The Danish National Diabetes Register until 1.1.2012

SDC
December 2014
http://bendixcarstensen.com/
Version 1.3 (still preliminary!)

Compiled Wednesday 10th December, 2014, 10:07
from: C:/Bendix/Steno/DM-register/NDR/2012/DMdemo2012.tex

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

Contents

1 The Danish National Diabetes Register **1**
1.1 The Danish National Diabetes Register . . . 1
1.2 Overview of included cases . . . 2
1.2.1 Change in inclusion criteria by time . . . 7
1.2.2 Effect of blood glucose criterion by sex and age . . . 11

2 Analysis based on original DM definition **15**
2.1 Register data — follow-up and deaths . . . 15
2.1.1 Population time . . . 17
2.1.2 Merging time . . . 18
2.1.3 Population deaths . . . 18
2.1.4 DM cases . . . 19
2.1.5 Saving it all for later analysis . . . 20
2.2 DM incidence . . . 22
2.2.1 No. of cases . . . 22
2.2.2 Age-Period-Cohort modelling . . . 23
2.2.3 Time-trends in rates . . . 25
2.2.4 Summary of the APC modelling . . . 26
2.2.5 Saving the fitted models . . . 28
2.3 Mortality . . . 31
2.3.1 Mortality in non-diabetics . . . 31
2.3.2 Mortality among DM patients . . . 33
2.3.3 Summary of the APC models for mortality . . . 35
2.3.4 Time-trends in mortality rates . . . 36
2.3.5 SMR . . . 39
2.3.6 Saving mortality rates . . . 40
2.4 Prevalence of diabetes . . . 42
2.4.1 Trends in prevalence . . . 44
2.4.2 Prevalence age-period interaction . . . 45
2.5 Components of prevalence . . . 48
2.5.1 Formalization . . . 49
2.5.2 Statistical framework . . . 49
2.5.3 Data for the calculations . . . 52
2.5.4 Prevalences . . . 52
2.5.5 Implementation of prevalence calculations . . . 55
2.5.6 Prevalences under different scenarios . . . 62
2.5.7 The actual numbers of diabetes patients in Denmark . . . 69

3 Analysis based on modified DM definition **75**
3.1 Register data — follow-up and deaths . . . 75
3.1.1 Population time . . . 77
3.1.2 Merging time . . . 78
3.1.3 Population deaths . . . 78
3.1.4 DM cases . . . 79
3.1.5 Saving it all for later analysis . . . 80
3.2 DM incidence . . . 82
3.2.1 No. of cases . . . 82
3.2.2 Age-Period-Cohort modelling . . . 83
3.2.3 Time-trends in rates . . . 85
3.2.4 Summary of the APC modelling . . . 86
3.2.5 Saving the fitted models . . . 88
3.3 Mortality . . . 91
3.3.1 Mortality in non-diabetics . . . 91
3.3.2 Mortality among DM patients . . . 93
3.3.3 Summary of the APC models for mortality . . . 95
3.3.4 Time-trends in mortality rates . . . 96
3.3.5 SMR . . . 99
3.3.6 Saving mortality rates . . . 100
3.4 Prevalence of diabetes . . . 102
3.4.1 Trends in prevalence . . . 104
3.4.2 Prevalence age-period interaction . . . 105
3.5 Components of prevalence . . . 108
3.5.1 Formalization . . . 109
3.5.2 Statistical framework . . . 109
3.5.3 Data for the calculations . . . 112
3.5.4 Prevalences . . . 112
3.5.5 Implementation of prevalence calculations . . . 115
3.5.6 Prevalences under different scenarios . . . 122
3.5.7 The actual numbers of diabetes patients in Denmark . . . 129

Chapter 1

The Danish National Diabetes Register

1.1 The Danish National Diabetes Register

First load the relevant package:

```
> options( width=95 )
> library( Epi )
> library( splines )
> memory.size( 3500 )
  [1] 3500

> print( sessionInfo(), l=F )

  R version 3.1.2 (2014-10-31)
  Platform: i386-w64-mingw32/i386 (32-bit)

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

  other attached packages:
  [1] Epi_1.1.68     foreign_0.8-61
```

Then we read the official version of the data (devoid of CPR-ids):

```
> load( file="./data/NDR2011.Rda" )
> str( dr )
  'data.frame':        497232 obs. of  11 variables:
   $ D_FODDTO    : Factor w/ 35518 levels "1889-04-06","1889-07-20",..: 646 34646 34646 893 34788 34
   $ C_SEX       : Factor w/ 2 levels "K","M": 1 1 1 1 2 2 2 2 2 1 ...
   $ D_INKLDTO   : Factor w/ 8938 levels "1941-09-15","1968-12-09",..: 931 6647 7703 2177 2542 6660
   $ C_INKLAARSAG: Factor w/ 6 levels "blod2i5","blod5i1",..: 3 5 5 3 6 5 5 5 5 5 ...
   $ D_DODSDTO   : Factor w/ 8057 levels "1971-05-01","1971-11-14",..: 564 NA NA 1525 NA NA 1643 421
   $ D_LPR       : Factor w/ 8942 levels "1941-09-15","1968-12-09",..: 1386 6650 7707 NA NA 6663 188
   $ D_FODT      : Factor w/ 397 levels "1990-01-10","1990-01-24",..: 2 NA NA 76 NA NA NA NA NA NA
   $ D_BLOD2I5   : Factor w/ 932 levels "1993-01-27","1993-03-03",..: NA NA NA NA NA NA NA NA NA NA
   $ D_BLOD5I1   : Factor w/ 1149 levels "1989-12-27","1990-01-03",..: NA NA NA NA NA NA NA 192 NA N
   $ D_INS       : Factor w/ 5796 levels "1994-01-01","1994-01-02",..: NA 3768 4721 NA NA 3804 NA NA
   $ D_OAD       : Factor w/ 5815 levels "1994-01-03","1994-01-04",..: NA NA NA NA 162 NA NA 1954 NA
```

We then groom it to a more readable format; first we transform the date variables to `cal.yr` format, and shorten and lower-case the variable names:

```
> dvar <- grep( "D_", names(dr) )
> for( i in dvar ) dr[,i] <- as.Date( dr[,i] )
> dr <- cal.yr( dr )
> names( dr ) <- tolower( substr( names(dr), 3, 10 ) )
> levels( dr$sex ) <- c("F","M")
> dr$sex <- relevel( dr$sex, 2 )
```

We also include a modified date of entry, namely that which emerges from excluding the blood glucose criteria as inclusion criteria, and save the groomed version of the NDR

```
> dr$doin   <- with( dr, pmin( lpr, fodt, ins, oad, na.rm=TRUE ) )
> dr$critin <- with( dr, ifelse(!is.na( ins) &  ins==doin, "ins",
+                        ifelse(!is.na( oad) &  oad==doin, "oad",
+                        ifelse(!is.na(fodt) & fodt==doin, "fodt",
+                        ifelse(!is.na( lpr) &  lpr==doin, "lpr", NA) ) ) ) )
> head( dr )
     foddto sex  inkldto inklaars  dodsdto      lpr     fodt blod2i5 blod5i1      ins      oad
1 1900.001   F 1990.063     fodt 1991.481 1991.309 1990.063      NA      NA       NA       NA
2 1999.999   F 2005.721      lpr       NA 2005.721       NA      NA      NA 2005.773       NA
3 1999.999   F 2008.615      lpr       NA 2008.615       NA      NA      NA 2008.689       NA
4 1901.001   F 1993.474     fodt 1994.112       NA 1993.474      NA      NA       NA       NA
5 2001.001   M 1994.474      oad       NA       NA       NA      NA      NA       NA 1994.474
6 2002.000   M 2005.756      lpr       NA 2005.756       NA      NA      NA 2005.880       NA
       doin critin
1 1990.063   fodt
2 2005.721    lpr
3 2008.615    lpr
4 1993.474   fodt
5 1994.474    oad
6 2005.756    lpr

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

1.2 Overview of included cases

We see that the new definition of inclusion removes some cases, and delays the entry for others:

```
> head( subset(dr,inklaars %in% levels(inklaars)[1:2] ) )
      foddto sex  inkldto inklaars  dodsdto      lpr    fodt  blod2i5  blod5i1      ins
13 1902.999   M 1990.715  blod5i1 1991.207       NA      NA       NA 1990.715       NA
33 1906.999   M 1990.810  blod5i1 1995.692 1990.901      NA 1994.126 1990.810 1994.238
34 1906.999   M 1991.155  blod5i1 2002.975 2002.906      NA 1995.755 1991.155       NA
35 1906.999   F 1990.427  blod5i1 1995.662       NA 1991.06       NA 1990.427 1994.452
40 1907.999   M 1991.884  blod5i1 1992.338       NA      NA       NA 1991.884       NA
45 1907.999   F 1990.408  blod5i1 1997.337 1991.654 1996.56 1994.471 1990.408       NA
         oad     doin critin
13        NA       NA   <NA>
33 1995.662 1990.901    lpr
34        NA 2002.906    lpr
35        NA 1991.060   fodt
40        NA       NA   <NA>
45 1994.334 1991.654    lpr
```

We now make a histogram of the included cases according to the variable `inkldto`, showing also how many would be missing if the glucose-criteria were not applied.

```
> par( mar=c(3,4,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> with( dr, hist( pmax(inkldto,1989.01),
+                 breaks=seq(1989,2012,1/12), main="",
+                 ylim=c(0,4000), col="black", xaxt="n",
+                 ylab="", xlab="Date of inclusion") )
> mtext( "No. of new cases per month", side=2, line=3, las=0 )
> text( 1989, -100, "<1989", cex=0.8 )
> text( seq(1990,2010,5)+0.5, rep(-100,5), paste(seq(1990,2010,5)), cex=0.8 )
> par( new=TRUE )
> with( subset(dr,is.na(doin)),
+           hist( pmax(inkldto,1989),
+                 breaks=seq(1989,2012,1/12), main="",
+                 ylim=c(0,4000), col="gray", border="gray", xaxt="n",
+                 ylab="", xlab="" ) )
> abline( v=1990:2012, col="red" )
```

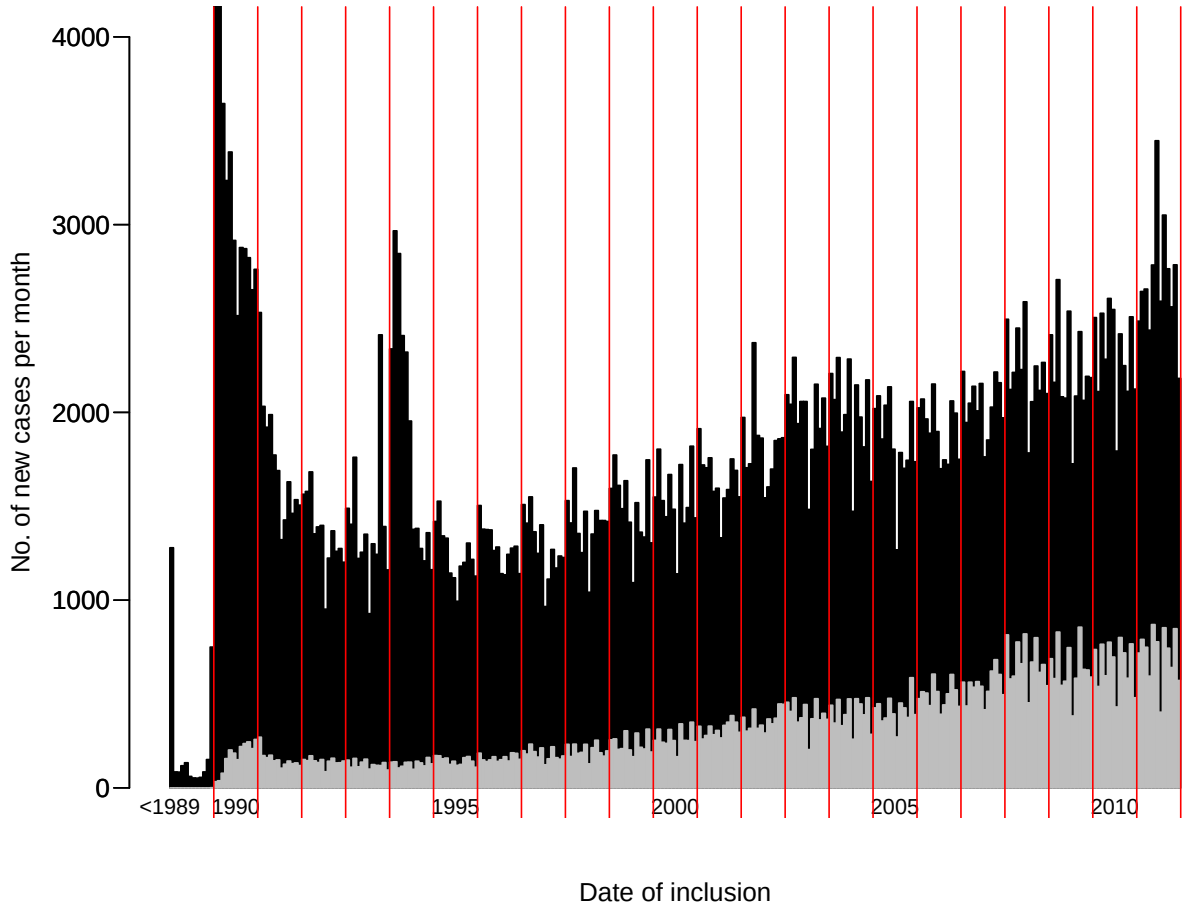

Figure 1.1: *Distribution of inclusions per month, using the original variable* `inkldto`*. The gray part of the bars are the persons that would not be in the register at all if the blood glucose-criteria were dropped.*

From figure 1.1 it seems that a lot of extra persons are diagnosed during 2011, but also that this increase is not reflected in the number of persons included on the blood glucose criteria. Also note that it would not be meaningful to take the black bars as the result of abandoning the glucose criteria; what would happen would be that some of the inclusion dates of the remaining persons would be later. We shall return to this subsequently.

We could also show the monthly distribution of inclusions if we had not used the glucose criterion; note that this is not just the black part of the bars in figure 1.1, since the dates of diagnosis for those included originally by the blood-glucose criteria, but included on another criterion, will have a later date of diagnosis in figure 1.2.

```
> str( dr )
  'data.frame':        497232 obs. of  13 variables:
   $ foddto  :Classes 'cal.yr', 'numeric'  num [1:497232] 1900 2000 2000 1901 2001 ...
   $ sex     : Factor w/ 2 levels "M","F": 2 2 2 2 1 1 1 1 1 2 ...
   $ inkldto :Classes 'cal.yr', 'numeric'  num [1:497232] 1990 2006 2009 1993 1994 ...
   $ inklaars: Factor w/ 6 levels "blod2i5","blod5i1",..: 3 5 5 3 6 5 5 5 5 5 ...
   $ dodsdto :Classes 'cal.yr', 'numeric'  num [1:497232] 1991 NA NA 1994 NA ...
   $ lpr     :Classes 'cal.yr', 'numeric'  num [1:497232] 1991 2006 2009 NA NA ...
   $ fodt    :Classes 'cal.yr', 'numeric'  num [1:497232] 1990 NA NA 1993 NA ...
   $ blod2i5 :Classes 'cal.yr', 'numeric'  num [1:497232] NA NA NA NA NA NA NA NA NA NA ...
   $ blod5i1 :Classes 'cal.yr', 'numeric'  num [1:497232] NA NA NA NA NA ...
   $ ins     :Classes 'cal.yr', 'numeric'  num [1:497232] NA 2006 2009 NA NA ...
   $ oad     :Classes 'cal.yr', 'numeric'  num [1:497232] NA NA NA NA 1994 ...
   $ doin    :Classes 'cal.yr', 'numeric'  num [1:497232] 1990 2006 2009 1993 1994 ...
   $ critin  : chr  "fodt" "lpr" "lpr" "fodt" ...

> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> with( dr, hist( pmax(doin,1989),
+                 breaks=seq(1989,2012,1/12), main="",
+                 ylim=c(0,4000), col="black", xaxt="n",
+                 ylab="No. of new cases per month",
+                 xlab="New date of inclusion") )
> text( 1989, -100, "<1989", cex=0.8 )
> text( seq(1990,2010,5)+0.5, rep(-100,5), paste(seq(1990,2010,5)), cex=0.8 )
> abline( v=1990:2012, col="red" )
> abline( v=2011.5, col="limegreen" )
```

If we tabulate the entire NDR by the relationship between the old (`inkldto`) and the new (`doin`) we get:

```
> with( dr, addmargins( table( earlier=doin<inkldto, later=doin>inkldto, exclude=NULL ) ) )
         later
  earlier  FALSE   TRUE   <NA>    Sum
    FALSE 309779  97520      0 407299
    <NA>       0      0  89933  89933
    Sum   309779  97520  89933 497232

> tt <- with( dr, addmargins( table( sex, later=doin>inkldto, useNA="ifany" ) ) )
> round( cbind( tt, tt[,1:3]/tt[,4]*100 ), 1 )

      FALSE  TRUE  <NA>    Sum FALSE TRUE <NA>
  M   168155 51041 38644 257840  65.2 19.8 15.0
  F   141624 46479 51289 239392  59.2 19.4 21.4
  Sum 309779 97520 89933 497232  62.3 19.6 18.1
```

The entries in this table shows that with the new definition, 62% keep the same inclusion date (`(FALSE)`), 20% has a later date of inclusion assigned (`(TRUE)`), and finally 18% of cases are excluded from the register entirely (`(<NA>)`).

It is noticeable that the fraction of persons not included by the modified criteria is larger among women (21%) than among men (15%), and the ditribution by age is also interesting:

```
> tt <- with( dr, addmargins( table( sex,
+                                    age=floor((inkldto-foddto)/10)*10,
+                                  later=doin>inkldto, useNA="ifany" ) ) )
> ftable( tt )
```

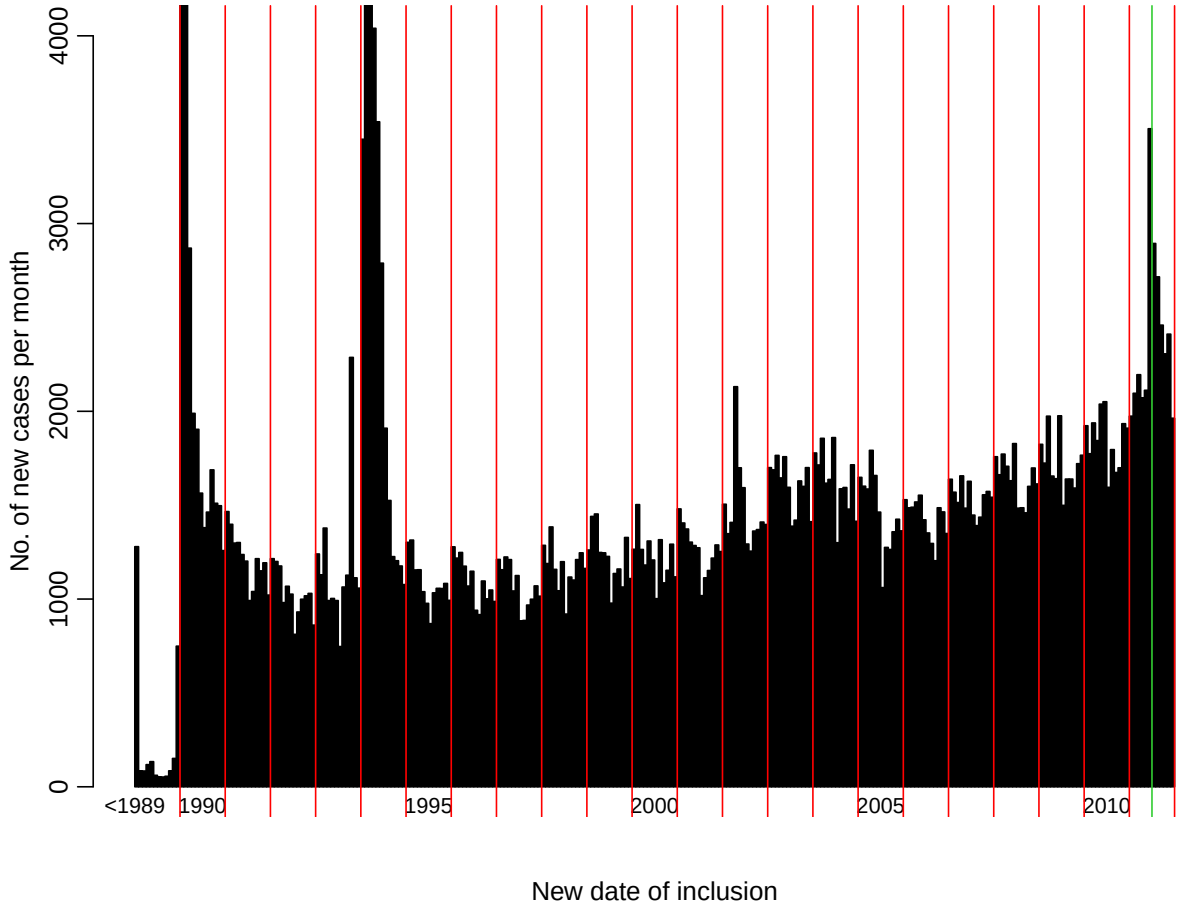

Figure 1.2: *Distribution of inclusions per month, using the new variable* `doin`.

	later	FALSE	TRUE	NA	Sum
sex	age				
M	-10	1	0	0	1
	0	1476	1	49	1526
	10	3177	33	168	3378
	20	4811	317	261	5389
	30	10480	1694	937	13111
	40	23084	6454	3204	32742
	50	38148	13088	7241	58477
	60	42653	15749	11656	70058
	70	30200	10793	10326	51319
	80	12820	2834	4445	20099
	90	1298	78	357	1733
	100	7	0	0	7
	Sum	168155	51041	38644	257840
F	-10	0	0	0	0
	0	1482	1	37	1520
	10	3041	49	319	3409
	20	4333	521	3026	7880
	30	7987	1907	3833	13727
	40	15025	5004	3898	23927
	50	23331	9510	7332	40173
	60	31415	13220	12047	56682
	70	31543	11964	12570	56077
	80	20262	4120	7447	31829
	90	3153	183	775	4111
	100	52	0	5	57

```
      Sum       141624  46479  51289 239392
  Sum -10            1      0      0      1
      0           2958      2     86   3046
      10          6218     82    487   6787
      20          9144    838   3287  13269
      30         18467   3601   4770  26838
      40         38109  11458   7102  56669
      50         61479  22598  14573  98650
      60         74068  28969  23703 126740
      70         61743  22757  22896 107396
      80         33082   6954  11892  51928
      90          4451    261   1132   5844
      100           59      0      5     64
      Sum       309779  97520  89933 497232
> ftable( round( tt[,,1:3]/tt[,,rep(4,3)]*100, 1 ) )
          later FALSE  TRUE    NA
  sex age
  M   -10       100.0   0.0   0.0
      0          96.7   0.1   3.2
      10         94.0   1.0   5.0
      20         89.3   5.9   4.8
      30         79.9  12.9   7.1
      40         70.5  19.7   9.8
      50         65.2  22.4  12.4
      60         60.9  22.5  16.6
      70         58.8  21.0  20.1
      80         63.8  14.1  22.1
      90         74.9   4.5  20.6
      100       100.0   0.0   0.0
      Sum        65.2  19.8  15.0
  F   -10         NaN   NaN   NaN
      0          97.5   0.1   2.4
      10         89.2   1.4   9.4
      20         55.0   6.6  38.4
      30         58.2  13.9  27.9
      40         62.8  20.9  16.3
      50         58.1  23.7  18.3
      60         55.4  23.3  21.3
      70         56.2  21.3  22.4
      80         63.7  12.9  23.4
      90         76.7   4.5  18.9
      100        91.2   0.0   8.8
      Sum        59.2  19.4  21.4
  Sum -10       100.0   0.0   0.0
      0          97.1   0.1   2.8
      10         91.6   1.2   7.2
      20         68.9   6.3  24.8
      30         68.8  13.4  17.8
      40         67.2  20.2  12.5
      50         62.3  22.9  14.8
      60         58.4  22.9  18.7
      70         57.5  21.2  21.3
      80         63.7  13.4  22.9
      90         76.2   4.5  19.4
      100        92.2   0.0   7.8
      Sum        62.3  19.6  18.1
```

We can plot the changes in dates of inclusion for a 10% random sample of the persons in the register:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> with( zz <- subset( dr, runif(nrow(dr))<0.1 ),
+       plot( pmin(doin,runif(nrow(zz),2013-0.7,2013+0.7),na.rm=TRUE), inkldto,
+             pch=16, cex=0.1, col=c("blue","red")[sex],
+             xlab="Modified date of inclusion", xlim=c(1989,2013),
```

```
+             ylab="Original date of inclusion", ylim=c(1989,2013), bty="n" ) )
> abline( v=1990:2012, h=1990:2012, col=gray(0.8) )
> axis( side=1, at=2013, "Not\nincl." )
```

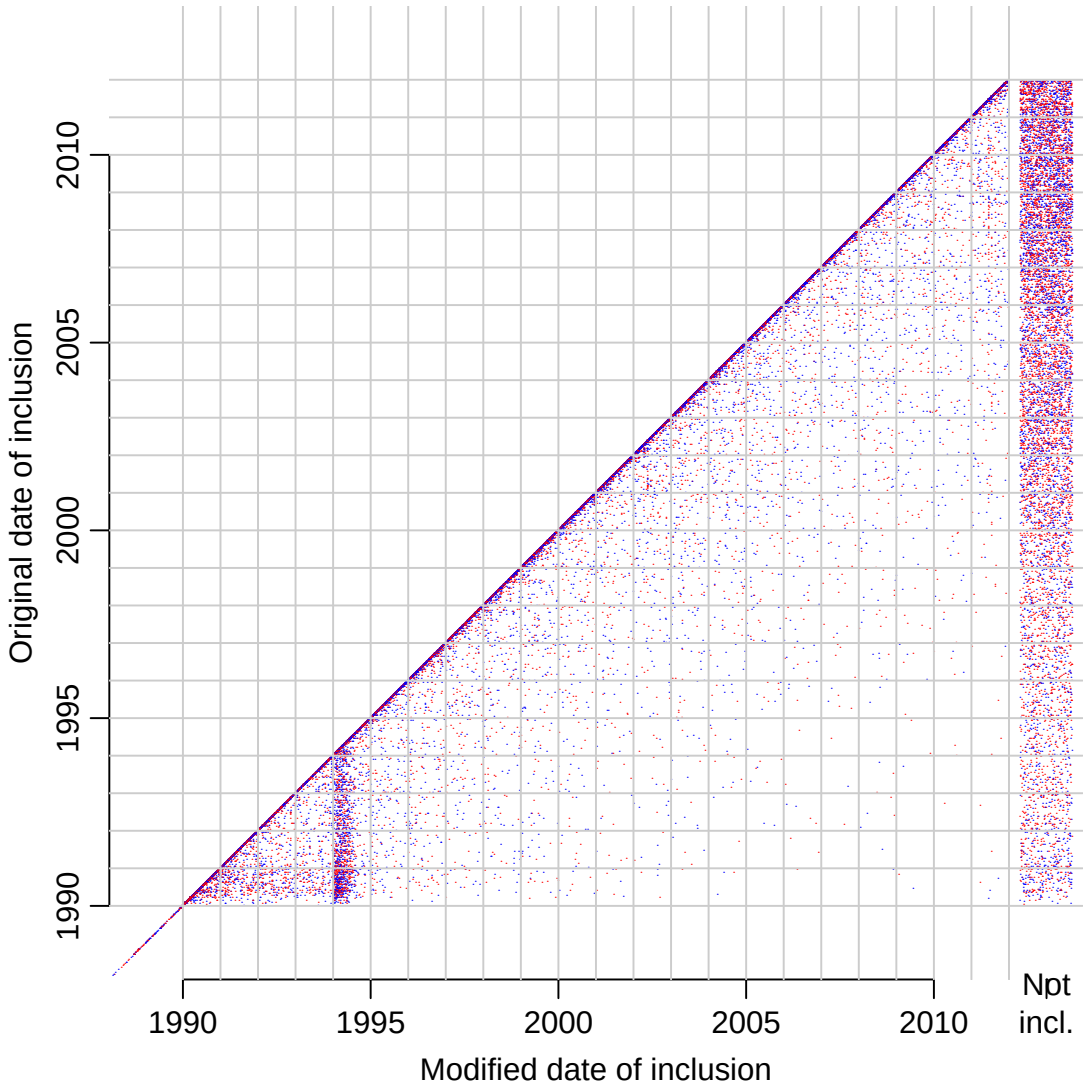

Figure 1.3: *Change in dates of inclusion from dropping the glucose criterion. Based on a 7% random sample of the NDR.*

From figure 1.3 we can see that the change in date of inclusion is quite substantial for many persons, but the general pattern is that the change in date of inclusion is generally less than 3 years. It is also apparent that those not included by the modified definition are more likely to be diagnosed later. Hardly surprising since the opportunity to fulfill any of the other inclusion criteria are shorter the later you are included (by the orginal definintion).

1.2.1 Change in inclusion criteria by time

First we will explore if any one of the inclusion criteria are behind the sudden increase in mid-2011, by colouring the cases by inclusion criterion, and we also make a more-long term

trend in inclusion criteria by year:

```
> mtab <- with( dr, table( inklaars, pmax(floor(inkldto*12)/12,1989) ) )
> mtab <- mtab[c(1,2,5,3,4,6),]
> t( mtab[,1:15+100] )
                   inklaars
                    blod2i5 blod5i1 lpr fodt ins oad
   1997.33333333333       0     475 404  112  14 244
   1997.41666666667       0     597 442  104  12 246
   1997.5                 0     336 343   86   9 192
   1997.58333333333       2     450 373   96   9 182
   1997.66666666667       0     548 386  122  19 195
   1997.75                1     429 386  109  15 231
   1997.83333333333       1     465 429  125  11 204
   1997.91666666667       1     487 355   95  12 274
   1998                   0     577 530  152   9 265
   1998.08333333333       0     525 472  130  11 273
   1998.16666666667       0     687 546  137  16 318
   1998.25                0     532 413  103  14 293
   1998.33333333333       0     521 383  113   9 227
   1998.41666666667       0     619 416  137   8 293
   1998.5                 0     360 353   97   9 223

> nl <- nrow( mtab )
> ytab <- with( dr, table( pmax(floor(inkldto),1989), sex, inklaars ) )
> ytab <- addmargins( ytab[,,c(1,2,5,3,4,6)], margin=2 )
> dimnames( ytab )[[1]] <- rep("",dim(ytab)[1])
> dimnames( ytab )[[1]][2+0:4*5] <- paste(1990+0:4*5)
> ftable( ytab, row.vars=1 )
       sex            M                                        F
       inklaars blod2i5 blod5i1   lpr  fodt   ins   oad blod2i5 blod5i1   lpr  fodt   ins   oad bloc

                      0       1  1513     0     0     0       0       2  1328     0     0     0
  1990                0    7299  7918  6217     0     0       0    7913  7151  9711     0     0
                      0    4900  4810  1053     0     0       0    4915  3830  1310     0     0
                      0    3807  3756   900     0     0       0    3780  2930  1076     0     0
                      0    3472  4940   784     0     0       0    3417  3213  1012     0     0
                      6    2706  3118   441  1948  3774      13    2781  2415   694  1727  3058
  1995                3    2746  2782   511   190  1524       4    2794  2106   775   302  1169
                      1    2898  2911   510   103  1593       2    3029  2153   717   148  1339
                      3    2875  2789   540    84  1637       2    3257  2208   748    66  1252
                      3    3143  3123   599    81  1870       2    3504  2241   857    62  1382
                      2    3439  3171   658    85  1959       3    3710  2370   960    75  1447
  2000                3    3610  3402   579    65  1961       4    3936  2525   858    55  1505
                      4    3924  3570   397    60  2260       2    4505  2678   578    54  1664
                      4    4098  3649  1036    52  2339       3    4641  2687  1699    73  1687
                      9    4355  3844  1368    71  2723      11    4579  2987  1850    48  1886
                      5    4282  3770  1371    65  2979       7    4561  2946  1858    50  2055
  2005                5    4091  3699   655    68  3101       6    4496  2853   953    72  2202
                      8    4476  3794   120    81  3615      15    5087  2960   199    68  2591
                     23    4552  3755   165   115  4109      15    5354  3053   270    96  2995
                     23    5098  3721   161   143  4859      19    5862  2707   265    95  3715
                     26    4755  3657   135   148  5577      28    5330  2785   211    82  3924
  2010               29    4597  3639   161   106  6442      34    5176  2745   228    94  4541
                     20    4611  3361  1599   126  7363      33    5278  2623  1969    97  5314

> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> clr <- c("black","limegreen","orange","red","lightblue","blue")
> barplot( mtab, beside=FALSE, ylim=c(0,4000),
+          col=clr, border="transparent", space=0, xaxt="n" )
> axis( side=1, at = (seq(1990,2010,5)-1989)*12,
+           labels =  seq(1990,2010,5) )
> axis( side=1, at = (1990:2013-1989)*12, rep("",24) )
> text( rep((1997-1989)*12,nl), 4000-(nl:1)*150, rownames(mtab),
+       col=clr, font=2, adj=0 )
> mosaicplot( ytab[-(1:5),3,6:1], off=0, color=clr[6:1], ylab="", main="" )
```

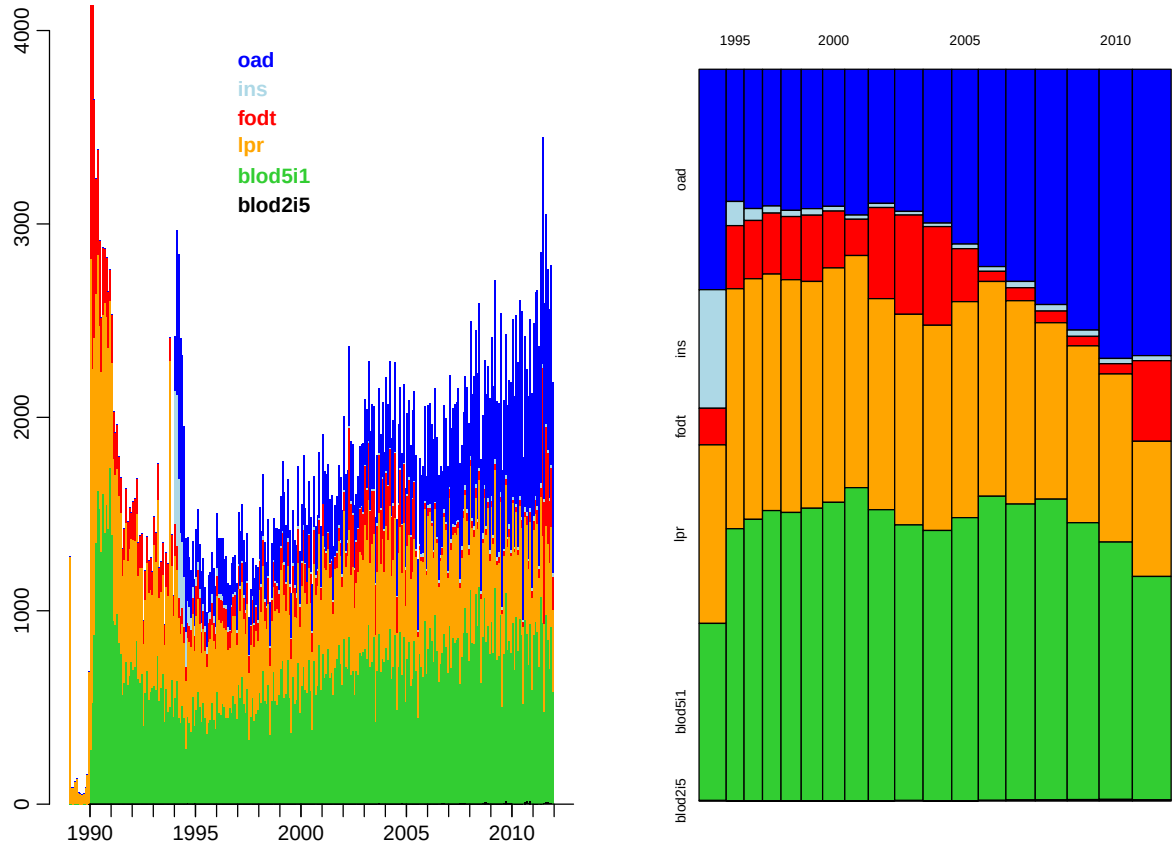

Figure 1.4: *Distribution of inclusions per month, using the original variable* `inkldto`*, colored by the inclusion criterion.*

From the figure 1.4 we can see that the inclusion criterion responsible for the increase in mid-2011 is the criterion `fodt`, chiropody (foot therapy).

It would be illustrative if we made the mosaicplot by sex (that is wuth the `ytab` was classified by sex too):

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1 )
> for( sx in c("M","F") )
+    { mosaicplot( ytab[-(1:5),sx,6:1],
+                  off=0, color=clr[6:1], ylab="", main="" )
+      text(0,1,sx,adj=c(1,0),font=2) }
```

From figure 1.5 we can see that by using the NDR from 1.1.1995 (regardless of inclusion algorithm) as accurate w.r.t. incidence date we will include a few hundred persons on the criterion "`ins`" which are presumably diagnosed earlier. To avoid this we would have to start follow-up only from 1.1.1997, but since it is only a phenomenon for the "`ins`" criterion, we maintain to use the register as complete and accurate for incidence from 1.1.1995.

From figure 1.2 we can see that there is a substantial increase in the middle of 2011. It is of course also because of the increase in the inclusions due to Chiropody — caused by some administrative details in the agreements between the national health authorities and the chiropodists.

Here are the cases distributed by inclusion criterion, both in the new and the old version:

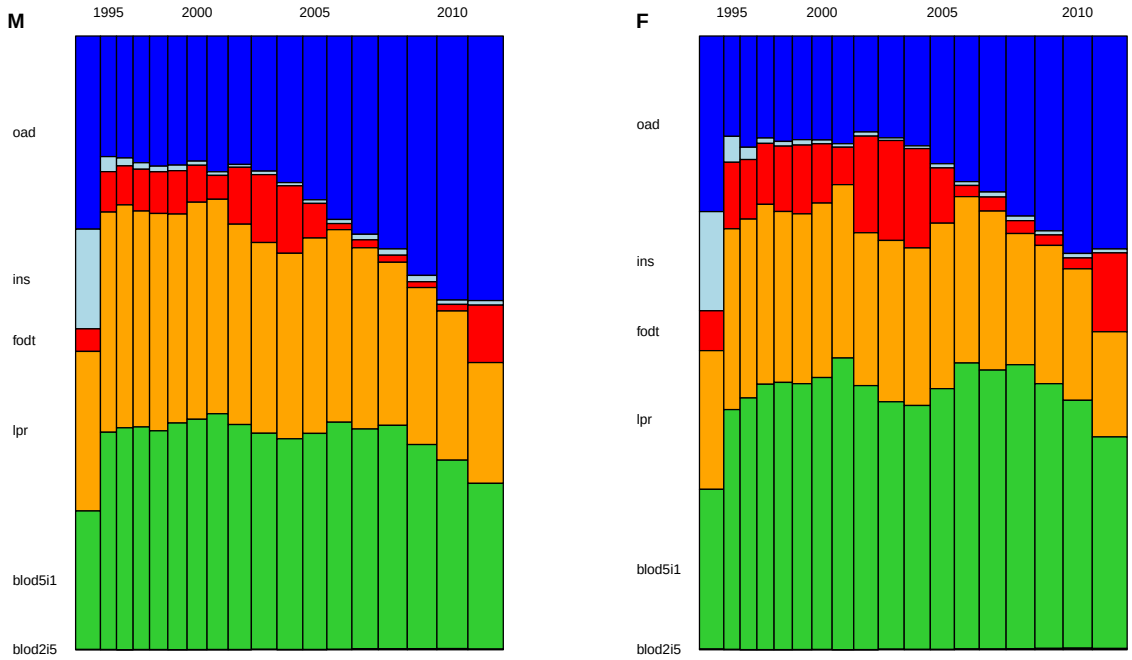

Figure 1.5: *Distribution of inclusion criteria in NDR by year of inclusion and sex. The width of the bars is proportional to the number of included persons each year, and height of the components is the relative contribution from each criterion, hence the* area *of each rectangle is proportional the the* number *of persons included in a given year by a given criterion.*

```
> with( dr, addmargins( table( floor(pmax(inkldto,1989)), inklaars ) ) )
```

	inklaars						
	blod2i5	blod5i1	fodt	ins	lpr	oad	Sum
1989	0	3	0	0	2841	0	2844
1990	0	15212	15928	0	15069	0	46209
1991	0	9815	2363	0	8640	0	20818
1992	0	7587	1976	0	6686	0	16249
1993	0	6889	1796	0	8153	0	16838
1994	19	5487	1135	3675	5533	6832	22681
1995	7	5540	1286	492	4888	2693	14906
1996	3	5927	1227	251	5064	2932	15404
1997	5	6132	1288	150	4997	2889	15461
1998	5	6647	1456	143	5364	3252	16867
1999	5	7149	1618	160	5541	3406	17879
2000	7	7546	1437	120	5927	3466	18503
2001	6	8429	975	114	6248	3924	19696
2002	7	8739	2735	125	6336	4026	21968
2003	20	8934	3218	119	6831	4609	23731
2004	12	8843	3229	115	6716	5034	23949
2005	11	8587	1608	140	6552	5303	22201
2006	23	9563	319	149	6754	6206	23014
2007	38	9906	435	211	6808	7104	24502
2008	42	10960	426	238	6428	8574	26668
2009	54	10085	346	230	6442	9501	26658
2010	63	9773	389	200	6384	10983	27792
2011	53	9889	3568	223	5984	12677	32394
Sum	380	187642	48758	6855	150186	103411	497232

```
> with( dr, addmargins( table( floor(pmax(   doin,1989)), critin   ) ) )
```

```
      critin
         fodt    ins    lpr    oad    Sum
1989        0      0   2843      0   2843
1990    16585      0  16320      0  32905
1991     3419      0  11094      0  14513
1992     3264      0   9052      0  12316
1993     3200      0  10830      0  14030
1994     1879   4811   6776  18831  32297
1995     1837    548   5524   5125  13034
1996     1814    293   5687   5331  13125
1997     1918    177   5592   5100  12787
1998     2148    166   6071   5636  14021
1999     2386    185   6297   5782  14650
2000     2034    140   6782   5741  14697
2001     1407    131   7195   6389  15122
2002     3974    155   7427   6254  17810
2003     4223    139   7836   7102  19300
2004     4245    142   7615   7549  19551
2005     2126    159   7454   7711  17450
2006      423    172   7823   8773  17191
2007      547    244   7753   9888  18432
2008      566    267   7407  11457  19697
2009      449    258   7377  12559  20643
2010      510    238   7245  14189  22182
2011     5395    266   6790  16252  28703
Sum     64349   8491 174790 159669 407299
```

1.2.2 Effect of blood glucose criterion by sex and age

It would however be substantially more interesting to see how the criterion change varies by sex, age and date of inclusion (that is the *original* exclusion date.

This is most easily done by using a purely descriptive continuous model for the probability of being excluded, resp. having a later diagnosis, as a function of age and date, separately for each sex.

```
> library( splines )
> dr <- transform( dr, ldat = pmax( doin-inkldto, 0, na.rm=TRUE )>0,
+                  inklage = inkldto-foddto )
> with( dr, table( ldat, doin-inkldto>0, useNA="ifany" ) )
  ldat     FALSE   TRUE   <NA>
    FALSE 309779      0  89933
    TRUE       0  97520      0
> system.time(
+ xF <- glm( is.na(doin) ~ Ns(inklage,knots=c(10,20,25,30,35,40,50,60,70,80,90))*
+                           Ns(inkldto,knots=seq(1996,2011,,4)),
+            family=binomial,
+            data=subset(dr,sex=="F") ) )
     user  system elapsed
    22.56    1.00   23.64
> lF <- update( xF, ldat ~ . )
> xM <- update( xF, data=subset(dr,sex=="M") )
> lM <- update( lF, data=subset(dr,sex=="M") )
> pxF <- pxM <-
+ plF <- plM <- NULL
> for( id in 1995:2012 )
+    {
+  nd <- data.frame( inklage = seq(10,90,,200), inkldto = id )
+ pxF <- cbind( pxF, predict( xF, newdata=nd, type="response" ) )
+ pxM <- cbind( pxM, predict( xM, newdata=nd, type="response" ) )
+ plF <- cbind( plF, predict( lF, newdata=nd, type="response" ) )
+ plM <- cbind( plM, predict( lM, newdata=nd, type="response" ) )
+    }
```

We now have the predicted probabilities for each year as a function of age in `pxM`, `pxF`, `plM`, and `plF`. We the plot the probabilities of either having a later diagnosis or not being included in the register at all

```
> bsq <- rgb(seq(200,0,,18),seq(200,0,,18),255,maxColorValue=255)
> rsq <- rgb(255,seq(200,0,,18),seq(200,0,,18),maxColorValue=255)
> par( mfcol=c(2,2), mar=c(0,0,0,0), oma=c(3,3,1,1), mgp=c(3,1,0)/1.6,
+     bty="n", las=1 )
> plot( nd$inklage, nd$inklage,
+       ylim=c(-3,72), yaxs="i", xlim=c(7,90), xaxs="i",
+       col="transparent", xaxt="n" )
> abline( v=1:9*10, h=0:7*10, col=gray(0.8) )
> rect( -5,70,100,75, col="white", border=gray(0.8) )
> matlines( nd$inklage, pxM*100,
+           type="l", lty=1, lwd=2, col=bsq )
> mtext("Men",side=3,line=0,col="blue", outer=FALSE, cex=0.83)
> mtext("P(Deletion if BG omitted) (%)",side=2,line=2, outer=FALSE, cex=0.83,las=0)
> plot( nd$inklage, nd$inklage,
+       ylim=c(-3,72), yaxs="i", xlim=c(7,90), xaxs="i",
+       col="transparent" )
> abline( v=1:9*10, h=0:7*10, col=gray(0.8) )
> matlines( nd$inklage, plM*100,
+          type="l", lty=1, lwd=2, col=bsq )
> mtext("P(Later dx if BG omitted) (%)",side=2,line=2, outer=FALSE, cex=0.83,las=0)
> axis(side=1,at=1:9*10,labels=NA)
> plot( nd$inklage, nd$inklage,
+       ylim=c(-3,72), yaxs="i", xlim=c(7,90), xaxs="i",
+       col="transparent", yaxt="n", xaxt="n" )
> abline( v=1:9*10, h=0:7*10, col=gray(0.8) )
> rect( -5,70,100,75, col="white", border=gray(0.8) )
> matlines( nd$inklage, pxF*100,
+           type="l", lty=1, lwd=2, col=rsq )
> mtext("Women",side=3,line=0,col="red", outer=FALSE, cex=0.83)
> plot( nd$inklage, nd$inklage,
+       ylim=c(-3,72), yaxs="i", xlim=c(7,90), xaxs="i",
+       col="transparent", yaxt="n" )
> abline( v=1:9*10, h=0:7*10, col=gray(0.8) )
> matlines( nd$inklage, plF*100,
+           type="l", lty=1, lwd=2, col=rsq )
> axis(side=1,at=1:9*10,labels=NA)
> mtext("Age at inclusion",side=1,line=1.5, outer=TRUE, cex=0.83)
```

From figure 1.6 it is clear that the probability of exclusion increases by time (not surprising since the observation time left to acquire a diagnosis on another criterion decreases), and the probability to acquire a later diagnosis decreases by time (which is not surprising either, for the same reason).

The most surprising feature however is that there is a peak in the fertile age for women in ages 20–40 approximately. In order to explore this a bit more we instead plot the probability of exclusion from the NDR as a function of calendar time for different ages, in women only though:

```
> pxF <- NULL
> for( ia in 20:40 )
+    {
+  nd <- data.frame( inkldto = seq(1995,2012,,200), inklage = ia )
+ pxF <- cbind( pxF, predict( xF, newdata=nd, type="response" ) )
+    }
```

We can now plot probabilities for different ages as a function of date of inclusion:

```
> par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,3,1,1), mgp=c(3,1,0)/1.6,
+     bty="n", las=1 )
```

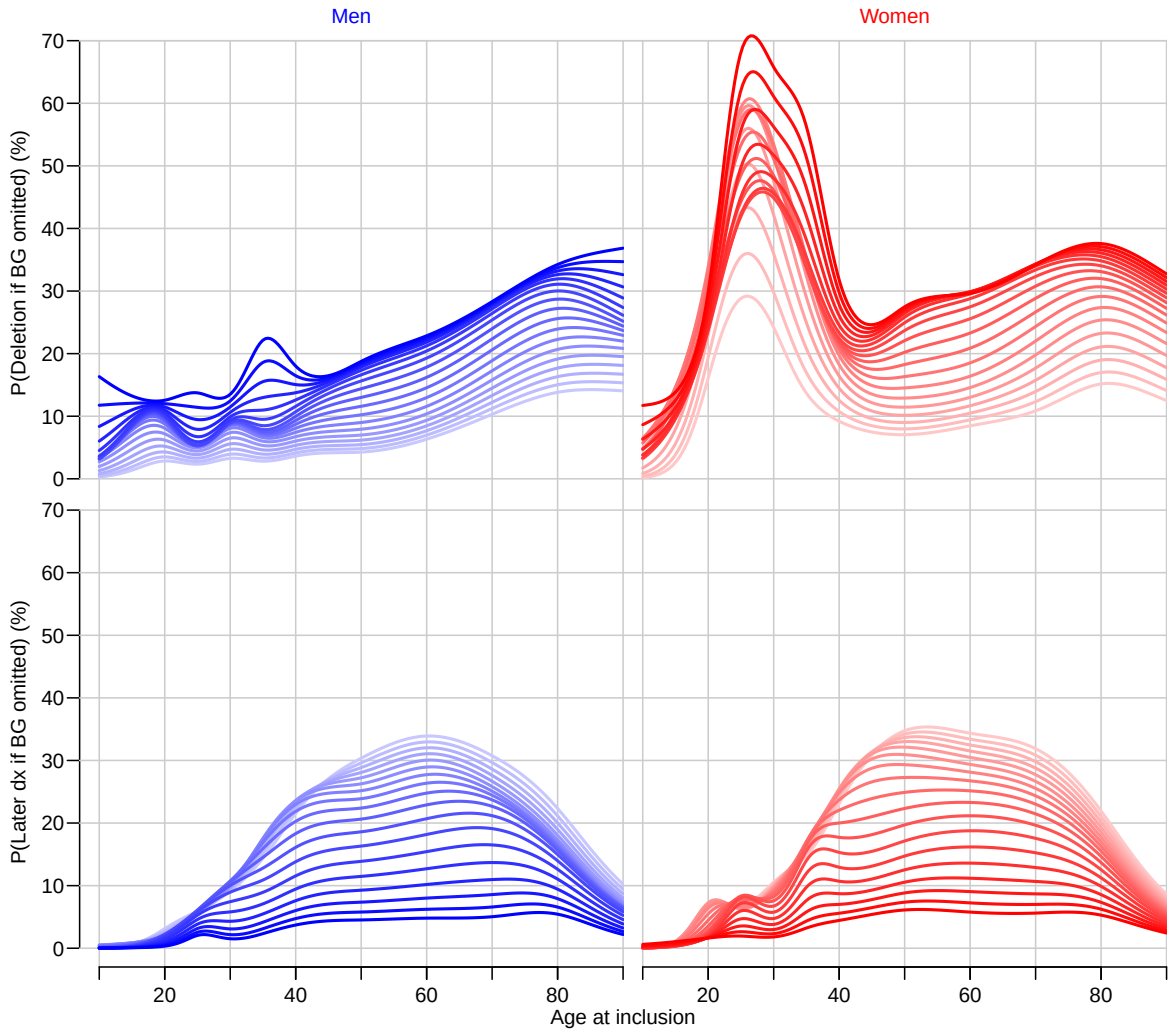

Figure 1.6: *Probabilities of either not having a diagnosis at all in the NDR (top panels) or having a later diagnosis (bottom panels). Colors refer to dates, from pale (1995) to full (2012).*

```
> plot( nd$inkldto, nd$inkldto, ylim=c(0,70), col="transparent" )
> abline( v=1995+0:5*5, h=0:7*10, col=gray(0.8) )
> matlines( nd$inkldto, pxF[,1:11]*100, ylim=c(0,70),
+           type="l", lty=1, lwd=2, col=rsq[1:11] )
> mtext("P(Deletion if BG omitted) (%)",side=2,line=2, outer=FALSE, cex=0.83,las=0)
> plot( nd$inkldto, nd$inkldto, ylim=c(0,70), col="transparent", yaxt="n" )
> abline( v=1995+0:5*5, h=0:7*10, col=gray(0.8) )
> matlines( nd$inkldto, pxF[,11:21]*100,
+          type="l", lty=1, lwd=2, col=rsq[1:11] )
> axis(side=1,at=1995+1:18,labels=NA)
> mtext("Date of inclusion",side=1,line=1.5, outer=TRUE, cex=0.83)
```

The specific shapes of the curves by calendar time (figure **??** shows that the largest fraction affceted by the blood-glucose citeria are women aged 24–34; for these ages the fraction only meeting glucose criteria is above 50% if included after 2010. However, the latest years may not be of greatest interest since persons included after 2010 only have had a limited time to be included by oter criteria. But it is interesting that there seems to be a peak for those diagnosed around 2001, decreasing a bit till 2005. So if we take the increase

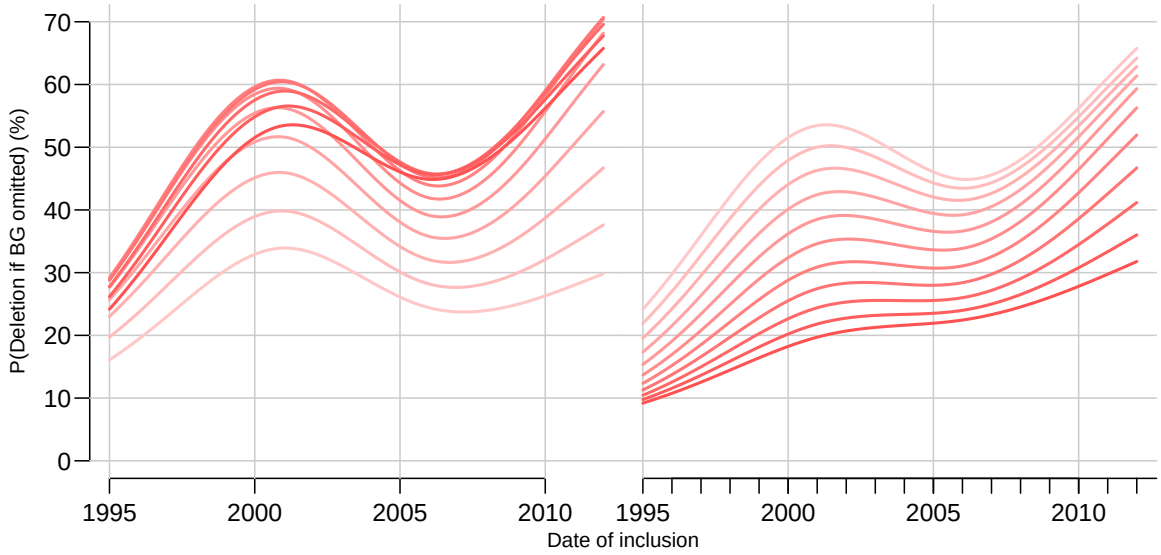

Figure 1.7: *Probabilities of not having a diagnosis at all in the NDR by modified (restrictive) criteria. Left panel: women aged 20–30 (pale to full), right panel: women age 30–40. Thus the darkest curve in let panel and the lightest cuorve in right panel are for 30-year old women.*

after 2005 as largely attributable to the effect of the diminishing observation time, it seems that there has been an increase until 2002 and a subsequent decrease in the fraction of women only identified by the blood glucose criteria.

Chapter 2

Analysis based on original DM definition

2.1 Register data — follow-up and deaths

First we load the register:

```
> load( file="./data/ndr.Rda" )
```

For setting up follow-up data we need convenience functions which maps `NA`s to either `FALSE` or `TRUE`:

```
> na2T <- function( x )     x | is.na(x)
> na2F <- function( x )  !(!x | is.na(x))
```

We now set up data as a Lexis object with three timescales: age, calendar time and diabetes duration. Note that we use the "original" definintion of diabetes, including the blood glucose criteria:

```
> dr$doDM <- dr$inkldto
> dr <- transform( dr, doe = pmax(doDM,1995),
+                      dox = pmin(2012,dodsdto,foddto+99,na.rm=TRUE) )
> Lx <- Lexis( entry = list( A = doe-foddto,
+                            P = doe,
+                          dur = doe-doDM ),
+              exit = list( P = dox ),
+       exit.status = factor( na2F(dodsdto==dox),
+                             labels=c("Alive","Dead") ),
+              data = subset( dr, doe<dox & doDM>foddto ) )
  NOTE: entry.status has been set to "Alive" for all.

> summary( Lx )

  Transitions:
       To
  From     Alive   Dead  Records:  Events: Risk time:  Persons:
    Alive 311404 158389    469793   158389    3219926    469793
```

There are fewer cases in `Lx` the in the entire register, but mostly because of persons that have died before 1995, or were included after age 98:

```
> addmargins( tt <- with( dr, table( dd=dodsdto<1995,
+                                    bb=inkldto>foddto+99,
+                                    exclude=NULL ) ) )
```

```
        bb
dd       FALSE   TRUE   <NA>    Sum
  FALSE 159205     90      0 159295
  TRUE   27049     18      0  27067
  <NA>  310858     12      0 310870
  Sum   497112    120      0 497232
```

```
> sum( c(tt[2,1],tt[,2]) )
```

```
[1] 27169
```

```
> nrow(dr) - nrow(Lx)
```

```
[1] 27439
```

The `Lexis` object `Lx` is now going to be used to construct a table of person-years among DM patients which we will subtract from the population person-years. Note that we also count the number of deaths, in order to construct a dataset also usable for mortality analyses.

So basically, we split the data along the age and period axis, and to avoid problems with memory overflow we do the splitting in smaller chunks.

```
> n.chunks <- 50
> lm <- round( seq(0,nrow(Lx),,n.chunks+1) )
> for( i in 1:n.chunks )
+ {
+ whr <- (lm[i]+1):(lm[i+1])
+ sP  <- splitLexis( Lx[whr,], 1995:2013, time.scale="P" )
+ sPA <- splitLexis( sP    ,      0:100 , time.scale="A" )
+ agg <- with( sPA, aggregate( cbind( y = lex.dur,
+                                      d = lex.Xst=="Dead" ),
+                               list( sex = sex,
+                                     A = floor(A),
+                                     P = floor(P),
+                                     U = floor(P)-floor(A)-floor(foddto) ),
+                               FUN = sum ) )
+ # Just to get the right structure of Agg, variables sx, A, P and U
+ # and UPPER-CASE Y and D to hold the aggregate person-time and events
+ if( i==1 ) Agg <- cbind( agg[1,1:4], Y=NA, D=NA )
+ Agg <- merge( Agg, agg, by=c("sex","A","P","U"), all=TRUE )
+ Agg <- transform( Agg, Y = pmax(Y,0,na.rm=TRUE) + pmax(y,0,na.rm=TRUE),
+                        D = pmax(D,0,na.rm=TRUE) + pmax(d,0,na.rm=TRUE) )[
+                   ,c("sex","A","P","U","Y","D")]
+ cat( "Merged in chunk", i, " at",
+      format(Sys.time(),format="%Y-%m-%d %H:%M:%S"), "\n" )
+ flush.console()
+ }
```

```
Merged in chunk 1  at 2013-08-30 09:10:08
Merged in chunk 2  at 2013-08-30 09:10:21
Merged in chunk 3  at 2013-08-30 09:10:35
Merged in chunk 4  at 2013-08-30 09:10:49
Merged in chunk 5  at 2013-08-30 09:11:02
Merged in chunk 6  at 2013-08-30 09:11:16
Merged in chunk 7  at 2013-08-30 09:11:30
Merged in chunk 8  at 2013-08-30 09:11:44
Merged in chunk 9  at 2013-08-30 09:11:59
Merged in chunk 10  at 2013-08-30 09:12:13
Merged in chunk 11  at 2013-08-30 09:12:27
Merged in chunk 12  at 2013-08-30 09:12:41
Merged in chunk 13  at 2013-08-30 09:12:54
Merged in chunk 14  at 2013-08-30 09:13:09
Merged in chunk 15  at 2013-08-30 09:13:23
Merged in chunk 16  at 2013-08-30 09:13:37
Merged in chunk 17  at 2013-08-30 09:13:51
Merged in chunk 18  at 2013-08-30 09:14:05
```

```
  Merged in chunk 19  at 2013-08-30 09:14:19
  Merged in chunk 20  at 2013-08-30 09:14:33
  Merged in chunk 21  at 2013-08-30 09:14:47
  Merged in chunk 22  at 2013-08-30 09:15:00
  Merged in chunk 23  at 2013-08-30 09:15:14
  Merged in chunk 24  at 2013-08-30 09:15:28
  Merged in chunk 25  at 2013-08-30 09:15:42
  Merged in chunk 26  at 2013-08-30 09:15:57
  Merged in chunk 27  at 2013-08-30 09:16:11
  Merged in chunk 28  at 2013-08-30 09:16:27
  Merged in chunk 29  at 2013-08-30 09:16:42
  Merged in chunk 30  at 2013-08-30 09:16:58
  Merged in chunk 31  at 2013-08-30 09:17:12
  Merged in chunk 32  at 2013-08-30 09:17:27
  Merged in chunk 33  at 2013-08-30 09:17:42
  Merged in chunk 34  at 2013-08-30 09:17:58
  Merged in chunk 35  at 2013-08-30 09:18:14
  Merged in chunk 36  at 2013-08-30 09:18:28
  Merged in chunk 37  at 2013-08-30 09:18:44
  Merged in chunk 38  at 2013-08-30 09:19:00
  Merged in chunk 39  at 2013-08-30 09:19:15
  Merged in chunk 40  at 2013-08-30 09:19:29
  Merged in chunk 41  at 2013-08-30 09:19:44
  Merged in chunk 42  at 2013-08-30 09:19:58
  Merged in chunk 43  at 2013-08-30 09:20:12
  Merged in chunk 44  at 2013-08-30 09:20:25
  Merged in chunk 45  at 2013-08-30 09:20:40
  Merged in chunk 46  at 2013-08-30 09:20:53
  Merged in chunk 47  at 2013-08-30 09:21:08
  Merged in chunk 48  at 2013-08-30 09:21:22
  Merged in chunk 49  at 2013-08-30 09:21:36
  Merged in chunk 50  at 2013-08-30 09:21:49
> summary( Agg )
   sex            A               P               U                Y
   M:3360   Min.   : 0.00   Min.   :1995   Min.   :0.0000   Min.   :   0.0144
   F:3354   1st Qu.:24.00   1st Qu.:1999   1st Qu.:0.0000   1st Qu.:  73.8552
            Median :49.00   Median :2003   Median :1.0000   Median : 309.5702
            Mean   :49.13   Mean   :2003   Mean   :0.5007   Mean   : 479.5838
            3rd Qu.:74.00   3rd Qu.:2007   3rd Qu.:1.0000   3rd Qu.: 743.7469
            Max.   :98.00   Max.   :2011   Max.   :1.0000   Max.   :3004.5955
       D
   Min.   :  0.00
   1st Qu.:  0.00
   Median :  6.00
   Mean   : 23.59
   3rd Qu.: 42.00
   Max.   :137.00
> head( Agg )
    sex A    P U         Y D
  1   M 0 1995 0 0.8062971 0
  2   M 0 1995 1 0.8596851 0
  3   M 0 1996 1 0.0403833 0
  4   M 0 1997 0 0.4572211 0
  5   M 0 1997 1 0.1731691 0
  6   M 0 1998 0 0.9185489 0
```

2.1.1 Population time

Now we need the population data. It can be obtained either from the `Y.dk` dataset in the `Epi` package or from the human mortality database. The data in the `Epi` package are more up-to-date which is what we need:

```
> data( Y.dk )
> head( Y.dk )
    sex A    P    C        Y upper
1   1 0 1971 1971 19195.00     0
2   1 0 1971 1970 17944.17     1
3   1 1 1971 1970 17968.83     0
4   1 1 1971 1969 18164.83     1
5   1 2 1971 1969 18178.67     0
6   1 2 1971 1968 18934.33     1
```

We want data from the population in the years 1995 trough 2011 and ages 0–98 (because the population data only has 98 as the last closed age-class):

```
> Y.dk <- transform( Y.dk, U = upper,
+                          sex = factor(sex,labels=c("M","F")) )
> Y.dk <- subset( Y.dk, A < 99   &
+                       P > 1994 &
+                       P < 2012 )[,c("sex","A","P","U","Y")]
```

2.1.2 Merging time

Now we merge the two data sets; we construct the risk time among DM patients is in the `Agg` dataset as `Y` and the risk time in the entire population is in the dataset `Y.dk`, also as `Y`, and hence in the merged dataset referred to as `Y.x` and `Y.y`, respectively. By that token we can construct `Y.DM` and `Y.nD` as the risk time among non-diabetics and among diabetes patients, respectively:

```
> YY <- merge( Agg, Y.dk, by=c("sex","A","P","U"), all.y=TRUE )
> YY <- transform( YY, Y.nD = Y.y-pmax(Y.x,0,na.rm=TRUE),
+                      Y.DM =     pmax(Y.x,0,na.rm=TRUE),
+                      D.DM =     pmax(  D,0,na.rm=TRUE) )[,c("sex","A","P","U","Y.nD","Y.DM","D.DM"
> str( YY )
  'data.frame':        6732 obs. of  7 variables:
   $ sex : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P   : num  1995 1995 1996 1996 1997 ...
   $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
   $ Y.nD: num  18027 17871 17427 18062 17387 ...
   $ Y.DM: num  0.8063 0.8597 0 0.0404 0.4572 ...
   $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...

> head( YY )
    sex A    P U     Y.nD      Y.DM D.DM
1   M 0 1995 0 18026.69 0.8062971    0
2   M 0 1995 1 17870.97 0.8596851    0
3   M 0 1996 0 17426.50 0.0000000    0
4   M 0 1996 1 18062.13 0.0403833    0
5   M 0 1997 0 17386.54 0.4572211    0
6   M 0 1997 1 17450.66 0.1731691    0
```

2.1.3 Population deaths

We can extract the number of deaths in Lexis-triangles from the Human mortality database, using the function

```
> require(RCurl)
> pth <- "http://www.mortality.org/hmd/DNK/STATS/Deaths_lexis.txt"
> upw <- "bxc@steno.dk:BxCPwd"
> txt <- getURL( pth, userpwd=upw )
> con <- textConnection( txt )
> mlx <- try( read.table( con, skip = 2, header = TRUE, na.strings = "."), TRUE)
> str( mlx )
  'data.frame':        39117 obs. of  6 variables:
   $ Year  : int  1835 1835 1835 1835 1835 1835 1835 1835 1835 1835 ...
   $ Age   : Factor w/ 111 levels "0","1","10","100",..: 1 1 2 2 24 24 35 35 46 46 ...
   $ Cohort: int  1835 1834 1834 1833 1833 1832 1832 1831 1831 1830 ...
   $ Female: num  2159 1156 502 364 293 ...
   $ Male  : num  2772 1604 562 402 332 ...
   $ Total : num  4930 2761 1064 766 626 ...
```

We then restrict and transform these data to be of the sam shape as the tabulated follow-up of the diabetes patients:

```
> mlx <- subset( mlx, Year>1994 & Year<2012 & Age!="110+" )
> mlx$A <- as.numeric(as.character(mlx$Age))
> mlx <- transform( mlx, P=Year,
+                        C=Cohort,
+                        U=Year-A-Cohort )
> mm <- data.frame( mlx[,c("A","P","U","Male")],
+                   sex=factor(1,levels=1:2,labels=c("M","F") ) )
> mf <- data.frame( mlx[,c("A","P","U","Female")],
+                   sex=factor(2,levels=1:2,labels=c("M","F") ) )
> names(mm)[4] <-
+ names(mf)[4] <- "D.nD"
> MM <- subset( rbind( mm, mf ), A < 99 )
> head( MM )
       A    P U D.nD sex
  35361 0 1995 0  179   M
  35362 0 1995 1   21   M
  35363 1 1995 0   13   M
  35364 1 1995 1    8   M
  35365 2 1995 0    2   M
  35366 2 1995 1    7   M
> save( MM, file="./data/MM.Rda" )
```

Now we have the total number of deaths in Lexis triangles for the relevant period, we can merge with the follow-up dataset, so we have the number of deaths and person-years by sex, age, period and diabetes status:

```
> TT <- transform( merge( YY, MM ), D.nD = D.nD - D.DM )
> head( TT )
    sex A    P U     Y.nD      Y.DM D.DM D.nD
  1   F 0 1995 0 17025.50 0.0000000    0  137
  2   F 0 1995 1 17100.54 0.1300479    0   16
  3   F 0 1996 0 16468.06 1.4401095    0  134
  4   F 0 1996 1 17067.30 1.8617385    0   23
  5   F 0 1997 0 16434.00 0.0000000    0  152
  6   F 0 1997 1 16499.84 1.9890486    0   14
```

2.1.4 DM cases

Finally we want to append the number of diabetes cases to the data frame, so we count the number of entries in the Lexis object `Lx`

```
> CC <- with( subset( Lx, P>1995 ),
+             table( sex, floor(A),
+                         floor(P),
+                         floor(P) - floor(A) - floor(P-A) ) )
> CC <- as.data.frame( CC )
> names( CC ) <- c("sex","A","P","U","X")
> for( i in 2:4 ) CC[,i] <- as.numeric(as.character(CC[,i]))
> str( CC )
  'data.frame':        6732 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
   $ P  : num  1995 1995 1995 1995 1995 ...
   $ U  : num  0 0 0 0 0 0 0 0 0 0 ...
   $ X  : int  1 0 4 2 5 1 3 1 5 1 ...
```

Now `CC` contains the number of incident cases of DM in per period 1995–2011 incl. in the column `X`.

2.1.5 Saving it all for later analysis

```
> TT <- merge( TT, CC )
> str( TT )
  'data.frame':        6732 obs. of  9 variables:
   $ sex : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P   : num  1995 1995 1996 1996 1997 ...
   $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
   $ Y.nD: num  17026 17101 16468 17067 16434 ...
   $ Y.DM: num  0 0.13 1.44 1.86 0 ...
   $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.nD: num  137 16 134 23 152 14 132 16 95 17 ...
   $ X   : int  0 2 4 4 0 2 0 0 0 1 ...
```

The data frame `TT` has the risk time in the states "No DM" (`Y.nD`) and "DM" (`Y.DM`) and the number of transitions from "No DM" to either "DM" (`X`) or "Death" (`D.nd`) and from "DM" to "Death" (`D.DM`).

We can now finally save the tabulated dataset which contains information for analysis of incidence rates of diabetes and mortality rates for both diabetes patients and non-patients. We just define an attribute which

```
> Vars <- matrix( c("Sex",
+                   "1-year age class",
+                   "1-year period",
+                   "Indicator of upper Lexis triangle",
+                   "P-Y among non-diabetics",
+                   "P-Y among diabetes patients",
+                   "Deaths among non-diabetics",
+                   "Deaths among diabetes patients",
+                   "Diabetes diagnoses among non-diabetics"), ncol(TT) )
> rownames( Vars ) <- names( TT )
> colnames( Vars ) <-
+     "Data frame using the original definintion of DM from NDR"
> attr( TT, "Variables" ) <- Vars
> str( TT )
  'data.frame':        6732 obs. of  9 variables:
   $ sex : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P   : num  1995 1995 1996 1996 1997 ...
   $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
   $ Y.nD: num  17026 17101 16468 17067 16434 ...
```

```
  $ Y.DM: num  0 0.13 1.44 1.86 0 ...
  $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...
  $ D.nD: num  137 16 134 23 152 14 132 16 95 17 ...
  $ X   : int  0 2 4 4 0 2 0 0 0 1 ...
  - attr(*, "Variables")= chr [1:9, 1] "Sex" "1-year age class" "1-year period" "Indicator of upper
   ..- attr(*, "dimnames")=List of 2
   .. ..$ : chr  "sex" "A" "P" "U" ...
   .. ..$ : chr "Data frame using the original definintion of DM from NDR"
> save( Lx, TT, file="./data/FU-o.Rda" )
```

2.2 DM incidence

In this chapter we use the original definition of DM for the NDR, so first we load the analysis data frame:

```
> library( Epi )
> load( file="./data/FU-o.Rda" )
> head( TT )
    sex A    P U     Y.nD      Y.DM D.DM D.nD X
1   F 0 1995 0 17025.50 0.0000000    0  137 0
2   F 0 1995 1 17100.54 0.1300479    0   16 2
3   F 0 1996 0 16468.06 1.4401095    0  134 4
4   F 0 1996 1 17067.30 1.8617385    0   23 4
5   F 0 1997 0 16434.00 0.0000000    0  152 0
6   F 0 1997 1 16499.84 1.9890486    0   14 2
> attr( TT, "Variables" )
      Data frame using the original definintion of DM from NDR
sex   "Sex"
A     "1-year age class"
P     "1-year period"
U     "Indicator of upper Lexis triangle"
Y.nD  "P-Y among non-diabetics"
Y.DM  "P-Y among diabetes patients"
D.DM  "Deaths among non-diabetics"
D.nD  "Deaths among diabetes patients"
X     "Diabetes diagnoses among non-diabetics"
```

2.2.1 No. of cases

We would like to see the number of prevalent cases as of 1.1.1995 and the number of new cases for each year after that and the prevalent number of cases at the end. These numbers are readily available from the Lexis object `Lx`:

```
> prnew <- rbind( with( subset( Lx, doDM<1995 & na2T(dodsdto>1995) ),
+                       table( sex ) ),
+                 with( subset( Lx, doDM>=1995 ),
+                       table( floor(doDM), sex ) ),
+                 with( subset( Lx, doDM<2012 & na2T(dodsdto>2012) ),
+                       table( sex ) ) )
> rownames( prnew )[1] <- "Prev 1.1.1995"
> rownames( prnew )[nrow(prnew)] <- "Prev 31.11.2011"
> addmargins( prnew, margin=2 )
                      M      F    Sum
Prev 1.1.1995     49438  49126  98564
1995               7743   7131  14874
1996               8008   7377  15385
1997               7916   7522  15438
1998               8808   8034  16842
1999               9300   8553  17853
2000               9608   8872  18480
2001              10206   9469  19675
2002              11170  10778  21948
2003              12364  11348  23712
2004              12462  11465  23927
2005              11613  10573  22186
2006              12090  10911  23001
2007              12709  11768  24477
2008              13994  12654  26648
2009              14290  12349  26639
2010              14962  12808  27770
2011              17072  15302  32374
Prev 31.11.2011 160383 150461 310844
```

2.2.2 Age-Period-Cohort modelling

We are going to use `X` and `Y.nD` as response variables in the analysis of diabetes incidence rates, however we first need to define the age and period properly:

```
> DD <- transform( TT, A = A + (1+U)/3,
+                      P = P + (2-U)/3,
+                      D = X,
+                      Y = Y.nD/1000 )[,c("sex","A","P","D","Y")]
```

With this groomed data frame in place we can fit separate models for men and women and plot the estimates together:

```
> acpM <- apc.fit( subset(DD,sex=="M"), ref.c=1950, parm="ACP", npar=c(18,5,12) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347    11206.3
  Age-drift              3346     4410.6   1   6795.6 < 2.2e-16
  Age-Cohort             3335     4302.8  11    107.9 < 2.2e-16
  Age-Period-Cohort      3331     3998.9   4    303.9 < 2.2e-16
  Age-Period             3342     4107.1 -11   -108.3 < 2.2e-16
  Age-drift              3346     4410.6  -4   -303.5 < 2.2e-16

> acpF <- apc.fit( subset(DD,sex=="F"), ref.c=1950, parm="ACP", npar=c(18,5,12) )

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347    11692.9
  Age-drift              3346     5242.5   1   6450.4 < 2.2e-16
  Age-Cohort             3335     5055.0  11    187.5 < 2.2e-16
  Age-Period-Cohort      3331     4632.3   4    422.7 < 2.2e-16
  Age-Period             3342     4837.9 -11   -205.6 < 2.2e-16
  Age-drift              3346     5242.5  -4   -404.6 < 2.2e-16

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( acpM, lty=1, ci=TRUE, col="blue", r.txt="Rate per 1000 PY" )

  cp.offset    RR.fac
       1790         1

> lines( acpF, lty=1, ci=TRUE, col="red" )



> apcM <- apc.fit( subset(DD,sex=="M"), ref.p=2000, parm="APC", npar=c(18,5,12) )

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347    11206.3
  Age-drift              3346     4410.6   1   6795.6 < 2.2e-16
  Age-Cohort             3335     4302.8  11    107.9 < 2.2e-16
  Age-Period-Cohort      3331     3998.9   4    303.9 < 2.2e-16
  Age-Period             3342     4107.1 -11   -108.3 < 2.2e-16
  Age-drift              3346     4410.6  -4   -303.5 < 2.2e-16

> apcF <- apc.fit( subset(DD,sex=="F"), ref.p=2000, parm="APC", npar=c(18,5,12) )
```

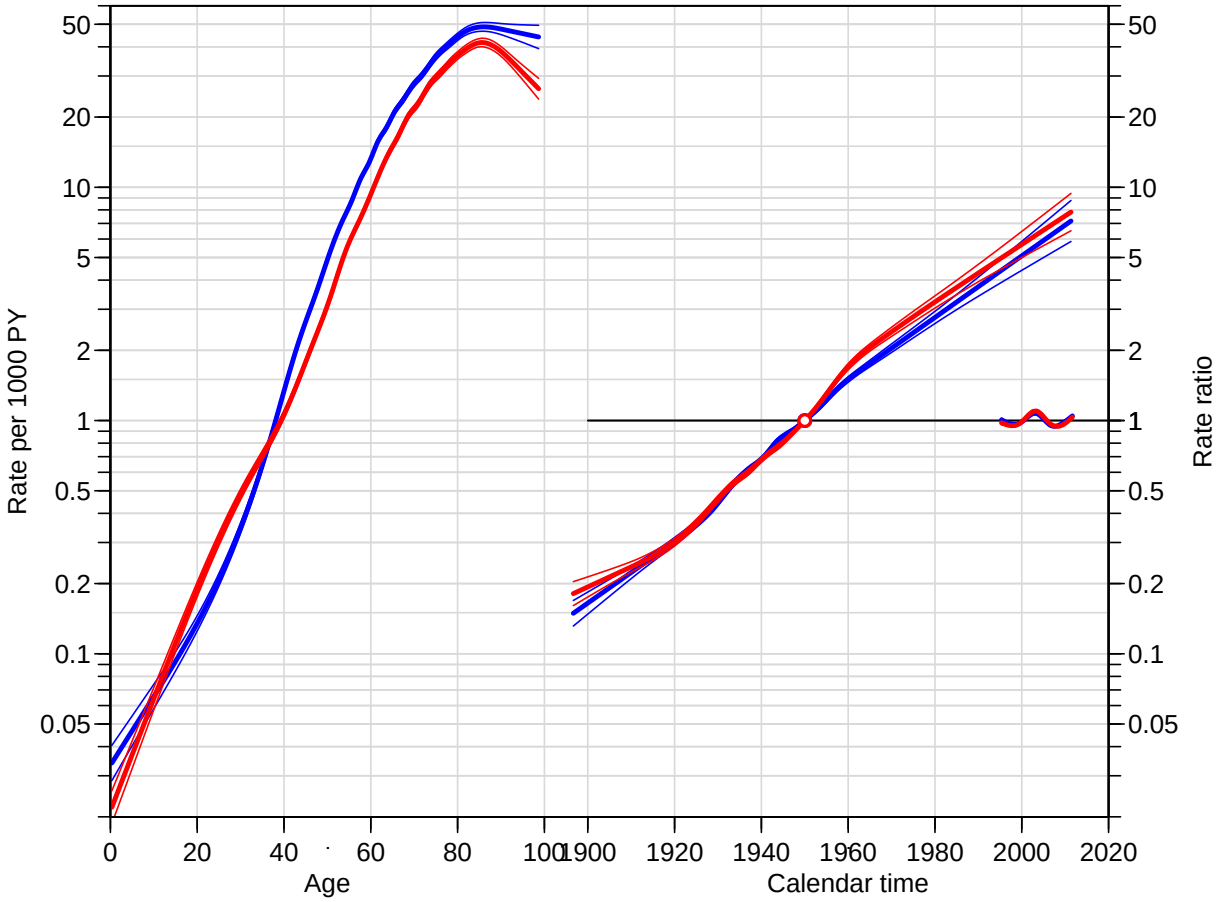

Figure 2.1: *Estimates from an APC-model for DM incidence rates in Denmark 1995–2011 (modified definition), cohort constrained to be 1 at 1950, period slope to be 0. Blue: Men, red: Women.*

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347    11692.9
  Age-drift              3346     5242.5   1   6450.4 < 2.2e-16
  Age-Cohort             3335     5055.0  11    187.5 < 2.2e-16
  Age-Period-Cohort      3331     4632.3   4    422.7 < 2.2e-16
  Age-Period             3342     4837.9 -11   -205.6 < 2.2e-16
  Age-drift              3346     5242.5  -4   -404.6 < 2.2e-16
> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( apcM, lty=1, ci=TRUE, col="blue", r.txt="Rate per 1000 PY" )
  cp.offset    RR.fac
       1790         1
> lines( apcF, lty=1, ci=TRUE, col="red" )
```

Both from figure **??** and **??** it is clear that there is some calendar-time effect at around 2005, where a downward change in incidence rates seem to occur. The major tendency is however the steady increase across cohort/period.

If we stick to the period-major parametrization as in figure **??**, we are essentially referring to cross-sectional rates, and they seem to have a peak around age 80. However since there

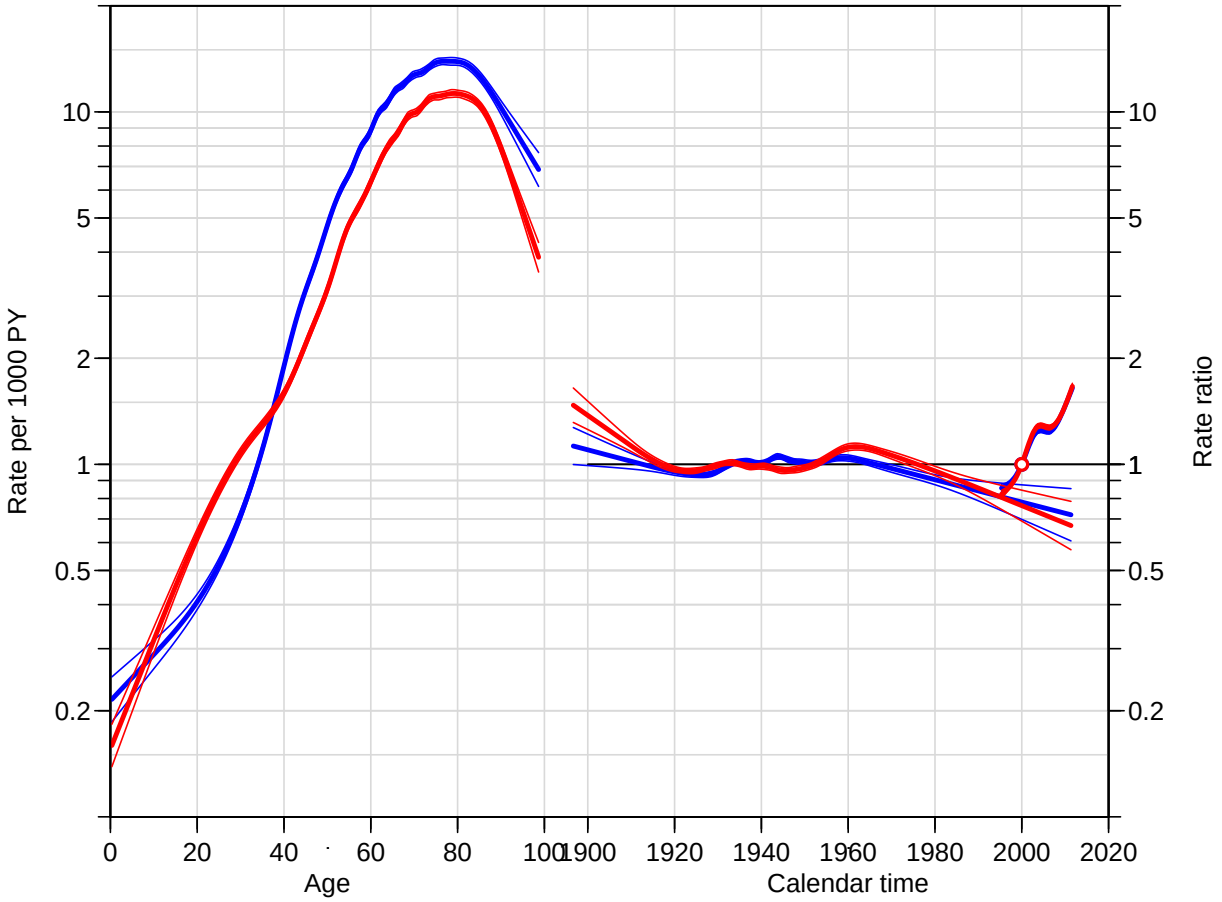

Figure 2.2: *Estimates from an APC-model for DM incidence rates in Denmark 1995–2011 (revised definition), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

is an increasing trend the peak incidence for a given generation is more likely at 85 years as shown in figure **??**, using the cohort major parametrization, the longitudinal approach.

2.2.3 Time-trends in rates

The overall time trend in the rates are in the `Drift` component of the `apc` object, here we give the average annual increase in incidence rates among men and women:

```
> pctchg <- (cbind( apcM$Drift, apcF$Drift )-1)*100
> colnames( pctchg ) <- c("Men","lo","up","Women","lo","up")
> round( pctchg, 2 )
```

```
      Men   lo   up Women   lo   up
APC 3.84 3.74 3.94  4.05 3.94 4.16
A-d 3.93 3.83 4.02  4.00 3.90 4.10
```

Thus we see that the average annual trend in rates is about 4% per year, slightly higher for women than for men.

2.2.4 Summary of the APC modelling

The deviance analysis of the model did not surprisingly show that both cohort and period have non-linear effects, however this formal significance is largely due to the large data base.

```
> par( mfrow=c(2,1), mar=c(0,4,0,4), oma=c(3,0,1,0), mgp=c(3,1,0)/1.6, las=1 )
> apc.frame( a.lab=seq(20,80,20), a.tic=c(5,seq(10,90,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=outer(c(1,2,5),10^(-1:1)), tic.fac=2,
+            r.tic=c(outer(c(5:9/10,1:5),10^(-1:1)),60), rr.ref=1,
+            gap=10, r.txt="DM incidence per 1000 PY", side=c(2,4) )
> lines( acpM, col="blue", ci=TRUE )
> lines( acpF, col="red" , ci=TRUE )
> apc.frame( a.lab=seq(20,80,20), a.tic=c(5,seq(10,90,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=outer(c(1,2,5),10^(-1:1)), tic.fac=2,
+            r.tic=c(outer(c(5:9/10,1:5),10^(-1:1)),60), rr.ref=1,
+            gap=10, r.txt="DM incidence per 1000 PY", side=c(1,2,4) )
> lines( apcM, col="blue", ci=TRUE )
> lines( apcF, col="red" , ci=TRUE )
```

In figure 2.3 is shown the same model in two different parametrizations, one with longitudinal and one with cross-sectional age-specific rates. Another way of visualizing the model is to show the estimated age-specific incidence rates for different birth cohorts.

To that end we use the model-objects returned by the `apc.fit` function to produce predicted rates. So we set up a prediction frame with ages for 15 different cohorts:

```
> prf <- data.frame( A = rep( c(NA,0:98), 8 ),
+                    C = rep( seq(1910,1980,10), each=100 ),
+                    Y = 1 )[-1,]
> prf <- transform( prf, P = C + A )
```

The we can make a fit of the models of relevance and make predictions based on this new frame. [1]

```
> Mapc <- glm( D ~ Ns(   A, kn=apcM$Knots$Age ) +
+                  Ns( P-A, kn=apcM$Knots$Coh ) +
+                  Ns( P  , kn=apcM$Knots$Per ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="M") )
> Map  <- glm( D ~ Ns( A, kn=apcM$Knots$Age ) +
+                  Ns( P, kn=apcM$Knots$Per ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="M") )
> Mac  <- glm( D ~ Ns(   A, kn=apcM$Knots$Age ) +
+                  Ns( P-A, kn=apcM$Knots$Coh ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="M") )
> Fapc <- glm( D ~ Ns(   A, kn=apcF$Knots$Age ) +
+                  Ns( P-A, kn=apcF$Knots$Coh ) +
+                  Ns( P  , kn=apcF$Knots$Per ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="F") )
> Fap  <- glm( D ~ Ns( A, kn=apcF$Knots$Age ) +
+                  Ns( P, kn=apcF$Knots$Per ),
```

[1] Note that we cannot use the returned model from the `apc` object since this is defined in terms specific matrices and *not* in terms of A, P and C:

```
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="F") )
> Fac  <- glm( D ~ Ns(   A, kn=apcF$Knots$Age ) +
+                  Ns( P-A, kn=apcF$Knots$Coh ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="F") )
> summary( fitted( apcM$Model ) - fitted( Mapc ) )
       Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
-1.535e-12 -7.105e-14  4.219e-15  2.242e-14  1.030e-13  1.137e-12
> summary( fitted( apcF$Model ) - fitted( Fapc ) )
       Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
-7.958e-13 -6.750e-14  7.105e-15  5.261e-14  1.776e-13  1.222e-12
```

From the last summary we see that the models are the same as those fitted by `apc.fit`, an moreover we can use this latter to make predictions, regardless of the overparametrization (we will get a warning, though). Recall that the `Y` was scaled to be person-millenia, so we get fitted values as rates per 1000 (namely the expected numbers based on the model for a data point where `Y` is equal to 1, as specified in `prf`):

```
> prr <- subset( prf, (P<2011 & P>1995) | is.na(P) )
> Mfit.apc <- predict( Mapc, newdata=prr )
> Mfit.ap  <- predict( Map , newdata=prr )
> Mfit.ac  <- predict( Mac , newdata=prr )
> Ffit.apc <- predict( Fapc, newdata=prr )
> Ffit.ap  <- predict( Fap , newdata=prr )
> Ffit.ac  <- predict( Fac , newdata=prr )
```

For comparison we overlay empirical rates, which we compute for the cohorts 1910 (born 1905–15), . . . , 1980 (born 1975–85) calculated in C-sets ( ); the `gc` and `gp` are the midpoints of the cohort and period in the C-sets:

```
> DD.x <- transform( DD,
+                    gc = floor(((P-A)-1905)/10)*10+1910,
+                    gp = floor(P)+0.5 )
> ee <- data.frame( xtabs( cbind(D,Y) ~ sex + gp + gc,
+                   data = subset( DD.x, gc>1905 & gc<1985 ) ) )
> ee <- reshape( ee, timevar = "Var4",
+                      idvar = c("sex","gp","gc"),
+                        dir = "wide" )
> names( ee )[4:5] <- c("D","Y")
> ee <- transform( ee, gp = as.numeric(as.character(gp)),
+                      gc = as.numeric(as.character(gc)) )
> str( ee )
  'data.frame':        272 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ gp : num  1996 1996 1996 1996 1998 ...
   $ gc : num  1910 1910 1910 1910 1910 1910 1910 1910 1910 1910 ...
   $ D  : num  579 966 473 853 402 747 369 717 323 618 ...
   $ Y  : num  52.4 103.7 45.1 93 38.4 ...
```

We then overlay the empirical over the fitted rates from the three different models, the age-period, the age-cohort and the apc-model:

```
> par( mfrow=c(2,1), mar=c(0,0,0,0), oma=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> matplot( prr$A, exp(Mfit.apc), type="l", lty=1,
+          log="y", ylim=c(0.2,25), lwd=3, xaxt="n", xlab="", ylab="" )
> matlines( prr$A, exp(Mfit.ap), type="l", lty="11", lwd=2 )
> matlines( prr$A, exp(Mfit.ac), type="l", lty=1, lwd=2 )
```

```
> with( subset(ee,sex=="M"),
+       points( gp-gc, D/Y, pch=16, col=rainbow(8)[factor(gc)], cex=0.8 ) )
> text( 20, 14, "Men", col="blue" )
> matplot( prr$A, exp(Ffit.apc), type="l", lty=1,
+          log="y", ylim=c(0.2,25), lwd=3, xlab="", ylab="" )
> matlines( prr$A, exp(Ffit.ap), type="l", lty="11", lwd=2 )
> matlines( prr$A, exp(Ffit.ac), type="l", lty=1, lwd=2 )
> with( subset(ee,sex=="F"),
+       points( gp-gc, D/Y, pch=16, col=rainbow(8)[factor(gc)], cex=0.8 ) )
> text( 20, 14, "Women", col="red" )
> mtext( "DM incidence rate per 1000 PY", side=2, outer=TRUE, line=2, las=0 )
> mtext( "Age (years)", side=1, outer=TRUE, line=2 )
```

From figure 2.4 we see that both the fitted and the empirical rates are indicative of a strong period effect with a characteristic dip around 2003–5, as seen in figure 2.3, so the significant non-linearity of the period effect is epidemiologically significant, not only statistically.

Note that the age-specific incidence rates in figure 2.3 are constructed gluing together the age-effects from the different cohorts, and aligning them to the 1950 cohort (the one with light-blue empirical rates).

2.2.5 Saving the fitted models

We then save these fitted APC-models with different parametrizations:

```
> save( Mapc, Mac, Fapc, Fac, file="./data/inc-o.Rda" )
```

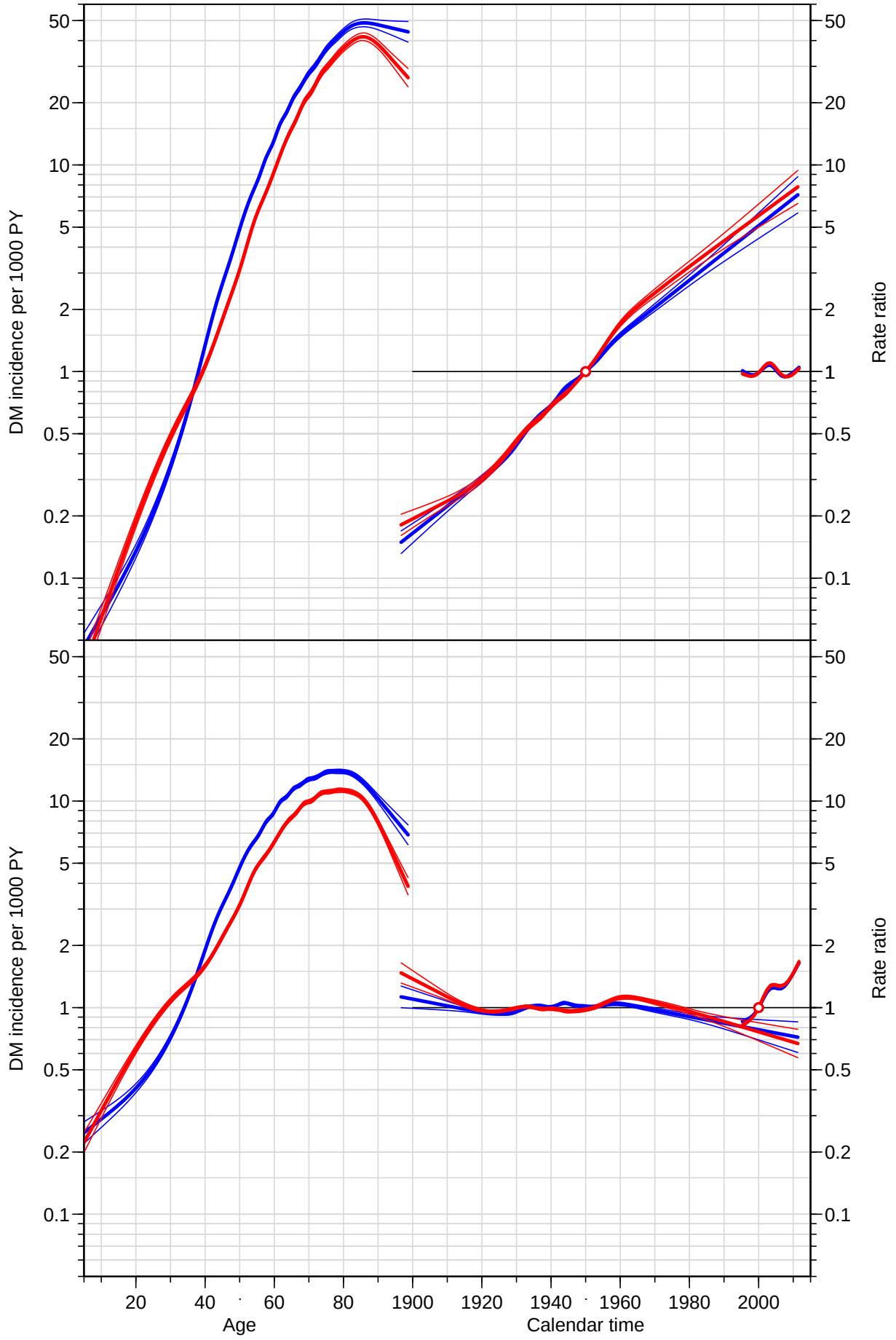

Figure 2.3: *Age-Period-Cohort models for DM incidence among men (blue) and women (red), using the same scaling in the two plots. The top panel is the parametrization with horizontal period effect and cohort reference 1950, bottom panel is the parametrization with horizontal*

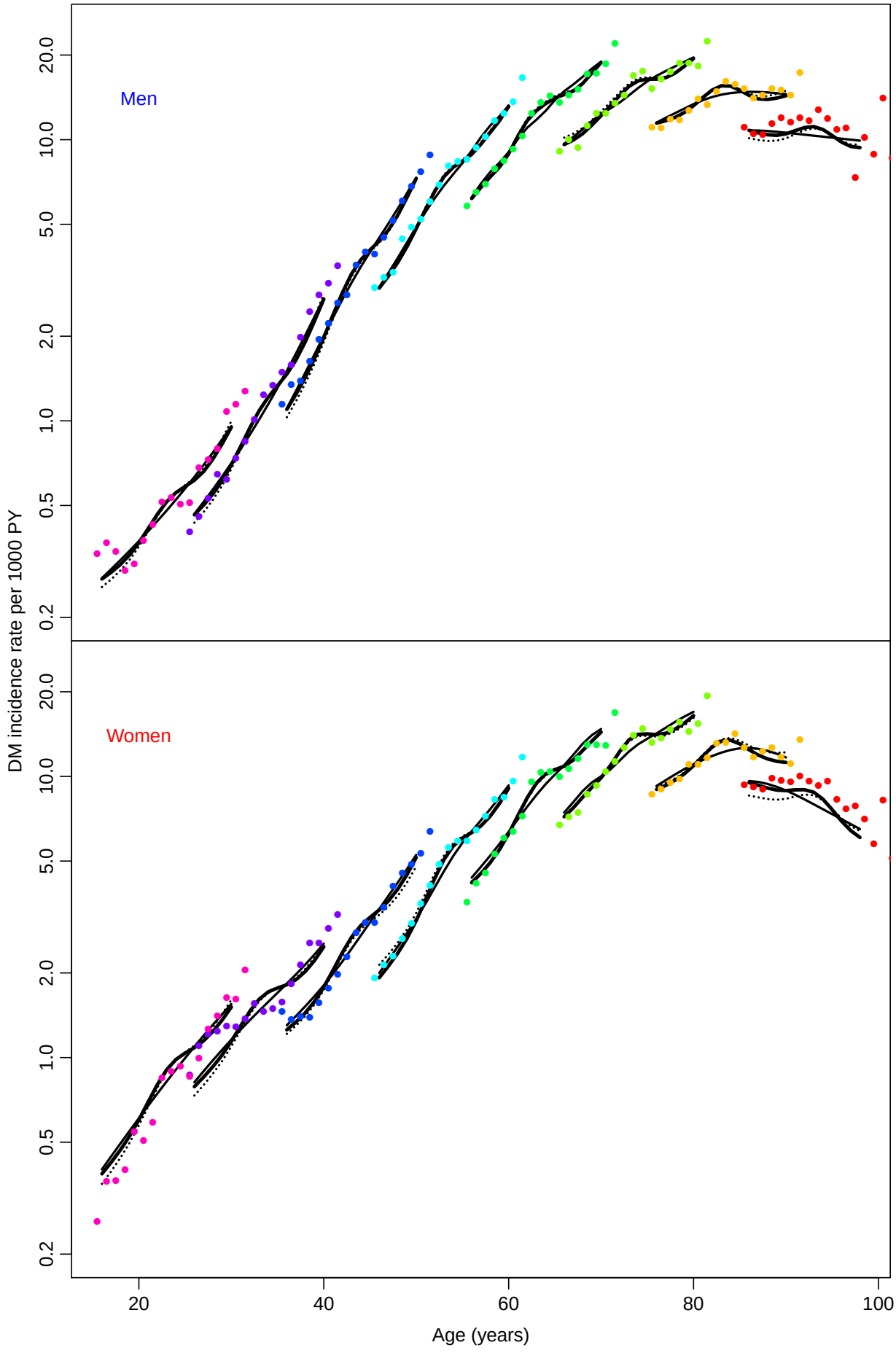

Figure 2.4: *Fitted age-specific incidence rates for the cohorts 1910,. . . ,1980: Full thick line: APC-model, broken line: AP model and full thin line: AC-model. Empirical age-specific*

2.3 Mortality

2.3.1 Mortality in non-diabetics

We are going to use `Y.nD` and `Y.nD` as response variables in the analysis of mortality rates, however we first need to define the age and period properly for analysis in Lexis triangles:

```
> nD <- subset( transform( TT, A = A + (1+U)/3,
+                              P = P + (2-U)/3,
+                              D = pmax(D.nD,0),
+                              Y = Y.nD/1000 )[,c("sex","A","P","D","Y")],
+               Y > 0 )
```

With this groomed data frame in place we can fit separate models for men and women and plot the estimates together:

```
> nDacpM <- apc.fit( subset(nD,sex=="M"),
+                    ref.c=1950,
+                     parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3346    13733.2
  Age-drift              3345     7586.8  1   6146.4 < 2.2e-16
  Age-Cohort             3336     6668.1  9    918.7 < 2.2e-16
  Age-Period-Cohort      3329     6623.3  7     44.8 1.504e-07
  Age-Period             3338     7552.5 -9   -929.1 < 2.2e-16
  Age-drift              3345     7586.8 -7    -34.3 1.485e-05

> nDacpF <- apc.fit( subset(nD,sex=="F"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3346    10525.1
  Age-drift              3345     7105.5  1   3419.5 < 2.2e-16
  Age-Cohort             3336     6207.0  9    898.6 < 2.2e-16
  Age-Period-Cohort      3329     6165.8  7     41.1 7.639e-07
  Age-Period             3338     7057.4 -9   -891.6 < 2.2e-16
  Age-drift              3345     7105.5 -7    -48.1 3.394e-08

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( nDacpF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )

  cp.offset    RR.fac
       1790        10

> lines( nDacpM, lty=1, ci=TRUE, col="blue" )
```

We also fit using the period-major parametrization:

```
> nDapcM <- apc.fit( subset(nD,sex=="M"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
```

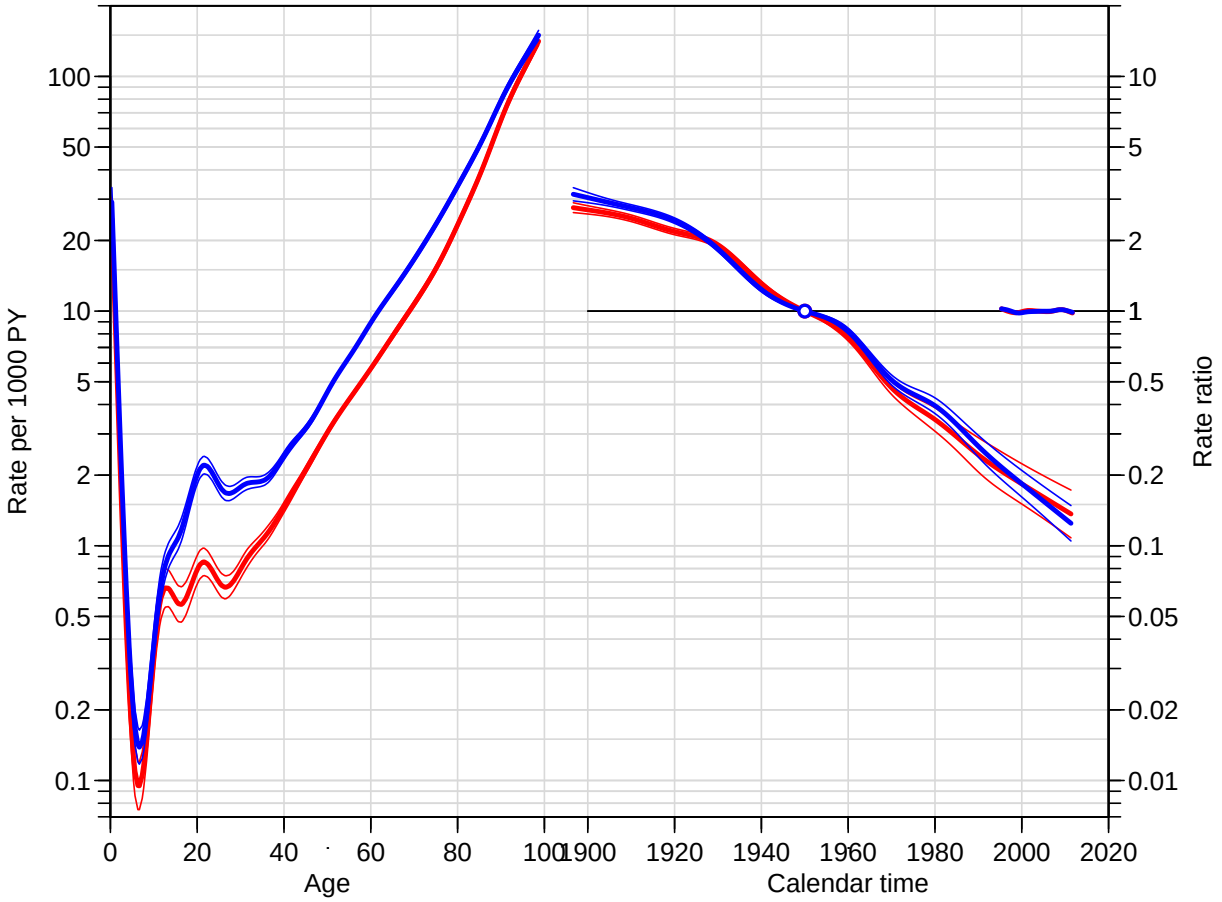

Figure 2.5: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), cohort effects constrained to be 1 at 1950, period slope to be 0. Blue: Men; red: Women.*

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

  Analysis of deviance for Age-Period-Cohort model

                     Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                     3346    13733.2
  Age-drift               3345     7586.8  1   6146.4 < 2.2e-16
  Age-Cohort              3336     6668.1  9    918.7 < 2.2e-16
  Age-Period-Cohort       3329     6623.3  7     44.8 1.504e-07
  Age-Period              3338     7552.5 -9   -929.1 < 2.2e-16
  Age-drift               3345     7586.8 -7    -34.3 1.485e-05
> nDapcF <- apc.fit( subset(nD,sex=="F"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                     Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                     3346    10525.1
  Age-drift               3345     7105.5  1   3419.5 < 2.2e-16
  Age-Cohort              3336     6207.0  9    898.6 < 2.2e-16
  Age-Period-Cohort       3329     6165.8  7     41.1 7.639e-07
```

```
  Age-Period             3338     7057.4 -9   -891.6 < 2.2e-16
  Age-drift              3345     7105.5 -7    -48.1 3.394e-08

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( nDapcF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )

  cp.offset     RR.fac
       1790        100

> lines( nDapcM, lty=1, ci=TRUE, col="blue" )
```

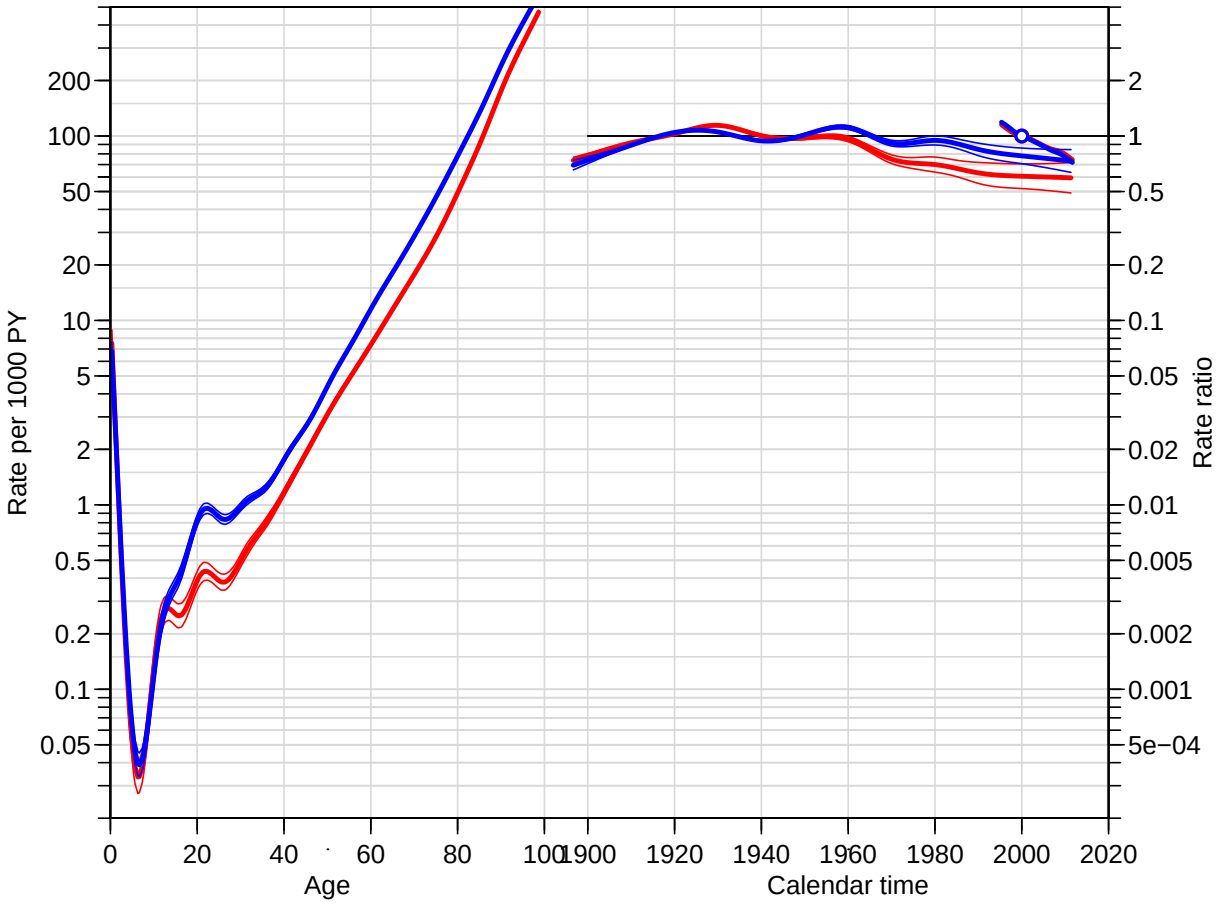

Figure 2.6: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

2.3.2 Mortality among DM patients

Here we use `D.DM` and `Y.DM` as response variables in the analysis of mortality rates among non-diabetics, and again we first need to define the age and period properly:

```
> DM <- subset( transform( TT, A = A + (1+U)/3,
+                              P = P + (2-U)/3,
+                              D = pmax(D.DM,0),
+                              Y = Y.DM/1000 )[,c("sex","A","P","D","Y")],
+               Y > 0 )
```

With this groomed data frame in place we can fit separate models for men and women and plot the estimates together:

```
> DMacpM <- apc.fit( subset(DM,sex=="M"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                    3340     5978.9
  Age-drift              3339     3070.8  1  2908.13  < 2e-16
  Age-Cohort             3330     2874.2  9   196.62  < 2e-16
  Age-Period-Cohort      3323     2855.8  7    18.39  0.01035
  Age-Period             3332     3056.2 -9  -200.39  < 2e-16
  Age-drift              3339     3070.8 -7   -14.61  0.04128

> DMacpF <- apc.fit( subset(DM,sex=="F"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3334     5020.2
  Age-drift              3333     2996.4  1  2023.77 < 2.2e-16
  Age-Cohort             3324     2807.3  9   189.11 < 2.2e-16
  Age-Period-Cohort      3317     2768.1  7    39.26 1.744e-06
  Age-Period             3326     2971.9 -9  -203.86 < 2.2e-16
  Age-drift              3333     2996.4 -7   -24.51 0.0009276

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( DMacpF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )

  cp.offset    RR.fac
       1790        10

> lines( DMacpM, lty=1, ci=TRUE, col="blue" )
```

We also fit using the period-major parametrization:

```
> DMapcM <- apc.fit( subset(DM,sex=="M"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                    3340     5978.9
  Age-drift              3339     3070.8  1  2908.13  < 2e-16
  Age-Cohort             3330     2874.2  9   196.62  < 2e-16
  Age-Period-Cohort      3323     2855.8  7    18.39  0.01035
  Age-Period             3332     3056.2 -9  -200.39  < 2e-16
  Age-drift              3339     3070.8 -7   -14.61  0.04128

> DMapcF <- apc.fit( subset(DM,sex=="F"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
```

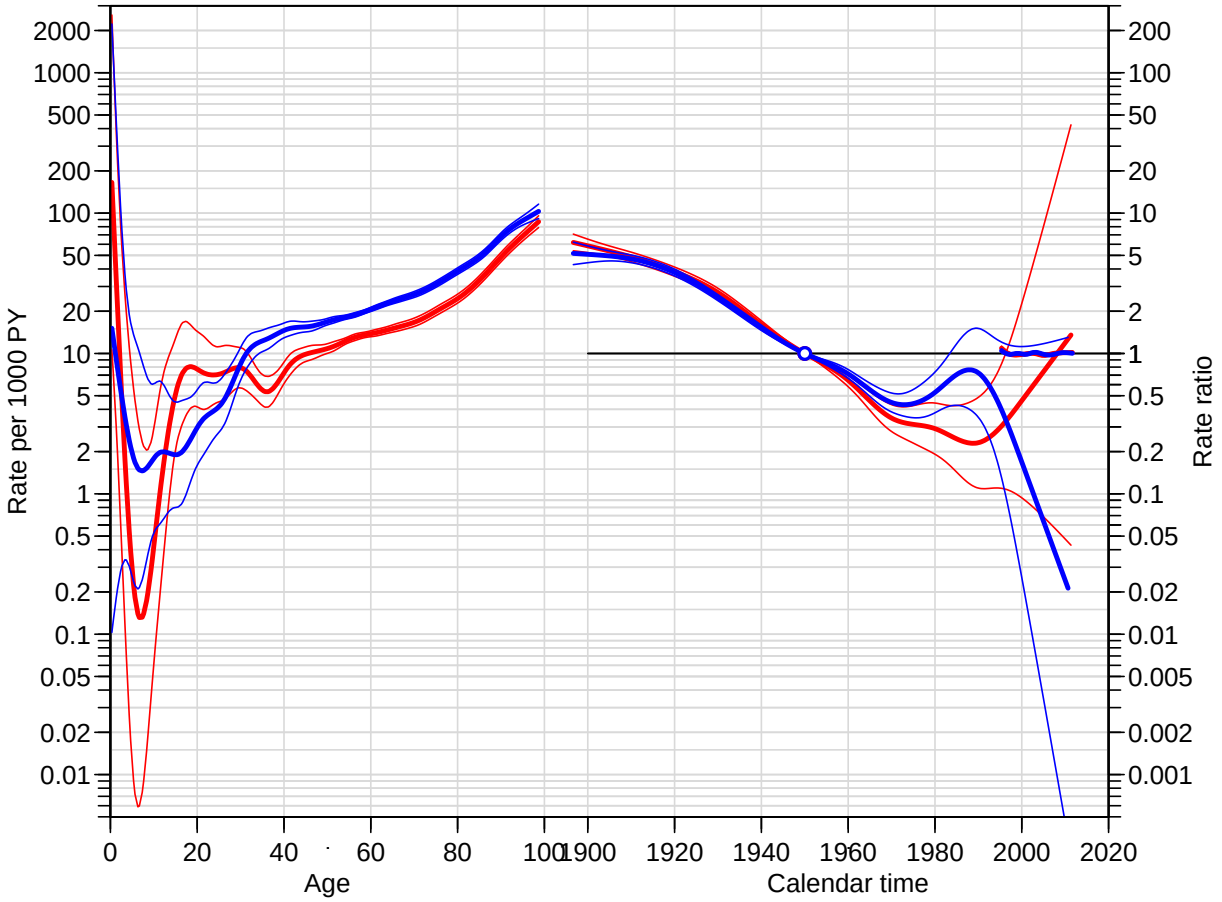

Figure 2.7: *Estimates from an APC-model for mortality among DM patients in Denmark 1995–2011 (original definition), cohort constrained to be 1 at 1950, period slope to be 0. Blue: Men, red: Women.*

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

  Analysis of deviance for Age-Period-Cohort model

                     Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                     3334     5020.2
  Age-drift               3333     2996.4  1  2023.77 < 2.2e-16
  Age-Cohort              3324     2807.3  9   189.11 < 2.2e-16
  Age-Period-Cohort       3317     2768.1  7    39.26 1.744e-06
  Age-Period              3326     2971.9 -9  -203.86 < 2.2e-16
  Age-drift               3333     2996.4 -7   -24.51 0.0009276
> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( DMapcF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )
  cp.offset    RR.fac
       1790         1
> lines( DMapcM, lty=1, ci=TRUE, col="blue" )
```

2.3.3 Summary of the APC models for mortality

The deviance analysis of the model did not surprisingly show that both cohort and period have non-linear effects, however this formal significance is largely due to the large data

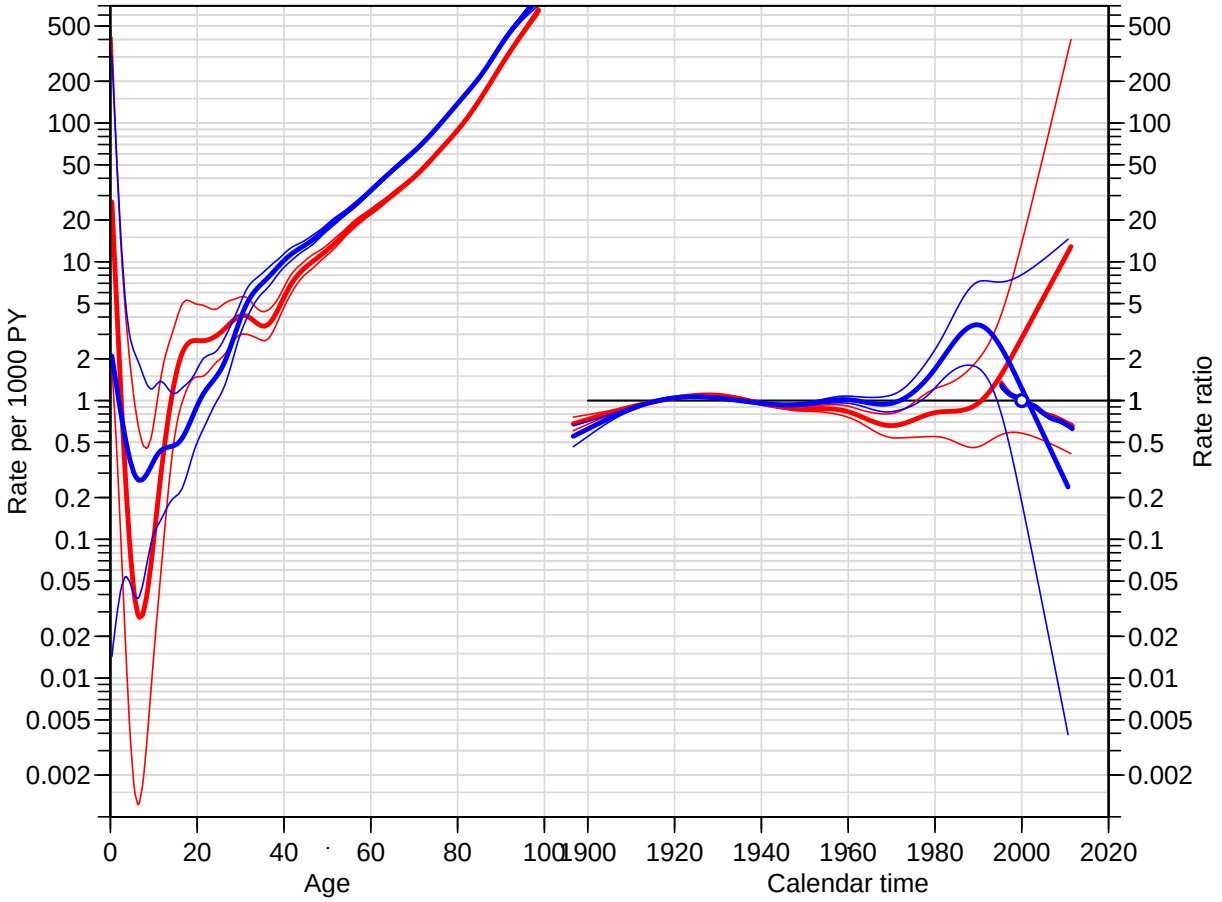

Figure 2.8: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

base, clearly there is no eidemiologically significant period-effect.

```
> par( mfrow=c(2,1), mar=c(0,4,0,4), oma=c(3,0,1,0), mgp=c(3,1,0)/1.6, las=1 )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+            cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+             r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+             r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=10,
+             gap=5, r.txt="Non-DM mortality per 1000 PY", side=c(2,4) )
> lines( nDacpM, col="blue", ci=TRUE )
> lines( nDacpF, col="red" , ci=TRUE )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+            cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+             r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+             r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=10,
+             gap=5, r.txt="DM mortality per 1000 PY", side=c(1,2,4) )
> lines( DMacpM, col="blue", ci=TRUE )
> lines( DMacpF, col="red" , ci=TRUE )
```

2.3.4 Time-trends in mortality rates

We can extract the timetrends for diabetics and non-diabetics by sex, and print the annual percentwise change:

```
> DA <- NArray( c( list( who = c("non-DM","DM"),
+                        sex = c("M","F") ),
+                   dimnames( nDacpM$Drift ) ) )
> DA["non-DM","M",,] <- nDacpM$Drift
> DA["non-DM","F",,] <- nDacpF$Drift
> DA[    "DM","M",,] <- DMacpM$Drift
> DA[    "DM","F",,] <- DMacpF$Drift
> round( ftable( (DA-1)*100, row.vars=1:2 ), 1 )
                    APC                  A-d
             exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
  who    sex
  non-DM M         -2.8 -2.9  -2.7      -2.5 -2.6  -2.5
         F         -2.4 -2.5  -2.3      -1.8 -1.9  -1.8
  DM     M         -4.0 -4.1  -3.9      -3.8 -3.9  -3.7
         F         -3.8 -4.0  -3.7      -3.4 -3.5  -3.2
```

We see that there is not much difference in the overall trend between man and women, but there seem to be a substantially steeper decrease in mortality among diabetes patienst than among non-diabetes patients.

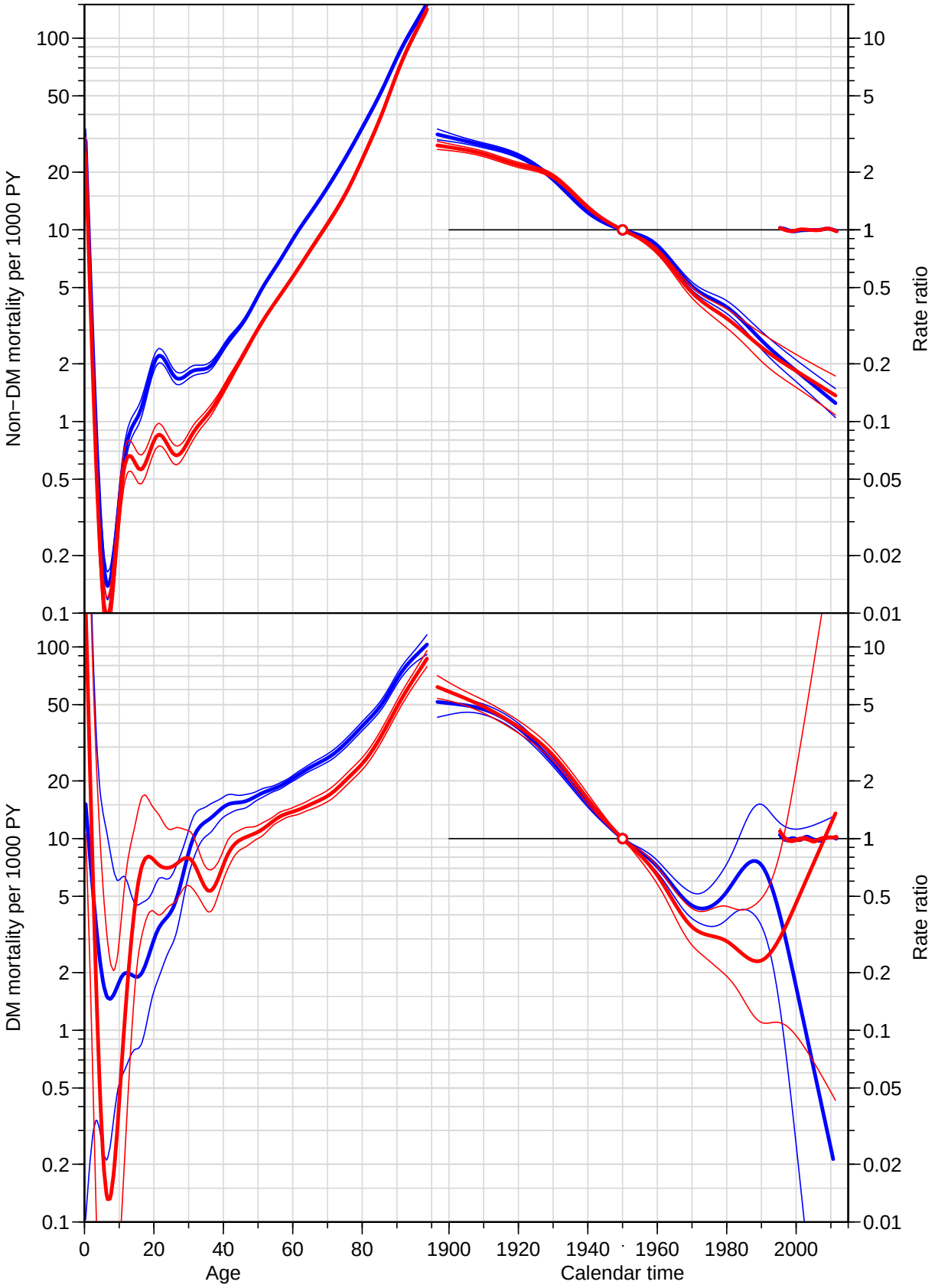

Figure 2.9: *Age-Period-Cohort models for mortality among men (blue) and women (red). Top panel is the mortality among non-diabetics and the lower panel is the mortality among diabetes patients.*

2.3.5 SMR

Since we have modelled both mortality rates by APC-models, and the analyses are done on (conditionally) independent datasets (follow-up in non-DM-, resp. DM-state), the ratio of the rates will also follow an APC-model, and the ratio of each set of effects will give three sets of RRs which wil mulitply to the overall RR. Since we have chosen the same reference cohort for both analyses, the cohort effect on the RR will also be with this reference. However, there is no *a priori* guarantee thate the period effect on the RR will be perfectly horizontal on average, even though it is going to be close.

However we will need a machinery to extract the RRs from the `apc` objects:

```
> make.RR.apc <-
+ function( a, b )
+ {
+   make.RR <-
+   function(A,B)
+   {
+   Z <- merge( A, B, by.x=1, by.y=1 )
+   lA <- log(Z[,2])
+   sA <- log(Z[,4]/Z[,3])/(2*1.96)
+   lB <- log(Z[,5])
+   sB <- log(Z[,7]/Z[,6])/(2*1.96)
+   RR <- cbind( A[,1], exp( lA-lB ),
+                       exp( lA-lB - 1.96*sqrt(sA^2+sB^2) ),
+                       exp( lA-lB + 1.96*sqrt(sA^2+sB^2) ) )
+   }
+ RR <- list( Age = make.RR( a$Age, b$Age ),
+             Per = make.RR( a$Per, b$Per ),
+             Coh = make.RR( a$Coh, b$Coh ),
+             Ref = a$Ref )
+ class( RR ) <- "apc"
+ RR
+ }
> SMR.M <- make.RR.apc( DMacpM, nDacpM )
> SMR.F <- make.RR.apc( DMacpF, nDacpF )
```

The two objects are not "real" `apc` objects, but they have the class attribute and the have the elements `Age`, `Per` and `Coh`, which are the only ones used by the `lines.apc` function. Hence we can plot the mortality rates for the DM patients together with the SMR relative to the non-diabetics.

```
> par( mfrow=c(2,1), mar=c(0,4,0,4), oma=c(3,0,1,0), mgp=c(3,1,0)/1.6, las=1 )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+            r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=10,
+            gap=5, r.txt="DM mortality per 1000 PY", side=c(2,4) )
> lines( DMacpM, col="blue", ci=TRUE )
> lines( DMacpF, col="red" , ci=TRUE )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+            r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=1,
+            gap=5, r.txt="SMR DM vs. non-DM", rr.txt="RR ratio", side=c(1,2,4) )
> abline( h=1 )
> lines( SMR.M, col="blue", ci=TRUE )
> lines( SMR.F, col="red" , ci=TRUE )
```

We see that the SMR is decreasing by age, and there seems to be no non-linear period effect, but an overall decreasing trend by period/birth cohort. Figure 2.10 shows a decrease in SMR from about 5 in age 40 to aound 1 in age 80 for the 1950 cohort. Note however that this is a bit of an extrapolation; the 1950 cohort has only been observed in ages 45 to 62.

2.3.6 Saving mortality rates

Finally, we save the `apc`-objects for subsequent use, however only the `ACP`-parametrized ones:

```
> save( nDacpM,
+       nDacpF,
+       DMacpM,
+       DMacpF, file="./data/APC-mort-o.Rda" )
```

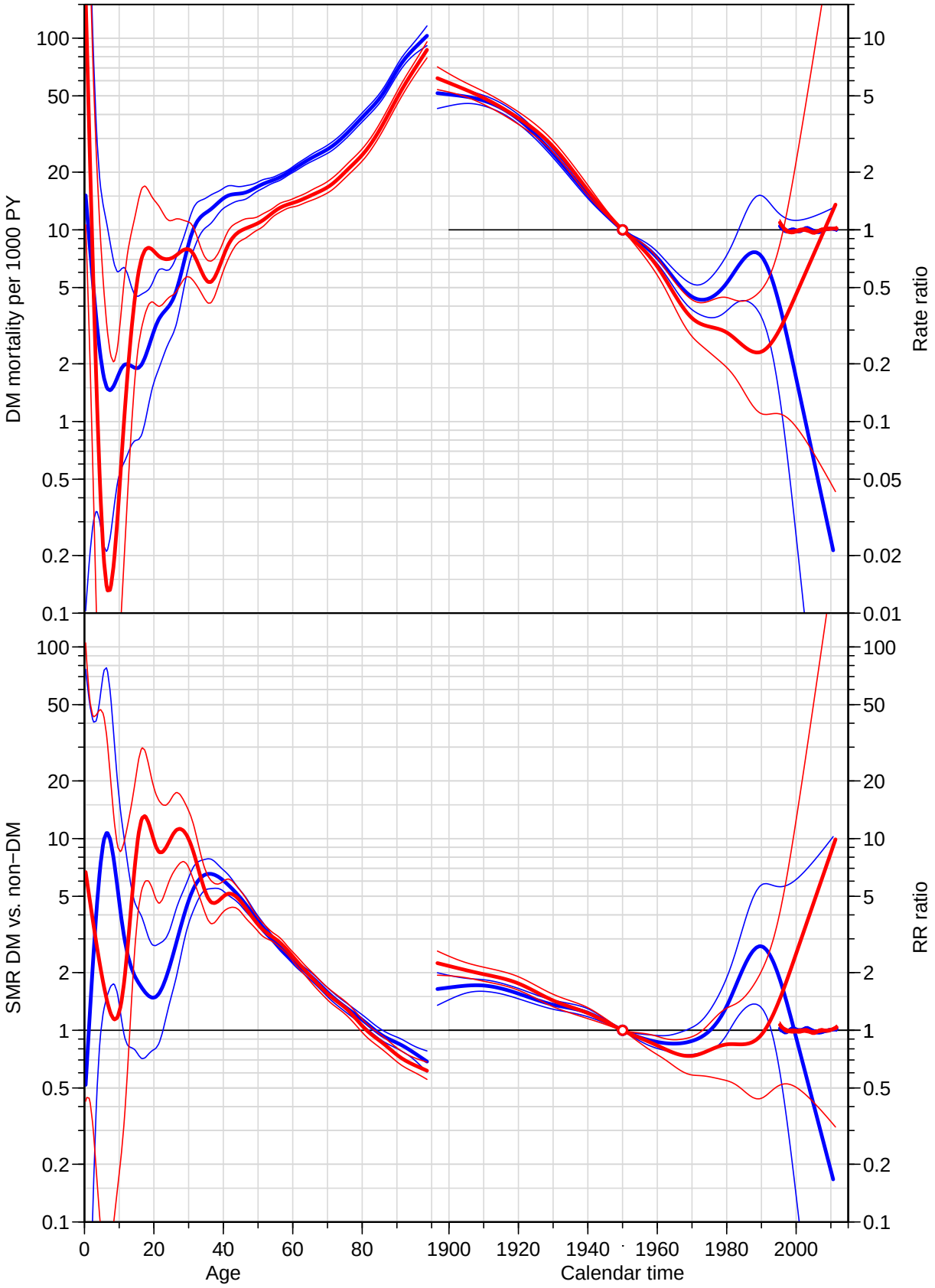

Figure 2.10: *Age-Period-Cohort models for mortality among men (blue) and women (red). Top panel is the mortality among diabetes patients, and the lower panel is the SMR versus the non-diabetic population.*

2.4 Prevalence of diabetes

We will analyze age-specific prevalence for each sex and each 1st January 1995—2012 separately, even though they are not independent.

First we set up a table of prevalent cases for each of the dates 1 January 1995–2012:

```
> pr <- NULL
> for( y in 1995:2012 )
+ pr <- rbind( pr,
+       cbind( with( subset( Lx, doDM < y & dox > (y-1/400) ),
+                    data.frame( table( sex, A=floor(y-foddto) ) ) ),
+              P = y ) )
> pr <- pr[,c(1,2,4,3)]
> pr$A <- as.numeric( as.character( pr$A ) )
> names( pr )[4] <- "X"
> str( pr )
  'data.frame':        3564 obs. of  4 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
   $ P  : int  1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
   $ X  : int  3 0 4 3 6 6 5 4 12 10 ...
```

Then we merge it with the population data:

```
> data( N.dk )
> head( N.dk )
    sex A    P     N
1   1 0 1971 35839
2   2 0 1971 34108
3   1 1 1971 36302
4   2 1 1971 34153
5   1 2 1971 37855
6   2 2 1971 35609
> N.dk <-  subset( N.dk, A<100 & P>1994 & P<2013 )
> N.dk$sex <- factor( N.dk$sex, labels=c("M","F") )
> str(N.dk)
  'data.frame':        3600 obs. of  4 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
   $ P  : num  1995 1995 1995 1995 1995 ...
   $ N  : num  35612 34094 34747 32967 35082 ...
> pr <- merge( pr, N.dk, all.y=TRUE )
> pr$X <- pmax( pr$X, 0, na.rm=TRUE )
> str( pr )
  'data.frame':        3600 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ A  : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P  : num  1995 1996 1997 1998 1999 ...
   $ X  : num  3 1 0 1 2 1 1 0 1 1 ...
   $ N  : num  35612 36055 34853 34774 34076 ...
```

We now have the empirical prevalences in the data frame `pr`, (`X`—no. of cases of DM, `N`—population size) for each of the 18 dates 1.1.1995 – 1.1.2012 in 1-year intervals.

```
> save( pr, file="./data/prev-o.Rda" )
```

So we model the prevalences as of 1 January each of the years 1995—2012, as a smooth function of age using a log-link binomial model with a smooth spline with 16 knots.

For the practical location of the spline knots we also define a small function which from the number of knots derives reasonable quantiles:

```
> qn <- function( nk, bd=2 ) seq( from =   1/(bd*nk),
+                                   to = 1-1/(bd*nk),
+                                length = nk )
> qn( 10, 2 )
 [1] 0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
> qn( 10, 5 )
 [1] 0.0200000 0.1266667 0.2333333 0.3400000 0.4466667 0.5533333 0.6600000
 [8] 0.7666667 0.8733333 0.9800000
```

Using this we get:

```
> ( kp.a <- c( 10, with( pr, quantile( rep(A,X), qn(15) ) ) ) )
          3.333333%       10% 16.66667% 23.33333%       30% 36.66667% 43.33333%
         10         28         40         47         52         56         59         62
        50% 56.66667% 63.33333%       70% 76.66667% 83.33333%       90% 96.66667%
         64         67         69         72         75         78         82         87
```

We now set up an array to hold the smoothed prevalences:

```
> a.pt <- 0:99
> p.pt <- 1995:2012
> pr.fit <- NArray( list( sex = c("M","F"),
+                            A = a.pt,
+                            P = p.pt ) )
```

So once we have set up the array to hold the smoothed empirical prevalences we can fill them into `pr.fit`:

```
> for( sx in dimnames(pr.fit)[["sex"]] )
+ for( dt in dimnames(pr.fit)[["P"]] )
+ pr.fit[sx,,dt] <- predict( glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ),
+                                  family = binomial(link="log"),
+                                    data = subset( pr, sex==sx & P==as.numeric(dt) ) ),
+                             newdata = data.frame( A=a.pt ),
+                                type = "response" )
```

We can plot how the age-specific prevalences have evolved over time:

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> matplot( a.pt, pr.fit["M",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", lty=1, col="blue", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt, pr.fit["M",,]*100, type="l", lty=1, col="blue", lwd=c(1,2) )
> text( 25, 22, "Men", adj=c(0,1), col="blue", cex=1.2 )
> text( 89, pr.fit["M","89","1995"]* 99, "1995", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M","80","2012"]*101, "2012", col="blue", adj=c(0,0) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, pr.fit["F",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", lty=1, col="red", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt, pr.fit["F",,]*100, type="l", lty=1, col="red", lwd=c(1,2) )
> text( 25, 22, "Women", adj=c(0,1), col="red", cex=1.2 )
> text( 89, pr.fit["F","89","1995"]* 99, "1995", col="red", adj=c(1,1) )
> text( 80, pr.fit["F","80","2012"]*101, "2012", col="red", adj=c(1,0) )
> axis( side=1 )
> mtext( "Age", side=1, line=1, outer=T )
> mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
```

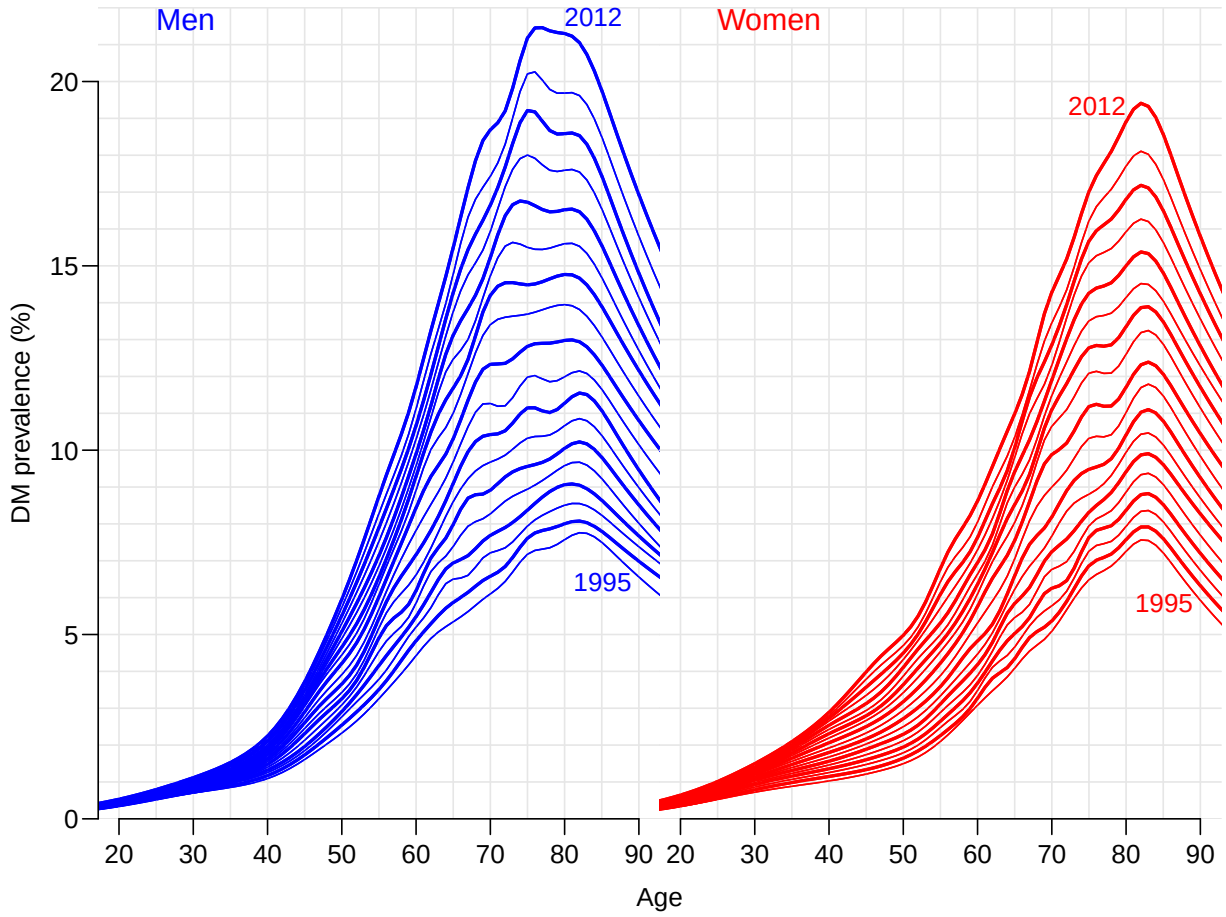

Figure 2.11: *Smoothed age-specific prevalences for the 17-year period 1995–2012. Blue is men, red is women.*

2.4.1 Trends in prevalence

A crude way of summarizing the prevalences is to assume that relative change is constant from year to year. So we set up a model that does this separately for men and women, and store the predicted values for comparison with those from the model with no assumption about the time evolution:

```
> pr.lfit <- pr.fit
> pr.chg <- NArray( list( dimnames(pr.fit)[["sex"]],
+                         c("% chg/y","lo","hi") ) )
> for( sx in dimnames(pr.fit)[["sex"]] )
+    {
+ lmod <- glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ) + P,
+               family = binomial(link="log"),
+                 data = subset( pr, sex==sx ) )
+ pr.chg[sx,] <- ( ci.exp( lmod, subset="P" ) - 1 ) * 100
+ pr.lfit[sx,,] <- predict( lmod,
+                           newdata = data.frame( A=rep(a.pt,     length(p.pt)),
+                                                 P=rep(p.pt,each=length(a.pt)) ),
+                              type = "response" )
+    }
```

This model is of course a simplification of the model above, with an arbitrary age-date interaction, so we can have a peep at how the predicted prevalences looks:

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> lblu <- rgb( 3,3,4,max=4 )
> lred <- rgb( 4,3,3,max=4 )
> matplot( a.pt, pr.fit["M",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="blue", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["M",,]*100, type="l", lty=1, col=lblu  , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["M",,]*100, type="l", lty=1, col="blue", lwd=c(2,3) )
> text( 25, 22, "Men", adj=c(0,1), col="blue", cex=1.2 )
> text( 89, pr.fit["M","89","1995"]* 99, "1995", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M","80","2012"]*101, "2012", col="blue", adj=c(0,0) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, pr.fit["F",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="red", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["F",,]*100, type="l", lty=1, col=lred , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["F",,]*100, type="l", lty=1, col="red", lwd=c(2,3) )
> text( 25, 22, "Women", adj=c(0,1), col="red", cex=1.2 )
> text( 89, pr.fit["F","89","1995"]* 99, "1995", col="red", adj=c(1,1) )
> text( 80, pr.fit["F","80","2012"]*101, "2012", col="red", adj=c(1,0) )
> axis( side=1 )
> mtext( "Age", side=1, line=1, outer=T )
> mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
```

From figure 2.12 we see that for men the summary using a constant relative change in prevalence is not a very good summary of the change in prevalences; it does not capture the change in the age of peak prevalence of men from 85 in 1995 to 75 in 2012. So the overall estimate of some 6% in relative annual increase of prevalences over the 17-year period 1995–2012, is not providing an adequate summary:

```
> round( pr.chg, 2 )
    % chg/y   lo   hi
M    5.71 5.68 5.74
F    6.01 5.98 6.04
```

2.4.2 Prevalence age-period interaction

Hence the relevant description of average changes per year would be using a model for the prevalences where we allowed the relative change to vary smoothly by age. This is done by including an interaction between a spline term in age and period, and the subsequently fishing out the relative change using a spline basis with a bit fewer knots to fish out the period multiplier.

It goes as follows, where we also as before extract the predicted values for comparison with the prevalence curves fitted separately for each year:

```
> ( kx.a <- c( 10, with( pr, quantile( rep(A,X), qn(5) ) ) ) )
      10% 30% 50% 70% 90%
   10  40  56  64  72  82
> CA <- Ns( 1:99, kn=kx.a, intercept=TRUE )
> A.chg <- NArray( list( A=1:99, c("Est","lo","hi"), sex=c("M","F") ) )
> for( sx in dimnames(pr.fit)[["sex"]] )
+    {
+ limod <- glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ) +
+                    I(P-2000):Ns( A, kn=kx.a, intercept=TRUE ),
```

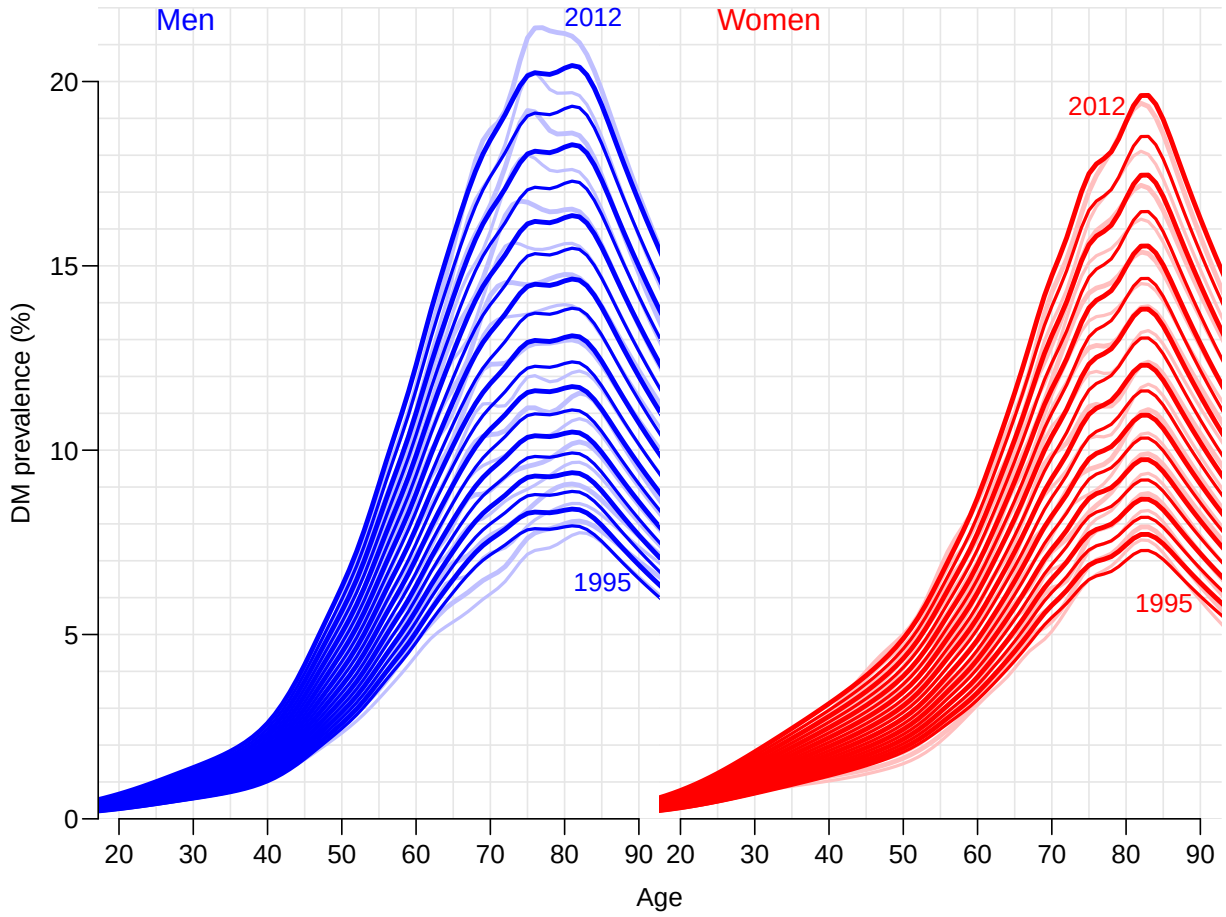

Figure 2.12: *Smoothed age-specific prevalences for the 17-year period 1995–2012 using a model with constant annual relative change in prevalences (full color) compared to the smoothing of the single years (pale color). Blue is men, red is women.*

```
+               family = binomial(link="log"),
+                 data = subset( pr, sex==sx ) )
+ A.chg[,,sx] <- ci.exp( limod, subset="P", ctr.mat=CA )
+ pr.lfit[sx,,] <- predict( limod,
+                           newdata = data.frame( A=rep(a.pt,     length(p.pt)),
+                                                 P=rep(p.pt,each=length(a.pt)) ),
+                              type = "response" )
+    }
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n" )
> matplot( 1:99, (cbind( A.chg[,,"M"], A.chg[,,"F"] )-1)*100,
+          col=rep(c("blue","red"),each=3), lwd=c(3,1,1), lty=1, type="l",
+          ylim=c(0,8), yaxs="i",
+          ylab="Annual change in DM prevalence (%)", xlab="Age" )
> abline( h=pr.chg[,1], col=c("blue","red") )
```

We can also as with the naïve linear change model show how the fitted values under this interaction model looks relative to the separate analyses by year (or full interaction model). The code is exactly as before, because we put the fitted values into the same structure as before:

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> lblu <- rgb( 3,3,4,max=4 )
```

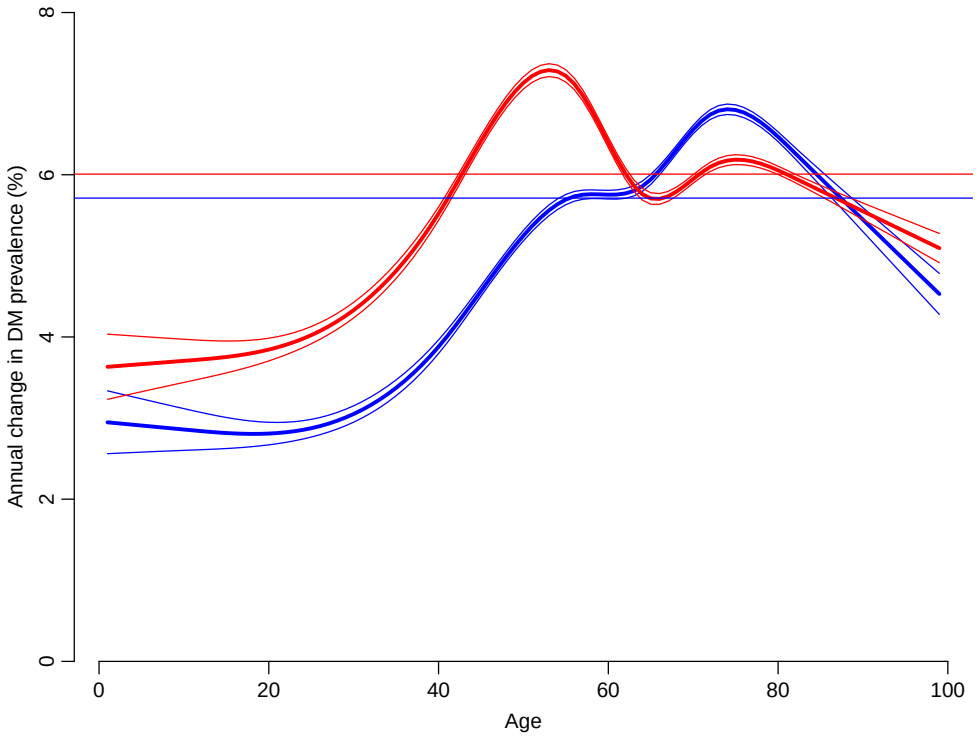

Figure 2.13: *The estimated change in prevalence in different ages, separately for men (blue) and women (red). The horizontal lines indicate the estimate from the naïve model with constant change for all ages.*

```
> lred <- rgb( 4,3,3,max=4 )
> matplot( a.pt, pr.fit["M",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="blue", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["M",,]*100, type="l", lty=1, col=lblu  , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["M",,]*100, type="l", lty=1, col="blue", lwd=c(2,3) )
> text( 25, 21.5, "Men", adj=0, col="blue", cex=1.2 )
> text( 89, pr.fit["M","89","1995"]* 99, "1995", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M","80","2012"]*101, "2012", col="blue", adj=c(0,0) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, pr.fit["F",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="red", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["F",,]*100, type="l", lty=1, col=lred , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["F",,]*100, type="l", lty=1, col="red", lwd=c(2,3) )
> text( 25, 21.5, "Women", adj=0, col="red", cex=1.2 )
> text( 89, pr.fit["F","89","1995"]* 99, "1995", col="red", adj=c(1,1) )
> text( 80, pr.fit["F","80","2012"]*101, "2012", col="red", adj=c(1,0) )
> axis( side=1 )
> mtext( "Age", side=1, line=1, outer=T )
> mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
```

From figure 2.14 it is seen that the model captures the actual pattern much better than the simple model with an annual change common across ages.

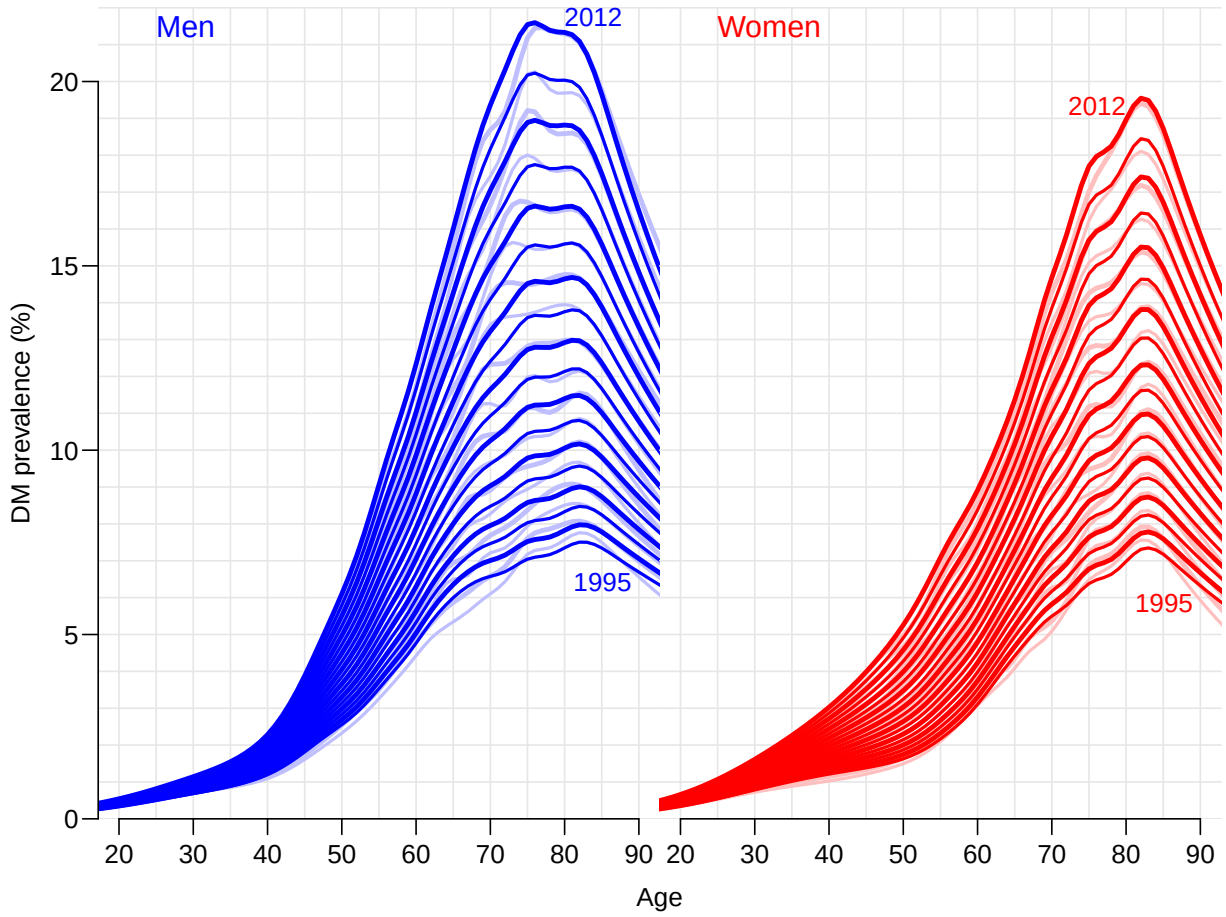

Figure 2.14: *Smoothed age-specific prevalences for the 17-year period 1995–2012 using a model with age-specific constant annual relative change in prevalences (full color) compared to the smoothing of the single years (pale color). Blue is men, red is women.*

2.5 Components of prevalence

The purpose of this chapter is to use the estimated transition rates to predict the prevalences at later (known) times.

This is in itself not an interesting endeavor, because we have the prevalence data available, but it will serve as an illustration that the rates are adequately modelled and that the degree of approximation is adequate when using a given interval length for probability calculations.

Specifically we address the problem of partitioning the changes in prevalence of diabetes in the Danish population over the last 17 years to:

1. changes in mortality rates among diabetes patients

2. changes in incidence rates of diabetes in the population

This measure will be sex- and age-specific, and hence independent of the demographic changes in the population.

2.5.1 Formalization

First we formalize the problem conceptually, then statistical, and finally outline the practical implementation based on analysis of rates.

2.5.1.1 Conceptual

The observed changes in prevalence of DM are a consequence of the changes in mortality and DM-incidence rates in the population and of the changes in the mortality rates in the DM population.

Of these the changes in population mortality presumably have the smaller role, but there is a connection, because they determine the available number of persons susceptible to a DM diagnosis.

Thus the starting point will be the population prevalence of DM as of 1.1.1995. The (age-specific) prevalence at any future point of time is obtained by applying the mortality rates in the two sub-strata of the population (DM / non-DM) and the DM-incidence rates to the non-DM part of the population.

The exercise consists in working out what the prevalence of diabetes would have been if:

1. mortality rates and diabetes rates had remained stable
2. only mortality rates had remained stable, but incidence rates had developed as observed
3. only incidence rates had remained stable, but mortality rates had developed as observed

The difference between observed prevalences and the predicted under scenario

1. 1 is the combined effect of changes in the rates as seen since the starting point chosen.
2. 2 is the effect of changing mortality rates alone. This could also be computed as the difference between scenarios 3 and 1.
3. 3 is the effect of changing incidence rates alone. This could also be computed as the difference between scenarios 2 and 1.

For the sake of completeness we shall compute both types of attribution of prevalences.

2.5.2 Statistical framework

First we consider the setup as outlined in figure 3.15:

```
> library( Epi )
> library( splines )
> tm <- matrix(NA,4,4)
> rownames(tm) <- colnames(tm) <- c("No DM","DM","Dead","Dead (DM)")
> tm[1,2] <- tm[1,3] <- tm[2,4] <- 1
> boxes( tm, boxpos = list( x=c(20,20,80,80),
+                           y=c(80,20,80,20) ),
+            wmult=1.3, hmult=4,
+            txt.arr = c( expression(lambda),
+                         expression(mu[W]),
+                         expression(mu[D][M]) ) )
```

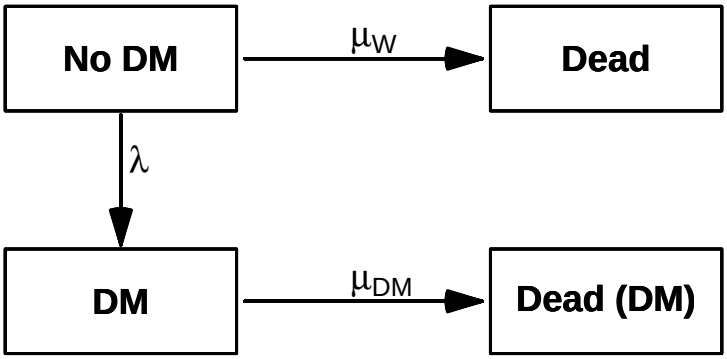

Figure 2.15: *The four states and transitions between them we consider*

The aim is to provide a precise formula for the age-specific prevalences at calendar time t, $p(a,t)$, given that we know the age-specific prevalence at some reference point t_0, $p(a,t_0)$ (in this case 1995), and the transition rates $\lambda(a,p)$, $\mu_{\rm W}(a,p)$ and $\mu_{\rm DM}(a,p)$.

We can in principle derive analytical expressions for this, but the easiest approach is to acquire parametric expressions for the transition rates and then update the age-specific prevalences by applying the transition probability matrix to a $A \times 2$ matrix of number of persons in each of the states no DM and DM.

For the given transition rates we can compute transition probabilities between states corresponding to a given (small) interval, δ, say, by first deriving the cumulative intensities for intervals of this length

$$\Lambda(a,p) = \lambda(a,p) \times \delta, \qquad M_{\rm W}(a,p) = \mu_{\rm W}(a,p) \times \delta, \qquad M_{\rm DM}(a,p) = \mu_{\rm DM}(a,p) \times \delta$$

and the the transition matrix $\mathbf{T}_{a,p}(\delta)$:

$$\mathbf{T}_{a,p}(\delta) = \begin{pmatrix} {\rm e}^{-\Lambda - M_{\rm W}} & \lambda {\rm e}^{-\Lambda - M_{\rm W}}\delta & \mu_{\rm W}{\rm e}^{-\Lambda - M_{\rm W}}\delta \\ 0 & {\rm e}^{-M_{\rm DM}} & \mu_{\rm DM}{\rm e}^{-\Lambda - M_{\rm DM}}\delta \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} {\rm e}^{-\Lambda - M_{\rm W}} & \Lambda {\rm e}^{-\Lambda - M_{\rm W}} & M_{\rm W}{\rm e}^{-\Lambda - M_{\rm W}} \\ 0 & {\rm e}^{-M_{\rm DM}} & M_{\rm DM}{\rm e}^{-\Lambda - M_{\rm DM}} \\ 0 & 0 & 1 \end{pmatrix}$$

So we see that the rates only enter via the cumulative rates over the intervals, so this is what we eventually must compute from models. For simplicity we left out the (a,p) qualification of all the terms in the expressions.

Now if we have the *number* of persons in age-class a and period p in states (W,DM,Dead) in the 3-vector $n(a,p)$ then:

$$n(a+\delta, p+\delta) = n(a,p)\mathbf{T}_{a,p}(\delta)$$

so updating the array of the number of persons in each state is merely a matter of matrix multiplication.

This updating machinery can be illustrated graphically in a Lexis diagram as in figure **??**:

```
> for( yy in 2000+0:3 )
+ for( aa in 40+0:3 )
+ {
+ pdf( paste("./graph/NDR-prup-",yy,"-",aa,".pdf", sep="" ),
```

```
+      height=7, width=7 )
+ par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
+ Lexis.diagram( age=40+c(-1,6), date=2000+c(-1,6), int=1 )
+ w <- 0.6
+ d <- 0.3
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1,1+w,NA,2,2+w), col="forestgreen", lwd=9, lend="butt", ljoin="bevel" )
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1+w,1+w+d,NA,2+w,2+w+d), col="red", lwd=9, lend="butt", ljoin="bevel" )
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1+w+d,2,NA,2+w+d,3), col="black", lwd=9, lend="butt", ljoin="bevel" )
+ for( an in 1:17 )
+ arrows( yy+1.1, aa+0.6, yy+1.9, aa+1.4, lwd=3, angle=an )
+ dev.off()
+ }
```

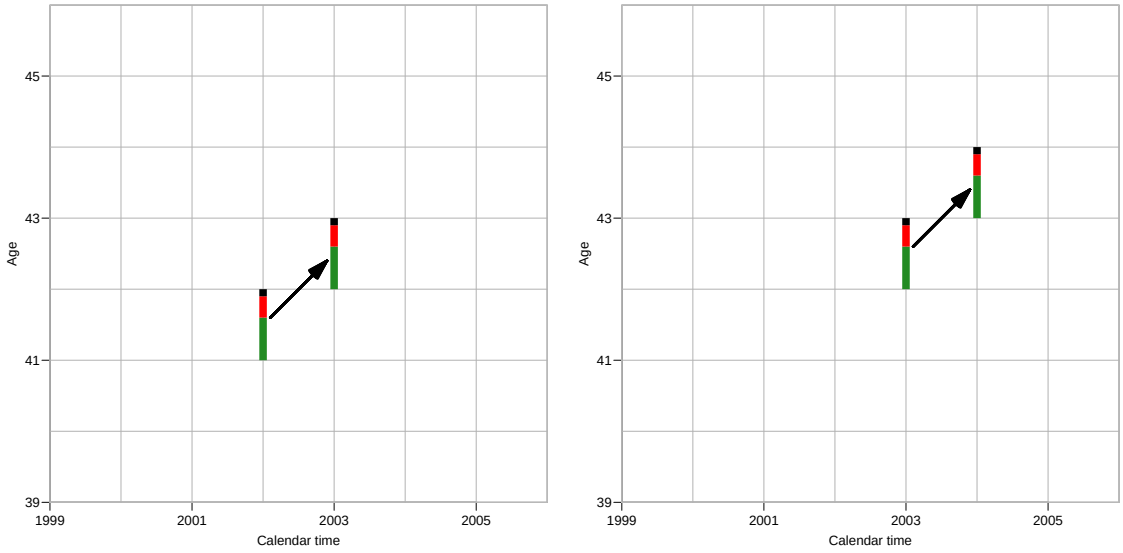

Figure 2.16: *Calculation of prevalences from one year to the next. Green are without diabetes, red with, and black dead.*

If we instead have the *fraction* of (living) persons in states (W,DM) in the vector $q(a,p)$ (which is now just a 2-vector) then:

$$\tilde{q}(a+\delta, p+\delta) = q(a,p)\mathbf{T}_{a,p}(\delta)[1:2,]$$

where we then will get the fraction of the persons in age a at time p who at time $p+\delta$ (and hence in age $a+\delta$) who are in states (W,DM,Dead). But since we are only interested in the progression of prevalences, then we instead use:

$$Q(a+\delta, p+\delta) = q(a,p)\mathbf{T}_{a,p}(\delta)[1:2, 1:2]$$
$$q(a+\delta, p+\delta) = Q(a+\delta, p+\delta) \Big/ \sum_{\text{W,DM}} Q(a+\delta, p+\delta)$$

so we update the prevalences at every step.

2.5.2.1 Births

Note that for every step in the updating we will lose estimates in an age-class; in order for this to work we need to feed in the number of births in each age-group with some assumption about the distribution between DM/non-DM; which we will assume is 0:1, that is we assume that no new-born diabetics enter.

2.5.3 Data for the calculations

We will use the models for the rates based on the 1-year data in Lexis triangles. There are two sets of models fitted to different datasets:

- Models for the prevalence of DM as a function of age. These will be based on a dataset with 1-year age-specific empirical prevalences, smoothed by a binomial model (with log-link), so producing a parametric age-prevalence curve for all combination of sex and dates 1 January 1995–2012.
- Models for rates, based on data for 1-year Lexis-triangles ($\triangledown$ and $\triangle$)
  - Incidence rates of DM among non-DM individuals
  - Mortality rates among non-DM individuals
  - Mortality rates among DM patients

  All data for these three sets of rates are in a single dataset.

The practical calculations will be based on quantities derived from these models. Calculations are made using intervals of length `int` as defined below, both in the age and the calendar time direction. The quantities that go into the calculations are:

1. Estimated prevalences at the midpoint of the age-intervals at 1.1.1995, as derived from the models for the prevalences.
2. Estimated incidence (DM) and mortality (non-DM, DM) rates evaluated at:
   (a) the midpoint of the updating periods, that is at times $1995 + n\texttt{int} + \texttt{int}/2, n = 0, \ldots$ and
   (b) the midpoint of the age at updating, that is updating age-class $(a, a + \texttt{int})$ to $(a + \texttt{int}, a + 2\texttt{int})$ we use the estimated rate at age $a + \texttt{int}$.

2.5.4 Prevalences

The observed prevalences and population size at the 1 January 1995–2012 available from a tabulation of the diabetes dome previously:

```
> load( file="./data/prev-o.Rda" )
> str( pr )
  'data.frame':        3600 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ A  : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P  : num  1995 1996 1997 1998 1999 ...
   $ X  : num  3 1 0 1 2 1 1 0 1 1 ...
   $ N  : num  35612 36055 34853 34774 34076 ...
```

```
> head( pr )
```

```
  sex A    P X     N
1   M 0 1995 3 35612
2   M 0 1996 1 36055
3   M 0 1997 0 34853
4   M 0 1998 1 34774
5   M 0 1999 2 34076
6   M 0 2000 1 33906
```

These are empirical prevalences (`X`—no. of cases of DM, `N`—population size) for each of the 18 dates 1.1.1995 – 1.1.2012 in 1-year intervals, but to get the machinery running we will need the prevalences as a continuous function of age.

So we model the prevalences as of 1 January each of the years 1995—2012, as a smooth function of age, with the intention of using the predicted prevalences at the midpoints each of the smaller age-classes that we use for the simulation.

So we collect the models for the prevalences So once we have set up the array to hold the smoothed empirical prevalences we can fill them into `pr.fit`; we use a log-link binomial model with a smooth spline with 15 knots.

```
> dnam <- list( sex = c("M","F"),
+                 t = sort(unique(pr$P)) )
> ( kp.a <- c( 10, with( pr, quantile( rep(A,X), (1:15-0.5)/15 ) ) ) )
```

```
           3.333333%       10% 16.66667% 23.33333%       30% 36.66667% 43.33333%
        10         28         40         47         52         56         59         62
       50% 56.66667% 63.33333%       70% 76.66667% 83.33333%       90% 96.66667%
        64         67         69         72         75         78         82         87
```

```
> pr.mod <- list()
>   length( pr.mod ) <- prod( sapply( dnam, length ) )
>      dim( pr.mod ) <-       sapply( dnam, length )
> dimnames( pr.mod ) <-              dnam
> for( dt in dimnames(pr.mod)[["t"]] )
+ for( sx in dimnames(pr.mod)[["sex"]] )
+ pr.mod[[sx,dt]] <- glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ),
+                         family = binomial(link="log"),
+                           data = subset( pr,
+                                          sex==sx & P==as.numeric(dt) ) )
```

For the calculations we shall only use the estimated prevalences as of 1.1.1995 as starting point for the simulation, that is from the models in `pr.mod[[sx,"1995"]]` for `sx` equal to either `M` or `F`.

2.5.4.1 Rates

First we the load the data for the models for incidence and mortality:

```
> load( file="./data/FU-o.Rda" )
> head( TT )
```

```
  sex A    P U     Y.nD      Y.DM D.DM D.nD X
1   F 0 1995 0 17025.50 0.0000000    0  137 0
2   F 0 1995 1 17100.54 0.1300479    0   16 2
3   F 0 1996 0 16468.06 1.4401095    0  134 4
4   F 0 1996 1 17067.30 1.8617385    0   23 4
5   F 0 1997 0 16434.00 0.0000000    0  152 0
6   F 0 1997 1 16499.84 1.9890486    0   14 2
```

```
> attr( TT, "Variables" )
```

```
      Data frame using the original definintion of DM from NDR
  sex  "Sex"
  A    "1-year age class"
  P    "1-year period"
  U    "Indicator of upper Lexis triangle"
  Y.nD "P-Y among non-diabetics"
  Y.DM "P-Y among diabetes patients"
  D.DM "Deaths among non-diabetics"
  D.nD "Deaths among diabetes patients"
  X    "Diabetes diagnoses among non-diabetics"
> DD <- transform( TT, A = A+(1+U)/3,
+                      P = P+(2-U)/3,
+                   D.nD = pmax(D.nD,0) )
> head( DD )
    sex         A        P U     Y.nD      Y.DM D.DM D.nD X
  1   F 0.3333333 1995.667 0 17025.50 0.0000000    0  137 0
  2   F 0.6666667 1995.333 1 17100.54 0.1300479    0   16 2
  3   F 0.3333333 1996.667 0 16468.06 1.4401095    0  134 4
  4   F 0.6666667 1996.333 1 17067.30 1.8617385    0   23 4
  5   F 0.3333333 1997.667 0 16434.00 0.0000000    0  152 0
  6   F 0.6666667 1997.333 1 16499.84 1.9890486    0   14 2
```

Then we can set up age-period-cohort models for the three types of rates of relevance; first we set up the knots for the period- and cohort-effects common for the three analyses, whereas we let the age-effect have knots depending on the position of the events on the age-scale:

```
> p.kn <- seq( 1996, 2011,, 5 )
> c.kn <- seq( 1900, 2010,, 8 )
```

Note that we name the vector of age-knots differently for the different models, because `predict.glm` apparently uses the global version of the knots vector and not the vector stored in the `glm` object.

2.5.4.1.1 Incidence rates

Here is the age-period cohort model for the rates of DM occurrence, using `(X,Y.nD)` as outcome variables:

```
> ( ai.kn <- with( DD, c(5,10,quantile( rep(A,X), probs=(1:10-0.5)/10 ) ) ) )
                          5%      15%      25%      35%      45%      55%
   5.00000 10.00000 31.66667 45.66667 52.33333 56.66667 60.66667 64.66667
       65%      75%      85%      95%
  68.33333 72.66667 77.66667 84.33333
> incM <- glm( X ~ Ns(   A, kn=ai.kn ) +
+                  Ns( P  , kn=p.kn ) +
+                  Ns( P-A, kn=c.kn ),
+              offset = log(Y.nD),
+              family = poisson,
+                data = subset(DD,sex=="M") )
> incF <- update( incM, data = subset(DD,sex=="F") )
```

2.5.4.1.2 Non-DM mortality rates

Here is the age-period cohort model for the mortality rates among non-diabetics, using `(D.nD,Y.nD)` as outcome variables:

```
> ( and.kn <- with( DD, c(5,15,quantile( rep(A,D.nD), probs=(1:10-0.5)/10 ) ) ) )
                          5%      15%      25%      35%      45%      55%
   5.00000 15.00000 45.66667 60.33333 67.33333 72.66667 76.66667 80.33333
       65%      75%      85%      95%
  83.33333 86.33333 89.66667 93.66667
```

```
> mndM <- glm( D.nD ~ Ns(   A, kn=and.kn ) +
+                     Ns( P  , kn=p.kn ) +
+                     Ns( P-A, kn=c.kn ),
+              offset = log(Y.nD),
+              family=poisson,
+              data = subset(DD,sex=="M") )
> mndF <- update( mndM, data = subset(DD,sex=="F") )
```

2.5.4.1.3 DM mortality rates Here is the age-period cohort model for the mortality rates among diabetes patients, using `(D.DM,Y.DM)` as outcome variables:

```
> ( adm.kn <- with( DD, c(25,quantile( rep(A,D.DM), probs=(1:11-0.5)/11 ) ) ) )
            4.545455% 13.63636% 22.72727% 31.81818% 40.90909%       50% 59.09091%
   25.00000  54.33333  63.66667  68.66667  72.66667  75.66667  78.33333  80.66667
  68.18182% 77.27273% 86.36364% 95.45455%
   83.33333  85.66667  88.66667  92.66667
> mdmM <- glm( D.DM ~ Ns(   A, kn=adm.kn ) +
+                     Ns( P  , kn=p.kn ) +
+                     Ns( P-A, kn=c.kn ),
+              offset = log(Y.DM),
+              family=poisson,
+              data = subset( DD, sex=="M" & Y.DM>0 ) )
> mdmF <- update( mdmM, data = subset(DD,sex=="F" & Y.DM>0) )
```

2.5.5 Implementation of prevalence calculations

We start by specifying the interval length for the updating, and then the points at which we want to predict. The transition rates are labeled by the midpoints of the Lexis squares (of width `int`) where we predict them (`a.pt` and `p.pt`), and the prevalences by the mid-points of the age-classes (`a.pt` and the time points `t.pt`

```
> int <- 0.5
> a.pt <- seq(int,100,int) - int/2
> t.pt <- seq(1995,2012,int)
> p.pt <- t.pt[-1] - int/2
```

All the predictions should be in units of the interval length chosen for calculations. We note from the calculations above that the quantities that enter the expressions for the transition probabilities are all cumulative rates over the intervals. Thus we use a prediction data frame with the person-years-variables set to `int`, and we use predicted rates at the period midpoints (`p.pt`), but we use the age-point at the *upper end* of the age-class, because we will be using the cumulative rates to predict transitions from the age-class $(a, a+\delta)$ to $(a+\delta, a+2\delta)$:

```
> nd <- data.frame( A = rep(a.pt+int/2,     length(p.pt)),
+                   P = rep(p.pt      ,each=length(a.pt)),
+                   Y.nD = int,
+                   Y.DM = int )
> str( nd )
  'data.frame':        6800 obs. of  4 variables:
   $ A   : num  0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 ...
   $ P   : num  1995 1995 1995 1995 1995 ...
   $ Y.nD: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
   $ Y.DM: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
> head( nd )
```

```
     A       P Y.nD Y.DM
1 0.5 1995.25  0.5  0.5
2 1.0 1995.25  0.5  0.5
3 1.5 1995.25  0.5  0.5
4 2.0 1995.25  0.5  0.5
5 2.5 1995.25  0.5  0.5
6 3.0 1995.25  0.5  0.5
```

```
> summary( nd )
```

```
       A                P             Y.nD          Y.DM
 Min.   :  0.50   Min.   :1995   Min.   :0.5   Min.   :0.5
 1st Qu.: 25.38   1st Qu.:1999   1st Qu.:0.5   1st Qu.:0.5
 Median : 50.25   Median :2004   Median :0.5   Median :0.5
 Mean   : 50.25   Mean   :2004   Mean   :0.5   Mean   :0.5
 3rd Qu.: 75.12   3rd Qu.:2008   3rd Qu.:0.5   3rd Qu.:0.5
 Max.   :100.00   Max.   :2012   Max.   :0.5   Max.   :0.5
```

2.5.5.1 Transition probabilities

We shall use the recursive scheme to predict the course of DM prevalence development in the population under various scenarios of mortality and incidence development. So we use the various structures to hold results and clarify calculations:

`Lambda,Mu.nD,Mu.DM` — arrays of cumulative rates over intervals of length `int`, evaluated at dates at the midpoint of calculation intervals, and at borders of age-intervals, corresponding to midpoints of C-sets of the Lexis diagram (⧄).

`pr.fit` — array of empirical prevalences at 1.1.1995–1.1.2012, smoothed by natural splines separately for each year.

`TR` — array of transition probabilities between states no DM, DM and Dead. Transition probabilities are computed under 4 different scenarios combining mortality and incidence rates either as they actually developed 1995–2012 or assuming they were constant at the 1995 level. These refer to intervals of length `int` years and are therefore labeled on the period dimension by the midpoint of these, a total of 17/`int`.

`prv` — array of predicted prevalences based on the initial prevalences at 1.1.1995 and the transition probabilities as put in `TR`. The scenario dimension refers to the 4 scenarios: "obs", "m-fix", "i-fix" and "all-f", but this dimension in the array is expanded by 3 extra levels "mort", "inc" and "const" that are to be filled with the part of the prevalences that are attributable to decrease in mortality, increase in incidence and the disequilibrium between rates and prevalence in 1995. Likewise the period dimension is expanded by one relative to that in `TR`, since this refer to points in time and not time intervals.

`prn` — array of predicted *number* of DM patients in one-year age classes at the 1 January each year. So the same structure as `prv`, but with substantially fewer entries along the age and period dimensions.

Thus, first we set up the arrays of the cumulative rates (note that the ages are at the midpoint of age-classes):

```
> Lambda <-
+ Mu.nD <-
+ Mu.DM <- NArray( list( a = a.pt,
+                        p = p.pt,
+                      sex = c("M","F") ) )
```

In order to compute the transition probabilities we need the cumulative incidences over intervals of length `int`. So first we predict these using the relevant points. Note that the person-years-variables are set to `int` in order to get cumulative rates over an interval of this length. Note that the compute fitted rates at `int/2` to the right of the labeling of the age-interval:

```
> nd <- data.frame( A = rep(a.pt+int/2,     length(p.pt)),
+                   P = rep(p.pt      ,each=length(a.pt)),
+                   Y.nD = int,
+                   Y.DM = int )
```

With this prediction frame in place we compute the cumulative rates:

```
> Lambda[,,"M"] <- predict.glm( incM, type="response", newdata=nd )
> Lambda[,,"F"] <- predict.glm( incF, type="response", newdata=nd )
>  Mu.nD[,,"M"] <- predict.glm( mndM, type="response", newdata=nd )
>  Mu.nD[,,"F"] <- predict.glm( mndF, type="response", newdata=nd )
>  Mu.DM[,,"M"] <- predict.glm( mdmM, type="response", newdata=nd )
>  Mu.DM[,,"F"] <- predict.glm( mdmF, type="response", newdata=nd )
```

Note that we get warning messages originating from the overparametrization of the age-period-cohort model.

In order to get the predicted prevalences by age, period and prediction type, we need the (1-step) transition matrices at all combinations of age (a) and date (p), this is put in array:

```
> states <- c("no DM","DM")
> TR <- NArray( c( dimnames(Lambda),
+                  list( from = states,
+                          to = states,
+                       scene = c("obs","m-fix","i-fix","all-f" ) ) ) )
> str( TR )
   logi [1:200, 1:34, 1:2, 1:2, 1:2, 1:4] NA NA NA NA NA NA ...
   - attr(*, "dimnames")=List of 6
    ..$ a    : chr [1:200] "0.25" "0.75" "1.25" "1.75" ...
    ..$ p    : chr [1:34] "1995.25" "1995.75" "1996.25" "1996.75" ...
    ..$ sex  : chr [1:2] "M" "F"
    ..$ from : chr [1:2] "no DM" "DM"
    ..$ to   : chr [1:2] "no DM" "DM"
    ..$ scene: chr [1:4] "obs" "m-fix" "i-fix" "all-f"
> prod( dim(TR) )
  [1] 217600
```

So we can now compute the one-`int`-step transition matrices for every combination of `a.pt` and `p.pt`, both in steps of `int` (in this case 0.5 year):

```
> TR[,,,"no DM","no DM","obs"] <- exp(-Lambda-Mu.nD)
> TR[,,,"no DM","DM"   ,"obs"] <- exp(-Lambda-Mu.nD)*Lambda
> TR[,,,"DM"   ,"no DM","obs"] <- 0
> TR[,,,"DM"   ,"DM"   ,"obs"] <- exp(-Mu.DM)
```

Note that we have not included the "Dead" state in the calculations, because we only bother about the fraction of diabetes patients in each age class at each time point. So the probabilities we compute do not sum to 1 within the "from" states.

The situation where both the mortality rates and incidence rates are fixed at the 1995 level is trivial, because transition probabilities in that case only depend on age and not on period.

When we fix the mortality or incidence at the 1995 level we just replace the expressions above with expressions where we replace the date dimension by `rep(1,np)`, where `np` is the number of periods:

```
> np <- dim(Lambda)["p"]
> TR[,,"no DM","no DM","m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])
> TR[,,"no DM","DM"   ,"m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])*Lambda
> TR[,,"DM"   ,"no DM","m-fix"] <- 0
> TR[,,"DM"   ,"DM"   ,"m-fix"] <- exp(-Mu.DM[,rep(1,np),])


> TR[,,"no DM","no DM","i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)
> TR[,,"no DM","DM"   ,"i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)*Lambda[,rep(1,np),]
> TR[,,"DM"   ,"no DM","i-fix"] <- 0
> TR[,,"DM"   ,"DM"   ,"i-fix"] <- exp(-Mu.DM)


> TR[,,"no DM","no DM","all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])
> TR[,,"no DM","DM"   ,"all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])*Lambda[,rep(1,np),]
> TR[,,"DM"   ,"no DM","all-f"] <- 0
> TR[,,"DM"   ,"DM"   ,"all-f"] <- exp(-Mu.DM[,rep(1,np),])
```

We have now collected the transition probabilities between "no DM" and "DM" as well as the probabilities of remaining in each of these, all referring to a duration of `int`.

2.5.5.2 Prediction of the observed prevalences

Note that we do not need to predict the population size; we only predict the prevalences as fractions. When we multiply the fraction of persons in states (no DM,DM) with the transition matrix, we get fraction of the persons in the previous state that are in states (no DM,DM), which does not sum to 1 (because of the dead ones), so we must rescale to prevalence age in each step.

When we do the predictions we need a starting point (and comparison points) for we predict the age-specific prevalences at 1 January each year at the midpoint of the age-intervals of length `int`, as stored in `a.pt`:

```
> pr.fit <- NArray( c( dimnames(Lambda)[c("a","sex")],
+                      dimnames(pr.mod)["t"] ) )
> for( sx in dimnames(pr.fit)[["sex"]] )
+ for( dt in dimnames(pr.fit)[["t"]] )
+ pr.fit[,sx,dt] <- predict( pr.mod[[sx,dt]],
+                            newdata = data.frame( A=as.numeric(dimnames(pr.fit)[["a"]]) ),
+                               type = "response" )
```

Then we set up an array to hold the predicted prevalences under different scenarios:

```
> dpr <- c( dimnames(Lambda)[c("a","p","sex")],
+           list( c(dimnames(TR)[["scene"]],"mort","inc","const") ) )
> names( dpr )[c(2,4)] <- c("t","what")
> dpr[["t"]] <- t.pt
> prv <- NArray( dpr )
```

To get the calculations started we insert the estimated prevalences at 1995 and assume the all newborns are without diabetes, that is the prevalence is 0 at age 0 (or rather at age `int/2`):

```
> ### Smoothed prevalences at 1.1.1995 - the starting values
> prv[,1,,] <- pr.fit[,,1]
> ### Prevalences at age 0 are set to 0
> prv[1,,,] <- 0
```

Then we can finally compute the prevalences at the desired points of the Lexis diagram:

```
> for( ip in 1:(dim(prv)["t"]-1) )
+ for( ia in 1:(dim(prv)["a"]-1) )
+ prv[ia+1,ip+1,,1:4] <-
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,] ) /
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","no DM",] )
```

Later we shall also compute the fraction of the prevalences that are attributable to trends in mortality and incidence as well as to the non-stationarity of the rates/prevalences as of 1995, so we put in three extra levels of the last dimension, and one extra levels of the period dimension because we want to predict to the end of the last period too (or, to put it differently, we need an extra first level to hold the starting prevalences as of 1.1.1995).

2.5.5.3 A function for the calculations

We now pack the previous into a function, `prcalc`, which takes the interval length (and the ending year) as arguments, and assumes that the smoothed prevalences (`pr.mod` as 2-dimensional list) and smoothed rates (`incM`, `incF`, `mndM`, `mndF`, `mdmM`, `mdmF`) are available in the workspace:

```
> prcalc <-
+ function( int=1, end=2012 )
+ {
+ # OBS: Assumes that the fitted prevalences pr.fit as well as the
+ # fitted models for rates, incM, incF, mndM, mndF, mdmM, mdmF are in
+ # the workspace
+ a.pt <- seq(int,100,int) - int/2
+ t.pt <- seq(1995,end,int)
+ p.pt <- t.pt[-1] - int/2
+ ### Prediction data frame
+ nd <- data.frame( A = rep(a.pt+int/2,     length(p.pt)),
+                   P = rep(p.pt      ,each=length(a.pt)),
+                   Y.nD = int,
+                   Y.DM = int )
+ ### Arrys to hold the rates at the relevant points, note that a.pt is
+ ### the first dimension, and p.pt the second so that predictions using
+ ### newdata=nd can be immediately put in the array, using the
+ ### column-major convention:
+ Lambda <-
+ Mu.nD <-
+ Mu.DM <- NArray( list( a = a.pt,
+                        p = p.pt,
+                      sex = c("M","F") ) )
+ ### Compute the cumulative rates over an interval
+ options( warn = -1 )
+ Lambda[,,"M"] <- predict.glm( incM, type="response", newdata=nd )
+ Lambda[,,"F"] <- predict.glm( incF, type="response", newdata=nd )
+  Mu.nD[,,"M"] <- predict.glm( mndM, type="response", newdata=nd )
+  Mu.nD[,,"F"] <- predict.glm( mndF, type="response", newdata=nd )
+  Mu.DM[,,"M"] <- predict.glm( mdmM, type="response", newdata=nd )
+  Mu.DM[,,"F"] <- predict.glm( mdmF, type="response", newdata=nd )
+ options( warn = 0 )
+ ### The fitted prevalences at ages a.pt but only at 1 Jan each year
```

```
+ pr.fit <- NArray( c( dimnames(Lambda)["a"],
+                       dimnames(pr.mod)[c("sex","t")] ) )
+ for( sx in dimnames(pr.fit)[["sex"]] )
+ for( dt in dimnames(pr.fit)[["t"]] )
+ pr.fit[,sx,dt] <- predict( pr.mod[[sx,dt]],
+                            newdata = data.frame( A=as.numeric(dimnames(pr.fit)[["a"]]) ),
+                               type = "response" )
+ ### Transition probabilities under various scenarios
+ states <- c("no DM","DM")
+ TR <- NArray( c( dimnames(Lambda),
+                  list( from = states,
+                          to = states,
+                       scene = c("obs","m-fix","i-fix","all-f" ) ) ) )
+ ### No of levels of the period-dimension
+ np <- dim(Lambda)[2]
+ ### Using observed rates throughout
+ TR[,,,"no DM","no DM","obs"  ] <- exp(-Lambda-Mu.nD)
+ TR[,,,"no DM","DM"   ,"obs"  ] <- exp(-Lambda-Mu.nD)*Lambda
+ TR[,,,"DM"   ,"no DM","obs"  ] <- 0
+ TR[,,,"DM"   ,"DM"   ,"obs"  ] <- exp(-Mu.DM)
+ ### Mortality rates fixed
+ TR[,,,"no DM","no DM","m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])
+ TR[,,,"no DM","DM"   ,"m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])*Lambda
+ TR[,,,"DM"   ,"no DM","m-fix"] <- 0
+ TR[,,,"DM"   ,"DM"   ,"m-fix"] <- exp(-Mu.DM[,rep(1,np),])
+ ### Incidence rates fixed
+ TR[,,,"no DM","no DM","i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)
+ TR[,,,"no DM","DM"   ,"i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)*Lambda[,rep(1,np),]
+ TR[,,,"DM"   ,"no DM","i-fix"] <- 0
+ TR[,,,"DM"   ,"DM"   ,"i-fix"] <- exp(-Mu.DM)
+ ### All rates fixed
+ TR[,,,"no DM","no DM","all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])
+ TR[,,,"no DM","DM"   ,"all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])*Lambda[,rep(1,np),]
+ TR[,,,"DM"   ,"no DM","all-f"] <- 0
+ TR[,,,"DM"   ,"DM"   ,"all-f"] <- exp(-Mu.DM[,rep(1,np),])
+ ### Array to hold the predicted prevalences
+ dpr <- c( dimnames(Lambda)[1:3],
+           list( c(dimnames(TR)[["scene"]],"mort","inc","const") ) )
+ names( dpr )[c(2,4)] <- c("t","what")
+ dpr[["t"]] <- t.pt
+ prv <- NArray( dpr )
+ ### Smoothed prevalences at 1.1.1995 - the starting values
+ prv[,1,,] <- pr.fit[,,1]
+ ### Prevalences at age 0 are set to 0
+ prv[1,,,] <- 0
+ ### Compute the prevalences
+ for( ip in 1:(dim(prv)["t"]-1) )
+ for( ia in 1:(dim(prv)["a"]-1) )
+ prv[ia+1,ip+1,,1:4] <-
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,] ) /
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","no DM",] )
+ ### ...and return them together with the observed
+ list( prv=prv, pr.fit=pr.fit )
+ }
```

Note in the last bit of the function definition that the reason that the last dimension, `scene`, is explicitly mentioned in the array `prv` is because this has dimension 7, but in `TR` only 4 — remember that `prv` also has thee extra levels to provide for the estimated part of the prevalences attributable to mortality change, incidence changes, and non-equilibrium at 1995.

2.5.5.4 Length of the calculation interval

In order to check whether the prediction using an interval length of 0.50 year is necessary we repeat the exercise using a 2-year interval for comparison

```
> system.time( prv  <- prcalc( int=0.1 ) )
     user  system elapsed
    19.81    1.52   21.33

> system.time( prvh <- prcalc( int=0.5 ) )

     user  system elapsed
     0.84    0.06    0.91

> system.time( prv1 <- prcalc( int=1.0 ) )

     user  system elapsed
     0.32    0.00    0.31

> system.time( prv2 <- prcalc( int=2.0 ) )

     user  system elapsed
     0.17    0.00    0.17
```

With these predictions in place we can now check whether we have made a reasonable approximation to the observed prevalences at 1.1.2012, and to which extent the calculation-interval influences this:

In the array `prv` are all the prevalences as predicted from the prevalence in 1995 using the estimated incidences and mortalities; predicted at intervals of `inc` whereas we have the smoothed empirical prevalences at 1 January 1995,...2012 in the array `pr.fit`:

```
> a.p2 <- as.numeric( dimnames(prv2$prv)[["a"]] )
> a.p1 <- as.numeric( dimnames(prv1$prv)[["a"]] )
> a.ph <- as.numeric( dimnames(prvh$prv)[["a"]] )
> a.pt <- as.numeric( dimnames(prv$prv )[["a"]] )
> wh <- c("1999","2005","2011")
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> matplot( a.pt, prv$pr.fit[,"M",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="blue", lty=1, lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"M","obs"]*100, lty="11", lwd=2, col="blue" )
> # matlines( a.ph, prvh$prv[,wh,"M","obs"]*100, lty="13", lwd=2, col="blue" )
> # matlines( a.p1, prv1$prv[,wh,"M","obs"]*100, lty="14", lwd=2, col="blue" )
> matlines( a.p2, prv2$prv[,wh,"M","obs"]*100, lty="22", lwd=2, col="blue" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> axis( side=1, at=1:9*10, labels=rep("",9) )
> axis( side=2 )
> matplot( a.pt, prv$pr.fit[,"F",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="red", lty=1, lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"F","obs"]*100, lty="11", lwd=2, col="red" )
> # matlines( a.ph, prvh$prv[,wh,"F","obs"]*100, lty="13", lwd=2, col="red" )
> # matlines( a.p1, prv1$prv[,wh,"F","obs"]*100, lty="14", lwd=2, col="red" )
> matlines( a.p2, prv2$prv[,wh,"F","obs"]*100, lty="22", lwd=2, col="red" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> axis( side=1, at=1:9*10, labels=rep("",9) )
> mtext( "Prevalence of DM (%)", side=2, line=2, las=0, outer=TRUE )
> mtext( "Age", side=1, line=2, las=0, outer=TRUE )
```

For presentation purposes we also just compare the observed and the predicted:

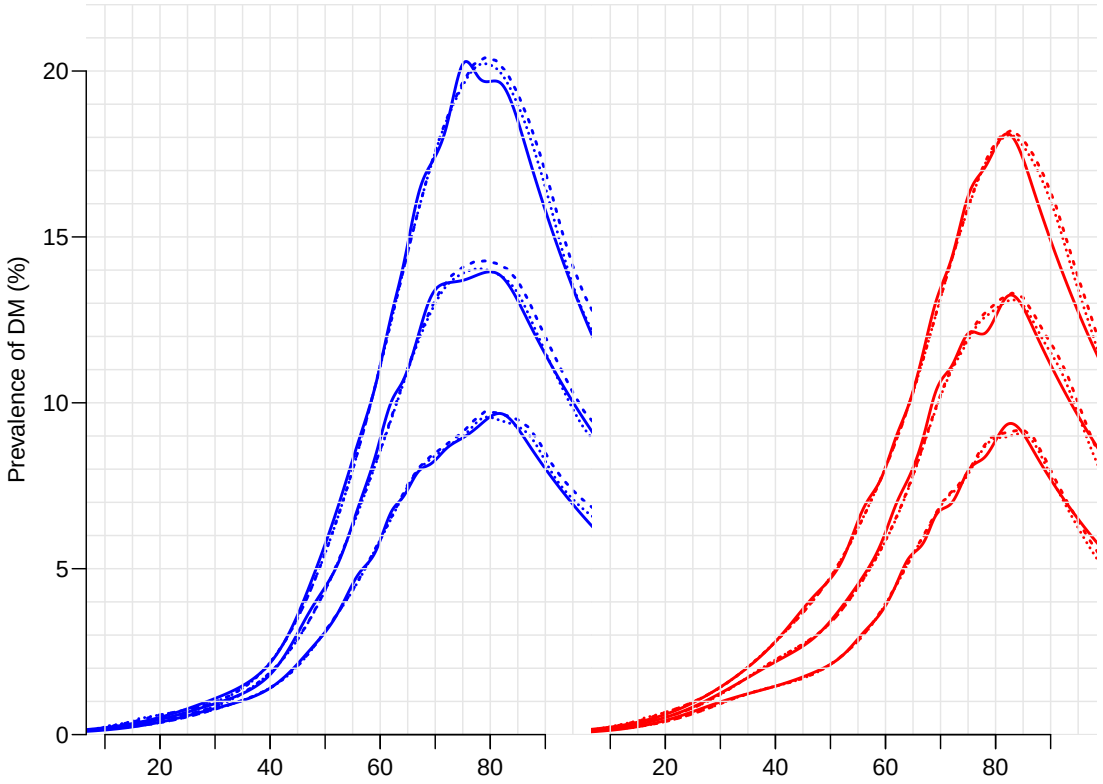

Figure 2.17: *Predicted prevalences at 1 January 1999, 2005 and 2011. Full (thin) lines: Smoothed empirical prevalences. Broken lines: Prediction using calculation intervals of 0.1 and 2 years (from dotted / broken).*

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> matplot( a.pt, prv$pr.fit[,"M",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="blue", lty=rep(1:2,each=3), lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"M","obs"]*100, lty="12", lwd=2, col="blue" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, prv$pr.fit[,"F",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="red", lty=rep(1:2,each=3), lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"F","obs"]*100, lty="12", lwd=2, col="red" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> mtext( "Prevalence of DM (%)", side=2, line=2, las=0, outer=TRUE )
> mtext( "Age", side=1, line=2, las=0, outer=TRUE )
```

2.5.6 Prevalences under different scenarios

We now compare the predicted prevalences under the four scenarios at 1.1.2012:

```
> np <- dim( prv$prv )[2]
> par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
> matplot( a.pt, cbind(prv$prv[,np,"M",],prv$prv[,1,"M",1])*100,
```

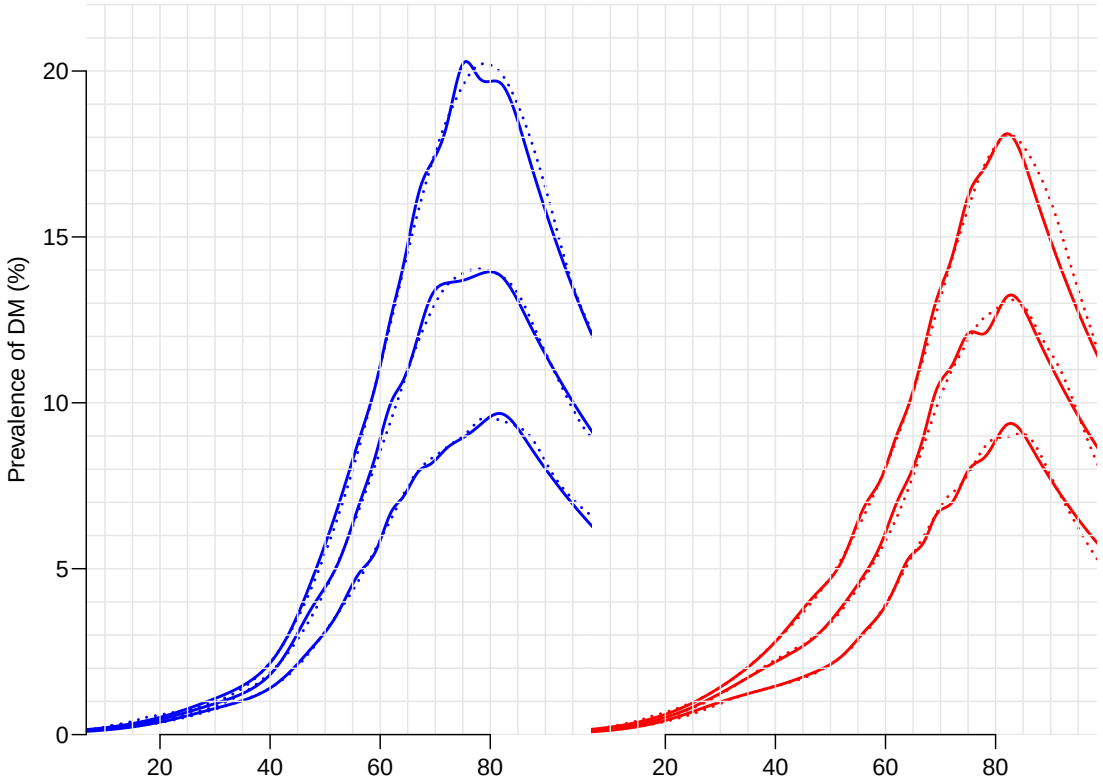

Figure 2.18: *Predicted prevalences at 1 January 1999, 2005 and 2011. Full (thin) lines: Smoothed empirical prevalences. Broken lines: Prediction using a calculation interval of 0.1 year.*

```
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
> matlines( a.pt, prv$prv[,np,"M",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="blue" )
> matlines( a.pt, prv$prv[,1,"M",]*100, type="l", lty=1, lwd=1, col="blue" )
> matplot( a.pt, cbind(prv$prv[,np,"F",],prv$prv[,1,"F",1])*100, yaxt="n",
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
> matlines( a.pt, prv$prv[,np,"F",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="red" )
> matlines( a.pt, prv$prv[,1,"F",]*100, type="l", lty=1, lwd=1, col="red" )
```

Here is a more elaborate graph, mainly for presentation purposes:

```
> scen <- c("Mort obs, Inc obs","Mort 1995, Inc obs","Mort obs, Inc 1995","Mort 1995, Inc 1995")
> c.a <- dimnames(prv$prv)[[1]][floor(dim(prv$prv)[1]/1.5)]
> n.a <- as.numeric( c.a )
> nt <- dim( prv$prv )[2]
> hts <- prv$prv[c.a,nt,"M",1:4]*100
> cau.exp <-
+ function( wh=1:4, fill=FALSE )
+ {
+ pdf( paste( "./graph/NDR-", paste(wh,collapse=""), if( fill ) "F",
+             "-o.pdf", sep="" ), height=8, width=11 )
+ par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
+ matplot( a.pt, cbind(prv$prv[,nt,"M",],prv$prv[,1,"M",1])*100, yaxs="i",
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="Prevalence (%)",
```

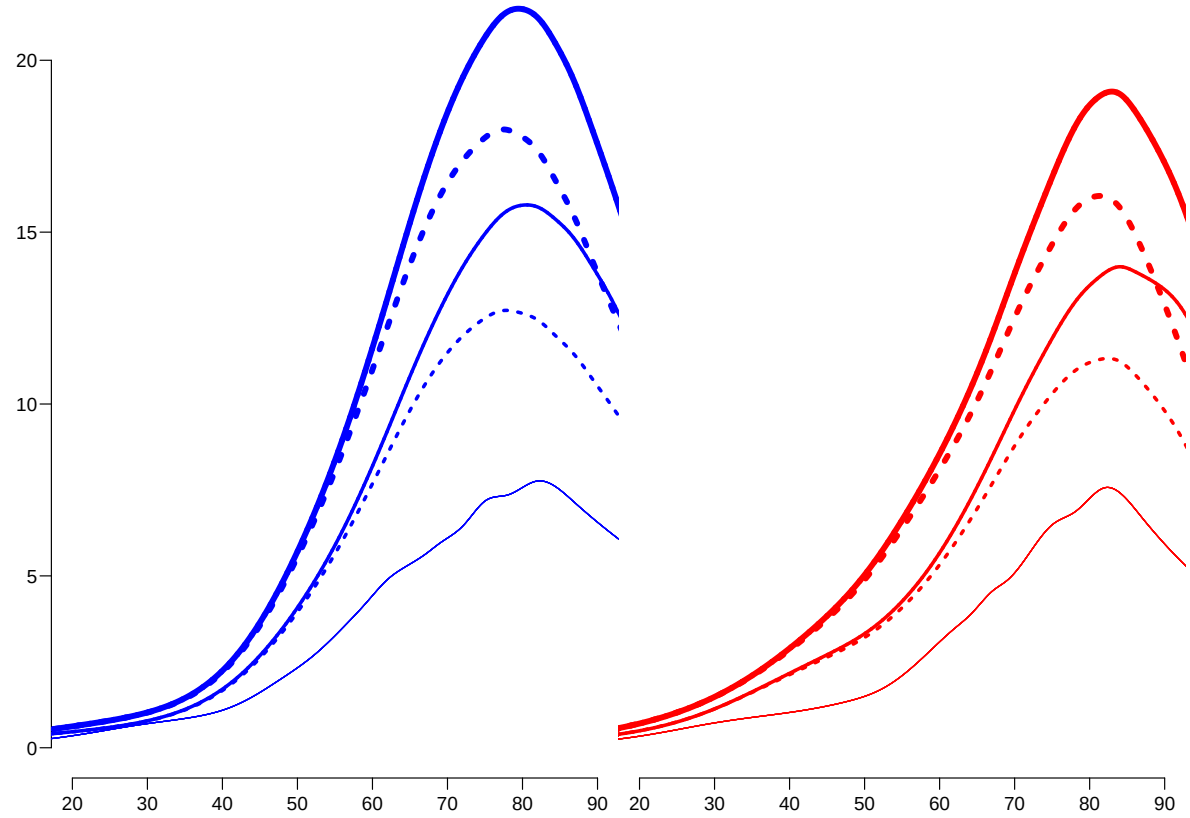

Figure 2.19: *The predicted prevalences at 1.1.2012 under different scenarios:*
Full lines: Mortality rates evolve as observed, Broken lines: Mortality rates remain as 1995.
Thick lines: Incidence rates evolve as observed, Thin lines: Incidence rates remain as in 1995.
The very thin lines lowest in the two displays are the observed prevalences in 1995.

```
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
+ matlines( a.pt, prv$prv[,nt,"M",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="blue" )
+ matlines( a.pt, prv$prv[,1,"M",]*100, type="l", lty=1, lwd=1, col="blue" )
+ mtext( "Age-specific DM prevalence (%)", side=2, line=2, las=0 )
+ text( rep(20,4)[wh], hts[wh], scen[wh], adj=0, col="blue", cex=1.2 )
+ for( i in 1:15 )
+ arrows( (20.20+strwidth(scen,cex=1.2))[wh], hts[wh], rep(n.a,4)[wh], hts[wh], col="blue",
+         angle=i, lwd=2 )
+ if( fill ) polygon( c(a.pt,rev(a.pt)),
+                     c(prv$prv[,nt,"M",wh[1]],rev(prv$prv[,nt,"M",wh[2]]))*100,
+                     col=rgb(0,0,1,0.3), border="transparent" )
+ matplot( a.pt, cbind(prv$prv[,nt,"F",],prv$prv[,1,"F",1])*100, yaxs="i",
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="", yaxt="n",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
+ matlines( a.pt, prv$prv[,nt,"F",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="red" )
+ matlines( a.pt, prv$prv[,1,"F",]*100, type="l", lty=1, lwd=1, col="red" )
+ if( fill ) polygon( c(a.pt,rev(a.pt)),
+                     c(prv$prv[,nt,"F",wh[1]],rev(prv$prv[,nt,"F",wh[2]]))*100,
+                     col=rgb(1,0,0,0.3), border="transparent")
+ dev.off()
+ }
```

```
> cau.exp(1:4)
  null device
            1
> for( ff in c(FALSE,TRUE) )
+    {
+ cau.exp(1:2,fill=ff)
+ cau.exp(3:4,fill=ff)
+ cau.exp(c(1,3),fill=ff)
+ cau.exp(c(2,4),fill=ff)
+    }
```

Figure 2.19 shows the predicted prevalences under 4 different scenarios compared to the observed prevalences as of 1.1.1995.

2.5.6.1 How much is attributable to what?

We can compute how much of the age-specific prevalences that are attributable to mortality changes and how much to changes in incidence rates.

The effect of mortality decline can be computed either as the difference between "obs" and "m-fix" or as the difference between "i-fix" and "all-f". But there is no guarantee that these two quantities are the same.

Similarly the effect of incidence increase can be computed either as the difference between "obs" and "i-fix" or as the difference between "m-fix" and "all-f". And there is no guarantee that these two are the same either.

Hence we explore how different these quantities are:

```
> dimnames( prv$prv )[4]
  $what
  [1] "obs"   "m-fix" "i-fix" "all-f" "mort"  "inc"   "const"
> par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
> matplot( a.pt, cbind( prv$prv[,nt,"M","obs"  ]-prv$prv[,nt,"M","m-fix"],
+                       prv$prv[,nt,"M","i-fix"]-prv$prv[,nt,"M","all-f"] )*100,
+          xlim=c(20,90), ylim=c(0,6), xlab="", ylab="",
+          type="l", lty=1, lwd=c(4,2)+1, col="blue" )
> matlines(a.pt, cbind( prv$prv[,nt,"M","obs"  ]-prv$prv[,nt,"M","i-fix"],
+                       prv$prv[,nt,"M","m-fix"]-prv$prv[,nt,"M","all-f"] )*100,
+          type="l", lty="22", lwd=c(4,2)+1, col="blue" )
> matplot( a.pt, cbind( prv$prv[,nt,"F","obs"  ]-prv$prv[,nt,"F","m-fix"],
+                       prv$prv[,nt,"F","i-fix"]-prv$prv[,nt,"F","all-f"] )*100,
+          xlim=c(20,90), ylim=c(0,6), xlab="", ylab="", yaxt="n",
+          type="l", lty=1, lwd=c(4,2)+1, col="red" )
> matlines(a.pt, cbind( prv$prv[,nt,"F","obs"  ]-prv$prv[,nt,"F","i-fix"],
+                       prv$prv[,nt,"F","m-fix"]-prv$prv[,nt,"F","all-f"] )*100,
+          type="l", lty="22", lwd=c(4,2)+1, col="red" )
> mtext( "Contribution to prevalence (%)", side=2, outer=TRUE, line=1.5, las=0)
> mtext( "Age (years)", side=1, outer=TRUE, line=1.5 )
```

From figure **??** we see that the two possible ways of computing the contribution give pretty much the same results — the differences never exceed some 0.3%. Therefore, if we want to attribute fractions of the prevalence in 2010 to decreasing mortality and increasing incidence, we would want two measures that had a sum equal the the difference between the scenario with observed mortality and incidence rates ("`obs`"), and the scenario with rates fixed to those from 1995 ("`all-f`").

The thin lines at the bottom of figure **??** represents the prevalence at 1.1.1995, so it is pretty clear that the incidence an mortality rates as observed by 1995 did not provide for at steady state.

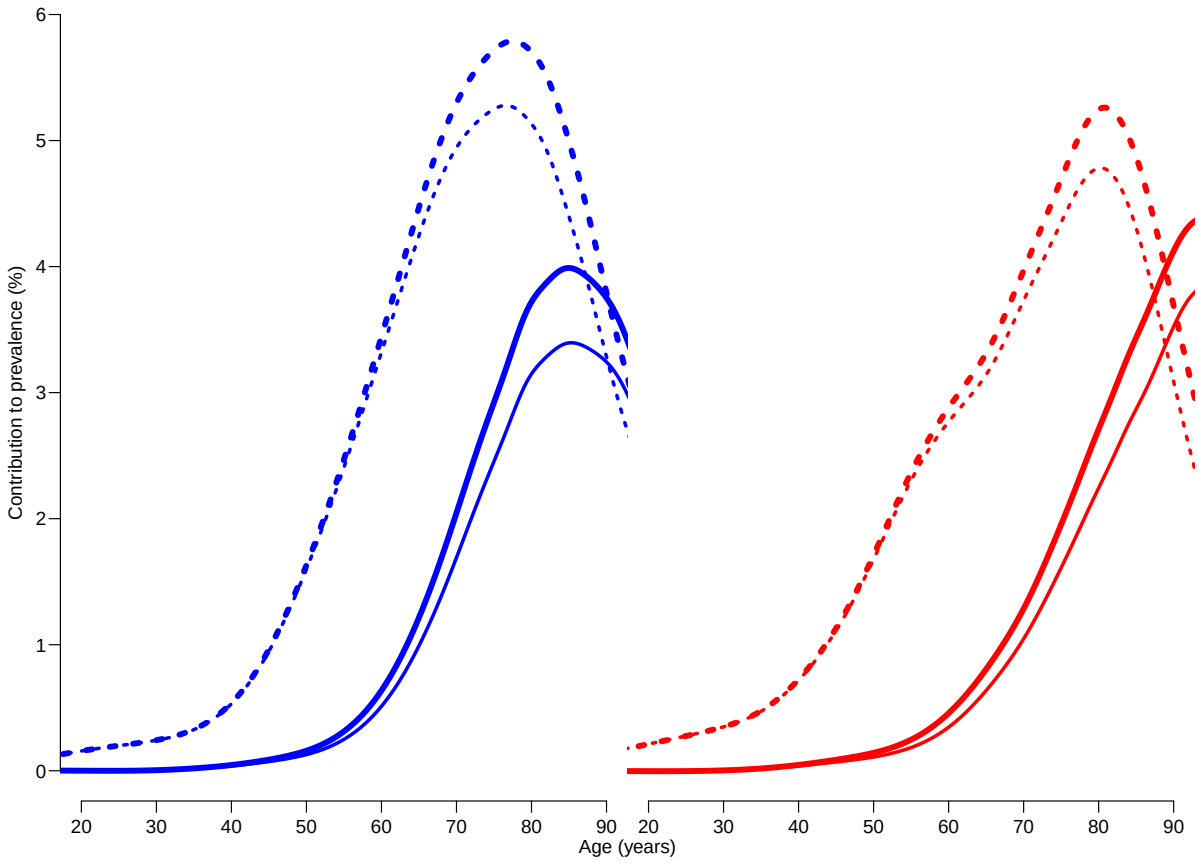

Figure 2.20: *Suggested contributions to age-specific prevalences from decreasing mortality over the last 17 years; the thick lines are obtained by subtracting the prediction based on fixing one rate from the one using the observed rates; thin lines based on subtracting the prediction based on fixing both rates from that where one is fixed. Full lines are for differences attributable to changes in mortality rates, broken lines are for changes attributable to changes in incidence rates.*

So basically we can subdivide the prevalence at any point in time into 4 components:

1. the "inherited" prevalences from 1995.
2. the prevalence attributable to rates of mortality and incidence as of 1995.
3. the prevalence attributable to the *increase* in the incidence rates.
4. the prevalence attributable to the *decrease* in the mortality rates.

So we now fill out the remaining 3 dimension of `prv`:

```
> prv$prv[,,,"mort" ] <- ( prv$prv[,,,"obs"  ]-prv$prv[,,,"m-fix"] +
+                          prv$prv[,,,"i-fix"]-prv$prv[,,,"all-f"] ) / 2
> prv$prv[,,,"inc"  ] <- ( prv$prv[,,,"obs"  ]-prv$prv[,,,"i-fix"] +
+                          prv$prv[,,,"m-fix"]-prv$prv[,,,"all-f"] ) / 2
> prv$prv[,,,"const"] <-   prv$prv[,,,"all-f"]-prv$prv[,rep(1,dim(prv$prv)[2]),,"obs"]
```

The components `obs`, `const`, `inc` and `mort` now together make up the total prevalence of diabetes for a given combinations of sex, age and date. Thus we can show these for each of the 17 dates 1996,. . . ,2012.

First we define a function to make the component plot, and then use this for men and women separately:

```
> poly.parts <-
+ function( x, crv, col, xlim, ylim, txt="" )
+ {
+ crv <- t(apply(cbind(0,crv),1,cumsum))
+ matplot( x, crv, type="n", xaxt="n", yaxt="n", xlab="", ylab="",
+          xlim=xlim, ylim=ylim, yaxs="i", bty="n" )
+ for( i in 2:ncol(crv) )
+ polygon( c(x,rev(x)), c(crv[,i],rev(crv[,i-1])),
+          col=col[i-1], border="transparent")
+ text( par("usr")[1:2]%*%c(0.1,0.9),
+       par("usr")[3:4]%*%c(0.9,0.1), txt, adj=c(1,0), font=2 )
+ }
```

We can now show the impact of changes in incidence and mortality on the age-specific prevalences:

```
> nt <- dim( prv$prv )[2]
> par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
> clr <- rgb(c(3,2,1.5,0)/3,c(3,2,1.5,0)/3,1)
> poly.parts( a.pt, cbind(prv$prv[,1 ,"M","obs"],
+                         prv$prv[,nt,"M",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,22) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> axis( side=2 )
> text( rep(25,3), 17:19+0.5,
+       c("Original","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=0.8 )
> # box(bty="c")
>
> clr <- rgb(1,c(3,2,1.5,0)/3,c(3,2,1.5,0)/3)
> poly.parts( a.pt, cbind(prv$prv[,1 ,"F","obs"],
+                         prv$prv[,nt,"F",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,22) )
> # axis( side=2 )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> text( rep(25,3), 17:19+0.5,
+       c("Original","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=0.8 )
> mtext( "Age", side=1, outer=TRUE, line=1.5, font=1, las=0 )
> mtext( "Prevalence of DM (%)", side=2, outer=TRUE, line=2, font=1, las=0 )
> # box(bty="]")
```

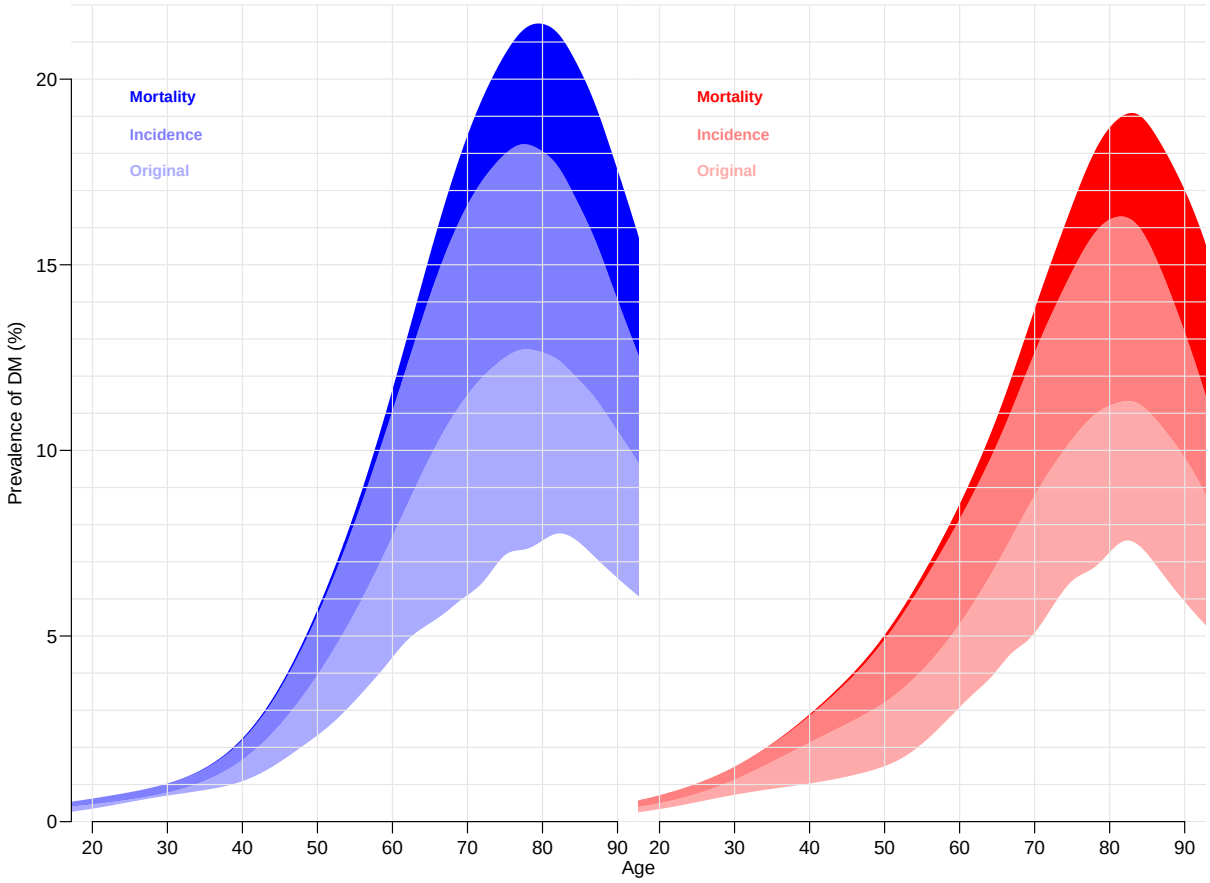

Figure 2.21: *Predicted age-specific prevalences of DM in Denmark 2012 among men (blue) and women (red), partitioned by the contribution from rates as they were in 1995 ("Original"), increases in incidence and decrease in mortality, respectively.*

2.5.7 The actual numbers of diabetes patients in Denmark

In the previous section we only looked at the age-specific prevalences, because these are the quantities that are driven by the incidence and mortality rates. However, it is also of interest to see how the actual number of diabetes patients would have looked under the different scenarios, specifically how the *number* of the current patients that can be attributed to the various components.

Also note that since the previous calculations were for age-specific prevalences we have a constant reference as the prevalences at 1995, but when we multiply by the population figures we would of course see differences in numbers and age-distribution of the diabetes population even if the age-specific prevalences were unchanged.

To show these effects we set up an array `prn` with structure like `prv$prv` to hold the number of diabetes patients by category, assuming the age-distribution in the population to be as actually observed (that is as extracted from Statistics Denmark, and as recorded in the data frame `pr`). However `prn` will have 100 age-classes rather than 100/`int`, and only 18 dates rater than 18/`int` as `prv$prv`.

This is done by selecting the relevant dates from `prv$prv` and then taking averages over age-classes.

```
> # The dates of the predicted prevalences as numerical
> prv.t <- as.numeric( dimnames(prv$prv)[["t"]] )
> # The dates where we want the prevalences
> prn.t <- 1995:2012
> # Find out where those are in prv.t
> nt <- length( prn.t )
> wh.t <- numeric( nt )
> for( it in 1:nt )
+    {
+    dd <- abs( prn.t[it]-prv.t )
+    wh.t[it] <- which(dd==min(dd))[1]
+    }
> # Take only prevalences at these dates
> prv.n <- data.frame( as.table( prv$prv[,wh.t,,] ) )
> str( prv.n )
  'data.frame':        252000 obs. of  5 variables:
   $ a   : Factor w/ 1000 levels "0.05","0.15",..: 1 2 3 4 5 6 7 8 9 10 ...
   $ t   : Factor w/ 18 levels "1995","1996",..: 1 1 1 1 1 1 1 1 1 1 ...
   $ sex : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ what: Factor w/ 7 levels "obs","m-fix",..: 1 1 1 1 1 1 1 1 1 1 ...
   $ Freq: num  0 0.000413 0.000417 0.000422 0.000426 ...
> # Round the ages
> prv.n$a <- floor( as.numeric( as.character(prv.n$a) ) )
> prn <- xtabs( Freq ~ a + t + sex + what,
+                      data = aggregate( prv.n[5], prv.n[-5], mean ) )
> str( prn )
   xtabs [1:100, 1:18, 1:2, 1:7] 0.000388 0.000479 0.000535 0.000598 0.000668 ...
   - attr(*, "dimnames")=List of 4
    ..$ a   : chr [1:100] "0" "1" "2" "3" ...
    ..$ t   : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex : chr [1:2] "M" "F"
    ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
   - attr(*, "class")= chr [1:2] "xtabs" "table"
   - attr(*, "call")= language xtabs(formula = Freq ~ a + t + sex + what, data = aggregate(prv.n[5],
> dimnames( prn )[[4]]
  [1] "obs"   "m-fix" "i-fix" "all-f" "mort"  "inc"   "const"
```

Now `prn` contains the prevalences components (as fractions) for 100 age classes and 18 dates. However, the components "`mort`", "`inc`" and "`const`", correspond to the prevalences

attributable to decline in mortality, increase in incidence and initial imbalance. But the first component is the prevalences predicted using the observed (well, fitted) rates. But would need the prevalences as of 1995 too, and the first 4 dimensions are really not needed.

So we restructure the 4th dimension, so we have the observed prevalences as of 1995, the three change-components, and finally the fitted total.

```
> prn <- prn[,,,c(1,5:7,1)]
> dimnames( prn )[[4]][1] <- "1995"
> prn[,,,"1995"] <- prn[,rep(1,dim(prn)[2]),,"obs"]
> str( prn )
 num [1:100, 1:18, 1:2, 1:5] 0.000388 0.000479 0.000535 0.000598 0.000668 ...
 - attr(*, "dimnames")=List of 4
  ..$ a   : chr [1:100] "0" "1" "2" "3" ...
  ..$ t   : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ what: chr [1:5] "1995" "mort" "inc" "const" ...
```

In principle we would now to multiply these prevalences by the population figures at these times, however for stability we multiply the `relative` size of the 4 components to the empirical prevalences observed. The population prevalence figures are in `pr`:

```
> head( pr )
  sex A    P X     N
1   M 0 1995 3 35612
2   M 0 1996 1 36055
3   M 0 1997 0 34853
4   M 0 1998 1 34774
5   M 0 1999 2 34076
6   M 0 2000 1 33906
> subset(pr,A<1 & P<1997)
      sex A    P X     N
1       M 0 1995 3 35612
2       M 0 1996 1 36055
1801    F 0 1995 0 34094
1802    F 0 1996 0 34051
> pop <- xtabs( N ~ A + P + sex, data=pr )
> dmp <- xtabs( X ~ A + P + sex, data=pr )
> str( pop )
 xtabs [1:100, 1:18, 1:2] 35612 34747 35082 33330 32974 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = N ~ A + P + sex, data = pr)
> str( dmp )
 xtabs [1:100, 1:18, 1:2] 3 4 6 5 12 21 22 34 29 29 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = X ~ A + P + sex, data = pr)
> str( prn )
 num [1:100, 1:18, 1:2, 1:5] 0.000388 0.000479 0.000535 0.000598 0.000668 ...
 - attr(*, "dimnames")=List of 4
  ..$ a   : chr [1:100] "0" "1" "2" "3" ...
  ..$ t   : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ what: chr [1:5] "1995" "mort" "inc" "const" ...
```

```
> prt <- apply( prn[,,,1:4], 1:3, sum )
> for( i in 1:4 )
+ prn[,,,i] <- (prn[,,,i]/prt) * dmp
```

First we draw a simple pyramid of the age-distribution of diabetes patients in Denmark:

```
> # Note: This uses the undocumented feature that if the first
> # number in a column is negative this is taken as the left endpoint of
> # the bar. So c(-m,m,f) is a bar starting at -m, and a division at
> # -m+m and an upper end at -m+m-f. Coloring is from the top, that is
> # the part stretching from -m+m to -m+m+f get the first color
> pp <- "2012"
> oo <- c("mort","inc","const","1995")
> lim <- 6
> clr <- c("red","blue")
> draw.dmp <-
+ function(pp)
+ {
+ par( mar=c(3,3,3,0), mgp=c(3,1,0)/1.6, las=1 )
+ barplot( height=t( cbind( -dmp[,pp,"M"],
+                           dmp[,pp,"M"],
+                           dmp[,pp,"F"] ) )/ 1000,
+          horiz=TRUE, col=clr,
+          border=NA,space=0,axes=FALSE,names.arg=rep("",dim(prn)[1]),
+          xlim=c(-1,1)*lim*1.05,xlab="Persons in 1 year class (1000s)",ylab="Age")
+ abline(h=seq(0,100,5),
+        v=seq(-lim,lim,0.5),
+        col="white")
+ axis( side=1, at=seq(-lim,lim,1), labels=abs(seq(-lim,lim,1)) )
+ axis( side=2, at=seq(0,100,20) )
+ mtext( pp, at=-lim, adj=1.4, cex=1.3, font=2 )
+ mtext( formatC(sum(dmp[,pp,"M"]),0,format="f",big.mark=","), at=-1, col="blue", line=0, cex=0.99 )
+ mtext( formatC(sum(dmp[,pp,"F"]),0,format="f",big.mark=","), at= 1, col="red" , line=0, cex=0.99 )
+ mtext( "N", at=0, line=0, cex=0.99 )
+ }
> pdf( "./graph/NDR-obs-film-o.pdf", width=8, height=6 )
> for( pp in paste(1995:2012) ) draw.dmp(pp)
> dev.off()
  null device
            1

> for( pp in paste(1995:2012) )
+ {
+ pdf( paste("./graph/NDR-obs-", pp, "-o.pdf", sep=""), width=8, height=6 )
+ draw.dmp(pp)
+ dev.off()
+ }
```

Using the same machinery we can also draw a population pyramid using colors that range from very light to full:

```
> shd <- c(0.0, 1.5, 2.0, 2.8) / 3
> een <- c(1,1,1,1)
> clr <- rgb( c(een,rev(shd)),
+             c(shd,rev(shd)),
+             c(shd,    een ) )
> clr
  [1] "#FF0000" "#FF8080" "#FFAAAA" "#FFEEEE" "#EEEEFF" "#AAAAFF" "#8080FF"
  [8] "#0000FF"

> # Note: The following uses the undocumented feature that if the first
> # number in a column is negative this is taken as the left endpoint of
> # the bar. So c(-m,m,f) is a bar starting at -m, and a division at
> # -m+m and an upper end at -m+m-f. Coloring is from the top, that is
> # the part stretching from -m+m to -m+m+f get the first color
```

```
> oo <- c("mort","inc","const","1995")
> draw.pyr <-
+ function(pp)
+ {
+ par( mar=c(3,3,3,0), mgp=c(3,1,0)/1.6, las=1 )
+ barplot( height=t( cbind( -apply( prn[,pp,"M",oo], 1, sum ),
+                                   prn[,pp,"M",oo],
+                                   prn[,pp,"F",rev(oo)] ) )/ 1000,
+          horiz=TRUE, col=clr[c(1,8:2)],
+          border=NA,space=0,axes=FALSE,names.arg=rep("",dim(prn)[1]),
+          xlim=c(-1,1)*lim*1.05,xlab="Persons in 1 year class (1000s)",ylab="Age" )
+ abline(h=seq(0,100,5),
+        v=seq(-lim,lim,0.5),
+        col="white")
+ axis( side=1, at=seq(-lim,lim,1), labels=abs(seq(-lim,lim,1)) )
+ axis( side=2, at=seq(0,100,20) )
+ tt <- addmargins( apply( prn[,pp,,oo],2:3, sum ), 2 )
+ nn <- tt / tt[,5] * 100
+ ppos <- 1:5-0.1
+ npos <- -rev(ppos)
+ mtext( pp, at=-lim, adj=1.8, line=2, cex=1.2, font=2 )
+ mtext( c(lg<- c("Mort","Inc","Const","Org","All"),rev(lg)),
+        at=c(npos,ppos), col="black", cex=0.99, line=2 )
+ mtext( formatC(tt["M",1:5],0,,"f",,,","), at=npos, col="blue", line=1, cex=0.99 )
+ mtext( formatC(tt["F",5:1],0,,"f",,,","), at=ppos, col="red" , line=1, cex=0.99 )
+ mtext( formatC(nn["M",1:4],1,4,"f"), at=npos[1:4], col="blue", line=0, cex=0.99 )
+ mtext( formatC(nn["F",4:1],1,4,"f"), at=ppos[2:5], col="red" , line=0, cex=0.99 )
+ mtext( "N", at=0, line=1, cex=0.99 )
+ mtext( "%", at=0, line=0, cex=0.99 )
+ }
> pdf( "./graph/NDR-film-o.pdf", width=9, height=6 )
> for( pp in paste(1995:2012) ) draw.pyr(pp)
> dev.off()
  null device
            1
> for( pp in paste(1996:2012) )
+ {
+ pdf( paste("./graph/NDR-", pp, "-o.pdf", sep=""), width=8, height=6 )
+ draw.pyr(pp)
+ dev.off()
+ }
```

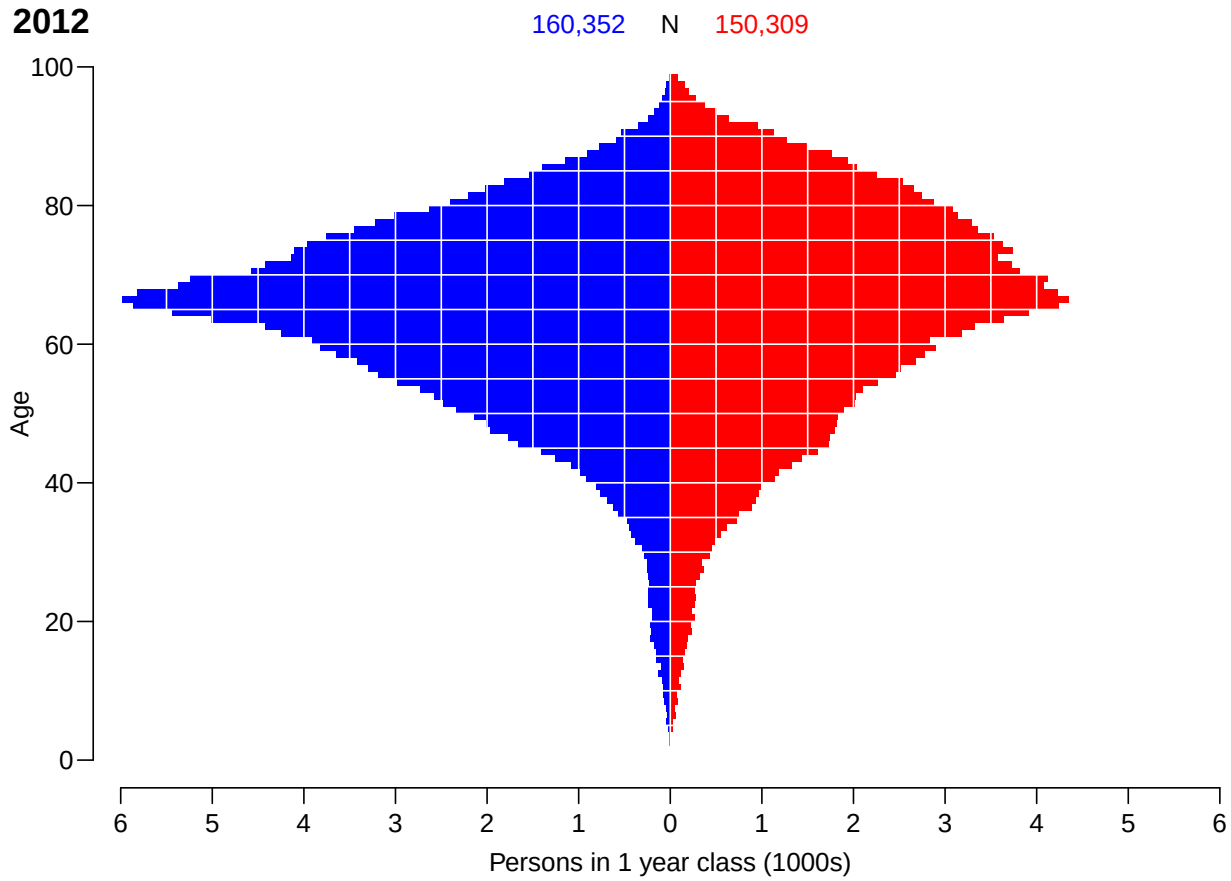

Figure 2.22: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2012.*

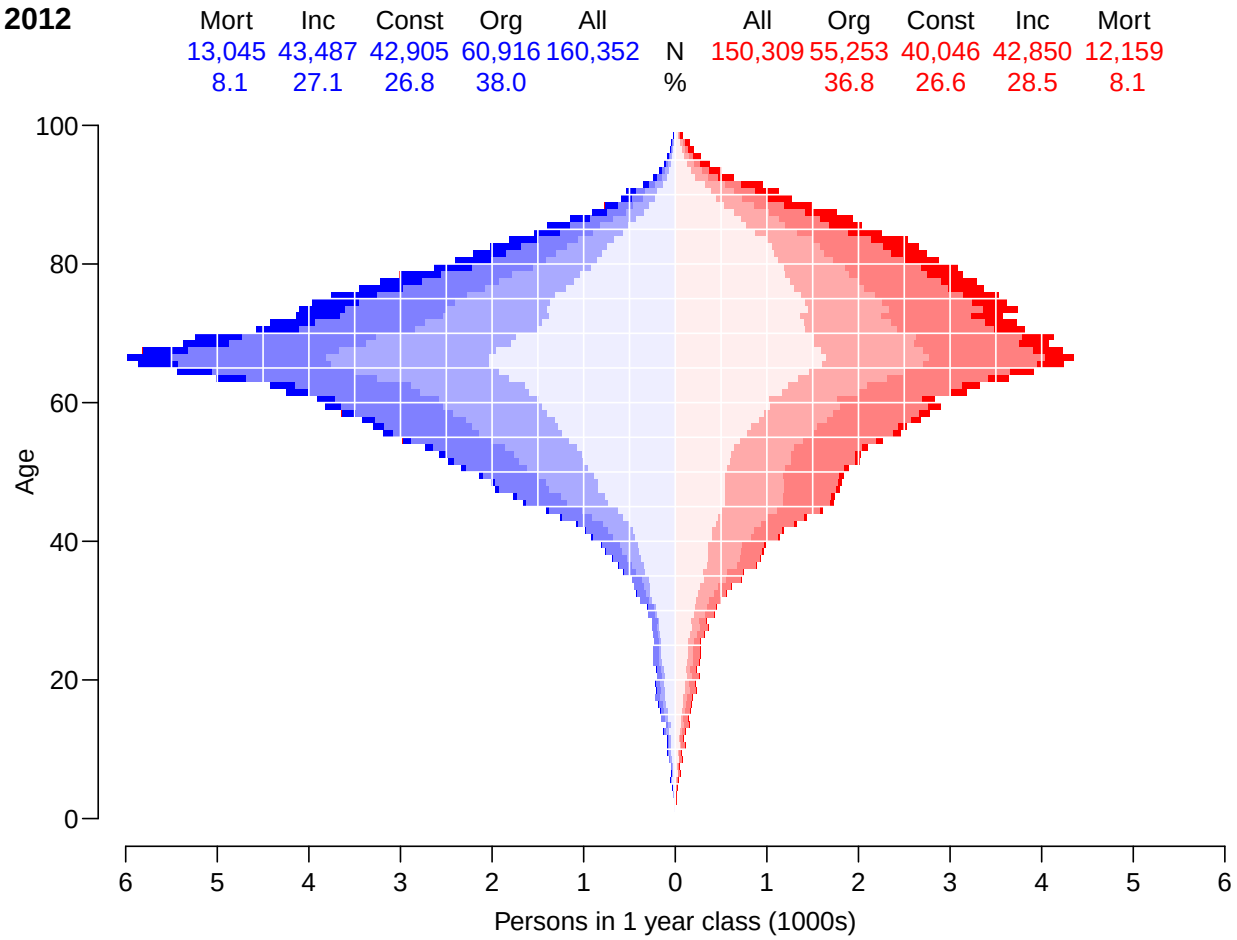

Figure 2.23: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2012, subdivided by the contribution from various causes: Mort: decrease in mortality, Inc: increase in incidence, Const: constant rates from 1995, Org: age-specific prevalence in 1995.*

Chapter 3

Analysis based on modified DM definition

3.1 Register data — follow-up and deaths

First we load the register:

```
> load( file="./data/ndr.Rda" )
```

For setting up follow-up data we need convenience functions which maps `NA`s to either `FALSE` or `TRUE`:

```
> na2T <- function( x )     x | is.na(x)
> na2F <- function( x )  !(!x | is.na(x))
```

We now set up data as a Lexis object with three timescales: age, calendar time and diabetes duration. Note that we use the modified definintion of diabetes, excluding the blood glucose criteria:

```
> dr$doDM <- dr$doin
> dr <- transform( dr, doe = pmax(doDM,1995),
+                      dox = pmin(2012,dodsdto,foddto+99,na.rm=TRUE) )
> Lx <- Lexis( entry = list( A = doe-foddto,
+                            P = doe,
+                          dur = doe-doDM ),
+              exit = list( P = dox ),
+       exit.status = factor( na2F(dodsdto==dox),
+                             labels=c("Alive","Dead") ),
+              data = subset( dr, doe<dox & doDM>foddto ) )
  NOTE: entry.status has been set to "Alive" for all.

> summary( Lx )

  Transitions:
       To
  From     Alive   Dead  Records:  Events: Risk time:  Persons:
    Alive 246426 136447    382873   136447    2611387    382873
```

There are fewer cases in `Lx` the in the entire register, but mostly because of persons that have died before 1995, or were included after age 98:

```
> addmargins( tt <- with( dr, table( dd=dodsdto<1995,
+                                    bb=inkldto>foddto+99,
+                                    exclude=NULL ) ) )
```

```
        bb
dd       FALSE   TRUE   <NA>    Sum
  FALSE 159205     90      0 159295
  TRUE   27049     18      0  27067
  <NA>  310858     12      0 310870
  Sum   497112    120      0 497232
> sum( c(tt[2,1],tt[,2]) )
  [1] 27169
> nrow(dr) - nrow(Lx)
  [1] 114359
```

The `Lexis` object `Lx` is now going to be used to construct a table of person-years among DM patients which we will subtract from the population person-years. Note that we also count the number of deaths, in order to construct a dataset also usable for mortality analyses.

So basically, we split the data along the age and period axis, and to avoid problems with memory overflow we do the splitting in smaller chunks.

```
> n.chunks <- 50
> lm <- round( seq(0,nrow(Lx),,n.chunks+1) )
> for( i in 1:n.chunks )
+ {
+ whr <- (lm[i]+1):(lm[i+1])
+ sP  <- splitLexis( Lx[whr,], 1995:2013, time.scale="P" )
+ sPA <- splitLexis( sP    ,      0:100 , time.scale="A" )
+ agg <- with( sPA, aggregate( cbind( y = lex.dur,
+                                      d = lex.Xst=="Dead" ),
+                               list( sex = sex,
+                                     A = floor(A),
+                                     P = floor(P),
+                                     U = floor(P)-floor(A)-floor(foddto) ),
+                               FUN = sum ) )
+ # Just to get the right structure of Agg, variables sx, A, P and U
+ # and UPPER-CASE Y and D to hold the aggregate person-time and events
+ if( i==1 ) Agg <- cbind( agg[1,1:4], Y=NA, D=NA )
+ Agg <- merge( Agg, agg, by=c("sex","A","P","U"), all=TRUE )
+ Agg <- transform( Agg, Y = pmax(Y,0,na.rm=TRUE) + pmax(y,0,na.rm=TRUE),
+                        D = pmax(D,0,na.rm=TRUE) + pmax(d,0,na.rm=TRUE) )[
+                   ,c("sex","A","P","U","Y","D")]
+ cat( "Merged in chunk", i, " at",
+      format(Sys.time(),format="%Y-%m-%d %H:%M:%S"), "\n" )
+ flush.console()
+ }
  Merged in chunk 1  at 2013-08-30 13:20:59
  Merged in chunk 2  at 2013-08-30 13:21:11
  Merged in chunk 3  at 2013-08-30 13:21:23
  Merged in chunk 4  at 2013-08-30 13:21:34
  Merged in chunk 5  at 2013-08-30 13:21:45
  Merged in chunk 6  at 2013-08-30 13:21:56
  Merged in chunk 7  at 2013-08-30 13:22:08
  Merged in chunk 8  at 2013-08-30 13:22:19
  Merged in chunk 9  at 2013-08-30 13:22:30
  Merged in chunk 10  at 2013-08-30 13:22:42
  Merged in chunk 11  at 2013-08-30 13:22:53
  Merged in chunk 12  at 2013-08-30 13:23:05
  Merged in chunk 13  at 2013-08-30 13:23:16
  Merged in chunk 14  at 2013-08-30 13:23:28
  Merged in chunk 15  at 2013-08-30 13:23:39
  Merged in chunk 16  at 2013-08-30 13:23:50
  Merged in chunk 17  at 2013-08-30 13:24:02
  Merged in chunk 18  at 2013-08-30 13:24:14
```

```
  Merged in chunk 19  at 2013-08-30 13:24:25
  Merged in chunk 20  at 2013-08-30 13:24:36
  Merged in chunk 21  at 2013-08-30 13:24:47
  Merged in chunk 22  at 2013-08-30 13:24:58
  Merged in chunk 23  at 2013-08-30 13:25:10
  Merged in chunk 24  at 2013-08-30 13:25:21
  Merged in chunk 25  at 2013-08-30 13:25:32
  Merged in chunk 26  at 2013-08-30 13:25:42
  Merged in chunk 27  at 2013-08-30 13:25:53
  Merged in chunk 28  at 2013-08-30 13:26:04
  Merged in chunk 29  at 2013-08-30 13:26:16
  Merged in chunk 30  at 2013-08-30 13:26:27
  Merged in chunk 31  at 2013-08-30 13:26:39
  Merged in chunk 32  at 2013-08-30 13:26:50
  Merged in chunk 33  at 2013-08-30 13:27:01
  Merged in chunk 34  at 2013-08-30 13:27:12
  Merged in chunk 35  at 2013-08-30 13:27:24
  Merged in chunk 36  at 2013-08-30 13:27:35
  Merged in chunk 37  at 2013-08-30 13:27:47
  Merged in chunk 38  at 2013-08-30 13:27:59
  Merged in chunk 39  at 2013-08-30 13:28:10
  Merged in chunk 40  at 2013-08-30 13:28:22
  Merged in chunk 41  at 2013-08-30 13:28:33
  Merged in chunk 42  at 2013-08-30 13:28:45
  Merged in chunk 43  at 2013-08-30 13:28:57
  Merged in chunk 44  at 2013-08-30 13:29:08
  Merged in chunk 45  at 2013-08-30 13:29:19
  Merged in chunk 46  at 2013-08-30 13:29:30
  Merged in chunk 47  at 2013-08-30 13:29:41
  Merged in chunk 48  at 2013-08-30 13:29:52
  Merged in chunk 49  at 2013-08-30 13:30:03
  Merged in chunk 50  at 2013-08-30 13:30:14
> summary( Agg )
   sex            A               P               U                Y
   M:3360   Min.   : 0.00   Min.   :1995   Min.   :0.0000   Min.   :   0.0144
   F:3354   1st Qu.:24.00   1st Qu.:1999   1st Qu.:0.0000   1st Qu.:  67.1376
            Median :49.00   Median :2003   Median :1.0000   Median : 244.8785
            Mean   :49.13   Mean   :2003   Mean   :0.5007   Mean   : 388.9465
            3rd Qu.:74.00   3rd Qu.:2007   3rd Qu.:1.0000   3rd Qu.: 611.8155
            Max.   :98.00   Max.   :2011   Max.   :1.0000   Max.   :2499.0219
       D
   Min.   :  0.00
   1st Qu.:  0.00
   Median :  5.00
   Mean   : 20.32
   3rd Qu.: 36.00
   Max.   :117.00
> head( Agg )
    sex A    P U         Y D
  1   M 0 1995 0 0.8062971 0
  2   M 0 1995 1 0.8596851 0
  3   M 0 1996 1 0.0403833 0
  4   M 0 1997 0 0.4572211 0
  5   M 0 1997 1 0.1731691 0
  6   M 0 1998 0 0.9185489 0
```

3.1.1 Population time

Now we need the population data. It can be obtained either from the `Y.dk` dataset in the `Epi` package or from the human mortality database. The data in the `Epi` package are more up-to-date which is what we need:

```
> data( Y.dk )
> head( Y.dk )
    sex A    P    C        Y upper
  1   1 0 1971 1971 19195.00     0
  2   1 0 1971 1970 17944.17     1
  3   1 1 1971 1970 17968.83     0
  4   1 1 1971 1969 18164.83     1
  5   1 2 1971 1969 18178.67     0
  6   1 2 1971 1968 18934.33     1
```

We want data from the population in the years 1995 trough 2011 and ages 0–98 (because the population data only has 98 as the last closed age-class):

```
> Y.dk <- transform( Y.dk, U = upper,
+                          sex = factor(sex,labels=c("M","F")) )
> Y.dk <- subset( Y.dk, A < 99   &
+                       P > 1994 &
+                       P < 2012 )[,c("sex","A","P","U","Y")]
```

3.1.2 Merging time

Now we merge the two data sets; we construct the risk time among DM patients is in the `Agg` dataset as `Y` and the risk time in the entire population is in the dataset `Y.dk`, also as `Y`, and hence in the merged dataset referred to as `Y.x` and `Y.y`, respectively. By that token we can construct `Y.DM` and `Y.nD` as the risk time among non-diabetics and among diabetes patients, respectively:

```
> YY <- merge( Agg, Y.dk, by=c("sex","A","P","U"), all.y=TRUE )
> YY <- transform( YY, Y.nD = Y.y-pmax(Y.x,0,na.rm=TRUE),
+                      Y.DM =     pmax(Y.x,0,na.rm=TRUE),
+                      D.DM =     pmax(  D,0,na.rm=TRUE) )[,c("sex","A","P","U","Y.nD","Y.DM","D.DM"
> str( YY )
  'data.frame':        6732 obs. of  7 variables:
   $ sex : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P   : num  1995 1995 1996 1996 1997 ...
   $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
   $ Y.nD: num  18027 17871 17427 18062 17387 ...
   $ Y.DM: num  0.8063 0.8597 0 0.0404 0.4572 ...
   $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...

> head( YY )

    sex A    P U     Y.nD      Y.DM D.DM
  1   M 0 1995 0 18026.69 0.8062971    0
  2   M 0 1995 1 17870.97 0.8596851    0
  3   M 0 1996 0 17426.50 0.0000000    0
  4   M 0 1996 1 18062.13 0.0403833    0
  5   M 0 1997 0 17386.54 0.4572211    0
  6   M 0 1997 1 17450.66 0.1731691    0
```

3.1.3 Population deaths

We can extract the number of deaths in Lexis-triangles from the Human mortality database, using the function

```
> require(RCurl)
> pth <- "http://www.mortality.org/hmd/DNK/STATS/Deaths_lexis.txt"
> upw <- "bxc@steno.dk:BxCPwd"
> txt <- getURL( pth, userpwd=upw )
> con <- textConnection( txt )
> mlx <- try( read.table( con, skip = 2, header = TRUE, na.strings = "."), TRUE)
> str( mlx )
  'data.frame':        39117 obs. of  6 variables:
   $ Year  : int  1835 1835 1835 1835 1835 1835 1835 1835 1835 1835 ...
   $ Age   : Factor w/ 111 levels "0","1","10","100",..: 1 1 2 2 24 24 35 35 46 46 ...
   $ Cohort: int  1835 1834 1834 1833 1833 1832 1832 1831 1831 1830 ...
   $ Female: num  2159 1156 502 364 293 ...
   $ Male  : num  2772 1604 562 402 332 ...
   $ Total : num  4930 2761 1064 766 626 ...
```

We then restrict and transform these data to be of the sam shape as the tabulated follow-up of the diabetes patients:

```
> mlx <- subset( mlx, Year>1994 & Year<2012 & Age!="110+" )
> mlx$A <- as.numeric(as.character(mlx$Age))
> mlx <- transform( mlx, P=Year,
+                        C=Cohort,
+                        U=Year-A-Cohort )
> mm <- data.frame( mlx[,c("A","P","U","Male")],
+                   sex=factor(1,levels=1:2,labels=c("M","F") ) )
> mf <- data.frame( mlx[,c("A","P","U","Female")],
+                   sex=factor(2,levels=1:2,labels=c("M","F") ) )
> names(mm)[4] <-
+ names(mf)[4] <- "D.nD"
> MM <- subset( rbind( mm, mf ), A < 99 )
> head( MM )
       A    P U D.nD sex
35361 0 1995 0  179   M
35362 0 1995 1   21   M
35363 1 1995 0   13   M
35364 1 1995 1    8   M
35365 2 1995 0    2   M
35366 2 1995 1    7   M
> save( MM, file="./data/MM.Rda" )
```

Now we have the total number of deaths in Lexis triangles for the relevant period, we can merge with the follow-up dataset, so we have the number of deaths and person-years by sex, age, period and diabetes status:

```
> TT <- transform( merge( YY, MM ), D.nD = D.nD - D.DM )
> head( TT )
    sex A    P U     Y.nD      Y.DM D.DM D.nD
1   F 0 1995 0 17025.50 0.0000000    0  137
2   F 0 1995 1 17100.54 0.1300479    0   16
3   F 0 1996 0 16468.06 1.4401095    0  134
4   F 0 1996 1 17067.30 1.8617385    0   23
5   F 0 1997 0 16434.00 0.0000000    0  152
6   F 0 1997 1 16499.84 1.9890486    0   14
```

3.1.4 DM cases

Finally we want to append the number of diabetes cases to the data frame, so we count the number of entries in the Lexis object `Lx`

```
> CC <- with( subset( Lx, P>1995 ),
+             table( sex, floor(A),
+                         floor(P),
+                         floor(P) - floor(A) - floor(P-A) ) )
> CC <- as.data.frame( CC )
> names( CC ) <- c("sex","A","P","U","X")
> for( i in 2:4 ) CC[,i] <- as.numeric(as.character(CC[,i]))
> str( CC )
  'data.frame':        6732 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
   $ P  : num  1995 1995 1995 1995 1995 ...
   $ U  : num  0 0 0 0 0 0 0 0 0 0 ...
   $ X  : int  1 0 4 2 5 1 3 1 5 1 ...
```

Now `CC` contains the number of incident cases of DM in per period 1995–2011 incl. in the column `X`.

3.1.5 Saving it all for later analysis

```
> TT <- merge( TT, CC )
> str( TT )
  'data.frame':        6732 obs. of  9 variables:
   $ sex : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P   : num  1995 1995 1996 1996 1997 ...
   $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
   $ Y.nD: num  17026 17101 16468 17067 16434 ...
   $ Y.DM: num  0 0.13 1.44 1.86 0 ...
   $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.nD: num  137 16 134 23 152 14 132 16 95 17 ...
   $ X   : int  0 2 4 4 0 2 0 0 0 1 ...
```

The data frame `TT` has the risk time in the states "No DM" (`Y.nD`) and "DM" (`Y.DM`) and the number of transitions from "No DM" to either "DM" (`X`) or "Death" (`D.nd`) and from "DM" to "Death" (`D.DM`).

We can now finally save the tabulated dataset which contains information for analysis of incidence rates of diabetes and mortality rates for both diabetes patients and non-patients. We just define an attribute which

```
> Vars <- matrix( c("Sex",
+                   "1-year age class",
+                   "1-year period",
+                   "Indicator of upper Lexis triangle",
+                   "P-Y among non-diabetics",
+                   "P-Y among diabetes patients",
+                   "Deaths among non-diabetics",
+                   "Deaths among diabetes patients",
+                   "Diabetes diagnoses among non-diabetics"), ncol(TT) )
> rownames( Vars ) <- names( TT )
> colnames( Vars ) <-
+     "Data frame using the original definintion of DM from NDR"
> attr( TT, "Variables" ) <- Vars
> str( TT )
  'data.frame':        6732 obs. of  9 variables:
   $ sex : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
   $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P   : num  1995 1995 1996 1996 1997 ...
   $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
   $ Y.nD: num  17026 17101 16468 17067 16434 ...
```

```
   $ Y.DM: num  0 0.13 1.44 1.86 0 ...
   $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...
   $ D.nD: num  137 16 134 23 152 14 132 16 95 17 ...
   $ X   : int  0 2 4 4 0 2 0 0 0 1 ...
   - attr(*, "Variables")= chr [1:9, 1] "Sex" "1-year age class" "1-year period" "Indicator of upper
    ..- attr(*, "dimnames")=List of 2
    .. ..$ : chr  "sex" "A" "P" "U" ...
    .. ..$ : chr "Data frame using the original definintion of DM from NDR"
> save( Lx, TT, file="./data/FU-m.Rda" )
```

3.2 DM incidence

In this chapter we use the original definition of DM for the NDR, so first we load the analysis data frame:

```
> library( Epi )
> load( file="./data/FU-m.Rda" )
> head( TT )
    sex A    P U     Y.nD      Y.DM D.DM D.nD X
1   F 0 1995 0 17025.50 0.0000000    0  137 0
2   F 0 1995 1 17100.54 0.1300479    0   16 2
3   F 0 1996 0 16468.06 1.4401095    0  134 4
4   F 0 1996 1 17067.30 1.8617385    0   23 4
5   F 0 1997 0 16434.00 0.0000000    0  152 0
6   F 0 1997 1 16499.84 1.9890486    0   14 2
> attr( TT, "Variables" )
      Data frame using the original definintion of DM from NDR
  sex  "Sex"
  A    "1-year age class"
  P    "1-year period"
  U    "Indicator of upper Lexis triangle"
  Y.nD "P-Y among non-diabetics"
  Y.DM "P-Y among diabetes patients"
  D.DM "Deaths among non-diabetics"
  D.nD "Deaths among diabetes patients"
  X    "Diabetes diagnoses among non-diabetics"
```

3.2.1 No. of cases

We would like to see the number of prevalent cases as of 1.1.1995 and the number of new cases for each year after that and the prevalent number of cases at the end. These numbers are readily available from the Lexis object `Lx`:

```
> prnew <- rbind( with( subset( Lx, doDM<1995 & na2T(dodsdto>1995) ),
+                       table( sex ) ),
+                 with( subset( Lx, doDM>=1995 ),
+                       table( floor(doDM), sex ) ),
+                 with( subset( Lx, doDM<2012 & na2T(dodsdto>2012) ),
+                       table( sex ) ) )
> rownames( prnew )[1] <- "Prev 1.1.1995"
> rownames( prnew )[nrow(prnew)] <- "Prev 31.11.2011"
> addmargins( prnew, margin=2 )
                       M      F    Sum
  Prev 1.1.1995    43211  41629  84840
  1995              6924   6079  13003
  1996              7052   6051  13103
  1997              6846   5919  12765
  1998              7742   6251  13993
  1999              7929   6696  14625
  2000              7988   6685  14673
  2001              8308   6794  15102
  2002              9465   8326  17791
  2003             10434   8842  19276
  2004             10539   8993  19532
  2005              9603   7832  17435
  2006              9707   7471  17178
  2007             10165   8243  18408
  2008             11015   8662  19677
  2009             11720   8905  20625
  2010             12599   9561  22160
  2011             15545  13142  28687
  Prev 31.11.2011 133935 112040 245975
```

3.2.2 Age-Period-Cohort modelling

We are going to use `X` and `Y.nD` as response variables in the analysis of diabetes incidence rates, however we first need to define the age and period properly:

```
> DD <- transform( TT, A = A + (1+U)/3,
+                      P = P + (2-U)/3,
+                      D = X,
+                      Y = Y.nD/1000 )[,c("sex","A","P","D","Y")]
```

With this groomed data frame in place we can fit separate models for men and women and plot the estimates together:

```
> acpM <- apc.fit( subset(DD,sex=="M"), ref.c=1950, parm="ACP", npar=c(18,5,12) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347     9923.2
  Age-drift              3346     5118.5   1   4804.8 < 2.2e-16
  Age-Cohort             3335     4950.7  11    167.8 < 2.2e-16
  Age-Period-Cohort      3331     4240.6   4    710.0 < 2.2e-16
  Age-Period             3342     4397.6 -11   -157.0 < 2.2e-16
  Age-drift              3346     5118.5  -4   -720.8 < 2.2e-16

> acpF <- apc.fit( subset(DD,sex=="F"), ref.c=1950, parm="ACP", npar=c(18,5,12) )

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347    10071.6
  Age-drift              3346     6504.6   1   3567.0 < 2.2e-16
  Age-Cohort             3335     6252.6  11    252.0 < 2.2e-16
  Age-Period-Cohort      3331     5106.7   4   1145.9 < 2.2e-16
  Age-Period             3342     5336.5 -11   -229.9 < 2.2e-16
  Age-drift              3346     6504.6  -4  -1168.0 < 2.2e-16

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( acpM, lty=1, ci=TRUE, col="blue", r.txt="Rate per 1000 PY" )

  cp.offset    RR.fac
       1790         1

> lines( acpF, lty=1, ci=TRUE, col="red" )



> apcM <- apc.fit( subset(DD,sex=="M"), ref.p=2000, parm="APC", npar=c(18,5,12) )

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347     9923.2
  Age-drift              3346     5118.5   1   4804.8 < 2.2e-16
  Age-Cohort             3335     4950.7  11    167.8 < 2.2e-16
  Age-Period-Cohort      3331     4240.6   4    710.0 < 2.2e-16
  Age-Period             3342     4397.6 -11   -157.0 < 2.2e-16
  Age-drift              3346     5118.5  -4   -720.8 < 2.2e-16

> apcF <- apc.fit( subset(DD,sex=="F"), ref.p=2000, parm="APC", npar=c(18,5,12) )
```

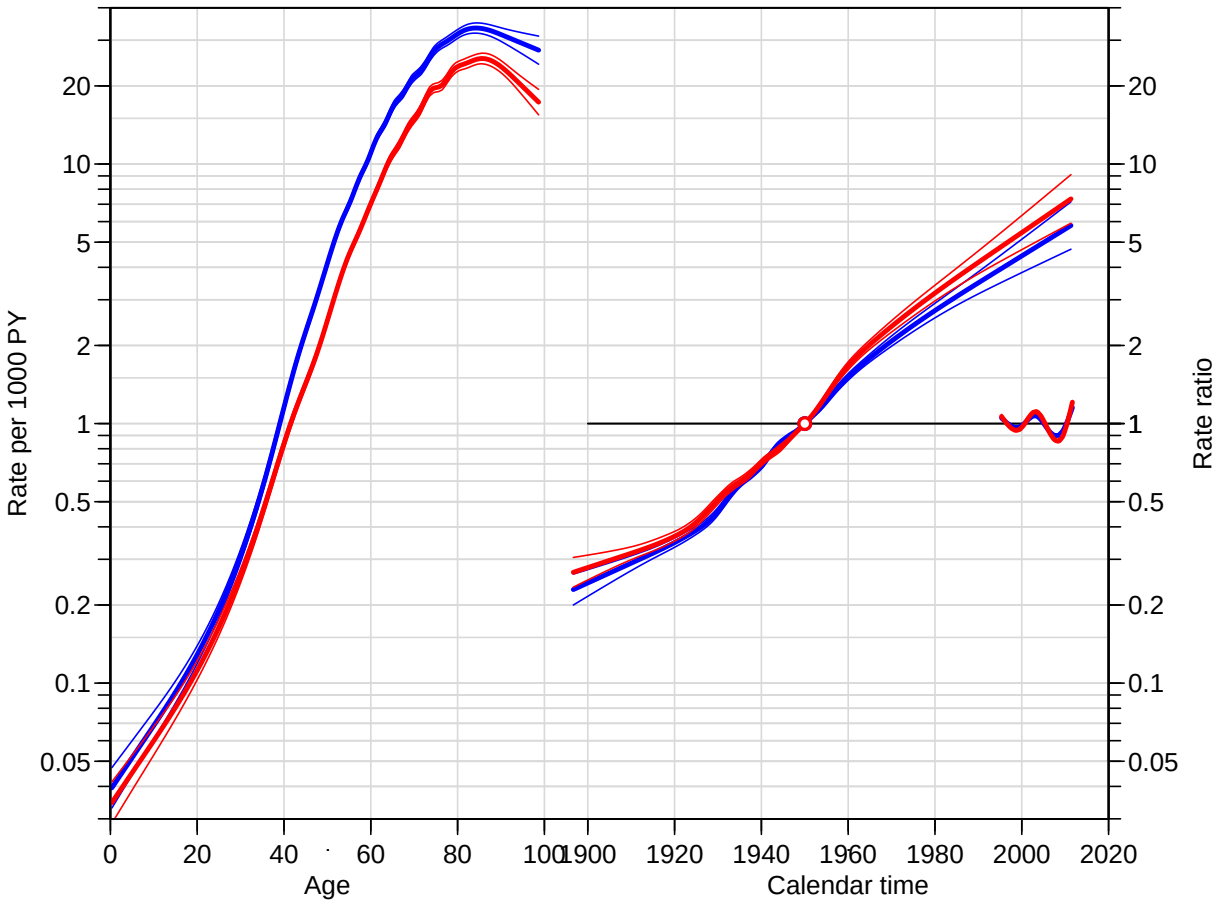

Figure 3.1: *Estimates from an APC-model for DM incidence rates in Denmark 1995–2011 (modified definition), cohort constrained to be 1 at 1950, period slope to be 0. Blue: Men, red: Women.*

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev  Df Deviance  Pr(>Chi)
  Age                    3347    10071.6
  Age-drift              3346     6504.6   1   3567.0 < 2.2e-16
  Age-Cohort             3335     6252.6  11    252.0 < 2.2e-16
  Age-Period-Cohort      3331     5106.7   4   1145.9 < 2.2e-16
  Age-Period             3342     5336.5 -11   -229.9 < 2.2e-16
  Age-drift              3346     6504.6  -4  -1168.0 < 2.2e-16
> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( apcM, lty=1, ci=TRUE, col="blue", r.txt="Rate per 1000 PY" )
  cp.offset    RR.fac
       1790         1
> lines( apcF, lty=1, ci=TRUE, col="red" )
```

Both from figure **??** and **??** it is clear that there is some calendar-time effect at around 2005, where a downward change in incidence rates seem to occur. The major tendency is however the steady increase across cohort/period.

If we stick to the period-major parametrization as in figure **??**, we are essentially referring to cross-sectional rates, and they seem to have a peak around age 80. However since there

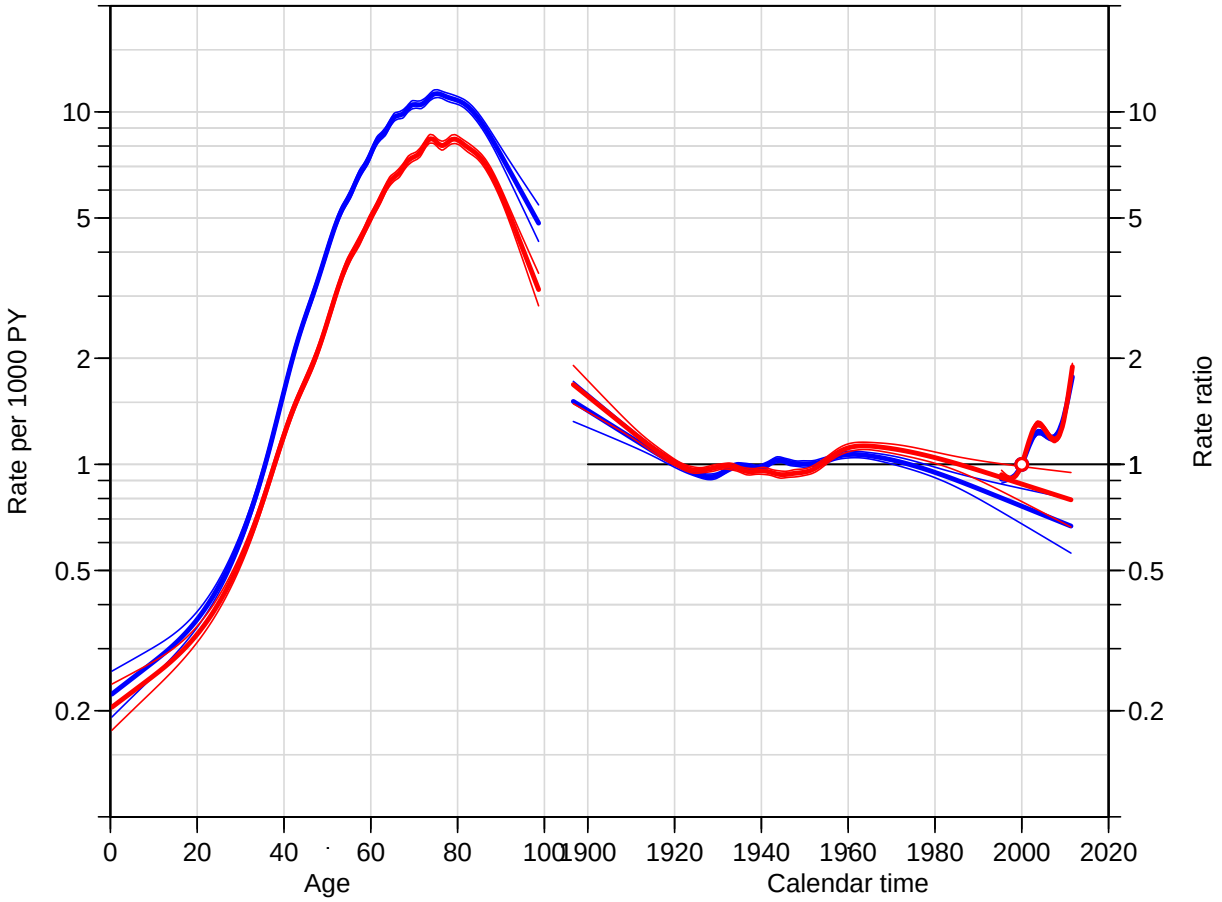

Figure 3.2: *Estimates from an APC-model for DM incidence rates in Denmark 1995–2011 (revised definition), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

is an increasing trend the peak incidence for a given generation is more likely at 85 years as shown in figure **??**, using the cohort major parametrization, the longitudinal approach.

3.2.3 Time-trends in rates

The overall time trend in the rates are in the `Drift` component of the `apc` object, here we give the average annual increase in incidence rates among men and women:

```
> pctchg <- (cbind( apcM$Drift, apcF$Drift )-1)*100
> colnames( pctchg ) <- c("Men","lo","up","Women","lo","up")
> round( pctchg, 2 )

      Men   lo   up Women   lo   up
APC 3.59 3.48 3.70  3.61 3.49 3.73
A-d 3.59 3.49 3.69  3.40 3.29 3.51
```

Thus we see that the average annual trend in rates is about 4% per year, slightly higher for women than for men.

3.2.4 Summary of the APC modelling

The deviance analysis of the model did not surprisingly show that both cohort and period have non-linear effects, however this formal significance is largely due to the large data base.

```
> par( mfrow=c(2,1), mar=c(0,4,0,4), oma=c(3,0,1,0), mgp=c(3,1,0)/1.6, las=1 )
> apc.frame( a.lab=seq(20,80,20), a.tic=c(5,seq(10,90,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=outer(c(1,2,5),10^(-1:1)), tic.fac=2,
+            r.tic=c(outer(c(5:9/10,1:5),10^(-1:1)),60), rr.ref=1,
+            gap=10, r.txt="DM incidence per 1000 PY", side=c(2,4) )
> lines( acpM, col="blue", ci=TRUE )
> lines( acpF, col="red" , ci=TRUE )
> apc.frame( a.lab=seq(20,80,20), a.tic=c(5,seq(10,90,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=outer(c(1,2,5),10^(-1:1)), tic.fac=2,
+            r.tic=c(outer(c(5:9/10,1:5),10^(-1:1)),60), rr.ref=1,
+            gap=10, r.txt="DM incidence per 1000 PY", side=c(1,2,4) )
> lines( apcM, col="blue", ci=TRUE )
> lines( apcF, col="red" , ci=TRUE )
```

In figure 3.3 is shown the same model in two different parametrizations, one with longitudinal and one with cross-sectional age-specific rates. Another way of visualizing the model is to show the estimated age-specific incidence rates for different birth cohorts.

To that end we use the model-objects returned by the `apc.fit` function to produce predicted rates. So we set up a prediction frame with ages for 15 different cohorts:

```
> prf <- data.frame( A = rep( c(NA,0:98), 8 ),
+                    C = rep( seq(1910,1980,10), each=100 ),
+                    Y = 1 )[-1,]
> prf <- transform( prf, P = C + A )
```

The we can make a fit of the models of relevance and make predictions based on this new frame. [1]

```
> Mapc <- glm( D ~ Ns(   A, kn=apcM$Knots$Age ) +
+                  Ns( P-A, kn=apcM$Knots$Coh ) +
+                  Ns( P  , kn=apcM$Knots$Per ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="M") )
> Map  <- glm( D ~ Ns( A, kn=apcM$Knots$Age ) +
+                  Ns( P, kn=apcM$Knots$Per ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="M") )
> Mac  <- glm( D ~ Ns(   A, kn=apcM$Knots$Age ) +
+                  Ns( P-A, kn=apcM$Knots$Coh ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="M") )
> Fapc <- glm( D ~ Ns(   A, kn=apcF$Knots$Age ) +
+                  Ns( P-A, kn=apcF$Knots$Coh ) +
+                  Ns( P  , kn=apcF$Knots$Per ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="F") )
> Fap  <- glm( D ~ Ns( A, kn=apcF$Knots$Age ) +
+                  Ns( P, kn=apcF$Knots$Per ),
```

[1] Note that we cannot use the returned model from the `apc` object since this is defined in terms specific matrices and *not* in terms of A, P and C:

```
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="F") )
> Fac  <- glm( D ~ Ns(   A, kn=apcF$Knots$Age ) +
+                  Ns( P-A, kn=apcF$Knots$Coh ),
+              offset = log( Y ),
+              family = poisson,
+                data = subset( DD, sex=="F") )
> summary( fitted( apcM$Model ) - fitted( Mapc ) )
       Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
-1.137e-12 -9.237e-14  1.332e-15  4.925e-15  8.438e-14  1.705e-12
> summary( fitted( apcF$Model ) - fitted( Fapc ) )
       Min.     1st Qu.      Median        Mean     3rd Qu.        Max.
-8.811e-13 -1.386e-13 -4.108e-15 -2.383e-14  7.816e-14  7.958e-13
```

From the last summary we see that the models are the same as those fitted by `apc.fit`, an moreover we can use this latter to make predictions, regardless of the overparametrization (we will get a warning, though). Recall that the `Y` was scaled to be person-millenia, so we get fitted values as rates per 1000 (namely the expected numbers based on the model for a data point where `Y` is equal to 1, as specified in `prf`):

```
> prr <- subset( prf, (P<2011 & P>1995) | is.na(P) )
> Mfit.apc <- predict( Mapc, newdata=prr )
> Mfit.ap  <- predict( Map , newdata=prr )
> Mfit.ac  <- predict( Mac , newdata=prr )
> Ffit.apc <- predict( Fapc, newdata=prr )
> Ffit.ap  <- predict( Fap , newdata=prr )
> Ffit.ac  <- predict( Fac , newdata=prr )
```

For comparison we overlay empirical rates, which we compute for the cohorts 1910 (born 1905–15), . . . , 1980 (born 1975–85) calculated in C-sets ( ); the `gc` and `gp` are the midpoints of the cohort and period in the C-sets:

```
> DD.x <- transform( DD,
+                    gc = floor(((P-A)-1905)/10)*10+1910,
+                    gp = floor(P)+0.5 )
> ee <- data.frame( xtabs( cbind(D,Y) ~ sex + gp + gc,
+                   data = subset( DD.x, gc>1905 & gc<1985 ) ) )
> ee <- reshape( ee, timevar = "Var4",
+                      idvar = c("sex","gp","gc"),
+                        dir = "wide" )
> names( ee )[4:5] <- c("D","Y")
> ee <- transform( ee, gp = as.numeric(as.character(gp)),
+                      gc = as.numeric(as.character(gc)) )
> str( ee )
  'data.frame':        272 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ gp : num  1996 1996 1996 1996 1998 ...
   $ gc : num  1910 1910 1910 1910 1910 1910 1910 1910 1910 1910 ...
   $ D  : num  540 867 427 727 368 615 308 592 293 500 ...
   $ Y  : num  53 104.9 45.6 94.2 38.9 ...
```

We then overlay the empirical over the fitted rates from the three different models, the age-period, the age-cohort and the apc-model:

```
> par( mfrow=c(2,1), mar=c(0,0,0,0), oma=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> matplot( prr$A, exp(Mfit.apc), type="l", lty=1,
+          log="y", ylim=c(0.2,25), lwd=3, xaxt="n", xlab="", ylab="" )
> matlines( prr$A, exp(Mfit.ap), type="l", lty="11", lwd=2 )
> matlines( prr$A, exp(Mfit.ac), type="l", lty=1, lwd=2 )
```

```
> with( subset(ee,sex=="M"),
+       points( gp-gc, D/Y, pch=16, col=rainbow(8)[factor(gc)], cex=0.8 ) )
> text( 20, 14, "Men", col="blue" )
> matplot( prr$A, exp(Ffit.apc), type="l", lty=1,
+          log="y", ylim=c(0.2,25), lwd=3, xlab="", ylab="" )
> matlines( prr$A, exp(Ffit.ap), type="l", lty="11", lwd=2 )
> matlines( prr$A, exp(Ffit.ac), type="l", lty=1, lwd=2 )
> with( subset(ee,sex=="F"),
+       points( gp-gc, D/Y, pch=16, col=rainbow(8)[factor(gc)], cex=0.8 ) )
> text( 20, 14, "Women", col="red" )
> mtext( "DM incidence rate per 1000 PY", side=2, outer=TRUE, line=2, las=0 )
> mtext( "Age (years)", side=1, outer=TRUE, line=2 )
```

From figure 3.4 we see that both the fitted and the empirical rates are indicative of a strong period effect with a characteristic dip around 2003–5, as seen in figure 3.3, so the significant non-linearity of the period effect is epidemiologically significant, not only statistically.

Note that the age-specific incidence rates in figure 3.3 are constructed gluing together the age-effects from the different cohorts, and aligning them to the 1950 cohort (the one with light-blue empirical rates).

3.2.5 Saving the fitted models

We then save these fitted APC-models with different parametrizations:

```
> save( Mapc, Mac, Fapc, Fac, file="./data/inc-m.Rda" )
```

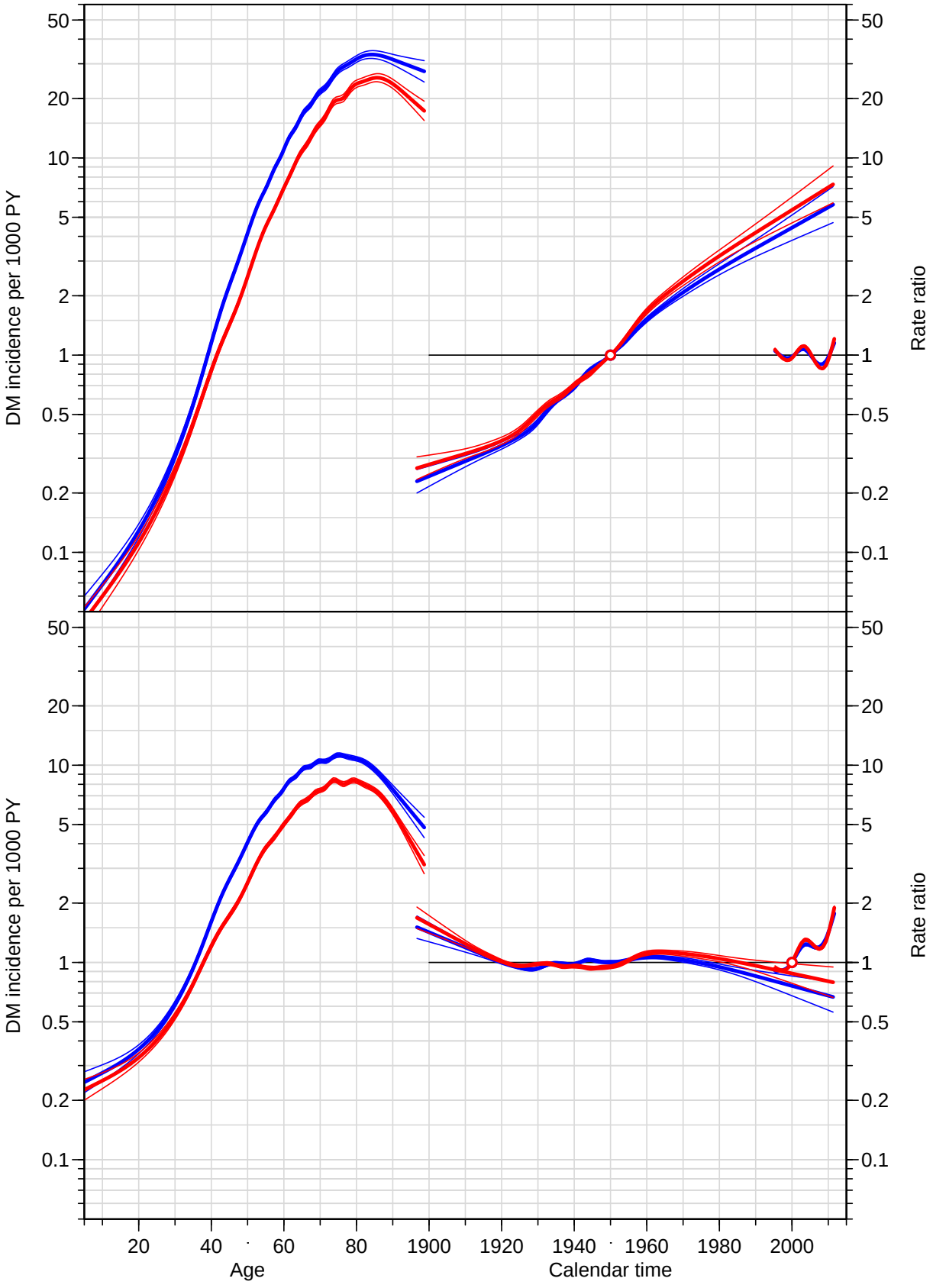

Figure 3.3: *Age-Period-Cohort models for DM incidence among men (blue) and women (red), using the same scaling in the two plots. The top panel is the parametrization with horizontal period effect and cohort reference 1950, bottom panel is the parametrization with horizontal cohort effect and period reference 2000.*

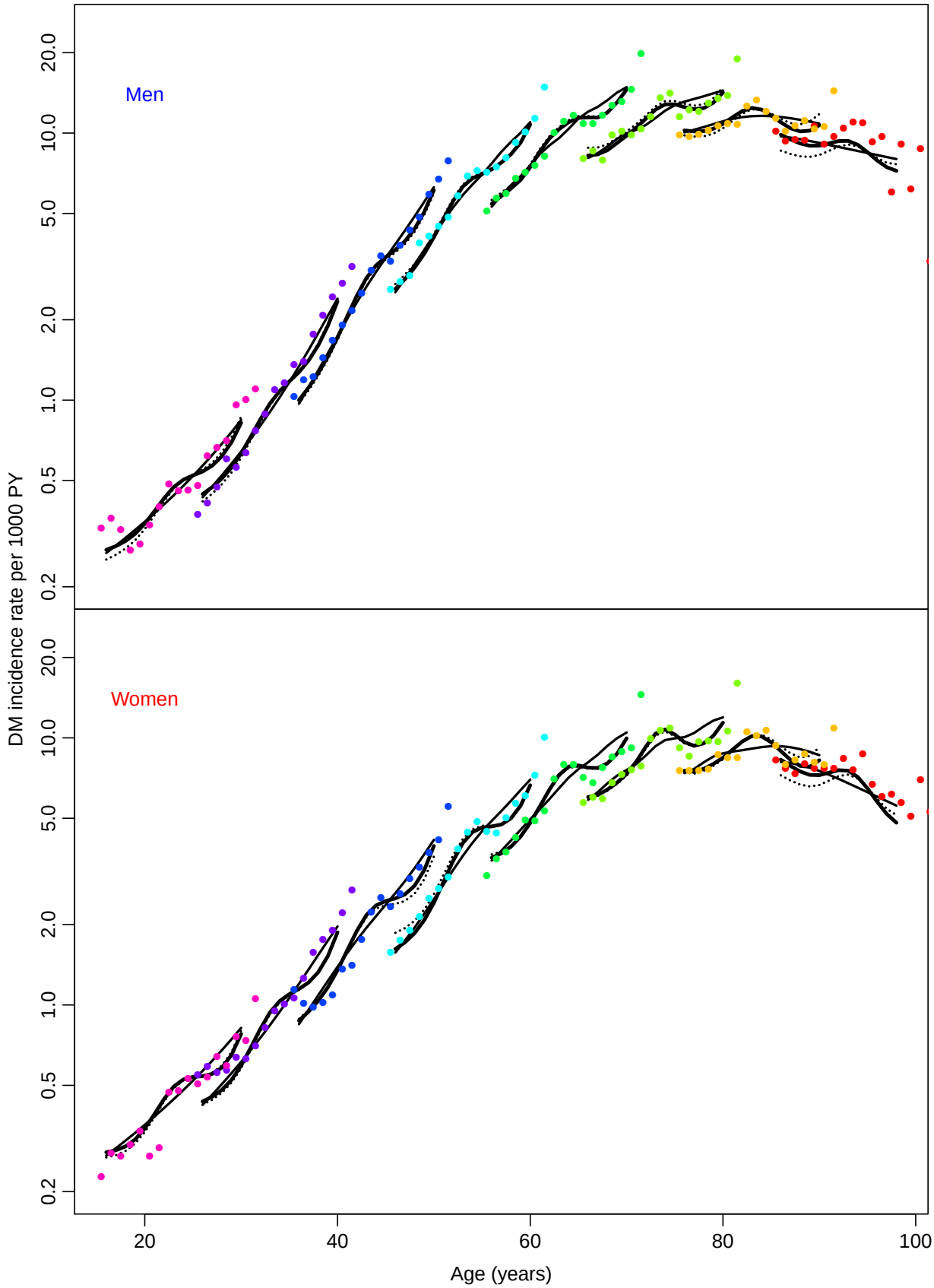

Figure 3.4: *Fitted age-specific incidence rates for the cohorts 1910,. . . ,1980: Full thick line: APC-model, broken line: AP model and full thin line: AC-model. Empirical age-specific rates from C-sets for 1-year period and 10-year cohorts are given as colored dots, colored separately for each cohort.*

3.3 Mortality

3.3.1 Mortality in non-diabetics

We are going to use `Y.nD` and `Y.nD` as response variables in the analysis of mortality rates, however we first need to define the age and period properly for analysis in Lexis triangles:

```
> nD <- subset( transform( TT, A = A + (1+U)/3,
+                              P = P + (2-U)/3,
+                              D = pmax(D.nD,0),
+                              Y = Y.nD/1000 )[,c("sex","A","P","D","Y")],
+               Y > 0 )
```

With this groomed data frame in place we can fit separate models for men and women and plot the estimates together:

```
> nDacpM <- apc.fit( subset(nD,sex=="M"),
+                    ref.c=1950,
+                     parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3346    13839.6
  Age-drift              3345     7573.4  1   6266.2 < 2.2e-16
  Age-Cohort             3336     6607.8  9    965.6 < 2.2e-16
  Age-Period-Cohort      3329     6565.1  7     42.8  3.69e-07
  Age-Period             3338     7539.7 -9   -974.6 < 2.2e-16
  Age-drift              3345     7573.4 -7    -33.7  1.96e-05

> nDacpF <- apc.fit( subset(nD,sex=="F"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3346    10510.3
  Age-drift              3345     7064.1  1   3446.1 < 2.2e-16
  Age-Cohort             3336     6140.8  9    923.3 < 2.2e-16
  Age-Period-Cohort      3329     6099.5  7     41.4 6.918e-07
  Age-Period             3338     7014.4 -9   -915.0 < 2.2e-16
  Age-drift              3345     7064.1 -7    -49.7 1.649e-08

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( nDacpF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )

  cp.offset    RR.fac
       1790        10

> lines( nDacpM, lty=1, ci=TRUE, col="blue" )
```

We also fit using the period-major parametrization:

```
> nDapcM <- apc.fit( subset(nD,sex=="M"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
```

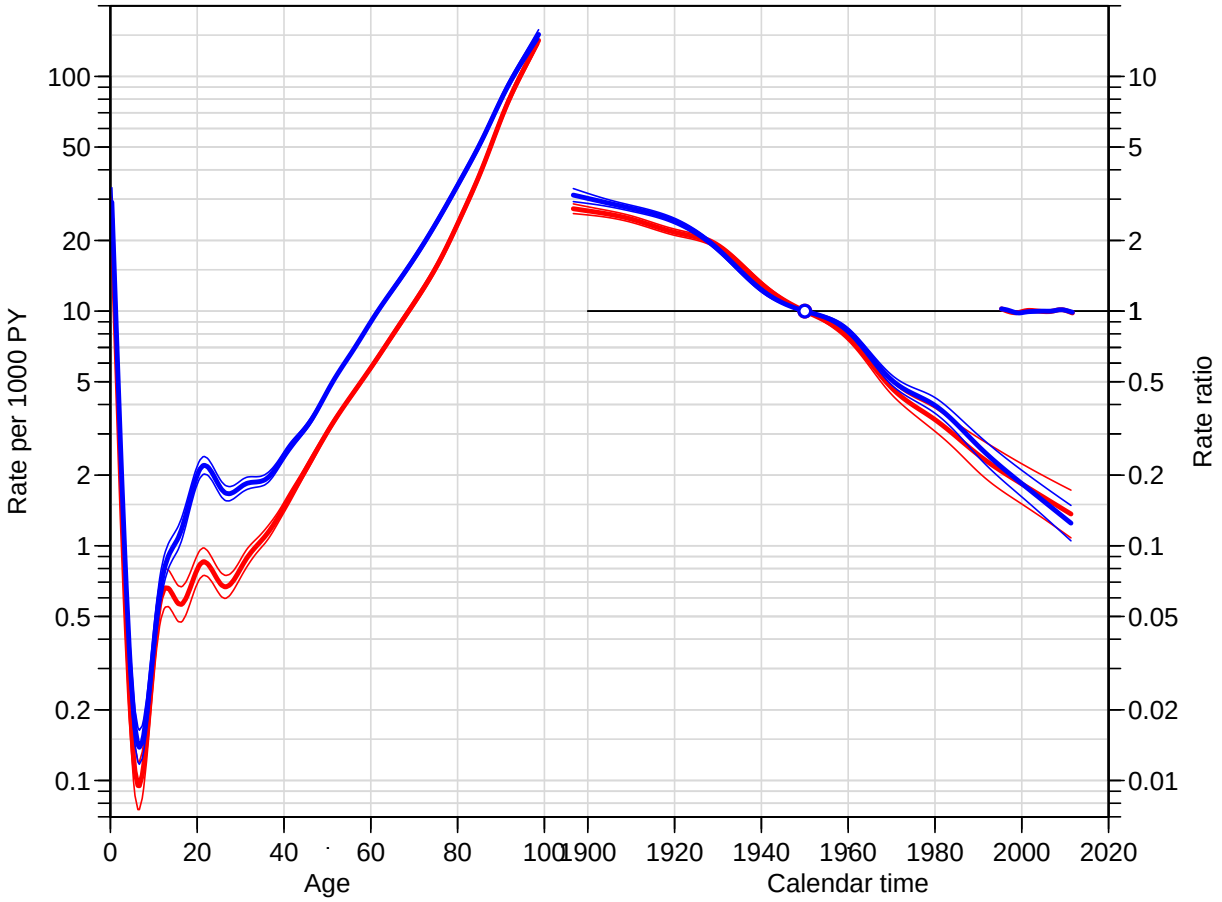

Figure 3.5: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), cohort effects constrained to be 1 at 1950, period slope to be 0. Blue: Men; red: Women.*

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

  Analysis of deviance for Age-Period-Cohort model

                     Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                     3346    13839.6
  Age-drift               3345     7573.4  1   6266.2 < 2.2e-16
  Age-Cohort              3336     6607.8  9    965.6 < 2.2e-16
  Age-Period-Cohort       3329     6565.1  7     42.8  3.69e-07
  Age-Period              3338     7539.7 -9   -974.6 < 2.2e-16
  Age-drift               3345     7573.4 -7    -33.7  1.96e-05
> nDapcF <- apc.fit( subset(nD,sex=="F"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                     Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                     3346    10510.3
  Age-drift               3345     7064.1  1   3446.1 < 2.2e-16
  Age-Cohort              3336     6140.8  9    923.3 < 2.2e-16
  Age-Period-Cohort       3329     6099.5  7     41.4 6.918e-07
```

```
  Age-Period             3338     7014.4 -9   -915.0 < 2.2e-16
  Age-drift              3345     7064.1 -7    -49.7 1.649e-08

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( nDapcF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )

  cp.offset     RR.fac
       1790        100

> lines( nDapcM, lty=1, ci=TRUE, col="blue" )
```

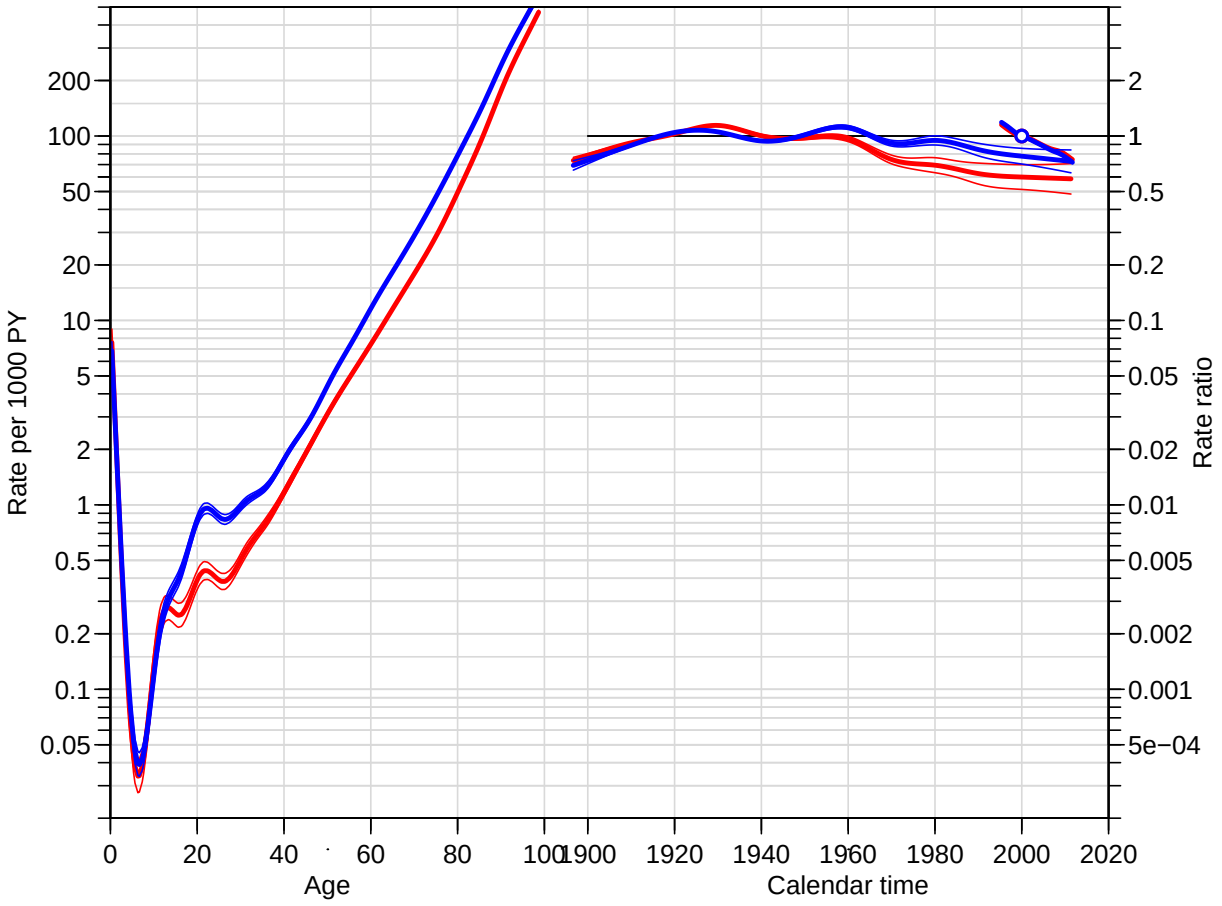

Figure 3.6: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

3.3.2 Mortality among DM patients

Here we use `D.DM` and `Y.DM` as response variables in the analysis of mortality rates among non-diabetics, and again we first need to define the age and period properly:

```
> DM <- subset( transform( TT, A = A + (1+U)/3,
+                              P = P + (2-U)/3,
+                              D = pmax(D.DM,0),
+                              Y = Y.DM/1000 )[,c("sex","A","P","D","Y")],
+               Y > 0 )
```

With this groomed data frame in place we can fit separate models for men and women and plot the estimates together:

```
> DMacpM <- apc.fit( subset(DM,sex=="M"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                    3340     5343.5
  Age-drift              3339     3032.4  1  2311.17  < 2e-16
  Age-Cohort             3330     2881.6  9   150.74  < 2e-16
  Age-Period-Cohort      3323     2859.8  7    21.86  0.00269
  Age-Period             3332     3015.2 -9  -155.45  < 2e-16
  Age-drift              3339     3032.4 -7   -17.14  0.01649

> DMacpF <- apc.fit( subset(DM,sex=="F"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )

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

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                    3334     4483.5
  Age-drift              3333     2962.3  1  1521.23 < 2.2e-16
  Age-Cohort             3324     2829.8  9   132.51 < 2.2e-16
  Age-Period-Cohort      3317     2789.5  7    40.34 1.081e-06
  Age-Period             3326     2935.9 -9  -146.42 < 2.2e-16
  Age-drift              3333     2962.3 -7   -26.43 0.0004213

> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( DMacpF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )

  cp.offset    RR.fac
       1790        10

> lines( DMacpM, lty=1, ci=TRUE, col="blue" )
```

We also fit using the period-major parametrization:

```
> DMapcM <- apc.fit( subset(DM,sex=="M"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
  [1] "ML of APC-model Poisson with log(Y) offset : ( APC ):\n"

  Analysis of deviance for Age-Period-Cohort model

                    Resid. Df Resid. Dev Df Deviance Pr(>Chi)
  Age                    3340     5343.5
  Age-drift              3339     3032.4  1  2311.17  < 2e-16
  Age-Cohort             3330     2881.6  9   150.74  < 2e-16
  Age-Period-Cohort      3323     2859.8  7    21.86  0.00269
  Age-Period             3332     3015.2 -9  -155.45  < 2e-16
  Age-drift              3339     3032.4 -7   -17.14  0.01649

> DMapcF <- apc.fit( subset(DM,sex=="F"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=seq(1,96,5),P=seq(1995,2011,2),C=seq(1900,2000,10)) )
```

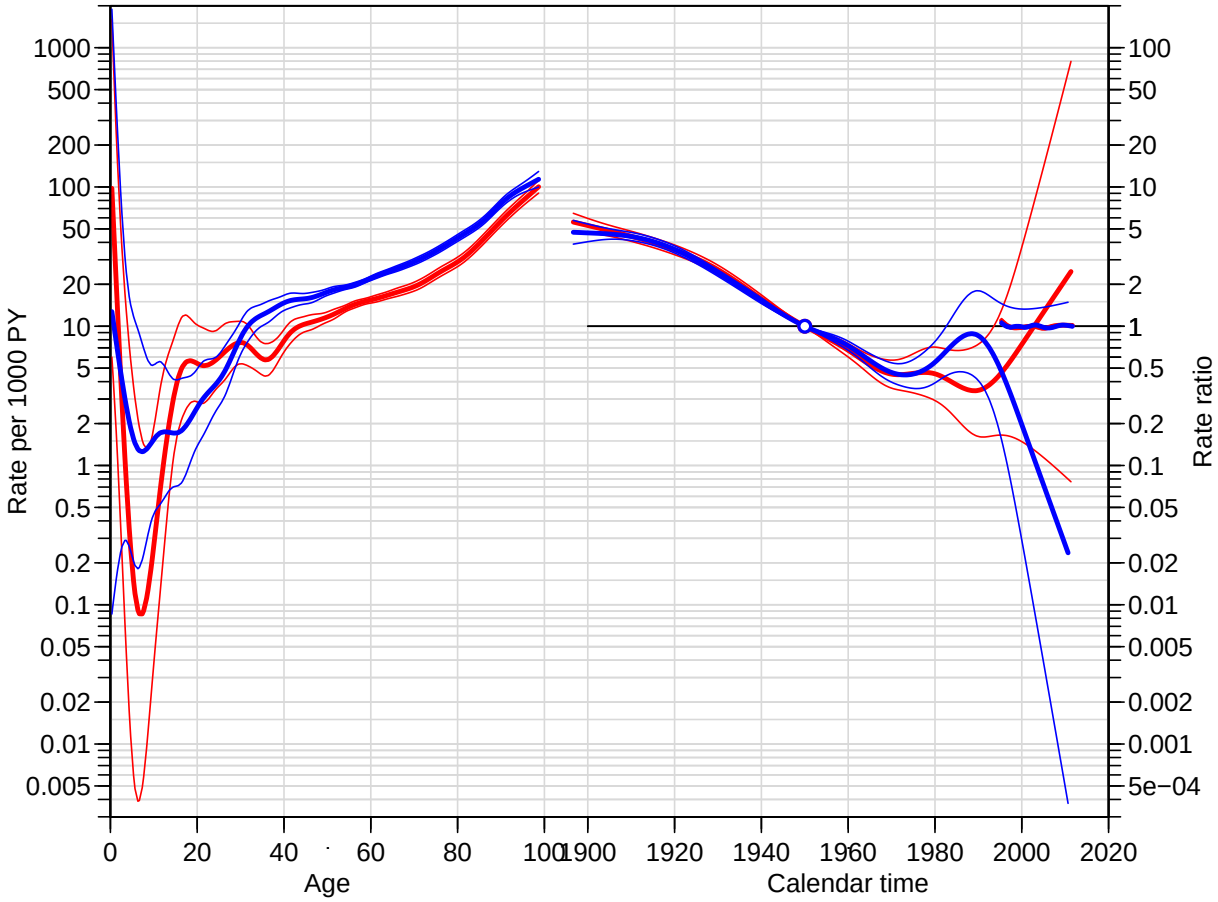

Figure 3.7: *Estimates from an APC-model for mortality among DM patients in Denmark 1995–2011 (original definition), cohort constrained to be 1 at 1950, period slope to be 0. Blue: Men, red: Women.*

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

  Analysis of deviance for Age-Period-Cohort model

                     Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
  Age                     3334     4483.5
  Age-drift               3333     2962.3  1  1521.23 < 2.2e-16
  Age-Cohort              3324     2829.8  9   132.51 < 2.2e-16
  Age-Period-Cohort       3317     2789.5  7    40.34 1.081e-06
  Age-Period              3326     2935.9 -9  -146.42 < 2.2e-16
  Age-drift               3333     2962.3 -7   -26.43 0.0004213
> par( mar=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> plot ( DMapcF, lty=1, ci=TRUE, col="red", r.txt="Rate per 1000 PY" )
  cp.offset    RR.fac
       1790         1
> lines( DMapcM, lty=1, ci=TRUE, col="blue" )
```

3.3.3 Summary of the APC models for mortality

The deviance analysis of the model did not surprisingly show that both cohort and period have non-linear effects, however this formal significance is largely due to the large data

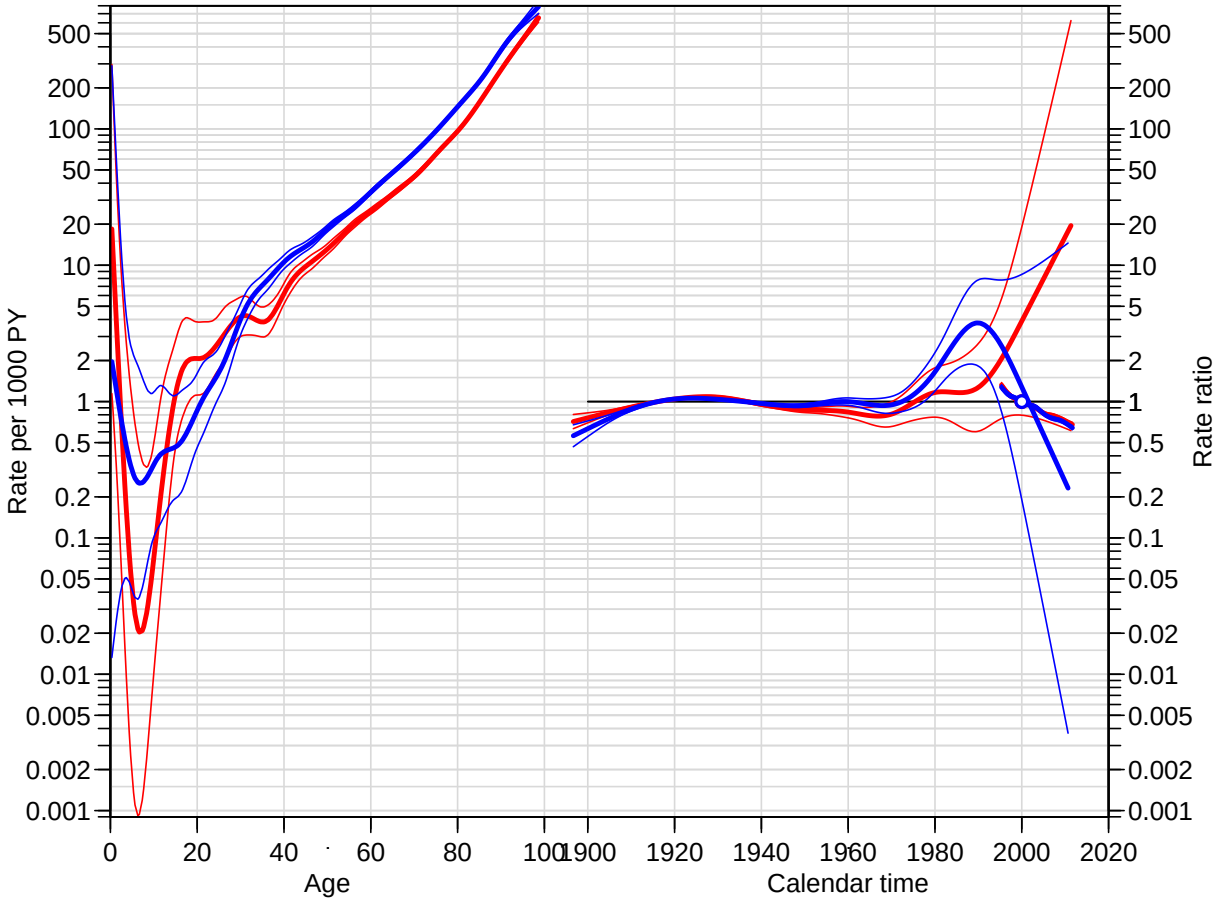

Figure 3.8: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

base, clearly there is no eidemiologically significant period-effect.

```
> par( mfrow=c(2,1), mar=c(0,4,0,4), oma=c(3,0,1,0), mgp=c(3,1,0)/1.6, las=1 )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+            r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=10,
+            gap=5, r.txt="Non-DM mortality per 1000 PY", side=c(2,4) )
> lines( nDacpM, col="blue", ci=TRUE )
> lines( nDacpF, col="red" , ci=TRUE )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+            r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=10,
+            gap=5, r.txt="DM mortality per 1000 PY", side=c(1,2,4) )
> lines( DMacpM, col="blue", ci=TRUE )
> lines( DMacpF, col="red" , ci=TRUE )
```

3.3.4 Time-trends in mortality rates

We can extract the timetrends for diabetics and non-diabetics by sex, and print the annual percentwise change:

```
> DA <- NArray( c( list( who = c("non-DM","DM"),
+                       sex = c("M","F") ),
+                  dimnames( nDacpM$Drift ) ) )
> DA["non-DM","M",,] <- nDacpM$Drift
> DA["non-DM","F",,] <- nDacpF$Drift
> DA[    "DM","M",,] <- DMacpM$Drift
> DA[    "DM","F",,] <- DMacpF$Drift
> round( ftable( (DA-1)*100, row.vars=1:2 ), 1 )
                     APC                       A-d
            exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
 who    sex
 non-DM M        -2.8 -2.9  -2.7      -2.5 -2.6  -2.4
        F        -2.4 -2.5  -2.3      -1.8 -1.9  -1.8
 DM     M        -3.8 -4.0  -3.6      -3.6 -3.8  -3.5
        F        -3.5 -3.7  -3.4      -3.1 -3.3  -3.0
```

We see that there is not much difference in the overall trend between man and women, but there seem to be a substantially steeper decrease in mortality among diabetes patienst than among non-diabetes patients.

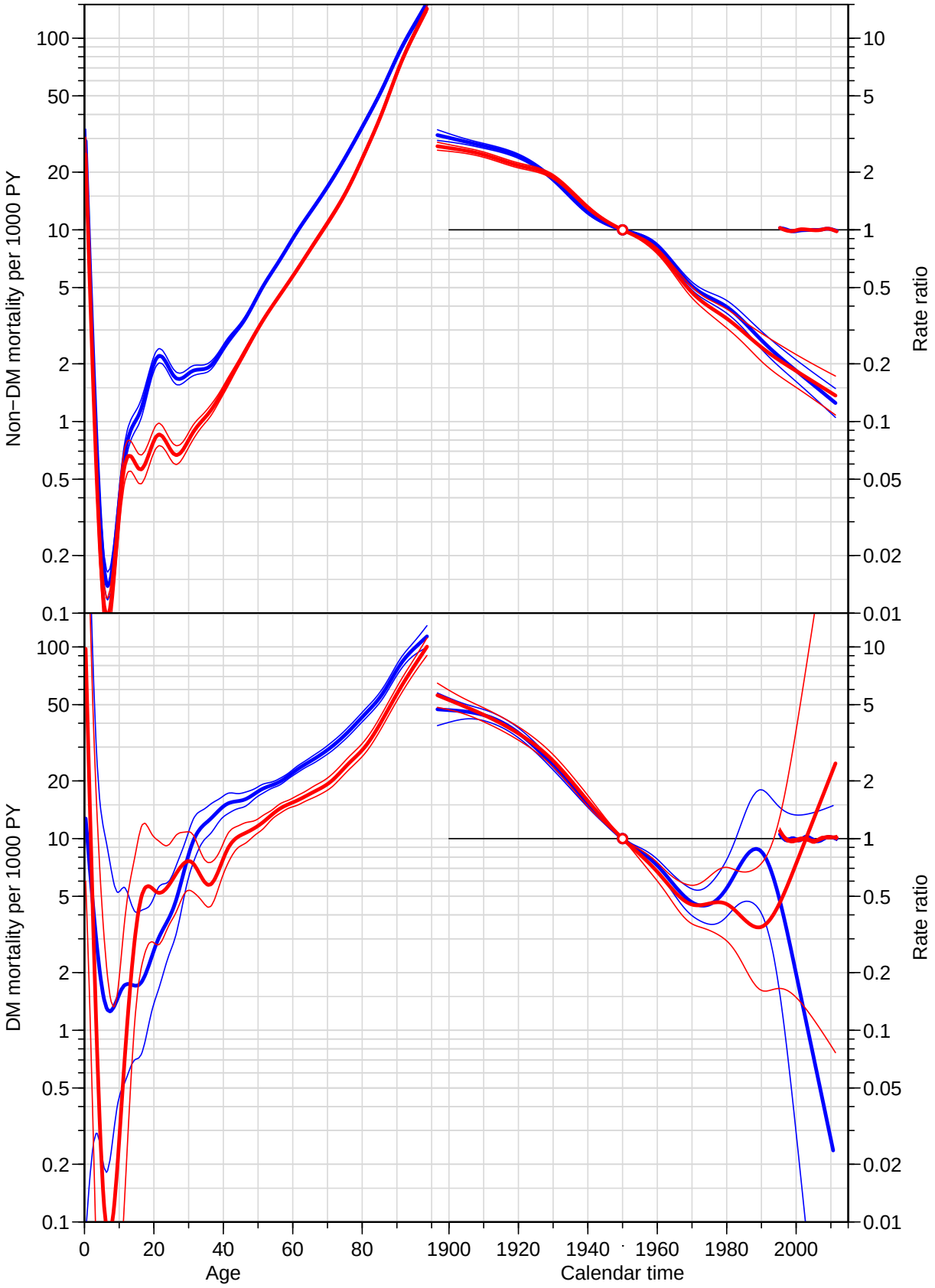

Figure 3.9: *Age-Period-Cohort models for mortality among men (blue) and women (red). Top panel is the mortality among non-diabetics and the lower panel is the mortality among diabetes patients.*

3.3.5 SMR

Since we have modelled both mortality rates by APC-models, and the analyses are done on (conditionally) independent datasets (follow-up in non-DM-, resp. DM-state), the ratio of the rates will also follow an APC-model, and the ratio of each set of effects will give three sets of RRs which wil mulitply to the overall RR. Since we have chosen the same reference cohort for both analyses, the cohort effect on the RR will also be with this reference. However, there is no *a priori* guarantee thate the period effect on the RR will be perfectly horizontal on average, even though it is going to be close.

However we will need a machinery to extract the RRs from the `apc` objects:

```
> make.RR.apc <-
+ function( a, b )
+ {
+   make.RR <-
+   function(A,B)
+   {
+   Z <- merge( A, B, by.x=1, by.y=1 )
+   lA <- log(Z[,2])
+   sA <- log(Z[,4]/Z[,3])/(2*1.96)
+   lB <- log(Z[,5])
+   sB <- log(Z[,7]/Z[,6])/(2*1.96)
+   RR <- cbind( A[,1], exp( lA-lB ),
+                       exp( lA-lB - 1.96*sqrt(sA^2+sB^2) ),
+                       exp( lA-lB + 1.96*sqrt(sA^2+sB^2) ) )
+   }
+ RR <- list( Age = make.RR( a$Age, b$Age ),
+             Per = make.RR( a$Per, b$Per ),
+             Coh = make.RR( a$Coh, b$Coh ),
+             Ref = a$Ref )
+ class( RR ) <- "apc"
+ RR
+ }
> SMR.M <- make.RR.apc( DMacpM, nDacpM )
> SMR.F <- make.RR.apc( DMacpF, nDacpF )
```

The two objects are not "real" `apc` objects, but they have the class attribute and the have the elements `Age`, `Per` and `Coh`, which are the only ones used by the `lines.apc` function. Hence we can plot the mortality rates for the DM patients together with the SMR relative to the non-diabetics.

```
> par( mfrow=c(2,1), mar=c(0,4,0,4), oma=c(3,0,1,0), mgp=c(3,1,0)/1.6, las=1 )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+            r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=10,
+            gap=5, r.txt="DM mortality per 1000 PY", side=c(2,4) )
> lines( DMacpM, col="blue", ci=TRUE )
> lines( DMacpF, col="red" , ci=TRUE )
> apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+           cp.lab=seq(1900,2000,20), cp.tic=c(seq(1900,2010,10),2015),
+            r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+            r.tic=c(outer(1:9,10^(-1:1)),100,150), rr.ref=1,
+            gap=5, r.txt="SMR DM vs. non-DM", rr.txt="RR ratio", side=c(1,2,4) )
> abline( h=1 )
> lines( SMR.M, col="blue", ci=TRUE )
> lines( SMR.F, col="red" , ci=TRUE )
```

We see that the SMR is decreasing by age, and there seems to be no non-linear period effect, but an overall decreasing trend by period/birth cohort. Figure 3.10 shows a decrease in SMR from about 5 in age 40 to aound 1 in age 80 for the 1950 cohort. Note however that this is a bit of an extrapolation; the 1950 cohort has only been observed in ages 45 to 62.

3.3.6 Saving mortality rates

Finally, we save the `apc`-objects for subsequent use, however only the `ACP`-parametrized ones:

```
> save( nDacpM,
+       nDacpF,
+       DMacpM,
+       DMacpF, file="./data/APC-mort-m.Rda" )
```

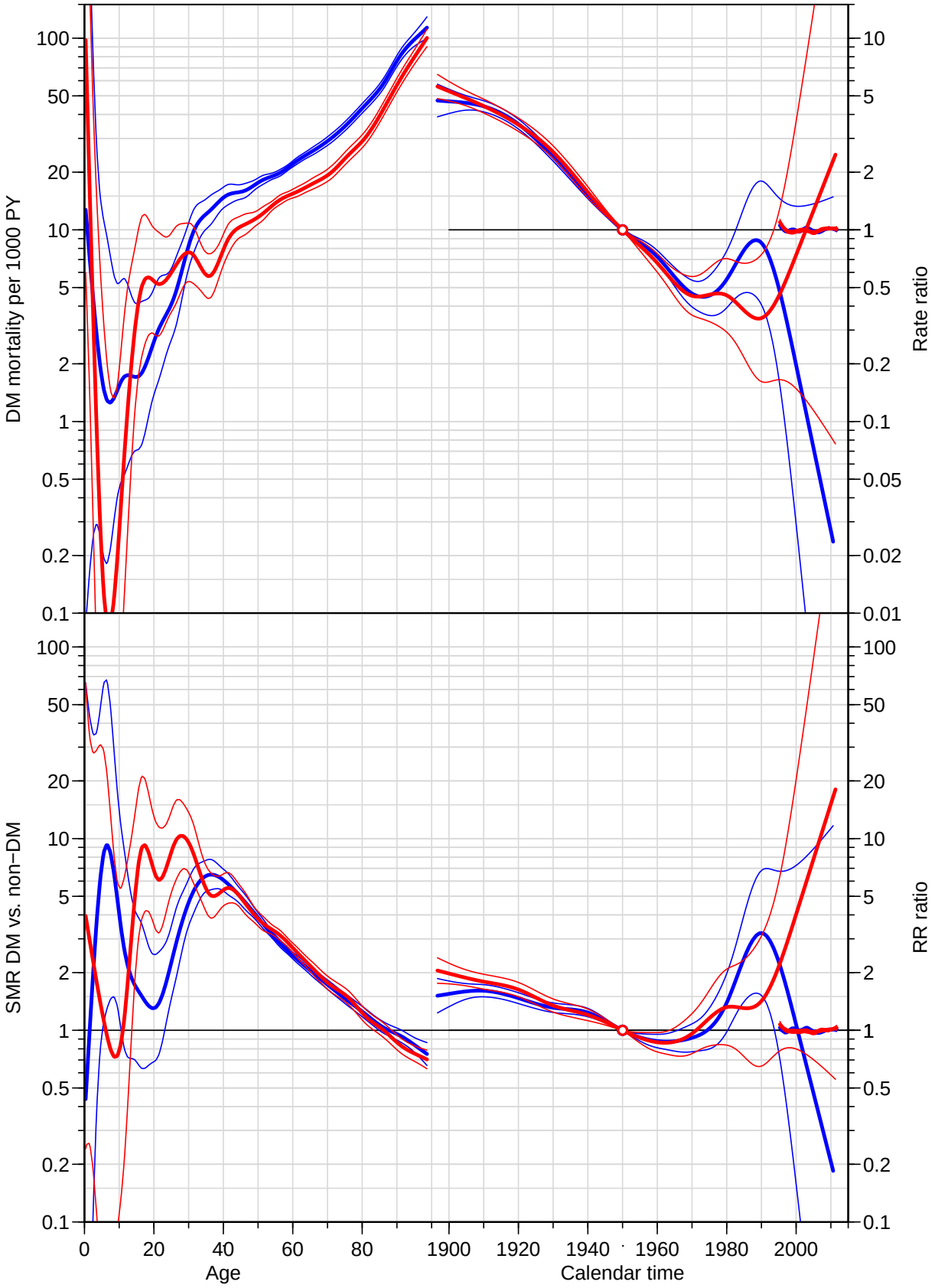

Figure 3.10: *Age-Period-Cohort models for mortality among men (blue) and women (red). Top panel is the mortality among diabetes patients, and the lower panel is the SMR versus the non-diabetic population.*

3.4 Prevalence of diabetes

We will analyze age-specific prevalence for each sex and each 1st January 1995—2012 separately, even though they are not independent.

First we set up a table of prevalent cases for each of the dates 1 January 1995–2012:

```
> pr <- NULL
> for( y in 1995:2012 )
+ pr <- rbind( pr,
+       cbind( with( subset( Lx, doDM < y & dox > (y-1/400) ),
+                    data.frame( table( sex, A=floor(y-foddto) ) ) ),
+              P = y ) )
> pr <- pr[,c(1,2,4,3)]
> pr$A <- as.numeric( as.character( pr$A ) )
> names( pr )[4] <- "X"
> str( pr )
  'data.frame':        3564 obs. of  4 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
   $ P  : int  1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
   $ X  : int  3 0 4 3 6 6 5 4 12 10 ...
```

Then we merge it with the population data:

```
> data( N.dk )
> head( N.dk )
    sex A    P     N
1   1 0 1971 35839
2   2 0 1971 34108
3   1 1 1971 36302
4   2 1 1971 34153
5   1 2 1971 37855
6   2 2 1971 35609
> N.dk <-  subset( N.dk, A<100 & P>1994 & P<2013 )
> N.dk$sex <- factor( N.dk$sex, labels=c("M","F") )
> str(N.dk)
  'data.frame':        3600 obs. of  4 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
   $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
   $ P  : num  1995 1995 1995 1995 1995 ...
   $ N  : num  35612 34094 34747 32967 35082 ...
> pr <- merge( pr, N.dk, all.y=TRUE )
> pr$X <- pmax( pr$X, 0, na.rm=TRUE )
> str( pr )
  'data.frame':        3600 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ A  : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P  : num  1995 1996 1997 1998 1999 ...
   $ X  : num  3 1 0 1 2 1 1 0 1 1 ...
   $ N  : num  35612 36055 34853 34774 34076 ...
```

We now have the empirical prevalences in the data frame `pr`, (`X`—no. of cases of DM, `N`—population size) for each of the 18 dates 1.1.1995 – 1.1.2012 in 1-year intervals.

```
> save( pr, file="./data/prev-m.Rda" )
```

So we model the prevalences as of 1 January each of the years 1995—2012, as a smooth function of age using a log-link binomial model with a smooth spline with 16 knots.

For the practical location of the spline knots we also define a small function which from the number of knots derives reasonable quantiles:

```
> qn <- function( nk, bd=2 ) seq( from =   1/(bd*nk),
+                                   to = 1-1/(bd*nk),
+                                length = nk )
> qn( 10, 2 )
 [1] 0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
> qn( 10, 5 )
 [1] 0.0200000 0.1266667 0.2333333 0.3400000 0.4466667 0.5533333 0.6600000
 [8] 0.7666667 0.8733333 0.9800000
```

Using this we get:

```
> ( kp.a <- c( 10, with( pr, quantile( rep(A,X), qn(15) ) ) ) )
          3.333333%       10% 16.66667% 23.33333%       30% 36.66667% 43.33333%
         10         27         40         47         52         55         58         61
        50% 56.66667% 63.33333%       70% 76.66667% 83.33333%       90% 96.66667%
         64         66         69         72         74         78         81         87
```

We now set up an array to hold the smoothed prevalences:

```
> a.pt <- 0:99
> p.pt <- 1995:2012
> pr.fit <- NArray( list( sex = c("M","F"),
+                            A = a.pt,
+                            P = p.pt ) )
```

So once we have set up the array to hold the smoothed empirical prevalences we can fill them into `pr.fit`:

```
> for( sx in dimnames(pr.fit)[["sex"]] )
+ for( dt in dimnames(pr.fit)[["P"]] )
+ pr.fit[sx,,dt] <- predict( glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ),
+                                  family = binomial(link="log"),
+                                    data = subset( pr, sex==sx & P==as.numeric(dt) ) ),
+                             newdata = data.frame( A=a.pt ),
+                                type = "response" )
```

We can plot how the age-specific prevalences have evolved over time:

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> matplot( a.pt, pr.fit["M",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", lty=1, col="blue", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt, pr.fit["M",,]*100, type="l", lty=1, col="blue", lwd=c(1,2) )
> text( 25, 22, "Men", adj=c(0,1), col="blue", cex=1.2 )
> text( 89, pr.fit["M","89","1995"]* 99, "1995", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M","80","2012"]*101, "2012", col="blue", adj=c(0,0) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, pr.fit["F",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", lty=1, col="red", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt, pr.fit["F",,]*100, type="l", lty=1, col="red", lwd=c(1,2) )
> text( 25, 22, "Women", adj=c(0,1), col="red", cex=1.2 )
> text( 89, pr.fit["F","89","1995"]* 99, "1995", col="red", adj=c(1,1) )
> text( 80, pr.fit["F","80","2012"]*101, "2012", col="red", adj=c(1,0) )
> axis( side=1 )
> mtext( "Age", side=1, line=1, outer=T )
> mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
```

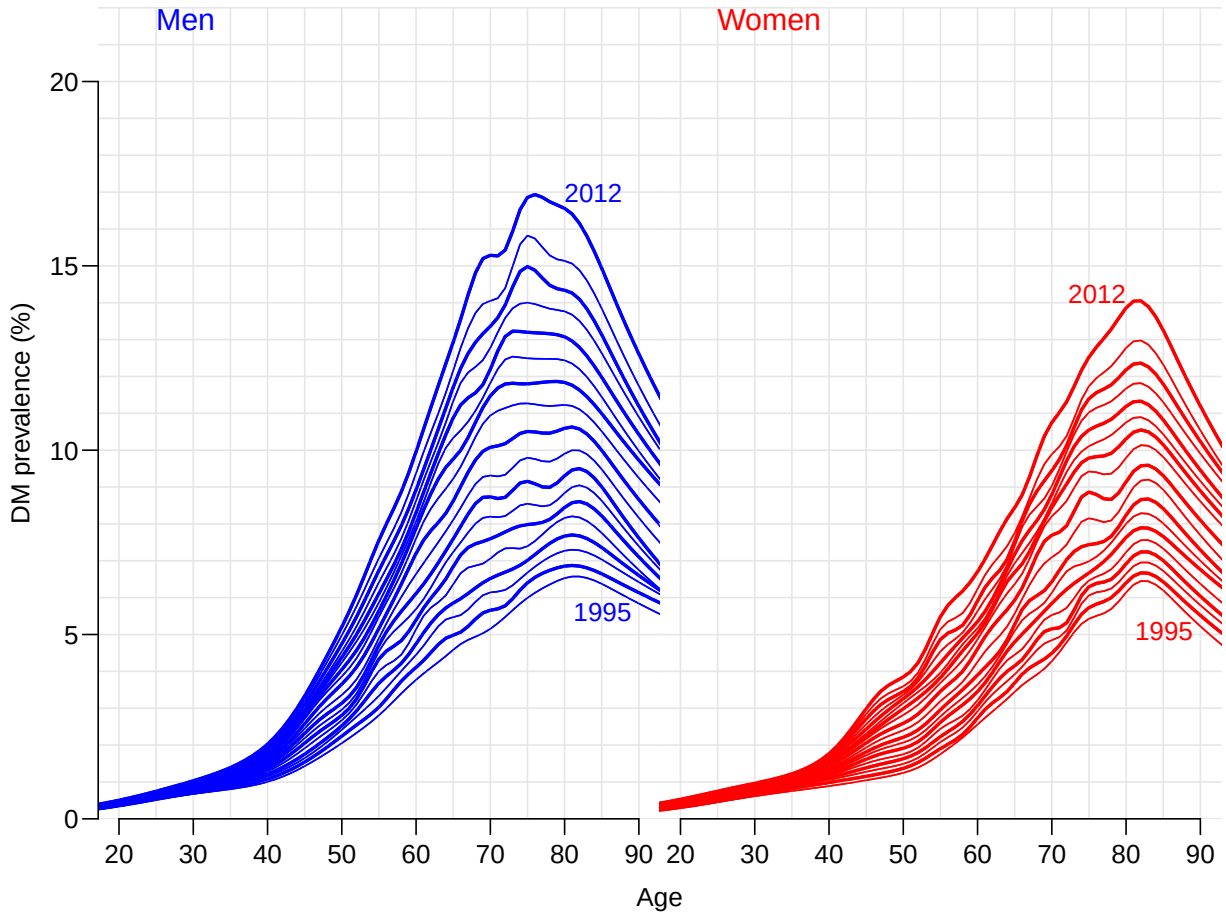

Figure 3.11: *Smoothed age-specific prevalences for the 17-year period 1995–2012. Blue is men, red is women.*

3.4.1 Trends in prevalence

A crude way of summarizing the prevalences is to assume that relative change is constant from year to year. So we set up a model that does this separately for men and women, and store the predicted values for comparison with those from the model with no assumption about the time evolution:

```
> pr.lfit <- pr.fit
> pr.chg <- NArray( list( dimnames(pr.fit)[["sex"]],
+                         c("% chg/y","lo","hi") ) )
> for( sx in dimnames(pr.fit)[["sex"]] )
+    {
+ lmod <- glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ) + P,
+               family = binomial(link="log"),
+                 data = subset( pr, sex==sx ) )
+ pr.chg[sx,] <- ( ci.exp( lmod, subset="P" ) - 1 ) * 100
+ pr.lfit[sx,,] <- predict( lmod,
+                           newdata = data.frame( A=rep(a.pt,     length(p.pt)),
+                                                 P=rep(p.pt,each=length(a.pt)) ),
+                              type = "response" )
+    }
```

This model is of course a simplification of the model above, with an arbitrary age-date interaction, so we can have a peep at how the predicted prevalences looks:

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> lblu <- rgb( 3,3,4,max=4 )
> lred <- rgb( 4,3,3,max=4 )
> matplot( a.pt, pr.fit["M",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="blue", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["M",,]*100, type="l", lty=1, col=lblu  , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["M",,]*100, type="l", lty=1, col="blue", lwd=c(2,3) )
> text( 25, 22, "Men", adj=c(0,1), col="blue", cex=1.2 )
> text( 89, pr.fit["M","89","1995"]* 99, "1995", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M","80","2012"]*101, "2012", col="blue", adj=c(0,0) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, pr.fit["F",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="red", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["F",,]*100, type="l", lty=1, col=lred , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["F",,]*100, type="l", lty=1, col="red", lwd=c(2,3) )
> text( 25, 22, "Women", adj=c(0,1), col="red", cex=1.2 )
> text( 89, pr.fit["F","89","1995"]* 99, "1995", col="red", adj=c(1,1) )
> text( 80, pr.fit["F","80","2012"]*101, "2012", col="red", adj=c(1,0) )
> axis( side=1 )
> mtext( "Age", side=1, line=1, outer=T )
> mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
```

From figure 3.12 we see that for men the summary using a constant relative change in prevalence is not a very good summary of the change in prevalences; it does not capture the change in the age of peak prevalence of men from 85 in 1995 to 75 in 2012. So the overall estimate of some 6% in relative annual increase of prevalences over the 17-year period 1995–2012, is not providing an adequate summary:

```
> round( pr.chg, 2 )
   % chg/y   lo   hi
M    5.35 5.31 5.38
F    5.09 5.06 5.13
```

3.4.2 Prevalence age-period interaction

Hence the relevant description of average changes per year would be using a model for the prevalences where we allowed the relative change to vary smoothly by age. This is done by including an interaction between a spline term in age and period, and the subsequently fishing out the relative change using a spline basis with a bit fewer knots to fish out the period multiplier.

It goes as follows, where we also as before extract the predicted values for comparison with the prevalence curves fitted separately for each year:

```
> ( kx.a <- c( 10, with( pr, quantile( rep(A,X), qn(5) ) ) ) )
    10% 30% 50% 70% 90%
 10  40  55  64  72  81
> CA <- Ns( 1:99, kn=kx.a, intercept=TRUE )
> A.chg <- NArray( list( A=1:99, c("Est","lo","hi"), sex=c("M","F") ) )
> for( sx in dimnames(pr.fit)[["sex"]] )
+   {
+ limod <- glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ) +
+                   I(P-2000):Ns( A, kn=kx.a, intercept=TRUE ),
```

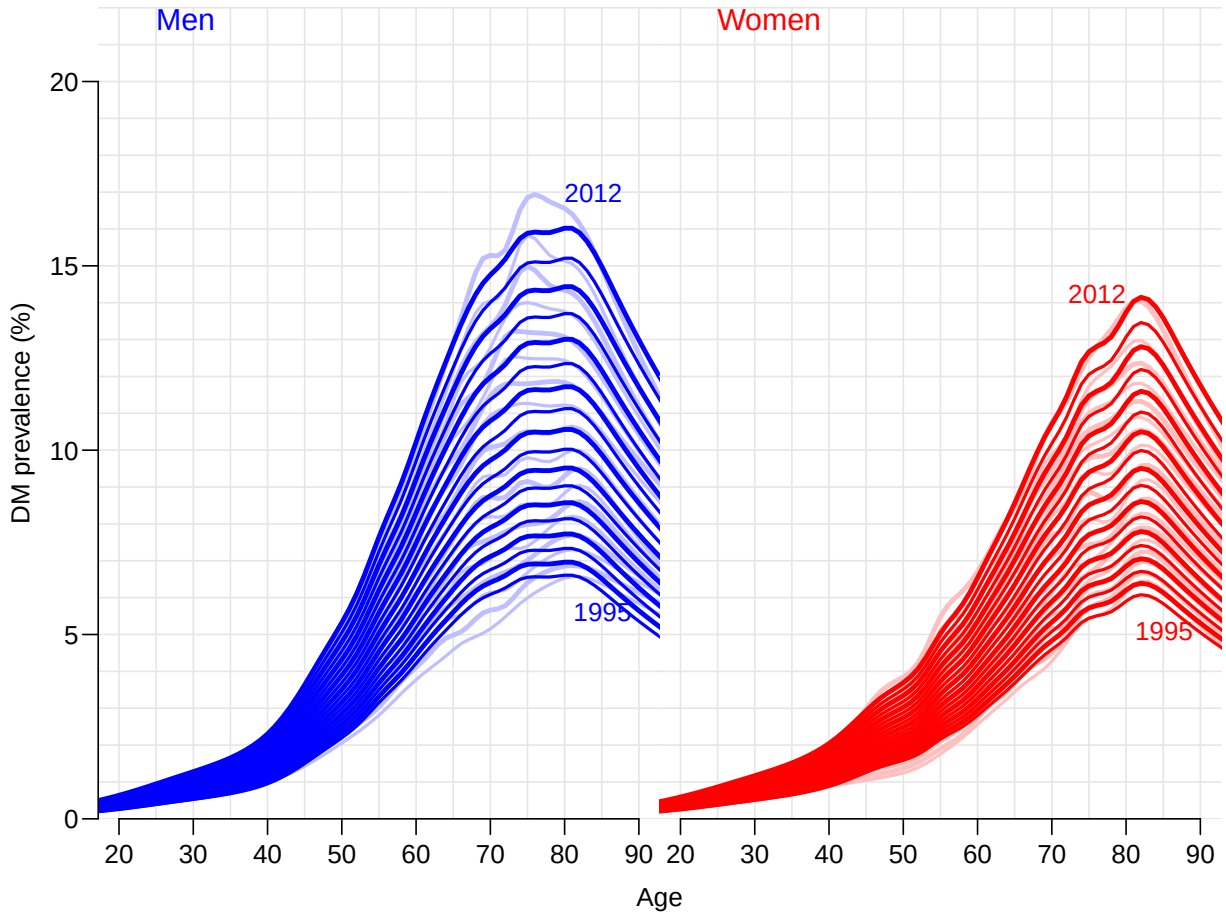

Figure 3.12: *Smoothed age-specific prevalences for the 17-year period 1995–2012 using a model with constant annual relative change in prevalences (full color) compared to the smoothing of the single years (pale color). Blue is men, red is women.*

```
+               family = binomial(link="log"),
+                 data = subset( pr, sex==sx ) )
+ A.chg[,,sx] <- ci.exp( limod, subset="P", ctr.mat=CA )
+ pr.lfit[sx,,] <- predict( limod,
+                           newdata = data.frame( A=rep(a.pt,     length(p.pt)),
+                                                 P=rep(p.pt,each=length(a.pt)) ),
+                              type = "response" )
+    }
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n" )
> matplot( 1:99, (cbind( A.chg[,,"M"], A.chg[,,"F"] )-1)*100,
+          col=rep(c("blue","red"),each=3), lwd=c(3,1,1), lty=1, type="l",
+          ylim=c(0,8), yaxs="i",
+          ylab="Annual change in DM prevalence (%)", xlab="Age" )
> abline( h=pr.chg[,1], col=c("blue","red") )
```

We can also as with the naïve linear change model show how the fitted values under this interaction model looks relative to the separate analyses by year (or full interaction model). The code is exactly as before, because we put the fitted values into the same structure as before:

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> lblu <- rgb( 3,3,4,max=4 )
```

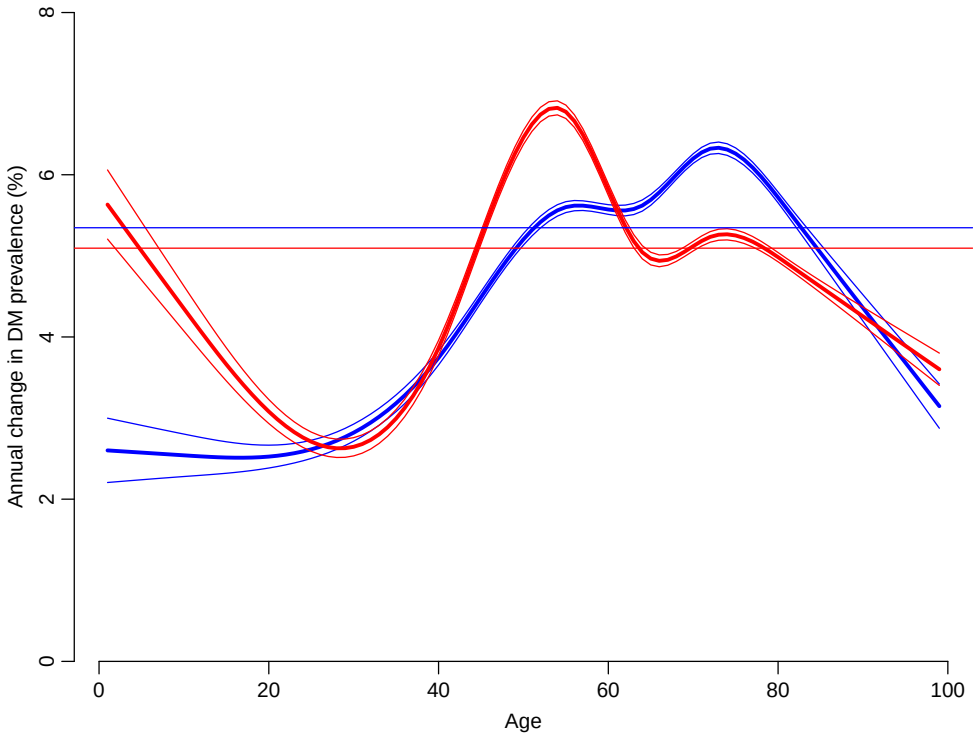

Figure 3.13: *The estimated change in prevalence in different ages, separately for men (blue) and women (red). The horizontal lines indicate the estimate from the naïve model with constant change for all ages.*

```
> lred <- rgb( 4,3,3,max=4 )
> matplot( a.pt, pr.fit["M",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="blue", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["M",,]*100, type="l", lty=1, col=lblu  , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["M",,]*100, type="l", lty=1, col="blue", lwd=c(2,3) )
> text( 25, 21.5, "Men", adj=0, col="blue", cex=1.2 )
> text( 89, pr.fit["M","89","1995"]* 99, "1995", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M","80","2012"]*101, "2012", col="blue", adj=c(0,0) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, pr.fit["F",,]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="n", lty=1, col="red", lwd=c(1,2) )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> matlines( a.pt,  pr.fit["F",,]*100, type="l", lty=1, col=lred , lwd=c(2,3) )
> matlines( a.pt, pr.lfit["F",,]*100, type="l", lty=1, col="red", lwd=c(2,3) )
> text( 25, 21.5, "Women", adj=0, col="red", cex=1.2 )
> text( 89, pr.fit["F","89","1995"]* 99, "1995", col="red", adj=c(1,1) )
> text( 80, pr.fit["F","80","2012"]*101, "2012", col="red", adj=c(1,0) )
> axis( side=1 )
> mtext( "Age", side=1, line=1, outer=T )
> mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
```

From figure 3.14 it is seen that the model captures the actual pattern much better than the simple model with an annual change common across ages.

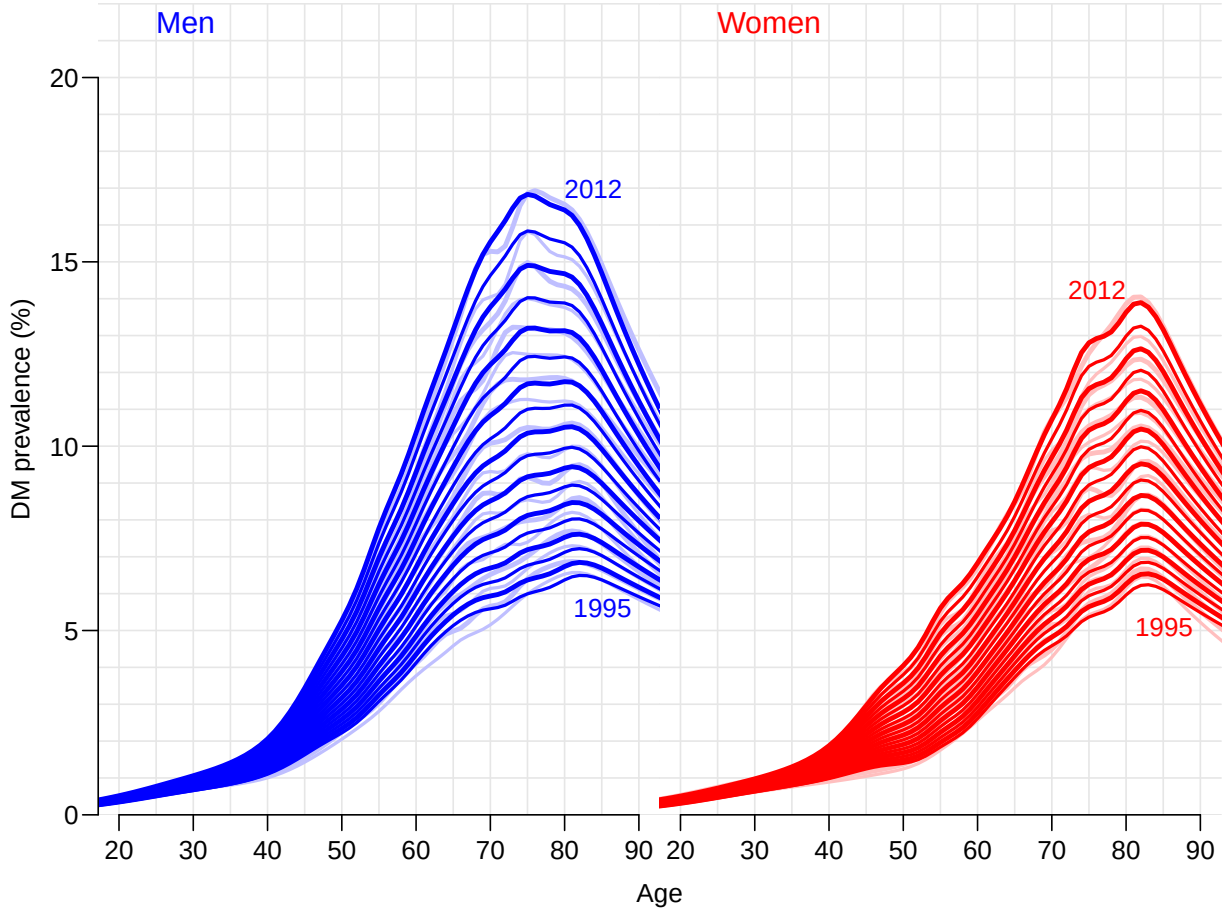

Figure 3.14: *Smoothed age-specific prevalences for the 17-year period 1995–2012 using a model with age-specific constant annual relative change in prevalences (full color) compared to the smoothing of the single years (pale color). Blue is men, red is women.*

3.5 Components of prevalence

The purpose of this chapter is to use the estimated transition rates to predict the prevalences at later (known) times.

This is in itself not an interesting endeavor, because we have the prevalence data available, but it will serve as an illustration that the rates are adequately modelled and that the degree of approximation is adequate when using a given interval length for probability calculations.

Specifically we address the problem of partitioning the changes in prevalence of diabetes in the Danish population over the last 17 years to:

1. changes in mortality rates among diabetes patients

2. changes in incidence rates of diabetes in the population

This measure will be sex- and age-specific, and hence independent of the demographic changes in the population.

3.5.1 Formalization

First we formalize the problem conceptually, then statistical, and finally outline the practical implementation based on analysis of rates.

3.5.1.1 Conceptual

The observed changes in prevalence of DM are a consequence of the changes in mortality and DM-incidence rates in the population and of the changes in the mortality rates in the DM population.

Of these the changes in population mortality presumably have the smaller role, but there is a connection, because they determine the available number of persons susceptible to a DM diagnosis.

Thus the starting point will be the population prevalence of DM as of 1.1.1995. The (age-specific) prevalence at any future point of time is obtained by applying the mortality rates in the two sub-strata of the population (DM / non-DM) and the DM-incidence rates to the non-DM part of the population.

The exercise consists in working out what the prevalence of diabetes would have been if:

1. mortality rates and diabetes rates had remained stable
2. only mortality rates had remained stable, but incidence rates had developed as observed
3. only incidence rates had remained stable, but mortality rates had developed as observed

The difference between observed prevalences and the predicted under scenario

1. 1 is the combined effect of changes in the rates as seen since the starting point chosen.
2. 2 is the effect of changing mortality rates alone. This could also be computed as the difference between scenarios 3 and 1.
3. 3 is the effect of changing incidence rates alone. This could also be computed as the difference between scenarios 2 and 1.

For the sake of completeness we shall compute both types of attribution of prevalences.

3.5.2 Statistical framework

First we consider the setup as outlined in figure 3.15:

```
> library( Epi )
> library( splines )
> tm <- matrix(NA,4,4)
> rownames(tm) <- colnames(tm) <- c("No DM","DM","Dead","Dead (DM)")
> tm[1,2] <- tm[1,3] <- tm[2,4] <- 1
> boxes( tm, boxpos = list( x=c(20,20,80,80),
+                           y=c(80,20,80,20) ),
+            wmult=1.3, hmult=4,
+            txt.arr = c( expression(lambda),
+                         expression(mu[W]),
+                         expression(mu[D][M]) ) )
```

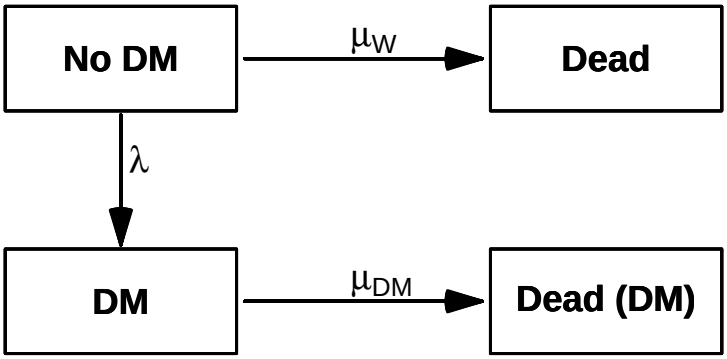

Figure 3.15: *The four states and transitions between them we consider*

The aim is to provide a precise formula for the age-specific prevalences at calendar time t, $p(a,t)$, given that we know the age-specific prevalence at some reference point t_0, $p(a,t_0)$ (in this case 1995), and the transition rates $\lambda(a,p)$, $\mu_{\mathrm{W}}(a,p)$ and $\mu_{\mathrm{DM}}(a,p)$.

We can in principle derive analytical expressions for this, but the easiest approach is to acquire parametric expressions for the transition rates and then update the age-specific prevalences by applying the transition probability matrix to a $A \times 2$ matrix of number of persons in each of the states no DM and DM.

For the given transition rates we can compute transition probabilities between states corresponding to a given (small) interval, δ, say, by first deriving the cumulative intensities for intervals of this length

$$\Lambda(a,p) = \lambda(a,p) \times \delta, \qquad M_{\mathrm{W}}(a,p) = \mu_{\mathrm{W}}(a,p) \times \delta, \qquad M_{\mathrm{DM}}(a,p) = \mu_{\mathrm{DM}}(a,p) \times \delta$$

and the the transition matrix $\mathbf{T}_{a,p}(\delta)$:

$$\mathbf{T}_{a,p}(\delta) = \begin{pmatrix} \mathrm{e}^{-\Lambda-M_{\mathrm{W}}} & \lambda\mathrm{e}^{-\Lambda-M_{\mathrm{W}}}\delta & \mu_{\mathrm{W}}\mathrm{e}^{-\Lambda-M_{\mathrm{W}}}\delta \\ 0 & \mathrm{e}^{-M_{\mathrm{DM}}} & \mu_{\mathrm{DM}}\mathrm{e}^{-\Lambda-M_{\mathrm{DM}}}\delta \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \mathrm{e}^{-\Lambda-M_{\mathrm{W}}} & \Lambda\mathrm{e}^{-\Lambda-M_{\mathrm{W}}} & M_{\mathrm{W}}\mathrm{e}^{-\Lambda-M_{\mathrm{W}}} \\ 0 & \mathrm{e}^{-M_{\mathrm{DM}}} & M_{\mathrm{DM}}\mathrm{e}^{-\Lambda-M_{\mathrm{DM}}} \\ 0 & 0 & 1 \end{pmatrix}$$

So we see that the rates only enter via the cumulative rates over the intervals, so this is what we eventually must compute from models. For simplicity we left out the (a,p) qualification of all the terms in the expressions.

Now if we have the *number* of persons in age-class a and period p in states (W,DM,Dead) in the 3-vector $n(a,p)$ then:

$$n(a+\delta, p+\delta) = n(a,p)\mathbf{T}_{a,p}(\delta)$$

so updating the array of the number of persons in each state is merely a matter of matrix multiplication.

This updating machinery can be illustrated graphically in a Lexis diagram as in figure **??**:

```
> for( yy in 2000+0:3 )
+ for( aa in 40+0:3 )
+ {
+ pdf( paste("./graph/NDR-prup-",yy,"-",aa,".pdf", sep="" ),
```

```
+      height=7, width=7 )
+ par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
+ Lexis.diagram( age=40+c(-1,6), date=2000+c(-1,6), int=1 )
+ w <- 0.6
+ d <- 0.3
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1,1+w,NA,2,2+w), col="forestgreen", lwd=9, lend="butt", ljoin="bevel" )
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1+w,1+w+d,NA,2+w,2+w+d), col="red", lwd=9, lend="butt", ljoin="bevel" )
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1+w+d,2,NA,2+w+d,3), col="black", lwd=9, lend="butt", ljoin="bevel" )
+ for( an in 1:17 )
+ arrows( yy+1.1, aa+0.6, yy+1.9, aa+1.4, lwd=3, angle=an )
+ dev.off()
+ }
```

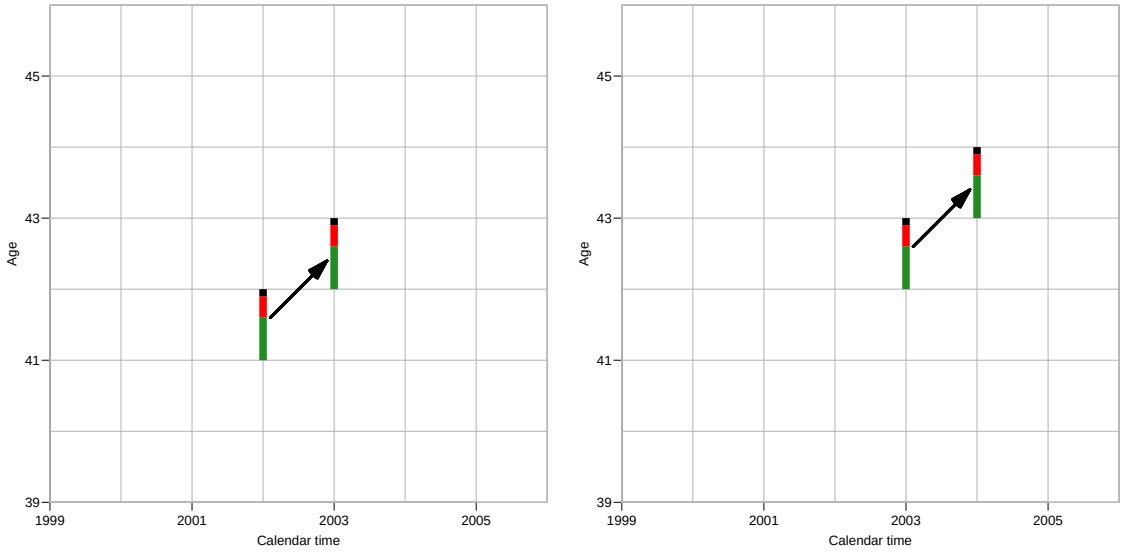

Figure 3.16: *Calculation of prevalences from one year to the next. Green are without diabetes, red with, and black dead.*

If we instead have the *fraction* of (living) persons in states (W,DM) in the vector $q(a,p)$ (which is now just a 2-vector) then:

$$\tilde{q}(a+\delta, p+\delta) = q(a,p)\mathbf{T}_{a,p}(\delta)[1:2,]$$

where we then will get the fraction of the persons in age a at time p who at time $p+\delta$ (and hence in age $a+\delta$) who are in states (W,DM,Dead). But since we are only interested in the progression of prevalences, then we instead use:

$$Q(a+\delta, p+\delta) = q(a,p)\mathbf{T}_{a,p}(\delta)[1:2,1:2]$$
$$q(a+\delta, p+\delta) = Q(a+\delta, p+\delta) \Big/ \sum_{\text{W,DM}} Q(a+\delta, p+\delta)$$

so we update the prevalences at every step.

3.5.2.1 Births

Note that for every step in the updating we will lose estimates in an age-class; in order for this to work we need to feed in the number of births in each age-group with some assumption about the distribution between DM/non-DM; which we will assume is 0:1, that is we assume that no new-born diabetics enter.

3.5.3 Data for the calculations

We will use the models for the rates based on the 1-year data in Lexis triangles. There are two sets of models fitted to different datasets:

- Models for the prevalence of DM as a function of age. These will be based on a dataset with 1-year age-specific empirical prevalences, smoothed by a binomial model (with log-link), so producing a parametric age-prevalence curve for all combination of sex and dates 1 January 1995–2012.
- Models for rates, based on data for 1-year Lexis-triangles ($\triangledown$ and $\triangle$)
  - Incidence rates of DM among non-DM individuals
  - Mortality rates among non-DM individuals
  - Mortality rates among DM patients

  All data for these three sets of rates are in a single dataset.

The practical calculations will be based on quantities derived from these models. Calculations are made using intervals of length `int` as defined below, both in the age and the calendar time direction. The quantities that go into the calculations are:

1. Estimated prevalences at the midpoint of the age-intervals at 1.1.1995, as derived from the models for the prevalences.
2. Estimated incidence (DM) and mortality (non-DM, DM) rates evaluated at:
   (a) the midpoint of the updating periods, that is at times $1995 + n\texttt{int} + \texttt{int}/2, n = 0, \ldots$ and
   (b) the midpoint of the age at updating, that is updating age-class $(a, a + \texttt{int})$ to $(a + \texttt{int}, a + 2\texttt{int})$ we use the estimated rate at age $a + \texttt{int}$.

3.5.4 Prevalences

The observed prevalences and population size at the 1 January 1995–2012 available from a tabulation of the diabetes dome previously:

```
> load( file="./data/prev-m.Rda" )
> str( pr )
  'data.frame':        3600 obs. of  5 variables:
   $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ A  : num  0 0 0 0 0 0 0 0 0 0 ...
   $ P  : num  1995 1996 1997 1998 1999 ...
   $ X  : num  3 1 0 1 2 1 1 0 1 1 ...
   $ N  : num  35612 36055 34853 34774 34076 ...
```

```
> head( pr )
```

```
   sex A    P X     N
1   M 0 1995 3 35612
2   M 0 1996 1 36055
3   M 0 1997 0 34853
4   M 0 1998 1 34774
5   M 0 1999 2 34076
6   M 0 2000 1 33906
```

These are empirical prevalences (`X`—no. of cases of DM, `N`—population size) for each of the 18 dates 1.1.1995 – 1.1.2012 in 1-year intervals, but to get the machinery running we will need the prevalences as a continuous function of age.

So we model the prevalences as of 1 January each of the years 1995—2012, as a smooth function of age, with the intention of using the predicted prevalences at the midpoints each of the smaller age-classes that we use for the simulation.

So we collect the models for the prevalences So once we have set up the array to hold the smoothed empirical prevalences we can fill them into `pr.fit`; we use a log-link binomial model with a smooth spline with 15 knots.

```
> dnam <- list( sex = c("M","F"),
+                 t = sort(unique(pr$P)) )
> ( kp.a <- c( 10, with( pr, quantile( rep(A,X), (1:15-0.5)/15 ) ) ) )
```

```
      3.333333%       10% 16.66667% 23.33333%       30% 36.66667% 43.33333%
10           27        40        47        52        55        58        61
     50% 56.66667% 63.33333%       70% 76.66667% 83.33333%       90% 96.66667%
64           66        69        72        74        78        81        87
```

```
> pr.mod <- list()
>   length( pr.mod ) <- prod( sapply( dnam, length ) )
>      dim( pr.mod ) <-       sapply( dnam, length )
> dimnames( pr.mod ) <-              dnam
> for( dt in dimnames(pr.mod)[["t"]] )
+ for( sx in dimnames(pr.mod)[["sex"]] )
+ pr.mod[[sx,dt]] <- glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ),
+                         family = binomial(link="log"),
+                           data = subset( pr,
+                                          sex==sx & P==as.numeric(dt) ) )
```

For the calculations we shall only use the estimated prevalences as of 1.1.1995 as starting point for the simulation, that is from the models in `pr.mod[[sx,"1995"]]` for `sx` equal to either `M` or `F`.

3.5.4.1 Rates

First we the load the data for the models for incidence and mortality:

```
> load( file="./data/FU-m.Rda" )
> head( TT )
```

```
   sex A    P U     Y.nD      Y.DM D.DM D.nD X
1   F 0 1995 0 17025.50 0.0000000    0  137 0
2   F 0 1995 1 17100.54 0.1300479    0   16 2
3   F 0 1996 0 16468.06 1.4401095    0  134 4
4   F 0 1996 1 17067.30 1.8617385    0   23 4
5   F 0 1997 0 16434.00 0.0000000    0  152 0
6   F 0 1997 1 16499.84 1.9890486    0   14 2
```

```
> attr( TT, "Variables" )
```

```
      Data frame using the original definintion of DM from NDR
  sex  "Sex"
  A    "1-year age class"
  P    "1-year period"
  U    "Indicator of upper Lexis triangle"
  Y.nD "P-Y among non-diabetics"
  Y.DM "P-Y among diabetes patients"
  D.DM "Deaths among non-diabetics"
  D.nD "Deaths among diabetes patients"
  X    "Diabetes diagnoses among non-diabetics"
> DD <- transform( TT, A = A+(1+U)/3,
+                      P = P+(2-U)/3,
+                   D.nD = pmax(D.nD,0) )
> head( DD )
    sex         A        P U     Y.nD      Y.DM D.DM D.nD X
  1   F 0.3333333 1995.667 0 17025.50 0.0000000    0  137 0
  2   F 0.6666667 1995.333 1 17100.54 0.1300479    0   16 2
  3   F 0.3333333 1996.667 0 16468.06 1.4401095    0  134 4
  4   F 0.6666667 1996.333 1 17067.30 1.8617385    0   23 4
  5   F 0.3333333 1997.667 0 16434.00 0.0000000    0  152 0
  6   F 0.6666667 1997.333 1 16499.84 1.9890486    0   14 2
```

Then we can set up age-period-cohort models for the three types of rates of relevance; first we set up the knots for the period- and cohort-effects common for the three analyses, whereas we let the age-effect have knots depending on the position of the events on the age-scale:

```
> p.kn <- seq( 1996, 2011,, 5 )
> c.kn <- seq( 1900, 2010,, 8 )
```

Note that we name the vector of age-knots differently for the different models, because `predict.glm` apparently uses the global version of the knots vector and not the vector stored in the `glm` object.

3.5.4.1.1 Incidence rates

Here is the age-period cohort model for the rates of DM occurrence, using `(X,Y.nD)` as outcome variables:

```
> ( ai.kn <- with( DD, c(5,10,quantile( rep(A,X), probs=(1:10-0.5)/10 ) ) ) )
                          5%      15%      25%      35%      45%      55%
   5.00000 10.00000 32.66667 45.66667 52.33333 56.66667 60.66667 64.33333
       65%      75%      85%      95%
  68.33333 72.66667 77.33333 84.66667
> incM <- glm( X ~ Ns(   A, kn=ai.kn ) +
+                  Ns( P  , kn=p.kn ) +
+                  Ns( P-A, kn=c.kn ),
+              offset = log(Y.nD),
+              family = poisson,
+                data = subset(DD,sex=="M") )
> incF <- update( incM, data = subset(DD,sex=="F") )
```

3.5.4.1.2 Non-DM mortality rates

Here is the age-period cohort model for the mortality rates among non-diabetics, using `(D.nD,Y.nD)` as outcome variables:

```
> ( and.kn <- with( DD, c(5,15,quantile( rep(A,D.nD), probs=(1:10-0.5)/10 ) ) ) )
                          5%      15%      25%      35%      45%      55%
   5.00000 15.00000 46.33333 60.33333 67.66667 72.66667 77.33333 80.33333
       65%      75%      85%      95%
  83.33333 86.33333 89.66667 93.66667
```

```
> mndM <- glm( D.nD ~ Ns(   A, kn=and.kn ) +
+                     Ns( P  , kn=p.kn ) +
+                     Ns( P-A, kn=c.kn ),
+              offset = log(Y.nD),
+              family=poisson,
+              data = subset(DD,sex=="M") )
> mndF <- update( mndM, data = subset(DD,sex=="F") )
```

3.5.4.1.3 DM mortality rates Here is the age-period cohort model for the mortality rates among diabetes patients, using `(D.DM,Y.DM)` as outcome variables:

```
> ( adm.kn <- with( DD, c(25,quantile( rep(A,D.DM), probs=(1:11-0.5)/11 ) ) ) )
            4.545455% 13.63636% 22.72727% 31.81818% 40.90909%       50% 59.09091%
   25.00000  53.66667  63.33333  68.33333  72.33333  75.33333  78.33333  80.66667
  68.18182% 77.27273% 86.36364% 95.45455%
   83.33333  85.66667  88.66667  92.66667
> mdmM <- glm( D.DM ~ Ns(   A, kn=adm.kn ) +
+                     Ns( P  , kn=p.kn ) +
+                     Ns( P-A, kn=c.kn ),
+              offset = log(Y.DM),
+              family=poisson,
+              data = subset( DD, sex=="M" & Y.DM>0 ) )
> mdmF <- update( mdmM, data = subset(DD,sex=="F" & Y.DM>0) )
```

3.5.5 Implementation of prevalence calculations

We start by specifying the interval length for the updating, and then the points at which we want to predict. The transition rates are labeled by the midpoints of the Lexis squares (of width `int`) where we predict them (`a.pt` and `p.pt`), and the prevalences by the mid-points of the age-classes (`a.pt` and the time points `t.pt`

```
> int <- 0.5
> a.pt <- seq(int,100,int) - int/2
> t.pt <- seq(1995,2012,int)
> p.pt <- t.pt[-1] - int/2
```

All the predictions should be in units of the interval length chosen for calculations. We note from the calculations above that the quantities that enter the expressions for the transition probabilities are all cumulative rates over the intervals. Thus we use a prediction data frame with the person-years-variables set to `int`, and we use predicted rates at the period midpoints (`p.pt`), but we use the age-point at the *upper end* of the age-class, because we will be using the cumulative rates to predict transitions from the age-class $(a, a+\delta)$ to $(a+\delta, a+2\delta)$:

```
> nd <- data.frame( A = rep(a.pt+int/2,     length(p.pt)),
+                   P = rep(p.pt      ,each=length(a.pt)),
+                   Y.nD = int,
+                   Y.DM = int )
> str( nd )
  'data.frame':        6800 obs. of  4 variables:
   $ A   : num  0.5 1 1.5 2 2.5 3 3.5 4 4.5 5 ...
   $ P   : num  1995 1995 1995 1995 1995 ...
   $ Y.nD: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
   $ Y.DM: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
> head( nd )
```

```
     A       P Y.nD Y.DM
1 0.5 1995.25  0.5  0.5
2 1.0 1995.25  0.5  0.5
3 1.5 1995.25  0.5  0.5
4 2.0 1995.25  0.5  0.5
5 2.5 1995.25  0.5  0.5
6 3.0 1995.25  0.5  0.5
```

```
> summary( nd )
```

```
       A                P             Y.nD          Y.DM
 Min.   :  0.50   Min.   :1995   Min.   :0.5   Min.   :0.5
 1st Qu.: 25.38   1st Qu.:1999   1st Qu.:0.5   1st Qu.:0.5
 Median : 50.25   Median :2004   Median :0.5   Median :0.5
 Mean   : 50.25   Mean   :2004   Mean   :0.5   Mean   :0.5
 3rd Qu.: 75.12   3rd Qu.:2008   3rd Qu.:0.5   3rd Qu.:0.5
 Max.   :100.00   Max.   :2012   Max.   :0.5   Max.   :0.5
```

3.5.5.1 Transition probabilities

We shall use the recursive scheme to predict the course of DM prevalence development in the population under various scenarios of mortality and incidence development. So we use the various structures to hold results and clarify calculations:

`Lambda,Mu.nD,Mu.DM` — arrays of cumulative rates over intervals of length `int`, evaluated at dates at the midpoint of calculation intervals, and at borders of age-intervals, corresponding to midpoints of C-sets of the Lexis diagram (⧄).

`pr.fit` — array of empirical prevalences at 1.1.1995–1.1.2012, smoothed by natural splines separately for each year.

`TR` — array of transition probabilities between states no DM, DM and Dead. Transition probabilities are computed under 4 different scenarios combining mortality and incidence rates either as they actually developed 1995–2012 or assuming they were constant at the 1995 level. These refer to intervals of length `int` years and are therefore labeled on the period dimension by the midpoint of these, a total of 17/`int`.

`prv` — array of predicted prevalences based on the initial prevalences at 1.1.1995 and the transition probabilities as put in `TR`. The scenario dimension refers to the 4 scenarios: "obs", "m-fix", "i-fix" and "all-f", but this dimension in the array is expanded by 3 extra levels "mort", "inc" and "const" that are to be filled with the part of the prevalences that are attributable to decrease in mortality, increase in incidence and the disequilibrium between rates and prevalence in 1995. Likewise the period dimension is expanded by one relative to that in `TR`, since this refer to points in time and not time intervals.

`prn` — array of predicted *number* of DM patients in one-year age classes at the 1 January each year. So the same structure as `prv`, but with substantially fewer entries along the age and period dimensions.

Thus, first we set up the arrays of the cumulative rates (note that the ages are at the midpoint of age-classes):

```
> Lambda <-
+ Mu.nD <-
+ Mu.DM <- NArray( list( a = a.pt,
+                        p = p.pt,
+                      sex = c("M","F") ) )
```

In order to compute the transition probabilities we need the cumulative incidences over intervals of length `int`. So first we predict these using the relevant points. Note that the person-years-variables are set to `int` in order to get cumulative rates over an interval of this length. Note that the compute fitted rates at `int/2` to the right of the labeling of the age-interval:

```
> nd <- data.frame( A = rep(a.pt+int/2,     length(p.pt)),
+                   P = rep(p.pt      ,each=length(a.pt)),
+                   Y.nD = int,
+                   Y.DM = int )
```

With this prediction frame in place we compute the cumulative rates:

```
> Lambda[,,"M"] <- predict.glm( incM, type="response", newdata=nd )
> Lambda[,,"F"] <- predict.glm( incF, type="response", newdata=nd )
>  Mu.nD[,,"M"] <- predict.glm( mndM, type="response", newdata=nd )
>  Mu.nD[,,"F"] <- predict.glm( mndF, type="response", newdata=nd )
>  Mu.DM[,,"M"] <- predict.glm( mdmM, type="response", newdata=nd )
>  Mu.DM[,,"F"] <- predict.glm( mdmF, type="response", newdata=nd )
```

Note that we get warning messages originating from the overparametrization of the age-period-cohort model.

In order to get the predicted prevalences by age, period and prediction type, we need the (1-step) transition matrices at all combinations of age (a) and date (p), this is put in array:

```
> states <- c("no DM","DM")
> TR <- NArray( c( dimnames(Lambda),
+                  list( from = states,
+                          to = states,
+                       scene = c("obs","m-fix","i-fix","all-f" ) ) ) )
> str( TR )
   logi [1:200, 1:34, 1:2, 1:2, 1:2, 1:4] NA NA NA NA NA NA ...
   - attr(*, "dimnames")=List of 6
    ..$ a    : chr [1:200] "0.25" "0.75" "1.25" "1.75" ...
    ..$ p    : chr [1:34] "1995.25" "1995.75" "1996.25" "1996.75" ...
    ..$ sex  : chr [1:2] "M" "F"
    ..$ from : chr [1:2] "no DM" "DM"
    ..$ to   : chr [1:2] "no DM" "DM"
    ..$ scene: chr [1:4] "obs" "m-fix" "i-fix" "all-f"
> prod( dim(TR) )
  [1] 217600
```

So we can now compute the one-`int`-step transition matrices for every combination of `a.pt` and `p.pt`, both in steps of `int` (in this case 0.5 year):

```
> TR[,,,"no DM","no DM","obs"] <- exp(-Lambda-Mu.nD)
> TR[,,,"no DM","DM"   ,"obs"] <- exp(-Lambda-Mu.nD)*Lambda
> TR[,,,"DM"   ,"no DM","obs"] <- 0
> TR[,,,"DM"   ,"DM"   ,"obs"] <- exp(-Mu.DM)
```

Note that we have not included the "Dead" state in the calculations, because we only bother about the fraction of diabetes patients in each age class at each time point. So the probabilities we compute do not sum to 1 within the "from" states.

The situation where both the mortality rates and incidence rates are fixed at the 1995 level is trivial, because transition probabilities in that case only depend on age and not on period.

When we fix the mortality or incidence at the 1995 level we just replace the expressions above with expressions where we replace the date dimension by `rep(1,np)`, where `np` is the number of periods:

```
> np <- dim(Lambda)["p"]
> TR[,,"no DM","no DM","m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])
> TR[,,"no DM","DM"   ,"m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])*Lambda
> TR[,,"DM"   ,"no DM","m-fix"] <- 0
> TR[,,"DM"   ,"DM"   ,"m-fix"] <- exp(-Mu.DM[,rep(1,np),])


> TR[,,"no DM","no DM","i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)
> TR[,,"no DM","DM"   ,"i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)*Lambda[,rep(1,np),]
> TR[,,"DM"   ,"no DM","i-fix"] <- 0
> TR[,,"DM"   ,"DM"   ,"i-fix"] <- exp(-Mu.DM)


> TR[,,"no DM","no DM","all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])
> TR[,,"no DM","DM"   ,"all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])*Lambda[,rep(1,np),]
> TR[,,"DM"   ,"no DM","all-f"] <- 0
> TR[,,"DM"   ,"DM"   ,"all-f"] <- exp(-Mu.DM[,rep(1,np),])
```

We have now collected the transition probabilities between "no DM" and "DM" as well as the probabilities of remaining in each of these, all referring to a duration of `int`.

3.5.5.2 Prediction of the observed prevalences

Note that we do not need to predict the population size; we only predict the prevalences as fractions. When we multiply the fraction of persons in states (no DM,DM) with the transition matrix, we get fraction of the persons in the previous state that are in states (no DM,DM), which does not sum to 1 (because of the dead ones), so we must rescale to prevalence age in each step.

When we do the predictions we need a starting point (and comparison points) for we predict the age-specific prevalences at 1 January each year at the midpoint of the age-intervals of length `int`, as stored in `a.pt`:

```
> pr.fit <- NArray( c( dimnames(Lambda)[c("a","sex")],
+                      dimnames(pr.mod)["t"] ) )
> for( sx in dimnames(pr.fit)[["sex"]] )
+ for( dt in dimnames(pr.fit)[["t"]] )
+ pr.fit[,sx,dt] <- predict( pr.mod[[sx,dt]],
+                            newdata = data.frame( A=as.numeric(dimnames(pr.fit)[["a"]]) ),
+                               type = "response" )
```

Then we set up an array to hold the predicted prevalences under different scenarios:

```
> dpr <- c( dimnames(Lambda)[c("a","p","sex")],
+           list( c(dimnames(TR)[["scene"]],"mort","inc","const") ) )
> names( dpr )[c(2,4)] <- c("t","what")
> dpr[["t"]] <- t.pt
> prv <- NArray( dpr )
```

To get the calculations started we insert the estimated prevalences at 1995 and assume the all newborns are without diabetes, that is the prevalence is 0 at age 0 (or rather at age `int/2`):

```
> ### Smoothed prevalences at 1.1.1995 - the starting values
> prv[,1,,] <- pr.fit[,,1]
> ### Prevalences at age 0 are set to 0
> prv[1,,,] <- 0
```

Then we can finally compute the prevalences at the desired points of the Lexis diagram:

```
> for( ip in 1:(dim(prv)["t"]-1) )
+ for( ia in 1:(dim(prv)["a"]-1) )
+ prv[ia+1,ip+1,,1:4] <-
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,] ) /
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","no DM",] )
```

Later we shall also compute the fraction of the prevalences that are attributable to trends in mortality and incidence as well as to the non-stationarity of the rates/prevalences as of 1995, so we put in three extra levels of the last dimension, and one extra levels of the period dimension because we want to predict to the end of the last period too (or, to put it differently, we need an extra first level to hold the starting prevalences as of 1.1.1995).

3.5.5.3 A function for the calculations

We now pack the previous into a function, `prcalc`, which takes the interval length (and the ending year) as arguments, and assumes that the smoothed prevalences (`pr.mod` as 2-dimensional list) and smoothed rates (`incM`, `incF`, `mndM`, `mndF`, `mdmM`, `mdmF`) are available in the workspace:

```
> prcalc <-
+ function( int=1, end=2012 )
+ {
+ # OBS: Assumes that the fitted prevalences pr.fit as well as the
+ # fitted models for rates, incM, incF, mndM, mndF, mdmM, mdmF are in
+ # the workspace
+ a.pt <- seq(int,100,int) - int/2
+ t.pt <- seq(1995,end,int)
+ p.pt <- t.pt[-1] - int/2
+ ### Prediction data frame
+ nd <- data.frame( A = rep(a.pt+int/2,     length(p.pt)),
+                   P = rep(p.pt      ,each=length(a.pt)),
+                   Y.nD = int,
+                   Y.DM = int )
+ ### Arrys to hold the rates at the relevant points, note that a.pt is
+ ### the first dimension, and p.pt the second so that predictions using
+ ### newdata=nd can be immediately put in the array, using the
+ ### column-major convention:
+ Lambda <-
+ Mu.nD <-
+ Mu.DM <- NArray( list( a = a.pt,
+                        p = p.pt,
+                      sex = c("M","F") ) )
+ ### Compute the cumulative rates over an interval
+ options( warn = -1 )
+ Lambda[,,"M"] <- predict.glm( incM, type="response", newdata=nd )
+ Lambda[,,"F"] <- predict.glm( incF, type="response", newdata=nd )
+  Mu.nD[,,"M"] <- predict.glm( mndM, type="response", newdata=nd )
+  Mu.nD[,,"F"] <- predict.glm( mndF, type="response", newdata=nd )
+  Mu.DM[,,"M"] <- predict.glm( mdmM, type="response", newdata=nd )
+  Mu.DM[,,"F"] <- predict.glm( mdmF, type="response", newdata=nd )
+ options( warn = 0 )
+ ### The fitted prevalences at ages a.pt but only at 1 Jan each year
```

```
+ pr.fit <- NArray( c( dimnames(Lambda)["a"],
+                       dimnames(pr.mod)[c("sex","t")] ) )
+ for( sx in dimnames(pr.fit)[["sex"]] )
+ for( dt in dimnames(pr.fit)[["t"]] )
+ pr.fit[,sx,dt] <- predict( pr.mod[[sx,dt]],
+                            newdata = data.frame( A=as.numeric(dimnames(pr.fit)[["a"]]) ),
+                               type = "response" )
+ ### Transition probabilities under various scenarios
+ states <- c("no DM","DM")
+ TR <- NArray( c( dimnames(Lambda),
+                  list( from = states,
+                          to = states,
+                       scene = c("obs","m-fix","i-fix","all-f" ) ) ) )
+ ### No of levels of the period-dimension
+ np <- dim(Lambda)[2]
+ ### Using observed rates throughout
+ TR[,,,"no DM","no DM","obs"  ] <- exp(-Lambda-Mu.nD)
+ TR[,,,"no DM","DM"   ,"obs"  ] <- exp(-Lambda-Mu.nD)*Lambda
+ TR[,,,"DM"   ,"no DM","obs"  ] <- 0
+ TR[,,,"DM"   ,"DM"   ,"obs"  ] <- exp(-Mu.DM)
+ ### Mortality rates fixed
+ TR[,,,"no DM","no DM","m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])
+ TR[,,,"no DM","DM"   ,"m-fix"] <- exp(-Lambda-Mu.nD[,rep(1,np),])*Lambda
+ TR[,,,"DM"   ,"no DM","m-fix"] <- 0
+ TR[,,,"DM"   ,"DM"   ,"m-fix"] <- exp(-Mu.DM[,rep(1,np),])
+ ### Incidence rates fixed
+ TR[,,,"no DM","no DM","i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)
+ TR[,,,"no DM","DM"   ,"i-fix"] <- exp(-Lambda[,rep(1,np),]-Mu.nD)*Lambda[,rep(1,np),]
+ TR[,,,"DM"   ,"no DM","i-fix"] <- 0
+ TR[,,,"DM"   ,"DM"   ,"i-fix"] <- exp(-Mu.DM)
+ ### All rates fixed
+ TR[,,,"no DM","no DM","all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])
+ TR[,,,"no DM","DM"   ,"all-f"] <- exp(-Lambda[,rep(1,np),]-Mu.nD[,rep(1,np),])*Lambda[,rep(1,np),]
+ TR[,,,"DM"   ,"no DM","all-f"] <- 0
+ TR[,,,"DM"   ,"DM"   ,"all-f"] <- exp(-Mu.DM[,rep(1,np),])
+ ### Array to hold the predicted prevalences
+ dpr <- c( dimnames(Lambda)[1:3],
+           list( c(dimnames(TR)[["scene"]],"mort","inc","const") ) )
+ names( dpr )[c(2,4)] <- c("t","what")
+ dpr[["t"]] <- t.pt
+ prv <- NArray( dpr )
+ ### Smoothed prevalences at 1.1.1995 - the starting values
+ prv[,1,,] <- pr.fit[,,1]
+ ### Prevalences at age 0 are set to 0
+ prv[1,,,] <- 0
+ ### Compute the prevalences
+ for( ip in 1:(dim(prv)["t"]-1) )
+ for( ia in 1:(dim(prv)["a"]-1) )
+ prv[ia+1,ip+1,,1:4] <-
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,] ) /
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"   ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"no DM","no DM",] )
+ ### ...and return them together with the observed
+ list( prv=prv, pr.fit=pr.fit )
+ }
```

Note in the last bit of the function definition that the reason that the last dimension, `scene`, is explicitly mentioned in the array `prv` is because this has dimension 7, but in `TR` only 4 — remember that `prv` also has thee extra levels to provide for the estimated part of the prevalences attributable to mortality change, incidence changes, and non-equilibrium at 1995.

3.5.5.4 Length of the calculation interval

In order to check whether the prediction using an interval length of 0.50 year is necessary we repeat the exercise using a 2-year interval for comparison

```
> system.time( prv  <- prcalc( int=0.1 ) )
     user  system elapsed
    20.18    1.26   21.45

> system.time( prvh <- prcalc( int=0.5 ) )

     user  system elapsed
     0.87    0.03    0.91

> system.time( prv1 <- prcalc( int=1.0 ) )

     user  system elapsed
     0.33    0.00    0.33

> system.time( prv2 <- prcalc( int=2.0 ) )

     user  system elapsed
     0.17    0.00    0.17
```

With these predictions in place we can now check whether we have made a reasonable approximation to the observed prevalences at 1.1.2012, and to which extent the calculation-interval influences this:

In the array `prv` are all the prevalences as predicted from the prevalence in 1995 using the estimated incidences and mortalities; predicted at intervals of `inc` whereas we have the smoothed empirical prevalences at 1 January 1995,...2012 in the array `pr.fit`:

```
> a.p2 <- as.numeric( dimnames(prv2$prv)[["a"]] )
> a.p1 <- as.numeric( dimnames(prv1$prv)[["a"]] )
> a.ph <- as.numeric( dimnames(prvh$prv)[["a"]] )
> a.pt <- as.numeric( dimnames(prv$prv )[["a"]] )
> wh <- c("1999","2005","2011")
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> matplot( a.pt, prv$pr.fit[,"M",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="blue", lty=1, lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"M","obs"]*100, lty="11", lwd=2, col="blue" )
> # matlines( a.ph, prvh$prv[,wh,"M","obs"]*100, lty="13", lwd=2, col="blue" )
> # matlines( a.p1, prv1$prv[,wh,"M","obs"]*100, lty="14", lwd=2, col="blue" )
> matlines( a.p2, prv2$prv[,wh,"M","obs"]*100, lty="22", lwd=2, col="blue" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> axis( side=1, at=1:9*10, labels=rep("",9) )
> axis( side=2 )
> matplot( a.pt, prv$pr.fit[,"F",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="red", lty=1, lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"F","obs"]*100, lty="11", lwd=2, col="red" )
> # matlines( a.ph, prvh$prv[,wh,"F","obs"]*100, lty="13", lwd=2, col="red" )
> # matlines( a.p1, prv1$prv[,wh,"F","obs"]*100, lty="14", lwd=2, col="red" )
> matlines( a.p2, prv2$prv[,wh,"F","obs"]*100, lty="22", lwd=2, col="red" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> axis( side=1, at=1:9*10, labels=rep("",9) )
> mtext( "Prevalence of DM (%)", side=2, line=2, las=0, outer=TRUE )
> mtext( "Age", side=1, line=2, las=0, outer=TRUE )
```

For presentation purposes we also just compare the observed and the predicted:

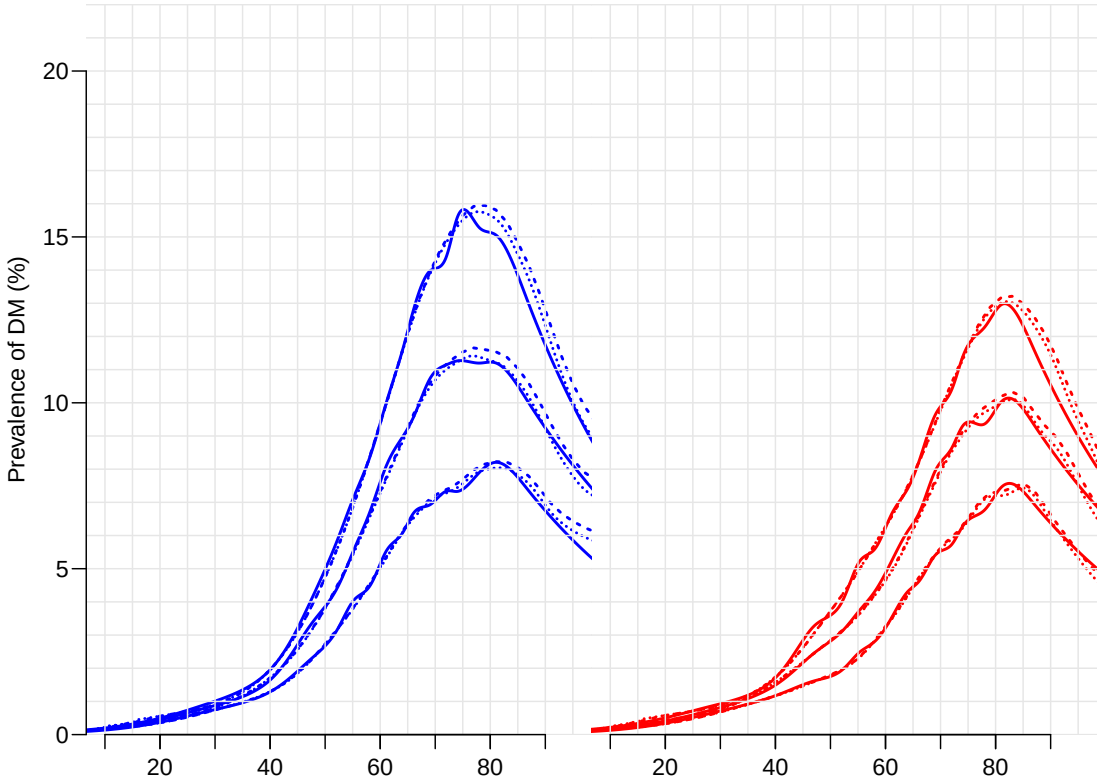

Figure 3.17: *Predicted prevalences at 1 January 1999, 2005 and 2011. Full (thin) lines: Smoothed empirical prevalences. Broken lines: Prediction using calculation intervals of 0.1 and 2 years (from dotted / broken).*

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> matplot( a.pt, prv$pr.fit[,"M",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="blue", lty=rep(1:2,each=3), lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"M","obs"]*100, lty="12", lwd=2, col="blue" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> axis( side=2 )
> matplot( a.pt, prv$pr.fit[,"F",wh]*100,
+          xlim=c(10,95), ylim=c(0,22), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", col="red", lty=rep(1:2,each=3), lwd=rep(c(2,3),each=3) )
> matlines( a.pt, prv$prv[,wh,"F","obs"]*100, lty="12", lwd=2, col="red" )
> abline( h=0:25, v=seq(0,100,5), col=gray(0.9) )
> axis( side=1 )
> mtext( "Prevalence of DM (%)", side=2, line=2, las=0, outer=TRUE )
> mtext( "Age", side=1, line=2, las=0, outer=TRUE )
```

3.5.6 Prevalences under different scenarios

We now compare the predicted prevalences under the four scenarios at 1.1.2012:

```
> np <- dim( prv$prv )[2]
> par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
> matplot( a.pt, cbind(prv$prv[,np,"M",],prv$prv[,1,"M",1])*100,
```

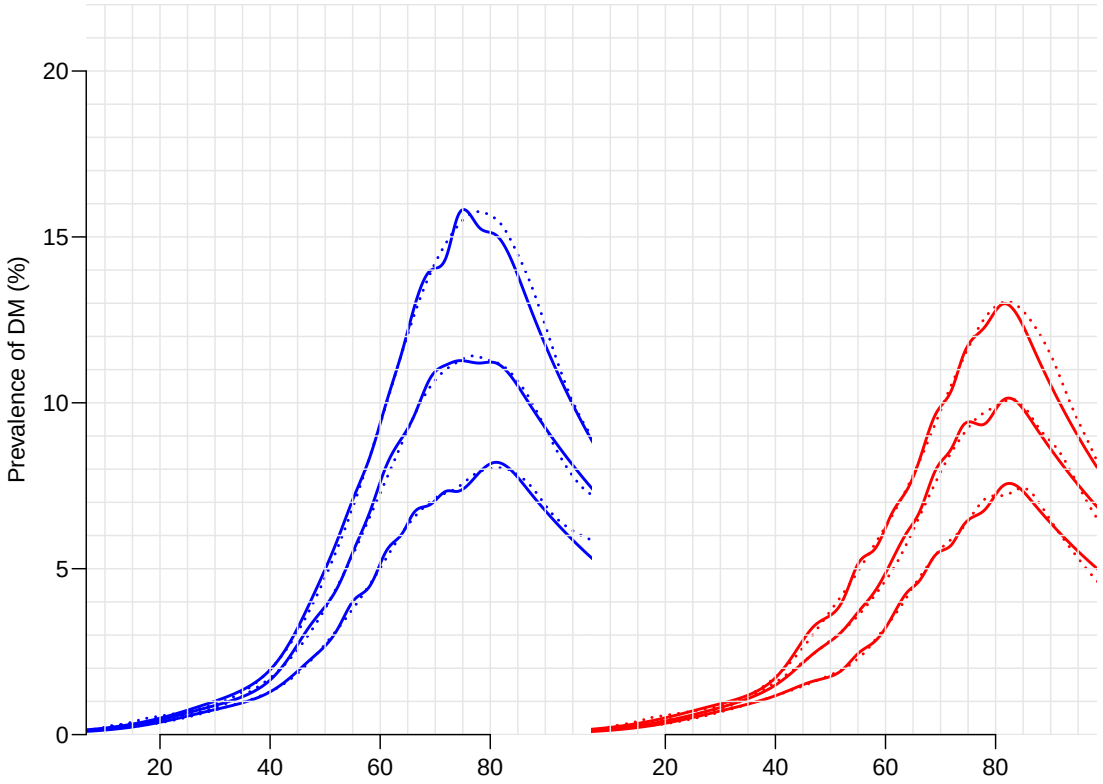

Figure 3.18: *Predicted prevalences at 1 January 1999, 2005 and 2011. Full (thin) lines: Smoothed empirical prevalences. Broken lines: Prediction using a calculation interval of 0.1 year.*

```
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
> matlines( a.pt, prv$prv[,np,"M",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="blue" )
> matlines( a.pt, prv$prv[,1,"M",]*100, type="l", lty=1, lwd=1, col="blue" )
> matplot( a.pt, cbind(prv$prv[,np,"F",],prv$prv[,1,"F",1])*100, yaxt="n",
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
> matlines( a.pt, prv$prv[,np,"F",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="red" )
> matlines( a.pt, prv$prv[,1,"F",]*100, type="l", lty=1, lwd=1, col="red" )
```

Here is a more elaborate graph, mainly for presentation purposes:

```
> scen <- c("Mort obs, Inc obs","Mort 1995, Inc obs","Mort obs, Inc 1995","Mort 1995, Inc 1995")
> c.a <- dimnames(prv$prv)[[1]][floor(dim(prv$prv)[1]/1.5)]
> n.a <- as.numeric( c.a )
> nt <- dim( prv$prv )[2]
> hts <- prv$prv[c.a,nt,"M",1:4]*100
> cau.exp <-
+ function( wh=1:4, fill=FALSE )
+ {
+ pdf( paste( "./graph/NDR-", paste(wh,collapse=""), if( fill ) "F",
+             "-m.pdf", sep="" ), height=8, width=11 )
+ par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
+ matplot( a.pt, cbind(prv$prv[,nt,"M",],prv$prv[,1,"M",1])*100, yaxs="i",
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="Prevalence (%)",
```

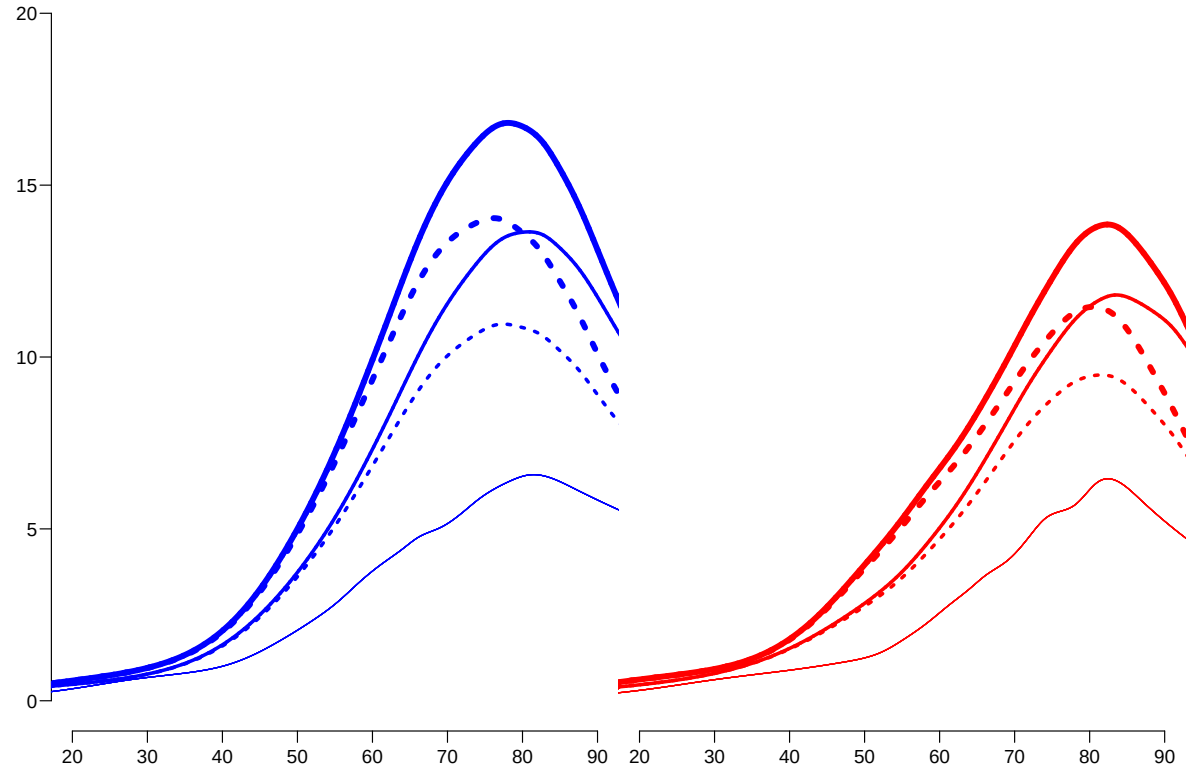

Figure 3.19: *The predicted prevalences at 1.1.2012 under different scenarios:*
Full lines: Mortality rates evolve as observed, Broken lines: Mortality rates remain as 1995.
Thick lines: Incidence rates evolve as observed, Thin lines: Incidence rates remain as in 1995.
The very thin lines lowest in the two displays are the observed prevalences in 1995.

```
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
+ matlines( a.pt, prv$prv[,nt,"M",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="blue" )
+ matlines( a.pt, prv$prv[,1,"M",]*100, type="l", lty=1, lwd=1, col="blue" )
+ mtext( "Age-specific DM prevalence (%)", side=2, line=2, las=0 )
+ text( rep(20,4)[wh], hts[wh], scen[wh], adj=0, col="blue", cex=1.2 )
+ for( i in 1:15 )
+ arrows( (20.20+strwidth(scen,cex=1.2))[wh], hts[wh], rep(n.a,4)[wh], hts[wh], col="blue",
+         angle=i, lwd=2 )
+ if( fill ) polygon( c(a.pt,rev(a.pt)),
+                     c(prv$prv[,nt,"M",wh[1]],rev(prv$prv[,nt,"M",wh[2]]))*100,
+                     col=rgb(0,0,1,0.3), border="transparent" )
+ matplot( a.pt, cbind(prv$prv[,nt,"F",],prv$prv[,1,"F",1])*100, yaxs="i",
+          xlim=c(20,90), ylim=c(0,22), xlab="Age", ylab="", yaxt="n",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
+ matlines( a.pt, prv$prv[,nt,"F",]*100,
+          type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="red" )
+ matlines( a.pt, prv$prv[,1,"F",]*100, type="l", lty=1, lwd=1, col="red" )
+ if( fill ) polygon( c(a.pt,rev(a.pt)),
+                     c(prv$prv[,nt,"F",wh[1]],rev(prv$prv[,nt,"F",wh[2]]))*100,
+                     col=rgb(1,0,0,0.3), border="transparent")
+ dev.off()
+ }
```

```
> cau.exp(1:4)
  null device
            1
> for( ff in c(FALSE,TRUE) )
+    {
+ cau.exp(1:2,fill=ff)
+ cau.exp(3:4,fill=ff)
+ cau.exp(c(1,3),fill=ff)
+ cau.exp(c(2,4),fill=ff)
+    }
```

Figure 3.19 shows the predicted prevalences under 4 different scenarios compared to the observed prevalences as of 1.1.1995.

3.5.6.1 How much is attributable to what?

We can compute how much of the age-specific prevalences that are attributable to mortality changes and how much to changes in incidence rates.

The effect of mortality decline can be computed either as the difference between "obs" and "m-fix" or as the difference between "i-fix" and "all-f". But there is no guarantee that these two quantities are the same.

Similarly the effect of incidence increase can be computed either as the difference between "obs" and "i-fix" or as the difference between "m-fix" and "all-f". And there is no guarantee that these two are the same either.

Hence we explore how different these quantities are:

```
> dimnames( prv$prv )[4]
  $what
  [1] "obs"   "m-fix" "i-fix" "all-f" "mort"  "inc"   "const"
> par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
> matplot( a.pt, cbind( prv$prv[,nt,"M","obs"  ]-prv$prv[,nt,"M","m-fix"],
+                       prv$prv[,nt,"M","i-fix"]-prv$prv[,nt,"M","all-f"] )*100,
+          xlim=c(20,90), ylim=c(0,6), xlab="", ylab="",
+          type="l", lty=1, lwd=c(4,2)+1, col="blue" )
> matlines(a.pt, cbind( prv$prv[,nt,"M","obs"  ]-prv$prv[,nt,"M","i-fix"],
+                       prv$prv[,nt,"M","m-fix"]-prv$prv[,nt,"M","all-f"] )*100,
+          type="l", lty="22", lwd=c(4,2)+1, col="blue" )
> matplot( a.pt, cbind( prv$prv[,nt,"F","obs"  ]-prv$prv[,nt,"F","m-fix"],
+                       prv$prv[,nt,"F","i-fix"]-prv$prv[,nt,"F","all-f"] )*100,
+          xlim=c(20,90), ylim=c(0,6), xlab="", ylab="", yaxt="n",
+          type="l", lty=1, lwd=c(4,2)+1, col="red" )
> matlines(a.pt, cbind( prv$prv[,nt,"F","obs"  ]-prv$prv[,nt,"F","i-fix"],
+                       prv$prv[,nt,"F","m-fix"]-prv$prv[,nt,"F","all-f"] )*100,
+          type="l", lty="22", lwd=c(4,2)+1, col="red" )
> mtext( "Contribution to prevalence (%)", side=2, outer=TRUE, line=1.5, las=0)
> mtext( "Age (years)", side=1, outer=TRUE, line=1.5 )
```

From figure **??** we see that the two possible ways of computing the contribution give pretty much the same results — the differences never exceed some 0.3%. Therefore, if we want to attribute fractions of the prevalence in 2010 to decreasing mortality and increasing incidence, we would want two measures that had a sum equal the the difference between the scenario with observed mortality and incidence rates ("`obs`"), and the scenario with rates fixed to those from 1995 ("`all-f`").

The thin lines at the bottom of figure **??** represents the prevalence at 1.1.1995, so it is pretty clear that the incidence an mortality rates as observed by 1995 did not provide for at steady state.

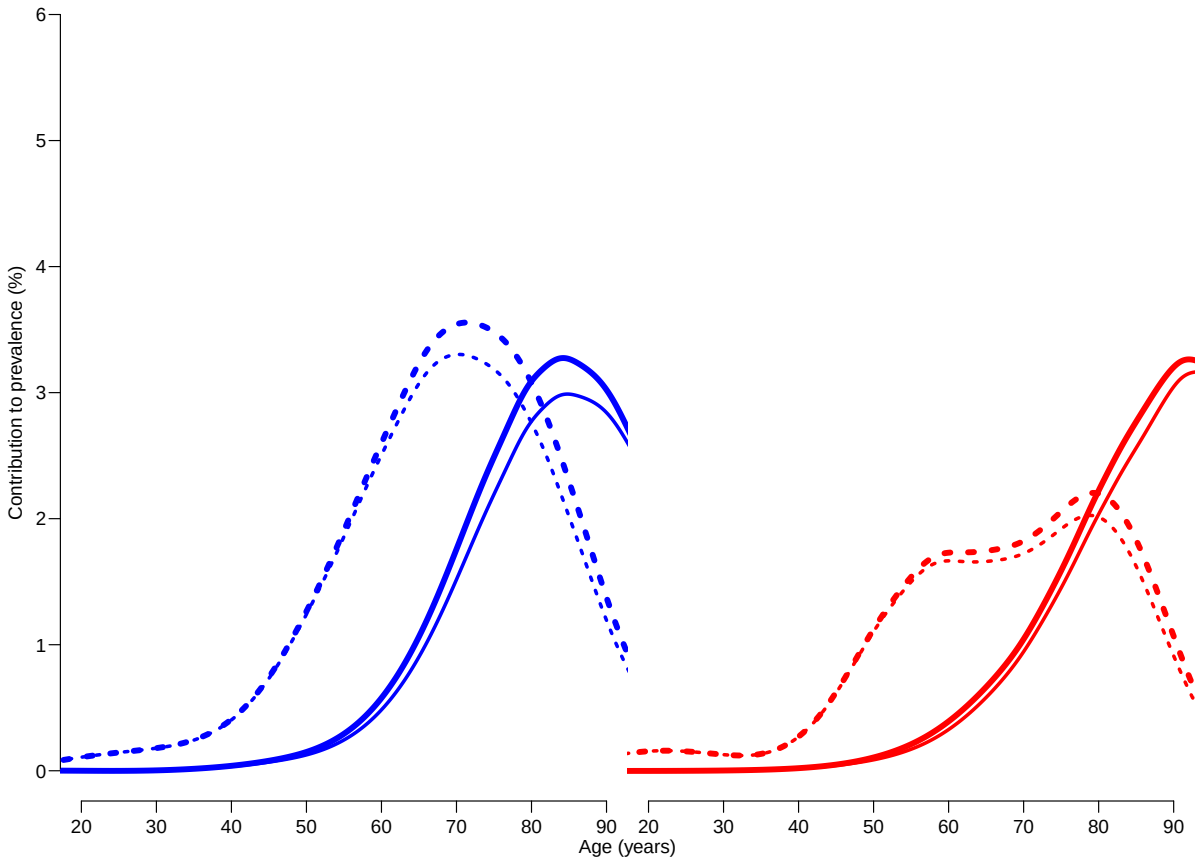

Figure 3.20: *Suggested contributions to age-specific prevalences from decreasing mortality over the last 17 years; the thick lines are obtained by subtracting the prediction based on fixing one rate from the one using the observed rates; thin lines based on subtracting the prediction based on fixing both rates from that where one is fixed. Full lines are for differences attributable to changes in mortality rates, broken lines are for changes attributable to changes in incidence rates.*

So basically we can subdivide the prevalence at any point in time into 4 components:

1. the "inherited" prevalences from 1995.
2. the prevalence attributable to rates of mortality and incidence as of 1995.
3. the prevalence attributable to the *increase* in the incidence rates.
4. the prevalence attributable to the *decrease* in the mortality rates.

So we now fill out the remaining 3 dimension of `prv`:

```
> prv$prv[,,,"mort" ] <- ( prv$prv[,,,"obs"  ]-prv$prv[,,,"m-fix"] +
+                          prv$prv[,,,"i-fix"]-prv$prv[,,,"all-f"] ) / 2
> prv$prv[,,,"inc"  ] <- ( prv$prv[,,,"obs"  ]-prv$prv[,,,"i-fix"] +
+                          prv$prv[,,,"m-fix"]-prv$prv[,,,"all-f"] ) / 2
> prv$prv[,,,"const"] <-   prv$prv[,,,"all-f"]-prv$prv[,rep(1,dim(prv$prv)[2]),,"obs"]
```

The components `obs`, `const`, `inc` and `mort` now together make up the total prevalence of diabetes for a given combinations of sex, age and date. Thus we can show these for each of the 17 dates 1996,. . . ,2012.

First we define a function to make the component plot, and then use this for men and women separately:

```
> poly.parts <-
+ function( x, crv, col, xlim, ylim, txt="" )
+ {
+ crv <- t(apply(cbind(0,crv),1,cumsum))
+ matplot( x, crv, type="n", xaxt="n", yaxt="n", xlab="", ylab="",
+          xlim=xlim, ylim=ylim, yaxs="i", bty="n" )
+ for( i in 2:ncol(crv) )
+ polygon( c(x,rev(x)), c(crv[,i],rev(crv[,i-1])),
+          col=col[i-1], border="transparent")
+ text( par("usr")[1:2]%*%c(0.1,0.9),
+       par("usr")[3:4]%*%c(0.9,0.1), txt, adj=c(1,0), font=2 )
+ }
```

We can now show the impact of changes in incidence and mortality on the age-specific prevalences:

```
> nt <- dim( prv$prv )[2]
> par( mfrow=c(1,2), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
> clr <- rgb(c(3,2,1.5,0)/3,c(3,2,1.5,0)/3,1)
> poly.parts( a.pt, cbind(prv$prv[,1 ,"M","obs"],
+                         prv$prv[,nt,"M",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,22) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> axis( side=2 )
> text( rep(25,3), 17:19+0.5,
+       c("Original","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=0.8 )
> # box(bty="c")
>
> clr <- rgb(1,c(3,2,1.5,0)/3,c(3,2,1.5,0)/3)
> poly.parts( a.pt, cbind(prv$prv[,1 ,"F","obs"],
+                         prv$prv[,nt,"F",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,22) )
> # axis( side=2 )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> text( rep(25,3), 17:19+0.5,
+       c("Original","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=0.8 )
> mtext( "Age", side=1, outer=TRUE, line=1.5, font=1, las=0 )
> mtext( "Prevalence of DM (%)", side=2, outer=TRUE, line=2, font=1, las=0 )
> # box(bty="]")
```

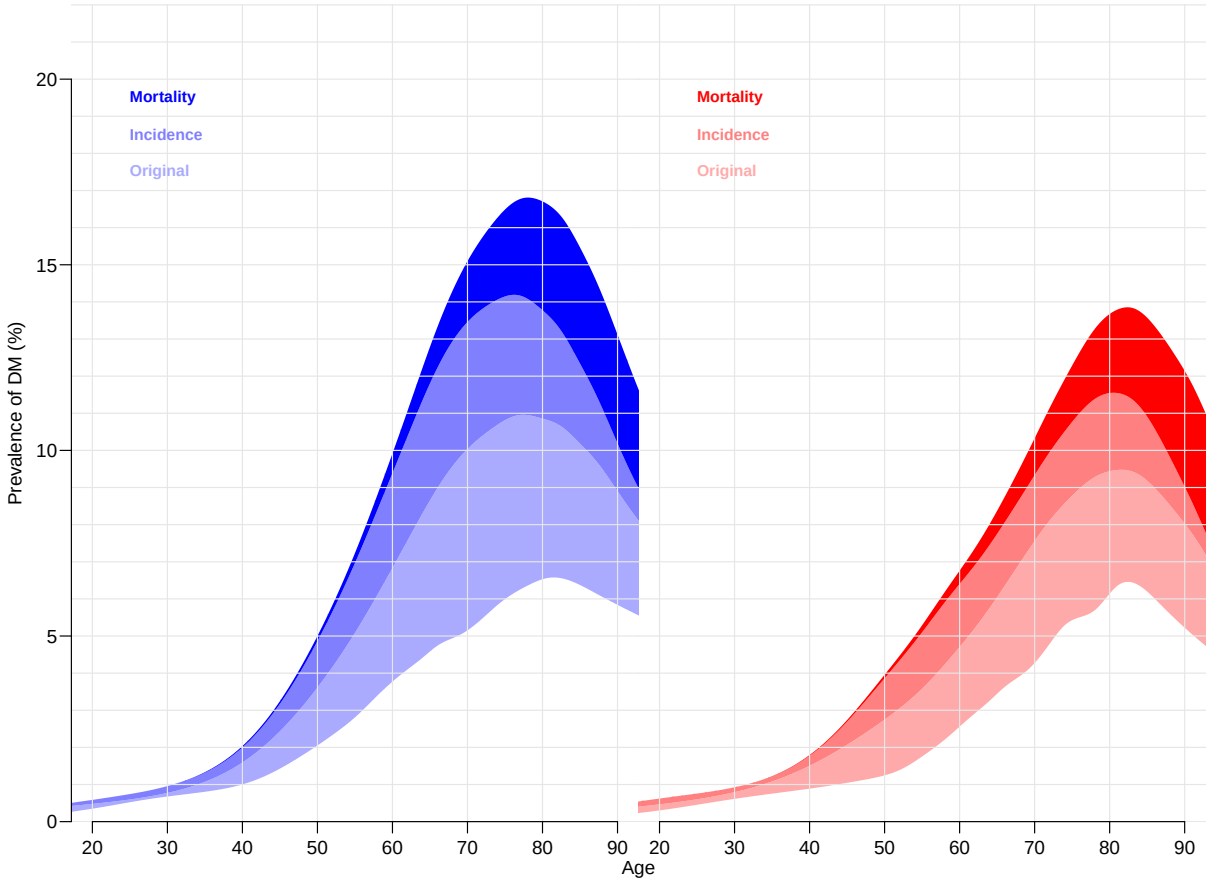

Figure 3.21: *Predicted age-specific prevalences of DM in Denmark 2012 among men (blue) and women (red), partitioned by the contribution from rates as they were in 1995 ("Original"), increases in incidence and decrease in mortality, respectively.*

3.5.7 The actual numbers of diabetes patients in Denmark

In the previous section we only looked at the age-specific prevalences, because these are the quantities that are driven by the incidence and mortality rates. However, it is also of interest to see how the actual number of diabetes patients would have looked under the different scenarios, specifically how the *number* of the current patients that can be attributed to the various components.

Also note that since the previous calculations were for age-specific prevalences we have a constant reference as the prevalences at 1995, but when we multiply by the population figures we would of course see differences in numbers and age-distribution of the diabetes population even if the age-specific prevalences were unchanged.

To show these effects we set up an array `prn` with structure like `prv$prv` to hold the number of diabetes patients by category, assuming the age-distribution in the population to be as actually observed (that is as extracted from Statistics Denmark, and as recorded in the data frame `pr`). However `prn` will have 100 age-classes rather than 100/`int`, and only 18 dates rater than 18/`int` as `prv$prv`.

This is done by selecting the relevant dates from `prv$prv` and then taking averages over age-classes.

```
> # The dates of the predicted prevalences as numerical
> prv.t <- as.numeric( dimnames(prv$prv)[["t"]] )
> # The dates where we want the prevalences
> prn.t <- 1995:2012
> # Find out where those are in prv.t
> nt <- length( prn.t )
> wh.t <- numeric( nt )
> for( it in 1:nt )
+    {
+    dd <- abs( prn.t[it]-prv.t )
+    wh.t[it] <- which(dd==min(dd))[1]
+    }
> # Take only prevalences at these dates
> prv.n <- data.frame( as.table( prv$prv[,wh.t,,] ) )
> str( prv.n )
  'data.frame':        252000 obs. of  5 variables:
   $ a   : Factor w/ 1000 levels "0.05","0.15",..: 1 2 3 4 5 6 7 8 9 10 ...
   $ t   : Factor w/ 18 levels "1995","1996",..: 1 1 1 1 1 1 1 1 1 1 ...
   $ sex : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
   $ what: Factor w/ 7 levels "obs","m-fix",..: 1 1 1 1 1 1 1 1 1 1 ...
   $ Freq: num  0 0.0004 0.000405 0.00041 0.000414 ...
> # Round the ages
> prv.n$a <- floor( as.numeric( as.character(prv.n$a) ) )
> prn <- xtabs( Freq ~ a + t + sex + what,
+                     data = aggregate( prv.n[5], prv.n[-5], mean ) )
> str( prn )
   xtabs [1:100, 1:18, 1:2, 1:7] 0.000377 0.000466 0.000522 0.000584 0.000654 ...
   - attr(*, "dimnames")=List of 4
    ..$ a   : chr [1:100] "0" "1" "2" "3" ...
    ..$ t   : chr [1:18] "1995" "1996" "1997" "1998" ...
    ..$ sex : chr [1:2] "M" "F"
    ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
   - attr(*, "class")= chr [1:2] "xtabs" "table"
   - attr(*, "call")= language xtabs(formula = Freq ~ a + t + sex + what, data = aggregate(prv.n[5],
> dimnames( prn )[[4]]
  [1] "obs"   "m-fix" "i-fix" "all-f" "mort"  "inc"   "const"
```

Now `prn` contains the prevalences components (as fractions) for 100 age classes and 18 dates. However, the components "`mort`", "`inc`" and "`const`", correspond to the prevalences

attributable to decline in mortality, increase in incidence and initial imbalance. But the first component is the prevalences predicted using the observed (well, fitted) rates. But would need the prevalences as of 1995 too, and the first 4 dimensions are really not needed.

So we restructure the 4th dimension, so we have the observed prevalences as of 1995, the three change-components, and finally the fitted total.

```
> prn <- prn[,,,c(1,5:7,1)]
> dimnames( prn )[[4]][1] <- "1995"
> prn[,,,"1995"] <- prn[,rep(1,dim(prn)[2]),,"obs"]
> str( prn )
 num [1:100, 1:18, 1:2, 1:5] 0.000377 0.000466 0.000522 0.000584 0.000654 ...
 - attr(*, "dimnames")=List of 4
  ..$ a   : chr [1:100] "0" "1" "2" "3" ...
  ..$ t   : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ what: chr [1:5] "1995" "mort" "inc" "const" ...
```

In principle we would now to multiply these prevalences by the population figures at these times, however for stability we multiply the `relative` size of the 4 components to the empirical prevalences observed. The population prevalence figures are in `pr`:

```
> head( pr )
  sex A    P X     N
1   M 0 1995 3 35612
2   M 0 1996 1 36055
3   M 0 1997 0 34853
4   M 0 1998 1 34774
5   M 0 1999 2 34076
6   M 0 2000 1 33906
> subset(pr,A<1 & P<1997)
      sex A    P X     N
1       M 0 1995 3 35612
2       M 0 1996 1 36055
1801    F 0 1995 0 34094
1802    F 0 1996 0 34051
> pop <- xtabs( N ~ A + P + sex, data=pr )
> dmp <- xtabs( X ~ A + P + sex, data=pr )
> str( pop )
 xtabs [1:100, 1:18, 1:2] 35612 34747 35082 33330 32974 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = N ~ A + P + sex, data = pr)
> str( dmp )
 xtabs [1:100, 1:18, 1:2] 3 4 6 5 12 21 22 34 29 29 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = X ~ A + P + sex, data = pr)
> str( prn )
 num [1:100, 1:18, 1:2, 1:5] 0.000377 0.000466 0.000522 0.000584 0.000654 ...
 - attr(*, "dimnames")=List of 4
  ..$ a   : chr [1:100] "0" "1" "2" "3" ...
  ..$ t   : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ what: chr [1:5] "1995" "mort" "inc" "const" ...
```

```
> prt <- apply( prn[,,,1:4], 1:3, sum )
> for( i in 1:4 )
+ prn[,,,i] <- (prn[,,,i]/prt) * dmp
```

First we draw a simple pyramid of the age-distribution of diabetes patients in Denmark:

```
> # Note: This uses the undocumented feature that if the first
> # number in a column is negative this is taken as the left endpoint of
> # the bar. So c(-m,m,f) is a bar starting at -m, and a division at
> # -m+m and an upper end at -m+m-f. Coloring is from the top, that is
> # the part stretching from -m+m to -m+m+f get the first color
> pp <- "2012"
> oo <- c("mort","inc","const","1995")
> lim <- 6
> clr <- c("red","blue")
> draw.dmp <-
+ function(pp)
+ {
+ par( mar=c(3,3,3,0), mgp=c(3,1,0)/1.6, las=1 )
+ barplot( height=t( cbind( -dmp[,pp,"M"],
+                           dmp[,pp,"M"],
+                           dmp[,pp,"F"] ) )/ 1000,
+          horiz=TRUE, col=clr,
+          border=NA,space=0,axes=FALSE,names.arg=rep("",dim(prn)[1]),
+          xlim=c(-1,1)*lim*1.05,xlab="Persons in 1 year class (1000s)",ylab="Age")
+ abline(h=seq(0,100,5),
+        v=seq(-lim,lim,0.5),
+        col="white")
+ axis( side=1, at=seq(-lim,lim,1), labels=abs(seq(-lim,lim,1)) )
+ axis( side=2, at=seq(0,100,20) )
+ mtext( pp, at=-lim, adj=1.4, cex=1.3, font=2 )
+ mtext( formatC(sum(dmp[,pp,"M"]),0,format="f",big.mark=","), at=-1, col="blue", line=0, cex=0.99 )
+ mtext( formatC(sum(dmp[,pp,"F"]),0,format="f",big.mark=","), at= 1, col="red" , line=0, cex=0.99 )
+ mtext( "N", at=0, line=0, cex=0.99 )
+ }
> pdf( "./graph/NDR-obs-film-m.pdf", width=8, height=6 )
> for( pp in paste(1995:2012) ) draw.dmp(pp)
> dev.off()
  null device
            1

> for( pp in paste(1995:2012) )
+ {
+ pdf( paste("./graph/NDR-obs-", pp, "-m.pdf", sep=""), width=8, height=6 )
+ draw.dmp(pp)
+ dev.off()
+ }
```

Using the same machinery we can also draw a population pyramid using colors that range from very light to full:

```
> shd <- c(0.0, 1.5, 2.0, 2.8) / 3
> een <- c(1,1,1,1)
> clr <- rgb( c(een,rev(shd)),
+             c(shd,rev(shd)),
+             c(shd,    een ) )
> clr
  [1] "#FF0000" "#FF8080" "#FFAAAA" "#FFEEEE" "#EEEEFF" "#AAAAFF" "#8080FF"
  [8] "#0000FF"

> # Note: The following uses the undocumented feature that if the first
> # number in a column is negative this is taken as the left endpoint of
> # the bar. So c(-m,m,f) is a bar starting at -m, and a division at
> # -m+m and an upper end at -m+m-f. Coloring is from the top, that is
> # the part stretching from -m+m to -m+m+f get the first color
```

```
> oo <- c("mort","inc","const","1995")
> draw.pyr <-
+ function(pp)
+ {
+ par( mar=c(3,3,3,0), mgp=c(3,1,0)/1.6, las=1 )
+ barplot( height=t( cbind( -apply( prn[,pp,"M",oo], 1, sum ),
+                                   prn[,pp,"M",oo],
+                                   prn[,pp,"F",rev(oo)] ) )/ 1000,
+          horiz=TRUE, col=clr[c(1,8:2)],
+          border=NA,space=0,axes=FALSE,names.arg=rep("",dim(prn)[1]),
+          xlim=c(-1,1)*lim*1.05,xlab="Persons in 1 year class (1000s)",ylab="Age" )
+ abline(h=seq(0,100,5),
+        v=seq(-lim,lim,0.5),
+        col="white")
+ axis( side=1, at=seq(-lim,lim,1), labels=abs(seq(-lim,lim,1)) )
+ axis( side=2, at=seq(0,100,20) )
+ tt <- addmargins( apply( prn[,pp,,oo],2:3, sum ), 2 )
+ nn <- tt / tt[,5] * 100
+ ppos <- 1:5-0.1
+ npos <- -rev(ppos)
+ mtext( pp, at=-lim, adj=1.8, line=2, cex=1.2, font=2 )
+ mtext( c(lg<- c("Mort","Inc","Const","Org","All"),rev(lg)),
+        at=c(npos,ppos), col="black", cex=0.99, line=2 )
+ mtext( formatC(tt["M",1:5],0,,"f",,,","), at=npos, col="blue", line=1, cex=0.99 )
+ mtext( formatC(tt["F",5:1],0,,"f",,,","), at=ppos, col="red" , line=1, cex=0.99 )
+ mtext( formatC(nn["M",1:4],1,4,"f"), at=npos[1:4], col="blue", line=0, cex=0.99 )
+ mtext( formatC(nn["F",4:1],1,4,"f"), at=ppos[2:5], col="red" , line=0, cex=0.99 )
+ mtext( "N", at=0, line=1, cex=0.99 )
+ mtext( "%", at=0, line=0, cex=0.99 )
+ }
> pdf( "./graph/NDR-film-m.pdf", width=9, height=6 )
> for( pp in paste(1995:2012) ) draw.pyr(pp)
> dev.off()
  null device
            1
> for( pp in paste(1996:2012) )
+ {
+ pdf( paste("./graph/NDR-", pp, "-m.pdf", sep=""), width=8, height=6 )
+ draw.pyr(pp)
+ dev.off()
+ }
```

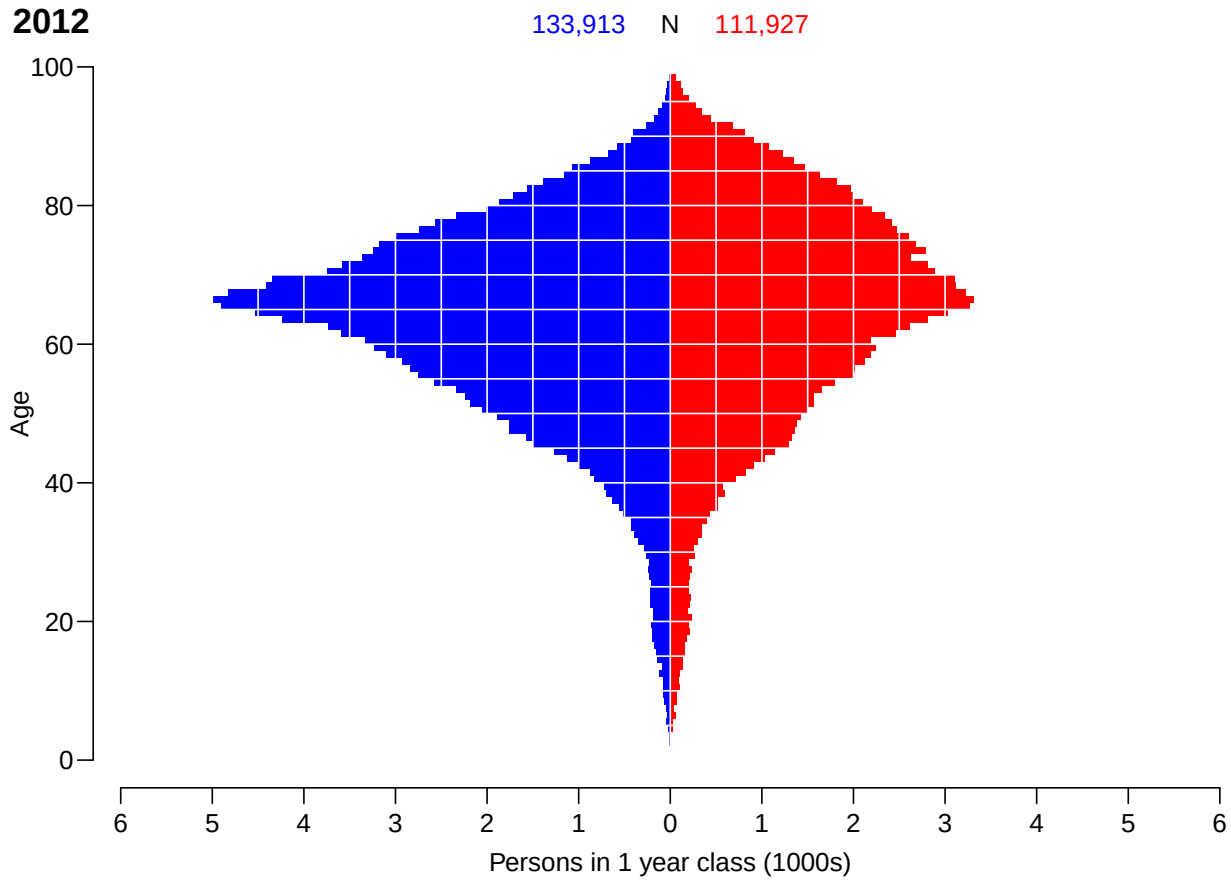

Figure 3.22: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2012.*

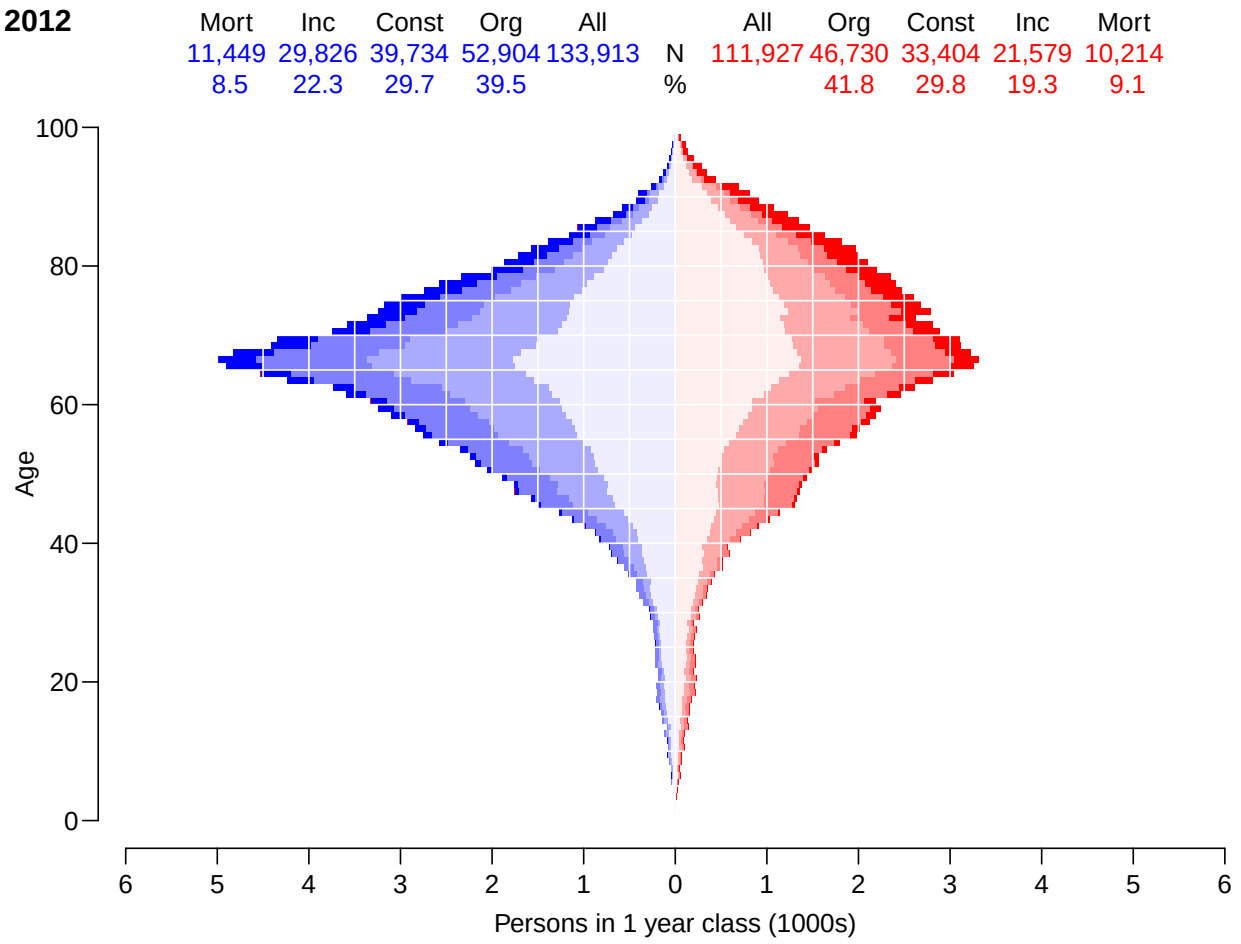

Figure 3.23: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2012, subdivided by the contribution from various causes: Mort: decrease in mortality, Inc: increase in incidence, Const: constant rates from 1995, Org: age-specific prevalence in 1995.*