-DC D-CD
    Well     .  2     .  4     .   5    .    .    .    .
    DM       .  .     4  .     .   .    9    .    .    .
    DM-Ca    .  .     .  .     .   .    .    .   30    .
    Ca       .  .     .  .     4   .    .   18    .    .
    Ca-DM    .  .     .  .     .   .    .    .    .   19
    D-W      .  .     .  .     .   .    .    .    .    .
    D-DM     .  .     .  .     .   .    .    .    .    .
    D-Ca     .  .     .  .     .   .    .    .    .    .
    D-DC     .  .     .  .     .   .    .    .    .    .
    D-CD     .  .     .  .     .   .    .    .    .    .
> print.table( round( ci2pr( TR[,,800,1,1] )*10^3 ), zero.print="." )
         to
  from    Well   DM DM-Ca   Ca Ca-DM  D-W D-DM D-Ca D-DC D-CD
    Well  990    2     .    4     .    5    .    .    .    .
```

```
    DM        .  987     4    .     .    .    9    .    .    .
    DM-Ca     .    .   971    .     .    .    .    .   29    .
    Ca        .    .     .  979     4    .    .   17    .    .
    Ca-DM     .    .     .    .   981    .    .    .    .   19
    D-W       .    .     .    .     . 1000    .    .    .    .
    D-DM      .    .     .    .     .    . 1000    .    .    .
    D-Ca      .    .     .    .     .    .    . 1000    .    .
    D-DC      .    .     .    .     .    .    .    . 1000    .
    D-CD      .    .     .    .     .    .    .    .    . 1000
> TRp <- apply( TR, 3:5, ci2pr )
> dim( TRp )
  [1]  100 1020    2    2
> # Note that apply does not recognize the dim attribute of FUN argument
> dim( TRp ) <- c(10,10,dim(TRp)[-1])
> dimnames( TRp ) <- dimnames( TR )
> print.table( round( TRp[,,800,1,1]*10^3 ), zero.print="." )
         to
  from    Well   DM DM-Ca   Ca Ca-DM  D-W D-DM D-Ca D-DC D-CD
    Well   990    2     .    4     .    5    .    .    .    .
    DM        .  987     4    .     .    .    9    .    .    .
    DM-Ca     .    .   971    .     .    .    .    .   29    .
    Ca        .    .     .  979     4    .    .   17    .    .
    Ca-DM     .    .     .    .   981    .    .    .    .   19
    D-W       .    .     .    .     . 1000    .    .    .    .
    D-DM      .    .     .    .     .    . 1000    .    .    .
    D-Ca      .    .     .    .     .    .    . 1000    .    .
    D-DC      .    .     .    .     .    .    .    . 1000    .
    D-CD      .    .     .    .     .    .    .    .    . 1000
```

The just printed matrix is the transition matrix (multiplied by 1000) from age 80 to 80.1, so in order to get the probability distribution at 80.1, we just multiply the state-distribution at time 80.0 (as a row vector) with the transition matrix; this must of course be looped over all the other dimensions of TR:

```
> names( dimnames( TRp ) )
  [1] "from"  "to"    "age"   "scene" "sex"
> for( sc in dimnames(TRp)[["scene"]] )
+ for( sx in dimnames(TRp)[["sex"]] )
+    {
+    # Initialize at age 0
+    PR[,1,sc,sx] <- c(1,rep(0,9))
+    # Compute distribution at endpoint of each interval
+    for( ag in 1:dim(TRp)[3] )
+    {
+    PR[,ag,sc,sx] <- PR[ ,max(ag-1,1),sc,sx] %*%
+                     TRp[,,    ag     ,sc,sx]
+    }
+    }
> summary( PR )
       Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  0.0000000 0.0003761 0.0116500 0.1000000 0.0746800 0.9997000
> summary( apply( PR, 2:4, sum ) )
     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
        1       1       1       1       1       1
```

Now we have the distribution of the persons in the different states under various scenarios, and we can plot the resulting distribution of the states as function of time; for eeach of the 4 combinations of scenario and sex we can plot the probabilities of being in each of the 10 states, but we must put them in the right order:

```
> round( t(PR[,600+1:5,1,1])*100, 1 )
```

```
         State
  age     Well   DM DM-Ca  Ca Ca-DM D-W D-DM D-Ca D-DC D-CD
    60.05 71.4 10.7   0.5 4.2   0.2 7.1  1.6  3.8  0.4  0.1
    60.15 71.2 10.8   0.5 4.2   0.2 7.1  1.7  3.9  0.4  0.1
    60.25 71.0 10.8   0.5 4.3   0.2 7.1  1.7  3.9  0.4  0.1
    60.35 70.8 10.9   0.5 4.3   0.2 7.2  1.7  4.0  0.4  0.1
    60.45 70.6 10.9   0.5 4.3   0.2 7.2  1.7  4.0  0.4  0.1
```

```
> perm <- c(2,3,5,4,1,6,8,10,9,7)
> round( t(PR[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
    60.05 10.7   0.5   0.2 4.2 71.4 7.1  3.8  0.1  0.4  1.6
    60.15 10.8   0.5   0.2 4.2 71.2 7.1  3.9  0.1  0.4  1.7
    60.25 10.8   0.5   0.2 4.3 71.0 7.1  3.9  0.1  0.4  1.7
    60.35 10.9   0.5   0.2 4.3 70.8 7.2  4.0  0.1  0.4  1.7
    60.45 10.9   0.5   0.2 4.3 70.6 7.2  4.0  0.1  0.4  1.7
```

```
> CR <- apply( PR[perm,,,], 2:4, cumsum )
> str( PR )
```

```
   num [1:10, 1:1020, 1:2, 1:2] 1.00 1.13e-05 0.00 2.72e-05 0.00 ...
   - attr(*, "dimnames")=List of 4
    ..$ State: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ age  : chr [1:1020] "0.05" "0.15" "0.25" "0.35" ...
    ..$ scene: chr [1:2] "Cross" "Long"
    ..$ sex  : chr [1:2] "M" "F"
```

```
> str( CR )
```

```
   num [1:10, 1:1020, 1:2, 1:2] 1.13e-05 1.13e-05 1.13e-05 3.85e-05 1.00 ...
   - attr(*, "dimnames")=List of 4
    ..$      : chr [1:10] "DM" "DM-Ca" "Ca-DM" "Ca" ...
    ..$ age  : chr [1:1020] "0.05" "0.15" "0.25" "0.35" ...
    ..$ scene: chr [1:2] "Cross" "Long"
    ..$ sex  : chr [1:2] "M" "F"
```

```
> ftable( round( apply( PR, c(1,3,4), max )*100, 1 ), col.vars=1 )
```

```
            State  Well    DM DM-Ca    Ca Ca-DM   D-W  D-DM  D-Ca  D-DC  D-CD
  scene sex
  Cross M         100.0  13.4   1.7   6.9   1.3  35.7  20.1  29.2   9.6   5.4
        F         100.0  12.7   1.9   9.1   1.7  36.5  19.8  29.7   7.9   5.6
  Long  M          99.8  15.5   5.4  11.1   6.4  21.6  12.9  25.5  19.7  11.3
        F          99.8  16.2   4.9  11.9   6.2  16.9  10.5  26.2  20.6  12.1
```

```
> ftable( round( apply( CR, c(1,3,4), max )*100, 1 ), col.vars=1 )
```

```
                DM DM-Ca Ca-DM    Ca  Well   D-W  D-Ca  D-CD  D-DC  D-DM
  scene sex
  Cross M     13.4  14.7  15.5  22.4 100.0 100.0 100.0 100.0 100.0 100.0
        F     12.7  14.4  15.9  24.9 100.0 100.0 100.0 100.0 100.0 100.0
  Long  M     15.5  19.2  23.8  34.6  99.8 100.0 100.0 100.0 100.0 100.0
        F     16.2  20.7  26.5  37.2  99.8 100.0 100.0 100.0 100.0 100.0
```

```
> round( t( CR[,400+1:5,"Cross","M"] )*100, 1 )
```

```
  age      DM DM-Ca Ca-DM  Ca Well  D-W D-Ca D-CD D-DC D-DM
    40.05 2.5   2.5   2.5 3.9 97.2 99.5 99.9 99.9 99.9  100
    40.15 2.5   2.5   2.5 3.9 97.2 99.5 99.9 99.9 99.9  100
    40.25 2.5   2.5   2.5 3.9 97.1 99.5 99.9 99.9 99.9  100
    40.35 2.5   2.6   2.6 4.0 97.1 99.5 99.9 99.9 99.9  100
    40.45 2.6   2.6   2.6 4.0 97.1 99.5 99.9 99.9 99.9  100
```

In order to plot the different probabilities we use the `polygon` trick, and in order to visualize the joint occurrence of diabetes and cancer we define semi-transparent colors

```
> nul <- rep( 0, dim(CR)[2] )
> sx <- 1
> sc <- 1
> aa <- as.numeric( dimnames(CR)[["age"]] )
> hred  <- rgb(0.9,0.3,0.3)
> hpur  <- rgb(0.8,0.0,0.8)
> hblue <- rgb(0.3,0.3,0.9)
> tred  <- rgb(99,0, 0,70,maxC=100)
> tblue <- rgb( 0,0,90,70,maxC=100)
> 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 dimnames(CR)[[3]][1] )
+ for( sx in dimnames(CR)[[4]] )
+    {
+ 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 )
+ text( 55, 50, sx, font=2 )
+
+ polygon( c(aa,rev(aa)), c(CR[3,,sc,sx],rev(nul))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[1,,sc,sx],
+                        rev(CR[4,,sc,sx]))*100,
+          col = tred, border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[6,,sc,sx],
+                        rev(CR[5,,sc,sx]))*100,
+          col = "gray", border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[7,,sc,sx],
+                        rev(CR[10,,sc,sx]))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[6,,sc,sx],
+                        rev(CR[9,,sc,sx]))*100,
+          col = tred, border="transparent")
+ matlines( aa, 100*t(CR[c(2,5,8),,sc,sx]),
+           lty=1, col=c("white","black")[c(1,2,1)], lwd=c(1,3,1), type="l" )
+ mtext( "Age (years)", side=1, outer=TRUE )
+    }
> mtext( "Fraction of persons (%)", side=2, outer=TRUE, las=0 )
```

We can of course also make the same exercise *conditional* being alive at age 50, 60 etc, but as is seen from figure 11 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 <-
+ PR50 <- PR60 <- PR70 <- PR*0
> dimnames( PR )[[2]][500]
  [1] "49.95"
> dimnames( PR )[[1]]
   [1] "Well"  "DM"    "DM-Ca" "Ca"    "Ca-DM" "D-W"   "D-DM"  "D-Ca"  "D-DC"  "D-CD"
> for( sc in dimnames(TR)[["scene"]] )
+ for( sx in dimnames(TR)[["sex"]] )
+    {
+    # Initialize to all being well at age 50, 60, 70
+    PR50[,500,sc,sx] <-
+    PR60[,600,sc,sx] <-
+    PR70[,700,sc,sx] <- c(1,rep(0,9))
+    # Initialize to all being DM at age 50, 60, 70
+    DM50[,500,sc,sx] <-
+    DM60[,600,sc,sx] <-
+    DM70[,700,sc,sx] <- c(0,1,rep(0,8))
```

```
+    for( ag in 501:1020 )
+       {
+                  PR50[,ag,sc,sx] <- PR50[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>600 ) PR60[,ag,sc,sx] <- PR60[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>700 ) PR70[,ag,sc,sx] <- PR70[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+                  DM50[,ag,sc,sx] <- DM50[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>600 ) DM60[,ag,sc,sx] <- DM60[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>700 ) DM70[,ag,sc,sx] <- DM70[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+       }
+    }
```

For further comparisons we print the distribution on states at age 102 years:

```
> round( ww <- cbind( PR[,1020,"Cross",],
+                     PR50[,1020,"Cross",],
+                     PR60[,1020,"Cross",],
+                     PR70[,1020,"Cross",] )*100, 1 )
           M    F    M    F    M    F    M    F
  Well   0.0  0.2  0.0  0.2  0.1  0.3  0.1  0.4
  DM     0.0  0.1  0.0  0.1  0.0  0.1  0.0  0.1
  DM-Ca  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
  Ca     0.0  0.1  0.0  0.1  0.1  0.1  0.1  0.1
  Ca-DM  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
  D-W   35.7 36.5 36.7 40.1 40.1 45.2 48.9 55.6
  D-DM  20.1 19.8 18.6 18.4 16.3 17.3 12.8 14.5
  D-Ca  29.2 29.7 30.1 29.0 30.5 26.9 28.5 22.6
  D-DC   9.6  7.9  8.6  6.4  6.8  5.2  4.1  3.2
  D-CD   5.4  5.6  5.9  5.6  6.1  4.9  5.5  3.5
```

We con compute the fraction of those without disease at different age and who eventually gets a DM diagnosis, who also have a cancer diagnosis:

```
> round( ww[c(7,9,10),], 1 )
          M    F    M    F    M    F    M    F
  D-DM 20.1 19.8 18.6 18.4 16.3 17.3 12.8 14.5
  D-DC  9.6  7.9  8.6  6.4  6.8  5.2  4.1  3.2
  D-CD  5.4  5.6  5.9  5.6  6.1  4.9  5.5  3.5

> round( apply(ww[    9:10 ,],2,sum)/
+        apply(ww[c(7,9:10),],2,sum)*100, 1 )

     M    F    M    F    M    F    M    F
  42.7 40.6 43.7 39.5 44.3 36.9 42.8 31.6
```

We can also compute the fractio that gets a cancer diagnosis, regradless of diabetes status

```
> round( addmargins(ww[c(8:10),],1), 1 )
          M    F    M    F    M    F    M    F
  D-Ca 29.2 29.7 30.1 29.0 30.5 26.9 28.5 22.6
  D-DC  9.6  7.9  8.6  6.4  6.8  5.2  4.1  3.2
  D-CD  5.4  5.6  5.9  5.6  6.1  4.9  5.5  3.5
  Sum  44.1 43.2 44.6 41.0 43.5 37.0 38.1 29.3
```

and we can see how that compares to the fraction among those with diabetes at a given age that contracts diabetes:

```
> round( cbind( PR[,1020,"Cross",],
+             DM50[,1020,"Cross",],
+             DM60[,1020,"Cross",],
+             DM70[,1020,"Cross",] )*100, 1 )
```

```
          M    F    M    F    M    F    M    F
Well    0.0  0.2  0.0  0.0  0.0  0.0  0.0  0.0
DM      0.0  0.1  0.0  0.1  0.0  0.1  0.0  0.1
DM-Ca   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
Ca      0.0  0.1  0.0  0.0  0.0  0.0  0.0  0.0
Ca-DM   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
D-W    35.7 36.5  0.0  0.0  0.0  0.0  0.0  0.0
D-DM   20.1 19.8 63.5 62.9 63.5 66.1 67.9 73.4
D-Ca   29.2 29.7  0.0  0.0  0.0  0.0  0.0  0.0
D-DC    9.6  7.9 36.5 37.0 36.4 33.8 32.0 26.5
D-CD    5.4  5.6  0.0  0.0  0.0  0.0  0.0  0.0
```

We can now plot the comparison between the life-long outloook of a person with and without diabetes:

```
> CRpl <-
+ function( PR, sc, sx, rm, sepcol="white" )
+   {
+ CR <- apply( PR[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[3,-rm,sc,sx],
+                                      rev(nul[-rm]))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[1,-rm,sc,sx],
+                                  rev(CR[4,-rm,sc,sx]))*100,
+          col = tred, border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[6,-rm,sc,sx],
+                                  rev(CR[5,-rm,sc,sx]))*100,
+          col = "gray", border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[7,-rm,sc,sx],
+                                  rev(CR[10,-rm,sc,sx]))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[6,-rm,sc,sx],
+                                  rev(CR[9,-rm,sc,sx]))*100,
+          col = tred, border="transparent")
+ matlines( aa[-rm], 100*t(CR[c(2,5,8),-rm,sc,sx]),
+           lty=1, col=c(sepcol,"black")[c(1,2,1)], lwd=c(1,3,1), type="l" )
+    }
```

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

```
> par( mfcol=c(3,4), mar=c(2,2,1,3), oma=c(2,2,2,0), mgp=c(3,1,0)/1.6, las=1 )
> for( sc in dimnames(CR)[[3]][1] )
+ for( sx in dimnames(CR)[[4]] )
+    {
+ CRpl( PR50, sc, sx, 1:500 )
+ CRpl( PR60, sc, sx, 1:600 )
+ CRpl( PR70, 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 )
```

9 Comparing cancer RRs

We have collected data form other studies that on population basis has esitmated the overall RR of specific cancers relative to

```
> library(Epi)
> # Read the keyed-in data
> oth <- read.table("./data/other.dat",header=T)
> str( oth )
  'data.frame':        216 obs. of  7 variables:
   $ sex   : Factor w/ 2 levels "F","M": 2 1 2 1 2 1 2 1 2 1 ...
   $ diag  : Factor w/ 24 levels "                          Brain",..: 22 22 12 12 7 7 2 2 20 20 .
   $ author: Factor w/ 6 levels "Adami","Coughlin",..: 6 6 6 6 6 6 6 6 6 6 ...
   $ RR    : num  1.1 1.1 1.3 1 1.2 1.1 1.3 1.1 1.1 1 ...
   $ lo    : num  1.1 1.1 1 0.7 1 1 1.1 1 0.9 0.9 ...
   $ hi    : num  1.1 1.1 1.6 1.5 1.3 1.4 1.4 1.2 1.2 1.2 ...
   $ N     : int  4666 4165 67 26 188 131 413 442 235 167 ...

> head( oth )

    sex                              diag    author  RR  lo  hi    N
  1   M           All malignant neoplasms Wiederoff 1.1 1.1 1.1 4666
  2   F           All malignant neoplasms Wiederoff 1.1 1.1 1.1 4165
  3   M                        Oesophagus Wiederoff 1.3 1.0 1.6   67
  4   F                        Oesophagus Wiederoff 1.0 0.7 1.5   26
  5   M                           Stomach Wiederoff 1.2 1.0 1.3  188
  6   F                           Stomach Wiederoff 1.1 1.0 1.4  131

> # Clean up the site names
> nn <- as.character( oth$diag )
> nn <- gsub( "  ", "", nn )
> nn <- ifelse( substr(nn,1,1)==substr( gsub(" ","",nn), 1, 1 ),
+               nn,
+               substr(nn,2,50) )
> oth$diag <- nn
> nn <- unique( oth$diag )
> nn

   [1] "All malignant neoplasms"    "Oesophagus"
   [3] "Stomach"                    "Colon"
   [5] "Rectum "                    "Liver"
   [7] "Pancreas"                   "Lung, bronchus and pleura"
   [9] "Melanoma of skin"           "Other skin"
  [11] "Breast"                     "Cervix uteri"
  [13] "Corpus uteri"               "Ovary, fallopian tube etc."
  [15] "Prostate"                   "Testis"
  [17] "Kidney"                     "Urinary bladder"
  [19] "Brain"                      "Thyroid"
  [21] "Hodgkins lymphoma"          "Non-Hodgkin lymphoma"
  [23] "Multiple myeloma"           "Leukaemia"

> # Make sure they are in the correct order
> oth$diag <- factor( oth$diag, levels=nn )
> levels( oth$diag )

   [1] "All malignant neoplasms"    "Oesophagus"
   [3] "Stomach"                    "Colon"
   [5] "Rectum "                    "Liver"
   [7] "Pancreas"                   "Lung, bronchus and pleura"
   [9] "Melanoma of skin"           "Other skin"
  [11] "Breast"                     "Cervix uteri"
  [13] "Corpus uteri"               "Ovary, fallopian tube etc."
  [15] "Prostate"                   "Testis"
  [17] "Kidney"                     "Urinary bladder"
  [19] "Brain"                      "Thyroid"
  [21] "Hodgkins lymphoma"          "Non-Hodgkin lymphoma"
  [23] "Multiple myeloma"           "Leukaemia"
```

```
> # Turn it into a 4-dimensional array for plotting
> xxx <- xtabs( cbind(RR,lo,hi) ~ author + diag + sex,
+               na.action = na.omit,
+               data= oth )[,-10,,]
> xxx[xxx==0] <- NA
> xxx <- xxx[c(1,3:5,2),,,]
> dimnames(xxx)[[1]]

  [1] "Adami"    "Johnson"  "Kajuter"  "Vecchia"  "Coughlin"

> dimnames(xxx)[[2]][8] <- paste( "     ", dimnames(xxx)[[2]][8] )
> dimnames(xxx)[[2]][8]

  [1] "      Lung, bronchus and pleura"

> nd <- dim(xxx)[2]
> # Get the data from the DK study
> load( file="../../data/ana1i.Rdata" )
> dimnames(res)[[1]] <- gsub(" \\(excl\\. anus)","",
+                            dimnames(res)[[1]])
> str( res )

   num [1:29, 1:2, 1:4, 1:3] 1.19 1.27 1.25 1.32 1.4 ...
   - attr(*, "dimnames")=List of 4
    ..$ diag: chr [1:29] "All malignant neoplasms" "Oesophagus" "Stomach" "Colon incl. rectosigmoide
    ..$ sex : chr [1:2] "M" "F"
    ..$ type: chr [1:4] "DM/noIns" "DM/Ins" "Ins vs. noIns" "DM"
    ..$ est : chr [1:3] "Est" "lo" "hi"

> data.frame( 1:29, dimnames(res)[[1]] )

     X1.29                dimnames.res...1..
  1      1           All malignant neoplasms
  2      2                       Oesophagus
  3      3                          Stomach
  4      4       Colon incl. rectosigmoideum
  5      5                  Ascending colon
  6      6                 Transverse colon
  7      7      Descending and sigmoid colon
  8      8 Other colon (unspec. or multiple)
  9      9                           Rectum
  10    10                 Colorectal cancer
  11    11                            Liver
  12    12                         Pancreas
  13    13         Lung, bronchus and pleura
  14    14                 Melanoma of skin
  15    15                           Breast
  16    16                      Cervix uteri
  17    17                      Corpus uteri
  18    18        Ovary, fallopian tube etc.
  19    19                         Prostate
  20    20                           Testis
  21    21                           Kidney
  22    22                  Urinary bladder
  23    23                            Brain
  24    24                          Thyroid
  25    25                Hodgkin's lymphoma
  26    26              Non-Hodgkin lymphoma
  27    27                 Multiple myeloma
  28    28                        Leukaemia
  29    29                            Other

> # Check which ones belong to the ones from the keyed-in data
> wh <- c(1:4,9,11:28)
> cbind( dimnames(res)[[1]][wh], dimnames(xxx)[[2]] )

       [,1]                          [,2]
  [1,] "All malignant neoplasms"     "All malignant neoplasms"
  [2,] "Oesophagus"                  "Oesophagus"
  [3,] "Stomach"                     "Stomach"
  [4,] "Colon incl. rectosigmoideum" "Colon"
```

```
 [5,] "Rectum"                      "Rectum "
 [6,] "Liver"                       "Liver"
 [7,] "Pancreas"                    "Pancreas"
 [8,] "Lung, bronchus and pleura"   "      Lung, bronchus and pleura"
 [9,] "Melanoma of skin"            "Melanoma of skin"
[10,] "Breast"                      "Breast"
[11,] "Cervix uteri"                "Cervix uteri"
[12,] "Corpus uteri"                "Corpus uteri"
[13,] "Ovary, fallopian tube etc."  "Ovary, fallopian tube etc."
[14,] "Prostate"                    "Prostate"
[15,] "Testis"                      "Testis"
[16,] "Kidney"                      "Kidney"
[17,] "Urinary bladder"             "Urinary bladder"
[18,] "Brain"                       "Brain"
[19,] "Thyroid"                     "Thyroid"
[20,] "Hodgkin's lymphoma"          "Hodgkins lymphoma"
[21,] "Non-Hodgkin lymphoma"        "Non-Hodgkin lymphoma"
[22,] "Multiple myeloma"            "Multiple myeloma"
[23,] "Leukaemia"                   "Leukaemia"
```

```
> # Extract the corresponding ones from data
> cwf <- res[wh,,,]
> # Nullify male breast cancer
> xxx[,"Breast","M",] <- NA
> cwf["Breast","M",,] <- NA
> # Combine to one array
> ( dnam <- dimnames( xxx ) )
```

```
$author
[1] "Adami"    "Johnson"  "Kajuter"  "Vecchia"  "Coughlin"

$diag
 [1] "All malignant neoplasms"         "Oesophagus"
 [3] "Stomach"                         "Colon"
 [5] "Rectum "                         "Liver"
 [7] "Pancreas"                        "      Lung, bronchus and pleura"
 [9] "Melanoma of skin"                "Breast"
[11] "Cervix uteri"                    "Corpus uteri"
[13] "Ovary, fallopian tube etc."      "Prostate"
[15] "Testis"                          "Kidney"
[17] "Urinary bladder"                 "Brain"
[19] "Thyroid"                         "Hodgkins lymphoma"
[21] "Non-Hodgkin lymphoma"            "Multiple myeloma"
[23] "Leukaemia"

$sex
[1] "F" "M"

[[4]]
[1] "RR" "lo" "hi"
```

```
> dnam[["author"]] <- c("Carstensen",dnam[["author"]])
> XXX <- NArray( dnam )
> XXX[-1,,,] <- xxx
> str( cwf )
```

```
 num [1:23, 1:2, 1:4, 1:3] 1.19 1.27 1.25 1.32 1.11 ...
 - attr(*, "dimnames")=List of 4
  ..$ diag: chr [1:23] "All malignant neoplasms" "Oesophagus" "Stomach" "Colon incl. rectosigmoide
  ..$ sex : chr [1:2] "M" "F"
  ..$ type: chr [1:4] "DM/noIns" "DM/Ins" "Ins vs. noIns" "DM"
  ..$ est : chr [1:3] "Est" "lo" "hi"
```

```
> str( XXX[1,,,] )
```

```
 num [1:23, 1:2, 1:3] NA NA NA NA NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ diag: chr [1:23] "All malignant neoplasms" "Oesophagus" "Stomach" "Colon" ...
  ..$ sex : chr [1:2] "F" "M"
  ..$     : chr [1:3] "RR" "lo" "hi"
```

```
> str( cwf[,2:1,4,] )

   num [1:23, 1:2, 1:3] 1.202 0.979 1.344 1.204 0.996 ...
   - attr(*, "dimnames")=List of 3
    ..$ diag: chr [1:23] "All malignant neoplasms" "Oesophagus" "Stomach" "Colon incl. rectosigmoide
    ..$ sex : chr [1:2] "F" "M"
    ..$ est : chr [1:3] "Est" "lo" "hi"

> XXX[1,,,] <- cwf[,2:1,4,]
> str( XXX )

   num [1:6, 1:23, 1:2, 1:3] 1.2 1.1 NA NA NA ...
   - attr(*, "dimnames")=List of 4
    ..$ author: chr [1:6] "Carstensen" "Adami" "Johnson" "Kajuter" ...
    ..$ diag  : chr [1:23] "All malignant neoplasms" "Oesophagus" "Stomach" "Colon" ...
    ..$ sex   : chr [1:2] "F" "M"
    ..$       : chr [1:3] "RR" "lo" "hi"

> RRCa <- XXX[,,2:1,]
> str( RRCa )

   num [1:6, 1:23, 1:2, 1:3] 1.21 1 NA NA NA ...
   - attr(*, "dimnames")=List of 4
    ..$ author: chr [1:6] "Carstensen" "Adami" "Johnson" "Kajuter" ...
    ..$ diag  : chr [1:23] "All malignant neoplasms" "Oesophagus" "Stomach" "Colon" ...
    ..$ sex   : chr [1:2] "M" "F"
    ..$       : chr [1:3] "RR" "lo" "hi"

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

With thsi array set up, we can make a comprehensive forest plot of estimates

```
> # Plot to compare the studies results to the previous ones.
> par( mar=c(3,1,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( RRCa[1,,"M",1:3], y=nd:1, col="transparent", lwd=1,
+           xlog=T, xlim=c(0.5,5), grid=c(5:19/10,4:10/2),
+           vref=1, ylim=c(1,nd),
+           xtic=c(0.5,0.7,1,1.5,2:5), xlab="RR, DM vs. non-DM" )
> nst <- dim(RRCa)[1]
> for( i in 1:nst )
+ {
+ pointsEst( RRCa[i,,"M",1:3], y=nd:1+(0+i)/(2.3*nst), col="blue",
+            lwd=1.5, cex=0.6 )
+ pointsEst( RRCa[i,,"F",1:3], y=nd:1+(i-nst+1)/(2.3*nst), col="red",
+            lwd=1.5, cex=0.6 )
+ }
```

10 Modelling of rates

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

```
> library( Epi )
> clear()
> load( file="./data/dcd.Rda" )
> addmargins( xtabs( cbind(D.dm,D.ca,D.dd) ~ state, data=dcd ), 1 )
  state     D.dm   D.ca   D.dd
    Well  289438 382959 431103
    DM         0  40854  93885
    DM-Ca      0      0  28648
    Ca     27958      0 289131
    Ca-DM      0      0  18465
    Dead       0      0      0
    Sum   317396 423813 861232
```

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 im modelling of the age-effects, we make histograms of the age-distribution of the events:

```
> par( mfrow=c(5,3), mar=c(3,1,1,1), mgp=c(3,1,0)/1.6 )
> par( mfg=c(1,1) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.dm), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="DM | Well" ) )
> par( mfg=c(1,2) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.ca), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Ca | Well" ) )
> par( mfg=c(1,3) ) ; with( subset( dcd, state=="Well" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | Well" ) )
> par( mfg=c(2,2) ) ; with( subset( dcd, state=="DM" ),
>                           hist( rep(A,D.ca), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Ca | DM" ) )
> par( mfg=c(2,3) ) ; with( subset( dcd, state=="DM" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | DM" ) )
> par( mfg=c(3,3) ) ; with( subset( dcd, state=="DM-Ca" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | DM-Ca" ) )
> par( mfg=c(4,1) ) ; with( subset( dcd, state=="Ca" ),
>                           hist( rep(A,D.dm), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="DM | Ca" ) )
> par( mfg=c(4,3) ) ; with( subset( dcd, state=="Ca" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | Ca" ) )
> par( mfg=c(5,3) ) ; with( subset( dcd, state=="Ca-DM" ),
>                           hist( rep(A,D.dd), breaks=0:100,
+                           col="black", main="", yaxt="n",
+                           ylab="", xlab="Dead | Ca-DM" ) )
```

10.1 Prerequisites for the natural splines

```
> library( Epi )
> library( splines )
> ( a.kn <- c(5,1:9*10) )

   [1]  5 10 20 30 40 50 60 70 80 90

> ( p.kn <- 1995.5 + 0:3*4.5 )

  [1] 1995.5 2000.0 2004.5 2009.0

> ( c.kn <- 1900 + 1:9*10 )

  [1] 1910 1920 1930 1940 1950 1960 1970 1980 1990
```

With this in place we can model the rates from each of the states, we fit either age-cohort or age-period models for the rates. Note that we are entering the person-years in units of 1/10 of a year, because we subsequently will compute cumulative probabilities over intervals of lenth 1/10 year:

```
> # Men
> cm.w2dm <- glm( D.dm ~ Ns(   A, knots=a.kn ) +
+                        Ns( P-A, knots=c.kn ),
+                        offset = log(Y*10),
+                        family = poisson,
+                        data = subset( dcd, state=="Well" & sex=="M" ) )
> cm.w2ca  <- update( cm.w2dm, D.ca ~ . )
> cm.w2dd  <- update( cm.w2dm, D.dd ~ . )
> cm.dm2ca <- update( cm.w2ca, data = subset( dcd, state=="DM" & sex=="M" ) )
> cm.dm2dd <- update( cm.w2dd, data = subset( dcd, state=="DM" & sex=="M" ) )
> cm.ca2dm <- update( cm.w2dm, data = subset( dcd, state=="Ca" & sex=="M" ) )
> cm.ca2dd <- update( cm.w2dd, . ~ . + state,
+                              data = subset( dcd, state %in% c("Ca","Ca-DM","DM-Ca" )
+                                                                  & sex=="M") )
> # Women
> cf.w2dm  <- update( cm.w2dm, data = subset( dcd, state=="Well" & sex=="F" ) )
> cf.w2ca  <- update( cf.w2dm, D.ca ~ . )
> cf.w2dd  <- update( cf.w2dm, D.dd ~ . )
> cf.dm2ca <- update( cf.w2ca, data = subset( dcd, state=="DM" & sex=="F" ) )
> cf.dm2dd <- update( cf.w2dd, data = subset( dcd, state=="DM" & sex=="F" ) )
> cf.ca2dm <- update( cf.w2dm, data = subset( dcd, state=="Ca" & sex=="F" ) )
> cf.ca2dd <- update( cf.w2dd, . ~ . + state,
+                              data = subset( dcd, state %in% c("Ca","Ca-DM","DM-Ca")
+                                                                  & sex=="F" ) )
```

The correponding models for crossectional rates are simply fitted:

```
> # Men
> pm.w2dm  <- update( cm.w2dm , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.w2ca  <- update( cm.w2ca , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.w2dd  <- update( cm.w2dd , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.dm2ca <- update( cm.dm2ca, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.dm2dd <- update( cm.dm2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.ca2dm <- update( cm.ca2dm, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pm.ca2dd <- update( cm.ca2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> # Women
> pf.w2dm  <- update( cf.w2dm , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.w2ca  <- update( cf.w2ca , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.w2dd  <- update( cf.w2dd , . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.dm2ca <- update( cf.dm2ca, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.dm2dd <- update( cf.dm2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.ca2dm <- update( cf.ca2dm, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )
> pf.ca2dd <- update( cf.ca2dd, . ~ . - Ns(P-A,knots=c.kn) + Ns(P,knots=p.kn) )


> round( ci.exp( pf.w2ca ), 3 )
```

```
                      exp(Est.)    2.5%   97.5%
(Intercept)               0.000   0.000   0.000
Ns(A, knots = a.kn)1      2.095   1.832   2.397
Ns(A, knots = a.kn)2      5.856   5.384   6.370
Ns(A, knots = a.kn)3     14.181  13.128  15.318
Ns(A, knots = a.kn)4     37.612  35.108  40.295
Ns(A, knots = a.kn)5     72.677  67.891  77.801
Ns(A, knots = a.kn)6    121.710 113.823 130.142
Ns(A, knots = a.kn)7    191.942 178.685 206.184
Ns(A, knots = a.kn)8     82.470  72.671  93.590
Ns(A, knots = a.kn)9    170.229 155.709 186.103
Ns(P, knots = p.kn)1      1.047   1.029   1.066
Ns(P, knots = p.kn)2      1.425   1.378   1.474
Ns(P, knots = p.kn)3      1.192   1.178   1.207
```

10.2 Computing the 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.

Since we have restricted ourselves to a scenery where we have only one time scale, namely age we can do the clculations in closed form by setting up the transition probability matrix for small intervals (of length `int` years. For the sake of completeness we also we also set up s asimilar matrix for the RR by period; mainly for showing the estimated RRs by period / cohort:

```
> int <- 0.1
> a.pt <- seq(0,102,int)[-1] - int/2
> p.pt <- seq(1995,2010,,100)
> c.pt <- seq(1900,2000,,100)
> ( 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 <-
+ anam <- list( from = states,
+                  to = states,
+                 age = a.pt,
+               scene = c("Cross","Long"),
+                 sex = c("M","F"),
+                what = c("Est","lo","hi") )
> pnam[3] <- list( p.pt ) ; names( pnam )[3] <- "per"
> cnam[3] <- list( c.pt ) ; names( cnam )[3] <- "coh"
> TR <- NArray( anam ) ; TR[is.na(TR)] <- 0
> PR <- TR[1,,,,,1]
> names( dimnames( PR ) )[1] <- "State"
> pRR <- NArray( pnam ) ; pRR[is.na(pRR)] <- 0
> cRR <- NArray( cnam ) ; cRR[is.na(cRR)] <- 0
```

First we fill in the estimated rates from the models just fitted, but we then must first have the relevant contrast matrices to extrcat the estimated rates:

```
> p.ref = 2005
> c.ref = 1950
> aM <- Ns(                  a.pt   , knots=a.kn )
> pR <- Ns( rep(p.ref,length(a.pt)), knots=p.kn )
> cR <- Ns( rep(c.ref,length(a.pt)), knots=c.kn )
> pA <- cbind( 1, aM, pR )
> cA <- cbind( 1, aM, cR )
> pM <- Ns( p.pt, knots=p.kn ) - Ns( rep(p.ref,length(p.pt)), knots=p.kn )
> cM <- Ns( c.pt, knots=c.kn ) - Ns( rep(c.ref,length(c.pt)), knots=c.kn )
```

First we fill in the age-specific rates that will later be used in the calculations of state occupancy probabilities:

```
> TR["Well" ,"DM"   ,,"Cross","M",] <- ci.exp( pm.w2dm , ctr.mat=pA )
> TR["Well" ,"Ca"   ,,"Cross","M",] <- ci.exp( pm.w2ca , ctr.mat=pA )
> TR["Well" ,"D-W"  ,,"Cross","M",] <- ci.exp( pm.w2dd , ctr.mat=pA )
> TR["DM"   ,"DM-Ca",,"Cross","M",] <- ci.exp( pm.dm2ca, ctr.mat=pA ) * 1.2
> TR["DM"   ,"D-DM" ,,"Cross","M",] <- ci.exp( pm.dm2dd, ctr.mat=pA )
> TR["Ca"   ,"Ca-DM",,"Cross","M",] <- ci.exp( pm.dm2ca, ctr.mat=pA )
> TR["Ca"   ,"D-Ca" ,,"Cross","M",] <- ci.exp( pm.ca2dd, ctr.mat=cbind(pA,1,0) )
> TR["DM-Ca","D-DC" ,,"Cross","M",] <- ci.exp( pm.ca2dd, ctr.mat=cbind(pA,0,0) )
> TR["Ca-DM","D-CD" ,,"Cross","M",] <- ci.exp( pm.ca2dd, ctr.mat=cbind(pA,0,1) )
> TR["Well" ,"DM"   ,,"Cross","F",] <- ci.exp( pf.w2dm , ctr.mat=pA )
> TR["Well" ,"Ca"   ,,"Cross","F",] <- ci.exp( pf.w2ca , ctr.mat=pA )
> TR["Well" ,"D-W"  ,,"Cross","F",] <- ci.exp( pf.w2dd , ctr.mat=pA )
> TR["DM"   ,"DM-Ca",,"Cross","F",] <- ci.exp( pf.dm2ca, ctr.mat=pA ) * 1.2
> TR["DM"   ,"D-DM" ,,"Cross","F",] <- ci.exp( pf.dm2dd, ctr.mat=pA )
> TR["Ca"   ,"Ca-DM",,"Cross","F",] <- ci.exp( pf.dm2ca, ctr.mat=pA )
> TR["Ca"   ,"D-Ca" ,,"Cross","F",] <- ci.exp( pf.ca2dd, ctr.mat=cbind(pA,1,0) )
> TR["DM-Ca","D-DC" ,,"Cross","F",] <- ci.exp( pf.ca2dd, ctr.mat=cbind(pA,0,0) )
> TR["Ca-DM","D-CD" ,,"Cross","F",] <- ci.exp( pf.ca2dd, ctr.mat=cbind(pA,0,1) )
> TR["Well" ,"DM"   ,,"Long" ,"M",] <- ci.exp( cm.w2dm , ctr.mat=cA )
> TR["Well" ,"Ca"   ,,"Long" ,"M",] <- ci.exp( cm.w2ca , ctr.mat=cA )
> TR["Well" ,"D-W"  ,,"Long" ,"M",] <- ci.exp( cm.w2dd , ctr.mat=cA )
> TR["DM"   ,"DM-Ca",,"Long" ,"M",] <- ci.exp( cm.dm2ca, ctr.mat=cA ) * 1.2
> TR["DM"   ,"D-DM" ,,"Long" ,"M",] <- ci.exp( cm.dm2dd, ctr.mat=cA )
> TR["Ca"   ,"Ca-DM",,"Long" ,"M",] <- ci.exp( cm.dm2ca, ctr.mat=cA )
> TR["Ca"   ,"D-Ca" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, ctr.mat=cbind(cA,1,0) )
> TR["DM-Ca","D-DC" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, ctr.mat=cbind(cA,0,0) )
> TR["Ca-DM","D-CD" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, ctr.mat=cbind(cA,0,1) )
> TR["Well" ,"DM"   ,,"Long" ,"F",] <- ci.exp( cf.w2dm , ctr.mat=cA )
> TR["Well" ,"Ca"   ,,"Long" ,"F",] <- ci.exp( cf.w2ca , ctr.mat=cA )
> TR["Well" ,"D-W"  ,,"Long" ,"F",] <- ci.exp( cf.w2dd , ctr.mat=cA )
> TR["DM"   ,"DM-Ca",,"Long" ,"F",] <- ci.exp( cf.dm2ca, ctr.mat=cA ) * 1.2
> TR["DM"   ,"D-DM" ,,"Long" ,"F",] <- ci.exp( cf.dm2dd, ctr.mat=cA )
> TR["Ca"   ,"Ca-DM",,"Long" ,"F",] <- ci.exp( cf.dm2ca, ctr.mat=cA )
> TR["Ca"   ,"D-Ca" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, ctr.mat=cbind(cA,1,0) )
> TR["DM-Ca","D-DC" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, ctr.mat=cbind(cA,0,0) )
> TR["Ca-DM","D-CD" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, ctr.mat=cbind(cA,0,1) )
```

We the use almost the same code to fill in the RRs associated with perios:

```
> pRR["Well" ,"DM"   ,,"Cross","M",] <- ci.exp( pm.w2dm , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"Ca"   ,,"Cross","M",] <- ci.exp( pm.w2ca , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"D-W"  ,,"Cross","M",] <- ci.exp( pm.w2dd , subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"DM-Ca",,"Cross","M",] <- ci.exp( pm.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"D-DM" ,,"Cross","M",] <- ci.exp( pm.dm2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"Ca-DM",,"Cross","M",] <- ci.exp( pm.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"D-Ca" ,,"Cross","M",] <- ci.exp( pm.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM-Ca","D-DC" ,,"Cross","M",] <- ci.exp( pm.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca-DM","D-CD" ,,"Cross","M",] <- ci.exp( pm.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"DM"   ,,"Cross","F",] <- ci.exp( pf.w2dm , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"Ca"   ,,"Cross","F",] <- ci.exp( pf.w2ca , subset="Ns\\(P", ctr.mat=pM )
> pRR["Well" ,"D-W"  ,,"Cross","F",] <- ci.exp( pf.w2dd , subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"DM-Ca",,"Cross","F",] <- ci.exp( pf.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM"   ,"D-DM" ,,"Cross","F",] <- ci.exp( pf.dm2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"Ca-DM",,"Cross","F",] <- ci.exp( pf.dm2ca, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca"   ,"D-Ca" ,,"Cross","F",] <- ci.exp( pf.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["DM-Ca","D-DC" ,,"Cross","F",] <- ci.exp( pf.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> pRR["Ca-DM","D-CD" ,,"Cross","F",] <- ci.exp( pf.ca2dd, subset="Ns\\(P", ctr.mat=pM )
> cRR["Well" ,"DM"   ,,"Long" ,"M",] <- ci.exp( cm.w2dm , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"Ca"   ,,"Long" ,"M",] <- ci.exp( cm.w2ca , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"D-W"  ,,"Long" ,"M",] <- ci.exp( cm.w2dd , subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"DM-Ca",,"Long" ,"M",] <- ci.exp( cm.dm2ca, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"D-DM" ,,"Long" ,"M",] <- ci.exp( cm.dm2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca"   ,"Ca-DM",,"Long" ,"M",] <- ci.exp( cm.dm2ca, subset="Ns\\(P", ctr.mat=cM )
```

```
> cRR["Ca"   ,"D-Ca" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM-Ca","D-DC" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca-DM","D-CD" ,,"Long" ,"M",] <- ci.exp( cm.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"DM"   ,,"Long" ,"F",] <- ci.exp( cf.w2dm , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"Ca"   ,,"Long" ,"F",] <- ci.exp( cf.w2ca , subset="Ns\\(P", ctr.mat=cM )
> cRR["Well" ,"D-W"  ,,"Long" ,"F",] <- ci.exp( cf.w2dd , subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"DM-Ca",,"Long" ,"F",] <- ci.exp( cf.dm2ca, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM"   ,"D-DM" ,,"Long" ,"F",] <- ci.exp( cf.dm2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca"   ,"Ca-DM",,"Long" ,"F",] <- ci.exp( cf.dm2ca, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca"   ,"D-Ca" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["DM-Ca","D-DC" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, subset="Ns\\(P", ctr.mat=cM )
> cRR["Ca-DM","D-CD" ,,"Long" ,"F",] <- ci.exp( cf.ca2dd, subset="Ns\\(P", ctr.mat=cM )
```

10.3 Transition rates

We have now estimates all 9 transition rates between states both using age-perido and age-cohort models. Of particular interest is the relationships between the cancer incidence rates between persons with and without diabetes, and the relationship between mortality rates between persons with and without DM and Cancer.

We would also like to see the ratio of the rates, so we need a small function that computes the ratio of two sets of estimates with the correct resluting ci:

```
> rates2RR <-
+ function( num, den )
+ {
+ # num and den are assumed to be 3-column-matrices with columns
+ # rate, lower, upper.
+ seln <- log( num[,3]/num[,1] )/1.96
+ seld <- log( den[,3]/den[,1] )/1.96
+ lrr  <- log( num[,1]/den[,1] )
+ exp( cbind( lrr, sqrt(seln^2+seld^2) ) %*% ci.mat() )
+ }
```

For the cross-sectional rates we make a plot of the age-specific cancer incidence rates along the RR by calendar time:

```
> clr <- c("limegreen","blue","red","black") # well, dm, ca, RR
> ylm <- c(5,5000)
> rrr <- 100
> par( mfrow=c(2,2), mar=c(3,2,1,3), oma=c(1,2,1,0),
+      mgp=c(3,1,0)/1.6, bty="n", las=1 )
> # Cancer incidence among men
> ciw <-  TR["Well","Ca"   ,,"Cross","M",] * 10^5
> cid <-  TR["DM"  ,"DM-Ca",,"Cross","M",] * 10^5
> rrw <- pRR["Well","Ca"   ,,"Cross","M",]
> rrd <- pRR["DM"  ,"DM-Ca",,"Cross","M",]
> plci <- function() {
+ matplot( a.pt, cbind( ciw, cid ),
+       lwd=rep(c(3,1,1),2), lty=1, col=rep(clr[1:2],each=3),type="l",
+       xlim=c(20,120), ylim=ylm/5, log="y", xaxt="n", xaxs="i",
+       ylab="", xlab="" )
+ matlines( a.pt, rates2RR( cid, ciw )*rrr,
+           lwd=c(3,1,1), lty=1, col=clr[4],type="l" )
+ abline( h=rrr )
+ matlines( p.pt-1890, cbind( rrw, rrd )*rrr,
+       lwd=rep(c(3,1,1),2), lty=1, col=clr[1:2],type="l" )
+ axis( side=4, at=5:20/10 * rrr, labels=FALSE, tcl=-0.3 )
+ axis( side=4, at=c(0.5,1,2) * rrr, labels=c(0.5,1,2) )
+ axis( side=4, at=3 * rrr, labels="RR", tcl=0 )
+ axis( side=1, at=seq( 20,100, 5), labels=F )
+ axis( side=1, at=seq( 20,100,20) )
```

```
+ axis( side=1, at=seq(105,120,5), labels=F )
+ axis( side=1, at=seq(110,120,10), labels=c(2000,2010) )
+ mtext( c("Age","Date"), at=c(60,112.5), side=1, line=3/1.6 )
+ }
> plci()
> # Cancer incidence among women
> ciw <-  TR["Well","Ca"   ,,"Cross","F",] * 10^5
> cid <-  TR["DM"  ,"DM-Ca",,"Cross","F",] * 10^5
> rrw <- pRR["Well","Ca"   ,,"Cross","F",]
> rrd <- pRR["DM"  ,"DM-Ca",,"Cross","F",]
> plci()
> # Mortality among men
> mtw  <-  TR["Well" ,"D-W" ,,"Cross","M",] * 10^5
> mtd  <-  TR["DM"   ,"D-DM",,"Cross","M",] * 10^5
> mtc  <-  TR["Ca"   ,"D-Ca",,"Cross","M",] * 10^5
> mtcd <-  TR["Ca-DM","D-CD",,"Cross","M",] * 10^5
> mtdc <-  TR["DM-Ca","D-DC",,"Cross","M",] * 10^5
> rrw  <- pRR["Well" ,"D-W" ,,"Cross","M",]
> rrd  <- pRR["DM"   ,"D-DM",,"Cross","M",]
> rrc  <- pRR["Ca"   ,"D-Ca",,"Cross","M",]
> rrcd <- pRR["Ca-DM","D-CD",,"Cross","M",]
> rrdc <- pRR["DM-Ca","D-DC",,"Cross","M",]
> plmt <- function() {
+ matplot( a.pt, cbind( mtw, mtd, mtc, mtcd, mtdc ),
+       lwd=rep(c(3,1,1),2), lty=rep(c(1:3),c(9,3,3)),
+       col=rep(clr[c(1:3,3,3)],each=3),type="l",
+       xlim=c(20,120), ylim=ylm, log="y", xaxt="n", xaxs="i",
+       ylab="", xlab="" )
+ # matlines( a.pt, rates2RR( mtd, mtw )*100,
+ #           lwd=c(3,1,1), lty=1, col=clr[4],type="l" )
+ lines( c(105,120), c(100,100) )
+ matlines( p.pt-1890, cbind( rrw, rrd, rrc, rrcd, rrdc )*rrr,
+       lwd=rep(c(3,1,1),2), lty=rep(c(1:3),c(9,3,3)),
+       col=rep(clr[c(1:3,3,3)],each=3),type="l" )
+ axis( side=4, at=5:20/10 * rrr, labels=FALSE, tcl=-0.3 )
+ axis( side=4, at=c(0.5,1,2) * rrr, labels=c(0.5,1,2) )
+ axis( side=4, at=3 * rrr, labels="RR", tcl=0 )
+ axis( side=1, at=seq( 20,100, 5), labels=F )
+ axis( side=1, at=seq( 20,100,20) )
+ axis( side=1, at=seq(105,120,5), labels=F )
+ axis( side=1, at=seq(110,120,10), labels=c(2000,2010) )
+ mtext( c("Age","Date"), at=c(60,112.5), side=1, line=3/1.6 )
+ }
> plmt()
> text( rep(23,3), ylm[2]*0.7^(3:1),
+       c("Well","DM","Cancer"),  col=clr[1:3], adj=0, font=2 )
> # Mortality among women
> mtw  <-  TR["Well" ,"D-W" ,,"Cross","F",] * 10^5
> mtd  <-  TR["DM"   ,"D-DM",,"Cross","F",] * 10^5
> mtc  <-  TR["Ca"   ,"D-Ca",,"Cross","F",] * 10^5
> mtcd <-  TR["Ca-DM","D-CD",,"Cross","F",] * 10^5
> mtdc <-  TR["DM-Ca","D-DC",,"Cross","F",] * 10^5
> rrw  <- pRR["Well" ,"D-W" ,,"Cross","F",]
> rrd  <- pRR["DM"   ,"D-DM",,"Cross","F",]
> rrc  <- pRR["Ca"   ,"D-Ca",,"Cross","F",]
> rrcd <- pRR["Ca-DM","D-CD",,"Cross","F",]
> rrdc <- pRR["DM-Ca","D-DC",,"Cross","F",]
> plmt()
> mtext( "Cancer incidence rates per 1000 PY",
+        side=2, line=1/1.6, at=0.75, las=0, outer=TRUE )
> mtext( "Mortality rates per 1000 PY",
+        side=2, line=1/1.6, at=0.25, las=0, outer=TRUE )
> mtext( "Men"  , side=3, line=-1, at=0.25, las=0, outer=TRUE )
> mtext( "Women", side=3, line=-1, at=0.75, las=0, outer=TRUE )
```

10.4 Transition probabilities

Now we have the transition rates corresponding to 1/10 year in the array `TR`, but we need to fill in the diagonals to get a proper transition matrix. To this ens we need a function that does this properly; note that the entries in `TR` are cumulative rates corresponding to a period of length 1/10 year (well, formally `int`). Thus if transition cumulatieve rates *from* a given state are, say, Λ_1, Λ_2 Λ_3, the 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 must be be 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 tehn be converted to 0 before the diagonal is
+ # filled with the survival probabilities
+ res[is.na(res)] <- 0
+ diag( res ) <- exp( -sm )
+ res
+ }
```

We can then convert the matrices of cumulative transition intensities to matrices of transition probabilities. From now on we do not need the cis any more, so we skip them:

```
> TR <- TR[,,,,,1]
> print.table( round(        TR[,,800,1,1]  *10^3 ), zero.print="." )
         to
  from    Well DM DM-Ca Ca Ca-DM D-W D-DM D-Ca D-DC D-CD
    Well     .  2     .  4     .   5    .    .    .    .
    DM       .  .     5  .     .   .    9    .    .    .
    DM-Ca    .  .     .  .     .   .    .    .   30    .
    Ca       .  .     .  .     4   .    .   18    .    .
    Ca-DM    .  .     .  .     .   .    .    .    .   19
    D-W      .  .     .  .     .   .    .    .    .    .
    D-DM     .  .     .  .     .   .    .    .    .    .
    D-Ca     .  .     .  .     .   .    .    .    .    .
    D-DC     .  .     .  .     .   .    .    .    .    .
    D-CD     .  .     .  .     .   .    .    .    .    .
> print.table( round( ci2pr( TR[,,800,1,1] )*10^3 ), zero.print="." )
         to
  from    Well   DM DM-Ca   Ca Ca-DM  D-W D-DM D-Ca D-DC D-CD
    Well  990    2     .    4     .    5    .    .    .    .
    DM       .  986     5    .     .    .    9    .    .    .
    DM-Ca    .    .   971    .     .    .    .    .   29    .
    Ca       .    .     .  979     4    .    .   17    .    .
    Ca-DM    .    .     .    .   981    .    .    .    .   19
    D-W      .    .     .    .     . 1000    .    .    .    .
    D-DM     .    .     .    .     .    . 1000    .    .    .
    D-Ca     .    .     .    .     .    .    . 1000    .    .
    D-DC     .    .     .    .     .    .    .    . 1000    .
    D-CD     .    .     .    .     .    .    .    .    . 1000
> TRp <- apply( TR, 3:5, ci2pr )
> dim( TRp )
  [1]  100 1020    2    2
> # Note that apply does not recognize the dim attribute of FUN argument
> dim( TRp ) <- c(10,10,dim(TRp)[-1])
> dimnames( TRp ) <- dimnames( TR )
> print.table( round( TRp[,,800,1,1]*10^3 ), zero.print="." )
```

```
        to
from     Well   DM DM-Ca   Ca Ca-DM  D-W D-DM D-Ca D-DC D-CD
  Well    990    2     .    4     .    5    .    .    .    .
  DM        .  986     5    .     .    .    9    .    .    .
  DM-Ca     .    .   971    .     .    .    .    .   29    .
  Ca        .    .     .  979     4    .    .   17    .    .
  Ca-DM     .    .     .    .   981    .    .    .    .   19
  D-W       .    .     .    .     . 1000    .    .    .    .
  D-DM      .    .     .    .     .    . 1000    .    .    .
  D-Ca      .    .     .    .     .    .    . 1000    .    .
  D-DC      .    .     .    .     .    .    .    . 1000    .
  D-CD      .    .     .    .     .    .    .    .    . 1000
```

The just printed matrix is the transition matrix (multiplied by 1000) from age 80 to 80.1, so in order to get the probability distribution at 80.1, we just multiply the state-distribution at time 80.0 (as a row vector) with the transition matrix; this must of course be looped over all the other dimensions of TR:

```
> names( dimnames( TRp ) )
  [1] "from"  "to"    "age"   "scene" "sex"
> for( sc in dimnames(TRp)[["scene"]] )
+ for( sx in dimnames(TRp)[["sex"]] )
+    {
+    # Initialize at age 0
+    PR[,1,sc,sx] <- c(1,rep(0,9))
+    # Compute distribution at endpoint of each interval
+    for( ag in 1:dim(TRp)[3] )
+    {
+    PR[,ag,sc,sx] <- PR[ ,max(ag-1,1),sc,sx] %*%
+                      TRp[,,    ag     ,sc,sx]
+    }
+    }
> summary( PR )
       Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
  0.0000000 0.0003898 0.0119700 0.1000000 0.0746400 0.9997000
> summary( apply( PR, 2:4, sum ) )
     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
        1       1       1       1       1       1
```

Now we have the distribution of the persons in the different states under various scenarios, and we can plot the resulting distribution of the states as function of time; for eeach of the 4 combinations of scenario and sex we can plot the probabilities of being in each of the 10 states, but we must put them in the right order:

```
> round( t(PR[,600+1:5,1,1])*100, 1 )
         State
  age     Well   DM DM-Ca  Ca Ca-DM D-W D-DM D-Ca D-DC D-CD
    60.05 71.4 10.6   0.5 4.2   0.2 7.1  1.6  3.8  0.5  0.1
    60.15 71.2 10.6   0.5 4.2   0.2 7.1  1.7  3.9  0.5  0.1
    60.25 71.0 10.7   0.6 4.3   0.2 7.1  1.7  3.9  0.5  0.1
    60.35 70.8 10.7   0.6 4.3   0.2 7.2  1.7  4.0  0.5  0.1
    60.45 70.6 10.8   0.6 4.3   0.2 7.2  1.7  4.0  0.5  0.1
> perm <- c(2,3,5,4,1,6,8,10,9,7)
> round( t(PR[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
    60.05 10.6   0.5   0.2 4.2 71.4 7.1  3.8  0.1  0.5  1.6
    60.15 10.6   0.5   0.2 4.2 71.2 7.1  3.9  0.1  0.5  1.7
    60.25 10.7   0.6   0.2 4.3 71.0 7.1  3.9  0.1  0.5  1.7
    60.35 10.7   0.6   0.2 4.3 70.8 7.2  4.0  0.1  0.5  1.7
    60.45 10.8   0.6   0.2 4.3 70.6 7.2  4.0  0.1  0.5  1.7
```

```
> CR <- apply( PR[perm,,,], 2:4, cumsum )
> str( PR )

   num [1:10, 1:1020, 1:2, 1:2] 1.00 1.13e-05 0.00 2.72e-05 0.00 ...
   - attr(*, "dimnames")=List of 4
    ..$ State: chr [1:10] "Well" "DM" "DM-Ca" "Ca" ...
    ..$ age  : chr [1:1020] "0.05" "0.15" "0.25" "0.35" ...
    ..$ scene: chr [1:2] "Cross" "Long"
    ..$ sex  : chr [1:2] "M" "F"

> str( CR )

   num [1:10, 1:1020, 1:2, 1:2] 1.13e-05 1.13e-05 1.13e-05 3.85e-05 1.00 ...
   - attr(*, "dimnames")=List of 4
    ..$      : chr [1:10] "DM" "DM-Ca" "Ca-DM" "Ca" ...
    ..$ age  : chr [1:1020] "0.05" "0.15" "0.25" "0.35" ...
    ..$ scene: chr [1:2] "Cross" "Long"
    ..$ sex  : chr [1:2] "M" "F"

> ftable( round( apply( PR, c(1,3,4), max )*100, 1 ), col.vars=1 )

            State  Well    DM DM-Ca    Ca Ca-DM   D-W  D-DM  D-Ca  D-DC  D-CD
  scene sex
  Cross M         100.0  12.9   1.9   6.9   1.3  35.7  18.8  29.2  10.9   5.4
        F         100.0  12.1   2.2   9.1   1.7  36.5  18.7  29.7   9.1   5.6
  Long  M          99.8  14.6   5.9  11.1   6.4  21.6  11.5  25.5  21.6  11.3
        F          99.8  15.0   5.4  11.9   6.2  16.9   9.4  26.2  22.7  12.1

> ftable( round( apply( CR, c(1,3,4), max )*100, 1 ), col.vars=1 )

                DM DM-Ca Ca-DM    Ca  Well   D-W  D-Ca  D-CD  D-DC  D-DM
  scene sex
  Cross M     12.9  14.4  15.2  22.0 100.0 100.0 100.0 100.0 100.0 100.0
        F     12.1  14.1  15.7  24.7 100.0 100.0 100.0 100.0 100.0 100.0
  Long  M     14.6  18.6  23.1  33.9  99.8 100.0 100.0 100.0 100.0 100.0
        F     15.0  20.0  25.7  36.5  99.8 100.0 100.0 100.0 100.0 100.0

> round( t( CR[,400+1:5,"Cross","M"] )*100, 1 )

  age      DM DM-Ca Ca-DM  Ca Well  D-W D-Ca D-CD D-DC D-DM
    40.05 2.5   2.5   2.5 3.9 97.2 99.5 99.9 99.9 99.9  100
    40.15 2.5   2.5   2.5 3.9 97.2 99.5 99.9 99.9 99.9  100
    40.25 2.5   2.5   2.5 3.9 97.1 99.5 99.9 99.9 99.9  100
    40.35 2.5   2.6   2.6 4.0 97.1 99.5 99.9 99.9 99.9  100
    40.45 2.5   2.6   2.6 4.0 97.1 99.5 99.9 99.9 99.9  100
```

In order to plot the different probabilities we use the `polygon` trick, and in order to visualize the joint occurrence of diabetes and cancer we define semi-transparent colors

```
> nul <- rep( 0, dim(CR)[2] )
> sx <- 1
> sc <- 1
> aa <- as.numeric( dimnames(CR)[["age"]] )
> hred  <- rgb(0.9,0.3,0.3)
> hpur  <- rgb(0.8,0.0,0.8)
> hblue <- rgb(0.3,0.3,0.9)
> tred  <- rgb(99,0, 0,70,maxC=100)
> tblue <- rgb( 0,0,90,70,maxC=100)
> 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 dimnames(CR)[[3]][1] )
+ for( sx in dimnames(CR)[[4]] )
+    {
+ 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 )
+ text( 55, 50, sx, font=2 )
```

```
+
+ polygon( c(aa,rev(aa)), c(CR[3,,sc,sx],rev(nul))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[1,,sc,sx],
+                       rev(CR[4,,sc,sx]))*100,
+          col = tred, border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[6,,sc,sx],
+                       rev(CR[5,,sc,sx]))*100,
+          col = "gray", border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[7,,sc,sx],
+                       rev(CR[10,,sc,sx]))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa,rev(aa)), c(CR[6,,sc,sx],
+                       rev(CR[9,,sc,sx]))*100,
+          col = tred, border="transparent")
+ matlines( aa, 100*t(CR[c(2,5,8),,sc,sx]),
+           lty=1, col=c("white","black")[c(1,2,1)], lwd=c(1,3,1), type="l" )
+ mtext( "Age (years)", side=1, outer=TRUE )
+    }
> mtext( "Fraction of persons (%)", side=2, outer=TRUE, las=0 )
```

We can of course also make the same exercise *conditional* being alive at age 50, 60 etc, but as is seen from figure 11 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 <-
+ PR50 <- PR60 <- PR70 <- PR*0
> dimnames( PR )[[2]][500]
  [1] "49.95"

> dimnames( PR )[[1]]

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

> for( sc in dimnames(TR)[["scene"]] )
+ for( sx in dimnames(TR)[["sex"]] )
+    {
+    # Initialize to all being well at age 50, 60, 70
+    PR50[,500,sc,sx] <-
+    PR60[,600,sc,sx] <-
+    PR70[,700,sc,sx] <- c(1,rep(0,9))
+    # Initialize to all being DM at age 50, 60, 70
+    DM50[,500,sc,sx] <-
+    DM60[,600,sc,sx] <-
+    DM70[,700,sc,sx] <- c(0,1,rep(0,8))
+    for( ag in 501:1020 )
+       {
+                 PR50[,ag,sc,sx] <- PR50[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>600 ) PR60[,ag,sc,sx] <- PR60[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>700 ) PR70[,ag,sc,sx] <- PR70[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+                 DM50[,ag,sc,sx] <- DM50[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>600 ) DM60[,ag,sc,sx] <- DM60[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+    if( ag>700 ) DM70[,ag,sc,sx] <- DM70[,ag-1,sc,sx] %*% TRp[,,ag,sc,sx]
+       }
+    }
```

For further comparisons we print the distribution on states at age 102 years:

```
> round( ww <- cbind( PR[,1020,"Cross",],
+                     PR50[,1020,"Cross",],
+                     PR60[,1020,"Cross",],
+                     PR70[,1020,"Cross",] )*100, 1 )
```

```
          M    F    M    F    M    F    M    F
  Well   0.0  0.2  0.0  0.2  0.1  0.3  0.1  0.4
  DM     0.0  0.0  0.0  0.1  0.0  0.1  0.0  0.1
  DM-Ca  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
  Ca     0.0  0.1  0.0  0.1  0.1  0.1  0.1  0.1
  Ca-DM  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
  D-W   35.7 36.5 36.7 40.1 40.1 45.2 48.9 55.6
  D-DM  18.8 18.7 17.4 17.4 15.3 16.5 12.2 14.0
  D-Ca  29.2 29.7 30.1 29.0 30.5 26.9 28.5 22.6
  D-DC  10.9  9.1  9.8  7.4  7.8  6.0  4.7  3.7
  D-CD   5.4  5.6  5.9  5.6  6.1  4.9  5.5  3.5
```

We con compute the fraction of those without disease at different age and who eventually gets a DM diagnosis, who also have a cancer diagnosis:

```
> round( ww[c(7,9,10),], 1 )
          M    F    M    F    M    F    M    F
  D-DM 18.8 18.7 17.4 17.4 15.3 16.5 12.2 14.0
  D-DC 10.9  9.1  9.8  7.4  7.8  6.0  4.7  3.7
  D-CD  5.4  5.6  5.9  5.6  6.1  4.9  5.5  3.5
> round( apply(ww[    9:10 ,],2,sum)/
+        apply(ww[c(7,9:10),],2,sum)*100, 1 )
     M    F    M    F    M    F    M    F
  46.5 44.1 47.4 42.7 47.7 39.9 45.6 34.1
```

We can also compute the fractio that gets a cancer diagnosis, regradless of diabetes status

```
> round( addmargins(ww[c(8:10),],1), 1 )
          M    F    M    F    M    F    M    F
  D-Ca 29.2 29.7 30.1 29.0 30.5 26.9 28.5 22.6
  D-DC 10.9  9.1  9.8  7.4  7.8  6.0  4.7  3.7
  D-CD  5.4  5.6  5.9  5.6  6.1  4.9  5.5  3.5
  Sum  45.5 44.4 45.8 42.0 44.5 37.8 38.8 29.8
```

and we can see how that compares to the fraction among those with diabetes at a given age that contracts diabetes:

```
> round( cbind( PR[,1020,"Cross",],
+             DM50[,1020,"Cross",],
+             DM60[,1020,"Cross",],
+             DM70[,1020,"Cross",] )*100, 1 )
          M    F    M    F    M    F    M    F
  Well   0.0  0.2  0.0  0.0  0.0  0.0  0.0  0.0
  DM     0.0  0.0  0.0  0.1  0.0  0.1  0.0  0.1
  DM-Ca  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
  Ca     0.0  0.1  0.0  0.0  0.0  0.0  0.0  0.0
  Ca-DM  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
  D-W   35.7 36.5  0.0  0.0  0.0  0.0  0.0  0.0
  D-DM  18.8 18.7 58.8 57.8 58.7 61.3 63.4 69.3
  D-Ca  29.2 29.7  0.0  0.0  0.0  0.0  0.0  0.0
  D-DC  10.9  9.1 41.2 42.1 41.3 38.7 36.6 30.6
  D-CD   5.4  5.6  0.0  0.0  0.0  0.0  0.0  0.0
```

We can now plot the comparison between the life-long outloook of a person with and without diabetes:

```
> CRpl <-
+ function( PR, sc, sx, rm, sepcol="white" )
+   {
+ CR <- apply( PR[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[3,-rm,sc,sx],
+                                      rev(nul[-rm]))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[1,-rm,sc,sx],
+                                  rev(CR[4,-rm,sc,sx]))*100,
+          col = tred, border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[6,-rm,sc,sx],
+                                  rev(CR[5,-rm,sc,sx]))*100,
+          col = "gray", border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[7,-rm,sc,sx],
+                                  rev(CR[10,-rm,sc,sx]))*100,
+          col = tblue, border="transparent")
+ polygon( c(aa[-rm],rev(aa[-rm])), c(CR[6,-rm,sc,sx],
+                                  rev(CR[9,-rm,sc,sx]))*100,
+          col = tred, border="transparent")
+ matlines( aa[-rm], 100*t(CR[c(2,5,8),-rm,sc,sx]),
+           lty=1, col=c(sepcol,"black")[c(1,2,1)], lwd=c(1,3,1), type="l" )
+    }
```

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

```
> par( mfcol=c(3,4), mar=c(2,2,1,3), oma=c(2,2,2,0), mgp=c(3,1,0)/1.6, las=1 )
> for( sc in dimnames(CR)[[3]][1] )
+ for( sx in dimnames(CR)[[4]] )
+    {
+ CRpl( PR50, sc, sx, 1:500 )
+ CRpl( PR60, sc, sx, 1:600 )
+ CRpl( PR70, 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 )
```

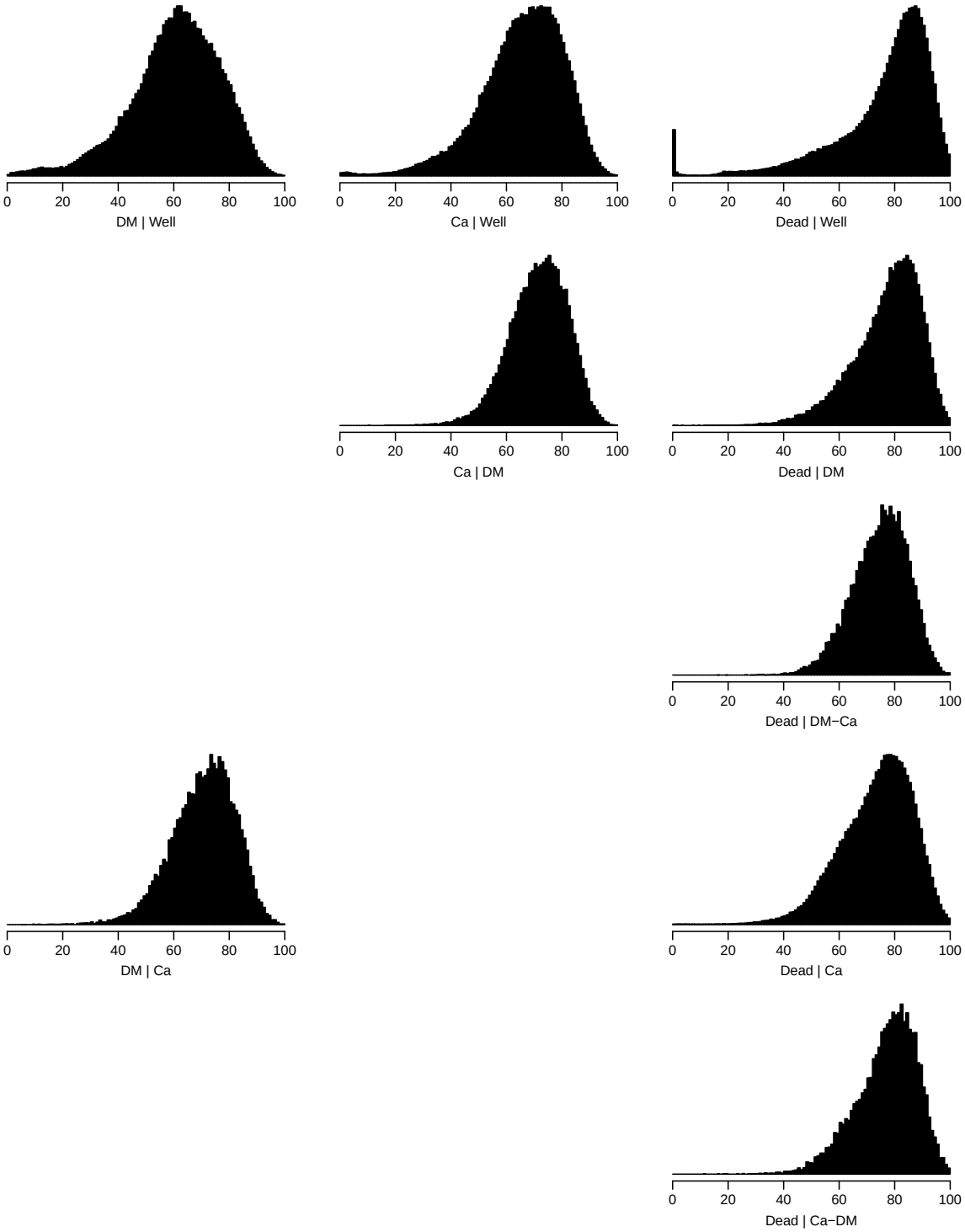

Figure 3: *Histograms of the age at event for the 9 possible transitions. Clearly, noting much is happening in the younger ages, so we shall have age-knots a little close in the older ages.*

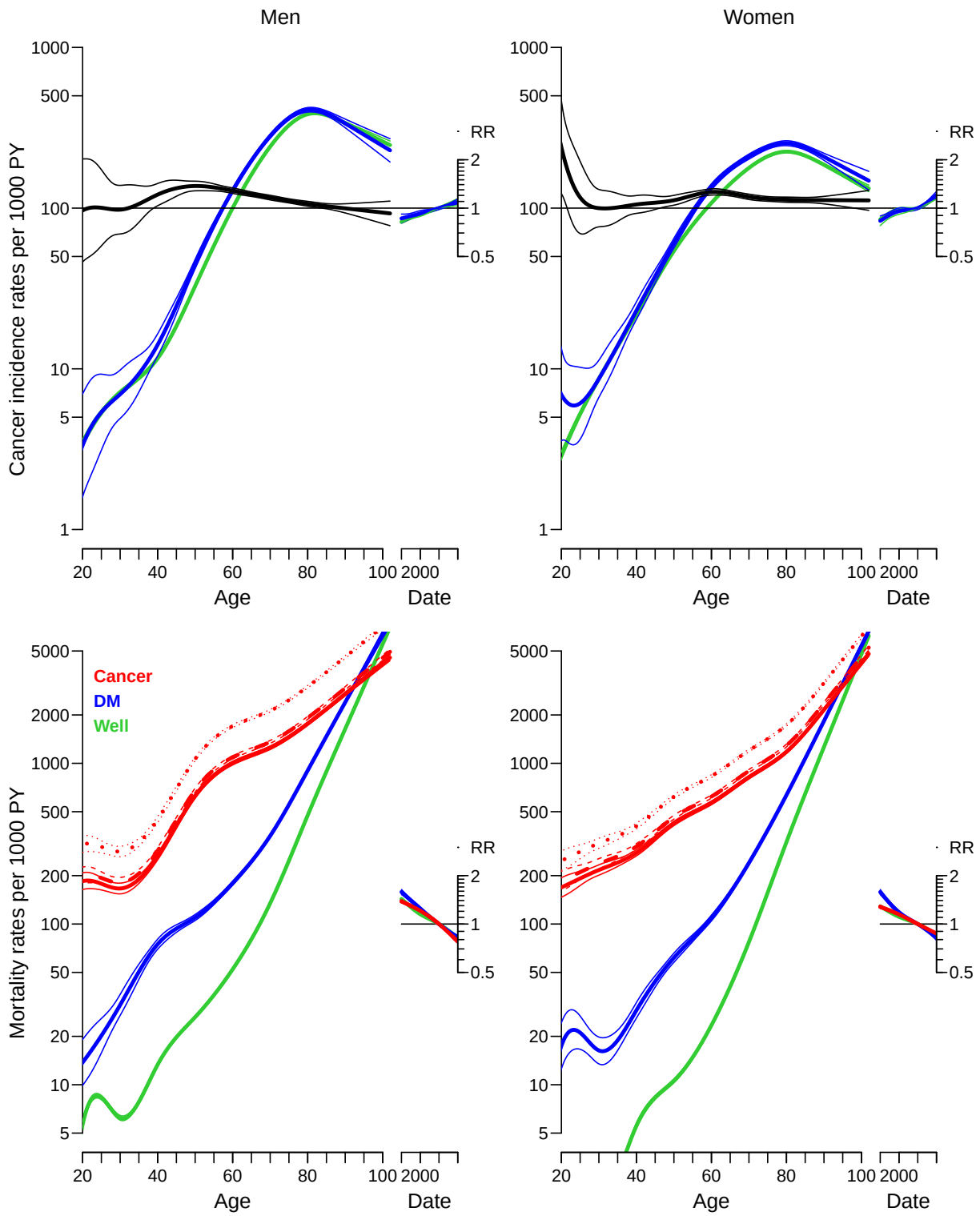

Figure 4: *Cancer incidence rates (top panels) and mortality rates (lower panels), among men (left panels) and women (right panels).*

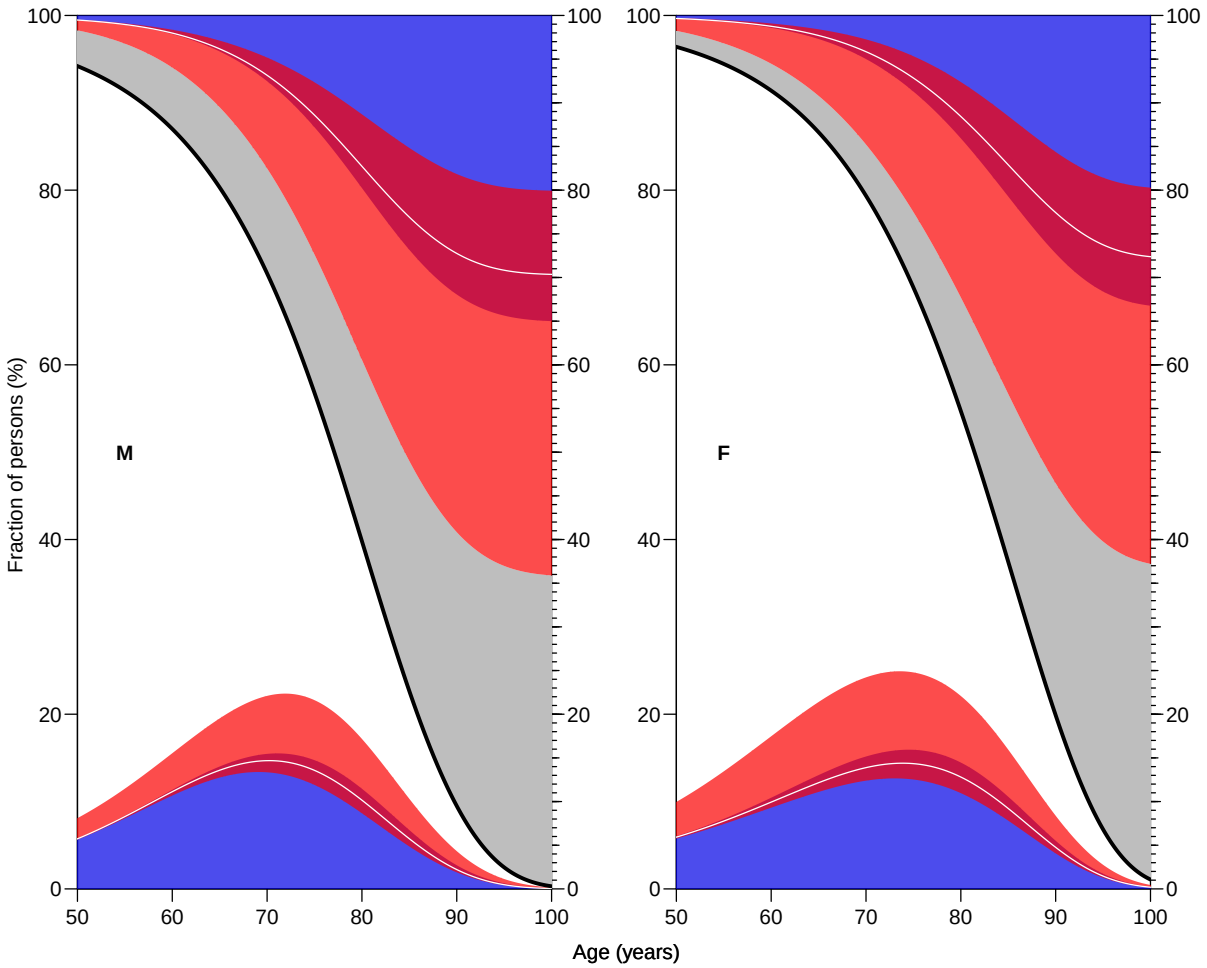

Figure 5: *Occupation probabilities for the various states in figure 1 at various ages, assuming all start out "well" at age 0.*
The thick black line is the overall survival curve, with "Dead" states above and "Alive below". The blue states are persons with a diagnosis of diabetes, the red states are persons with a cancer diagnosis. The maroon areas are persons with both diagnoses and the white line a separates those that have a DM diagnosis first (adjacent to the DM area) from thos with a cancer diagnosis first (adjacent to the cancer area). The white and gray areas are those who do not have any of two diseases.

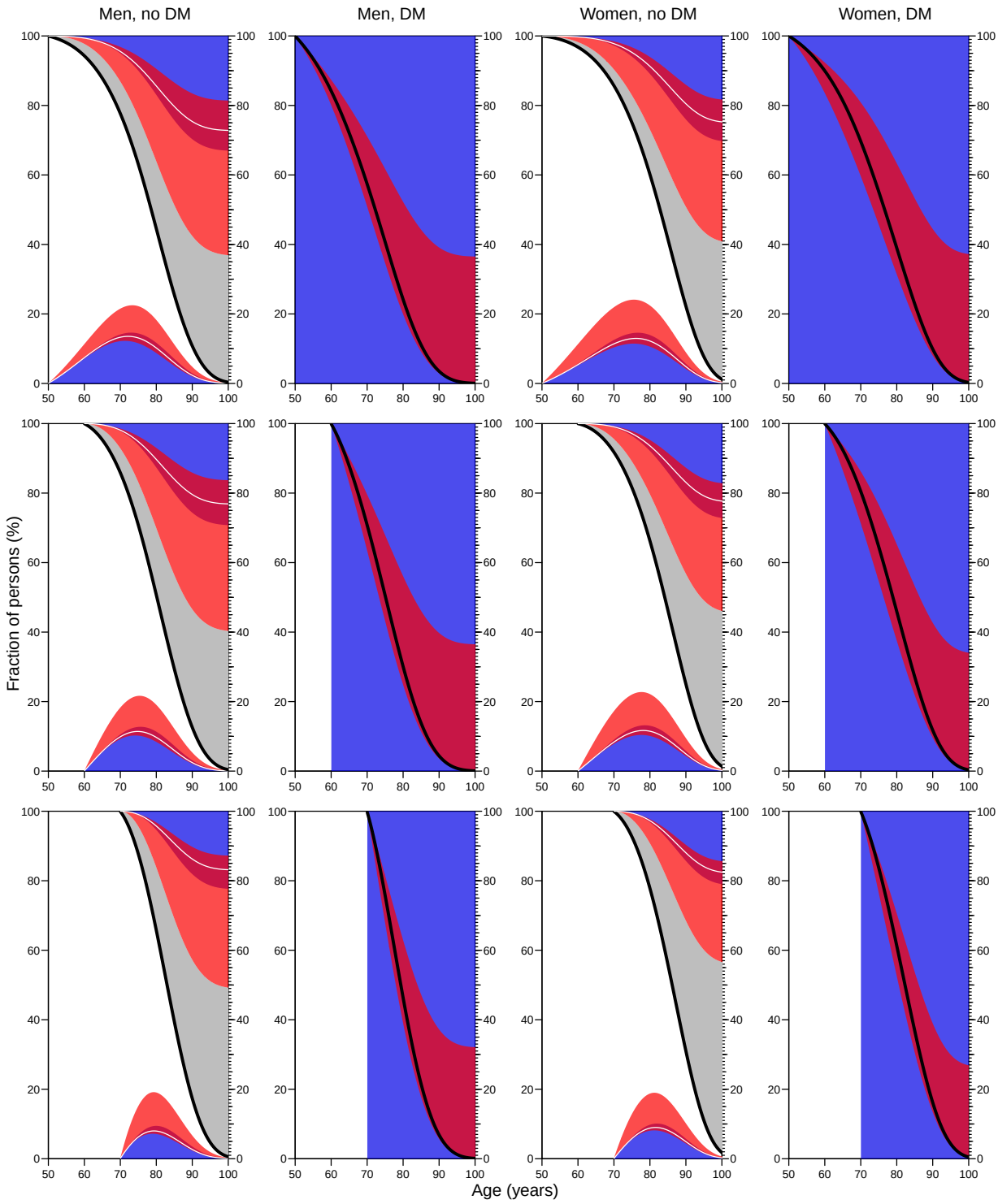

Figure 6: *Occupation probabilities for the various states*

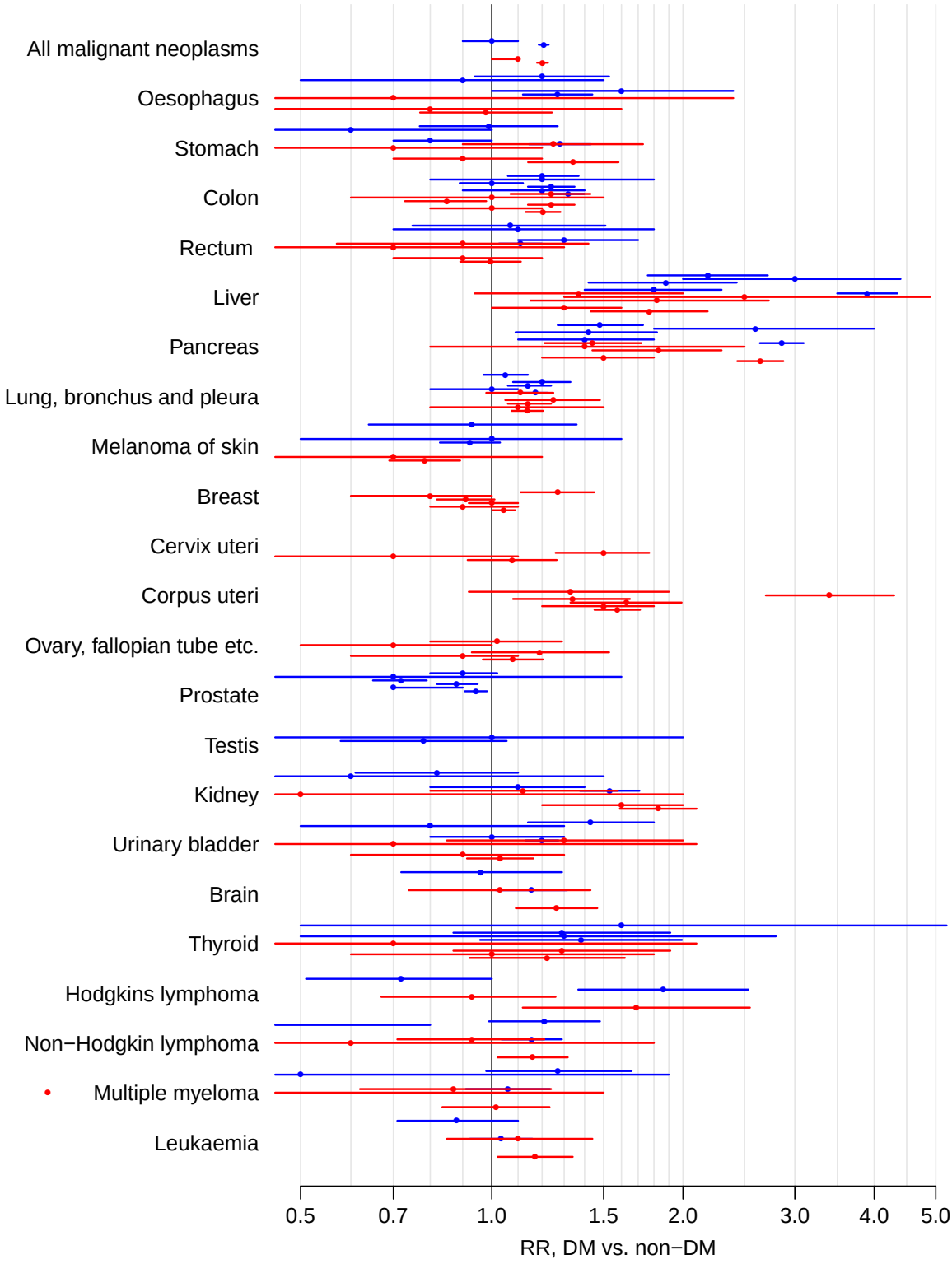

Figure 7: *Estimated RRs from different studies. Blue lines for men, red lines for women. Within each group of estimates they are from the studies (in order): []*

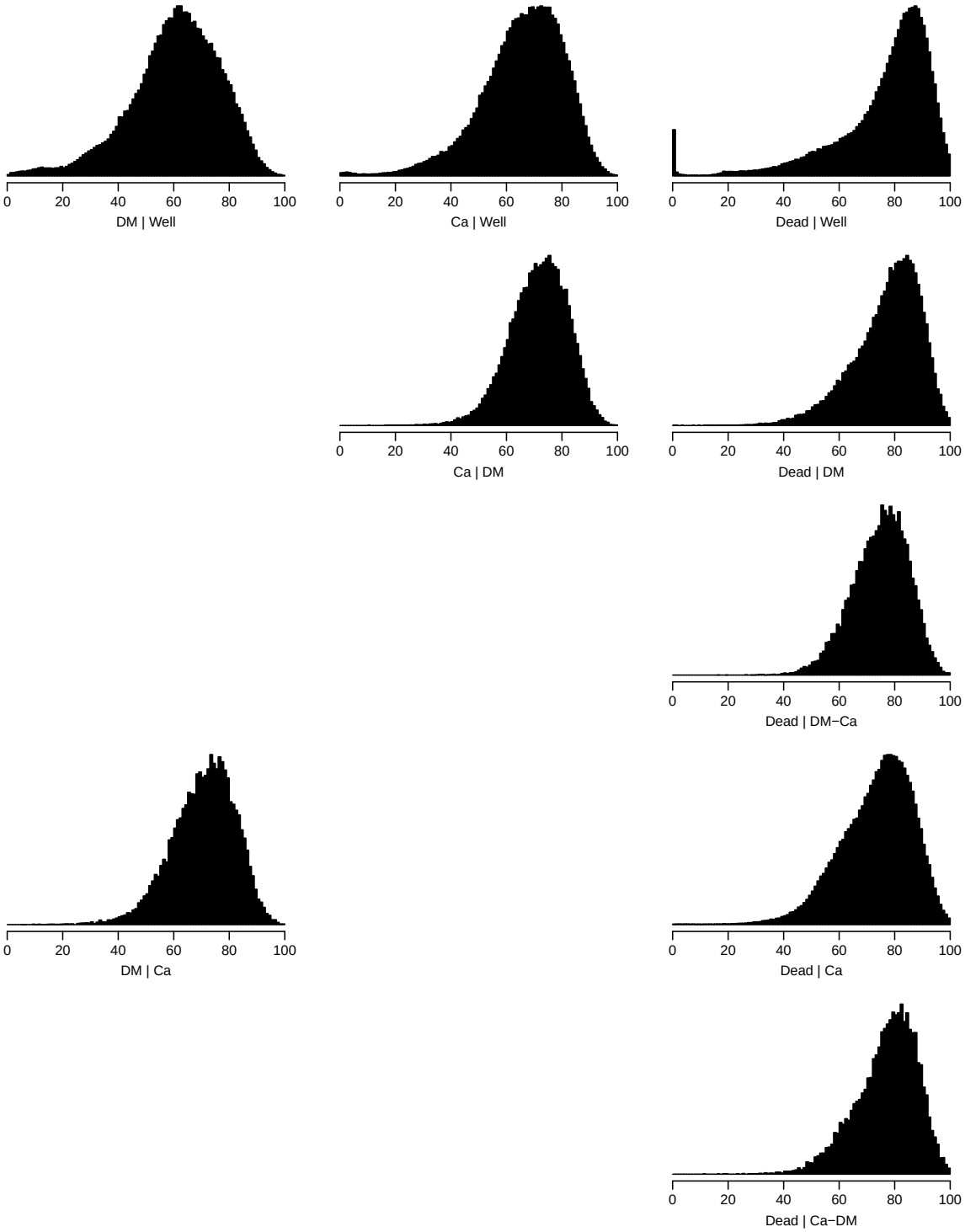

Figure 8: *Histograms of the age at event for the 9 possible transitions. Clearly, noting much is happening in the younger ages, so we shall have age-knots a little close in the older ages.*

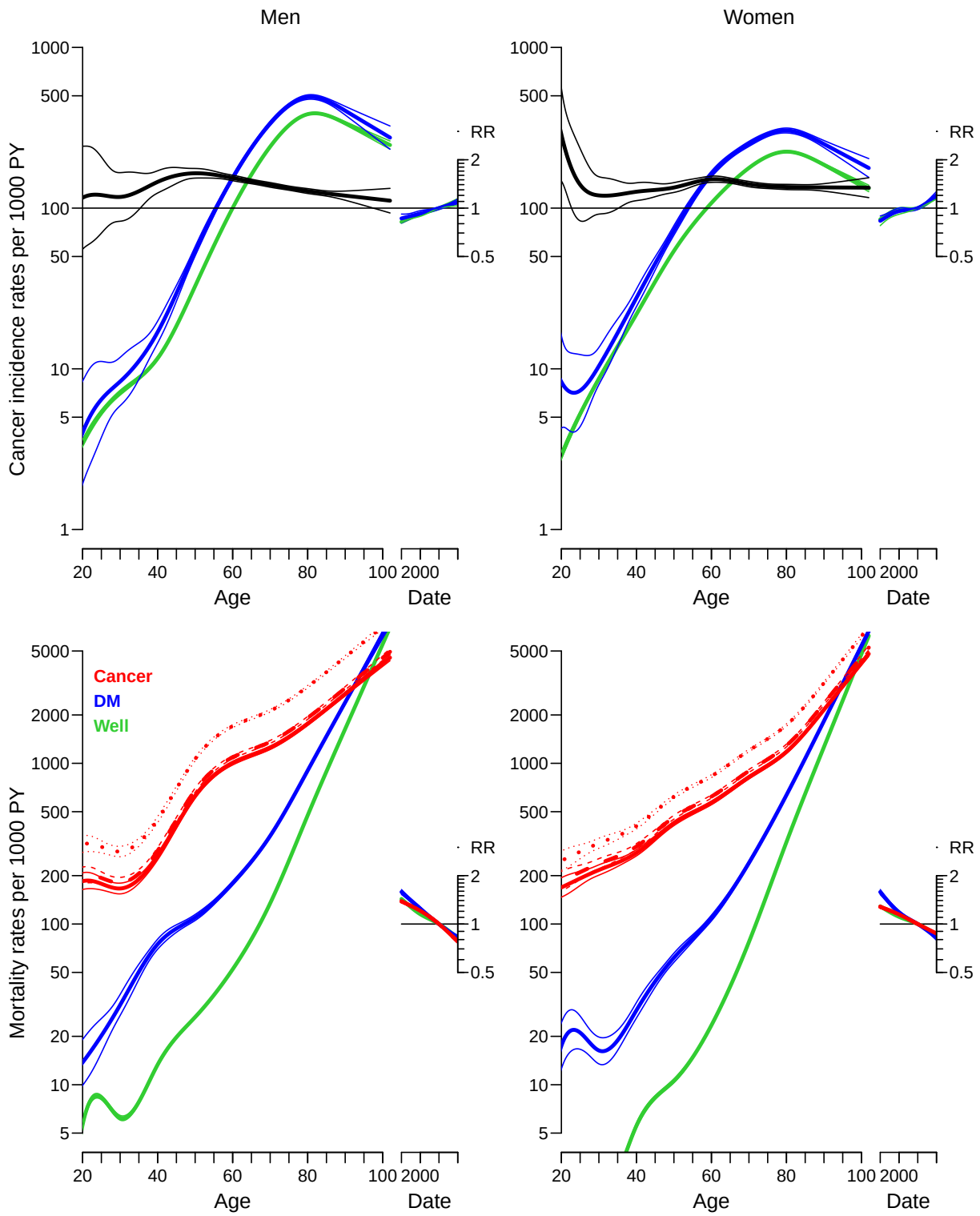

Figure 9: *Cancer incidence rates (top panels) and mortality rates (lower panels), among men (left panels) and women (right panels).*

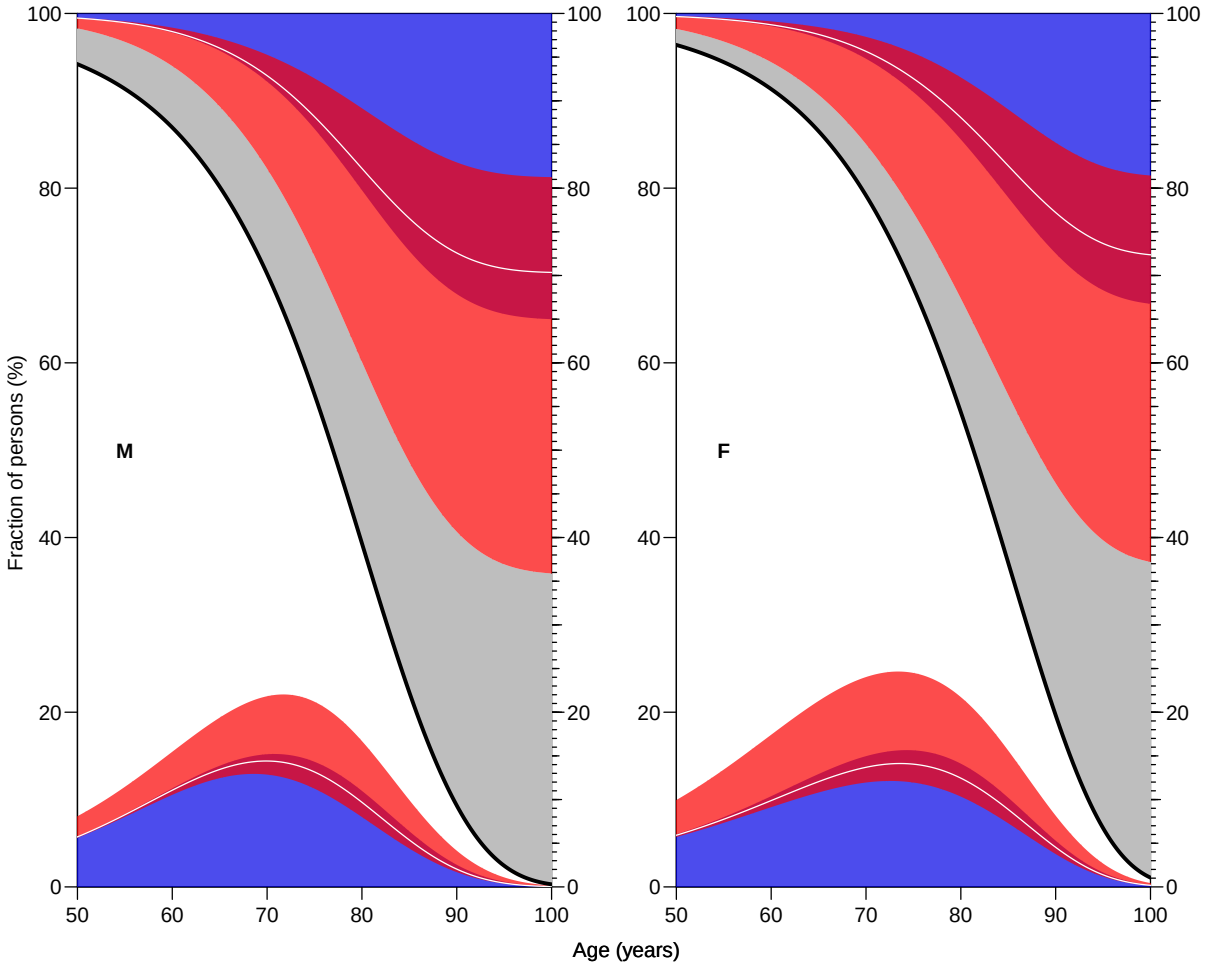

Figure 10: *Occupation probabilities for the various states in figure 1 at various ages, assuming all start out "well" at age 0.*
The thick black line is the overall survival curve, with "Dead" states above and "Alive below". The blue states are persons with a diagnosis of diabetes, the red states are persons with a cancer diagnosis. The maroon areas are persons with both diagnoses and the white line a separates those that have a DM diagnosis first (adjacent to the DM area) from thos with a cancer diagnosis first (adjacent to the cancer area). The white and gray areas are those who do not have any of two diseases.

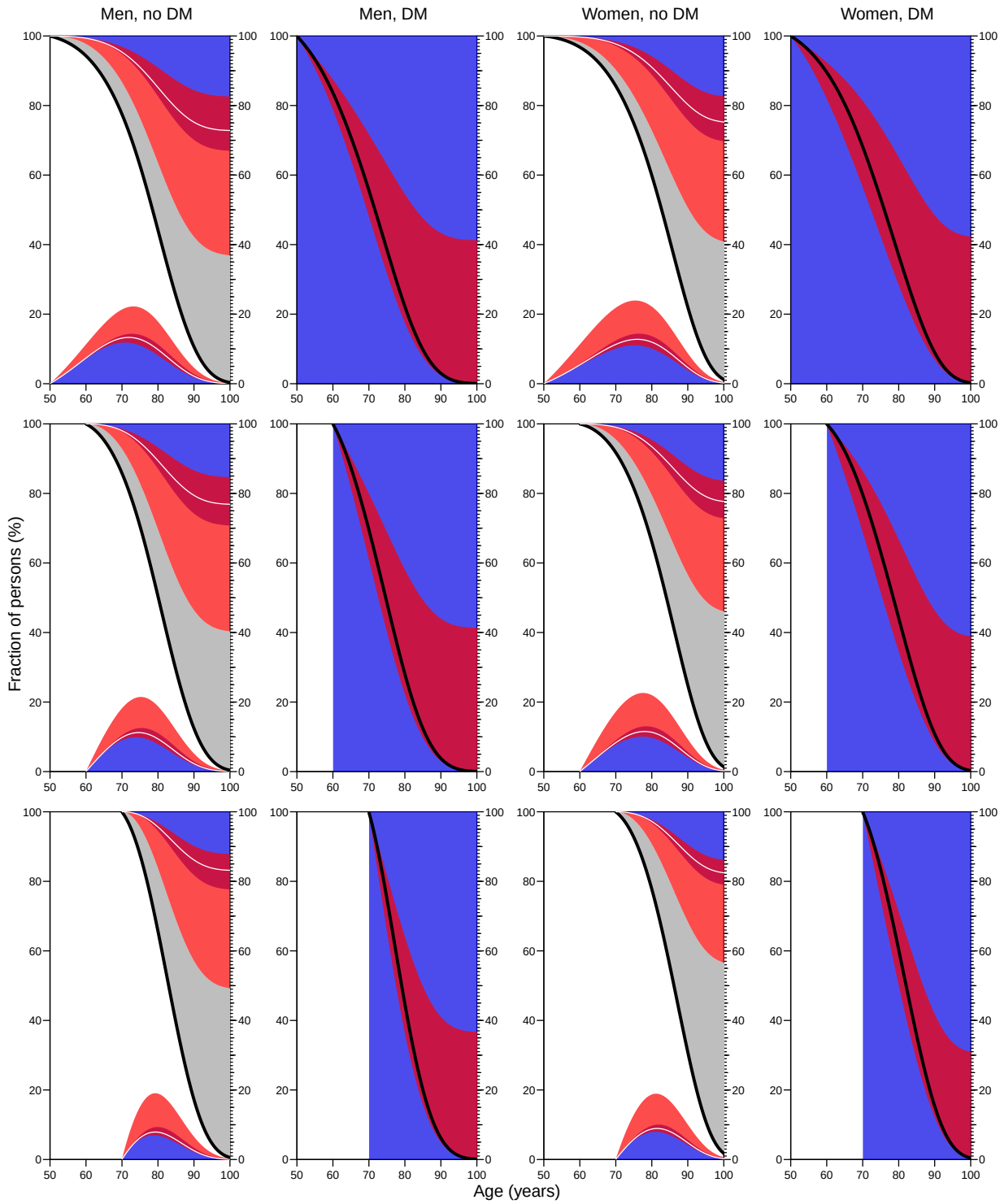

Figure 11: *Occupation probabilities for the various states*