Incidence, mortality and drug initiation in Danish T2 diabetes patients

SDC
March 2016
http://bendixcarstensen.com/DMreg/daffodil
Version 1.1

Compiled Tuesday 8th March, 2016, 19:19
from: /home/bendix/sdc/proj/daffodil/r/dafBxC.tex

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

Contents

Abstract:
Incidence, drugs and lifetime lost in T2D in Denmark

Aims

We intend to describe the trends in incidence of diabetes in Denmark, and in particular the changes in the fraction of patients initiated on glucose-lowering treatment at diagnosis.

Methods

Data

From Danish registers (national patient register, prescription register, health services register and adult diabetes clinical register), we defined T2 diabetes patients as persons who had at least one registration related to diabetes, the first after age 30. Persons with diagnosis before age 30 were excluded. Patients were subdivided by initiation on drug treatment or not at the time of diagnosis. The registers covered the period 1995–2014 incl. Population data were obtained from Statistics Denmark. All follow-up were tabulated by age and calendar time in 1 year intervals by age and calendar time.

Analyses

Incidence rates of the two types of T2 diabetes as well as total diabetes were analyzed by age-period-cohort (APC) models [1], implying that the rate-ratio between drug treatment and none at initiation (equal to the conditional odds) also followed an APC-model. Mortality was modeled similarly for persons with and without diabetes separately, and used to compute years of life lost to diabetes. Models were fitted using Poisson-models with smooth terms as restricted cubic splines. All analyses were done separately for men and women.

Results

The average annual increase in the total incidence rates over the period was 2% for men and 3% for women, while there was an annual decrease of 4% in the incidence of non-drug-treated T2 diabetes for both sexes, and 5% and 6% (men, resp. women) increase in incidence rates of drug-initiated diabetes.

The odds of being initiated on drugs at diagnosis was about 2 for both men and women in 1995, increasing to 10 for men and 15 for women — see figure 2.5 (page 31). The increase has been most predominant since 2007. The odds of drug initiation is increasing steeply by age for men, being twice as high in age 80 as in age 30, while the influence of age on initiation odds is much less for women.

The years of life lost to DM decreased over the period; for men diagnosed at age 50 from 9.2 to 5.7 years, for women from 10.2 to 5.4; in age 70 it was from 4.1 and 5.1 form men and women, bot falling to 2.9 years in 2015.

Conclusion

The incidence rates of T2 diabetes is increasing in Denmark, both for men and women, and the odds of having drug treatment at DM diagnosis is increasing by time. The decrease in the incidence rate of non-pharmacologically treated T2D indicates a change in the attitude among physicians towards a more aggressive pharmacological intervention at diagnosis. Mortality has decreased sharply, resulting in a 40% decrease in years of life lost to DM over the period.

Chapter 1

Overview and data structure

1.1 Aims

The aim of this study is to assess the occurrence of T2 diabetes (T2D), as well as the mortality of T2D patients, with a particular view to the occurrence and effect of medication of patients as observed at population level.

Specifically we shall use register data to assess

- incidence of T2D in Denmark:
  - non-pharmacologically treated
  - pharmacologically treated
- incidence of pharmacological treatment
- mortality of diabetes patients
- (for comparison) mortality of non-diabetic persons

Once these quantities have been described as functions of sex, (attained) age, calendar time, date of birth and — where relevant — duration of diabetes and treatment, we shall compute derived measures:

- prevalence of T2 diabetes
- life-time lost to diabetes
- expected life-time spent with diabetes without pharmacological treatment
- expected life-time spent with diabetes with pharmacological treatment

These quantities will be quantified as functions of age and calendar time of diagnosis of diabetes. Also, the time-trends in these quantities will be described.

1.2 Concepts

In order to formalize and illustrate the quantities we are concerned with, we show the transitions between the states of disease and death as defined above:

```
> library(Epi)
> TM <- matrix(NA,4,4)
> rownames(TM) <- colnames(TM) <- c("Well","DM-none","DM-drug","Dead")
> TM[   "Well","DM-none"] <-
+ TM[   "Well","DM-drug"] <-
+ TM[   "Well","Dead"   ] <-
+ TM["DM-none","DM-drug"] <-
+ TM["DM-none","Dead"   ] <-
+ TM["DM-drug","Dead"   ] <- 1
> wh <- list( x=c(10,85,85,60), y=c(50,90,10,50) )
> boxes( TM, boxpos=wh, hmult=3, wmult=1.5, col.bg=gray(c(1,1,1,0.8)) )
```

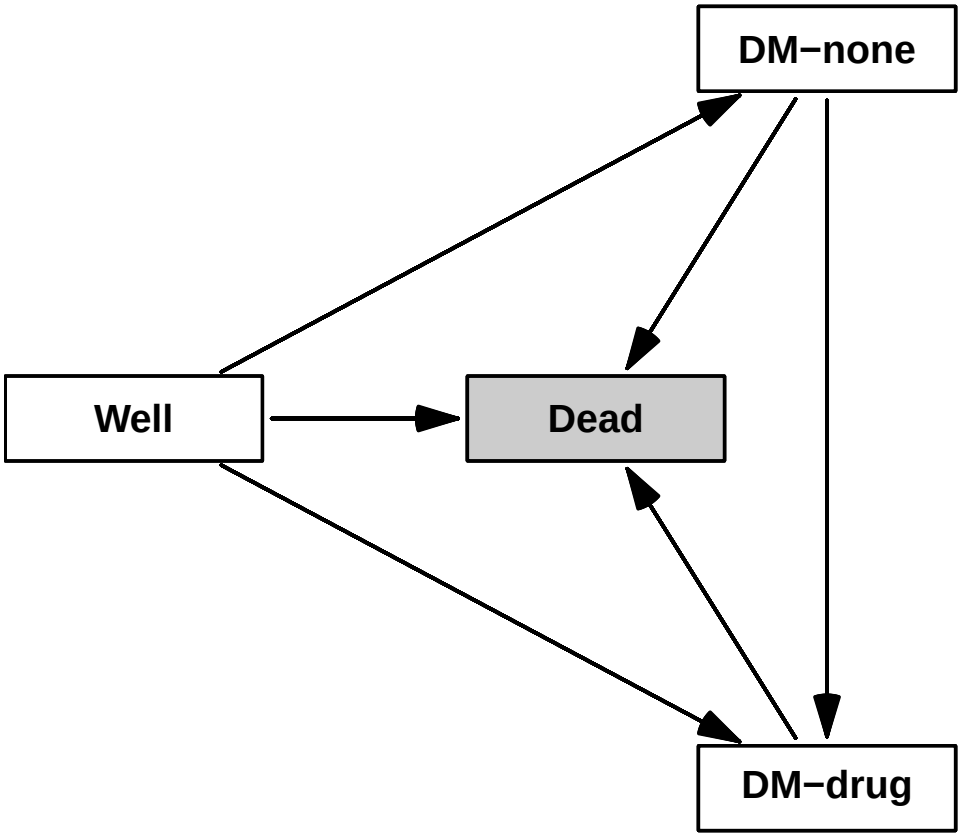

Figure 1.1: *Incidence and mortality transitions in the description of rates in diabetes occurrence. "DM-none" refers to patients diagnosed with diabetes but not on any antidiabetic drug, and "DM-drug" to patients that have filled at least one prescription for an antidiabetic drug.*

It is not necessarily of interest to model the incidence rates of "DM-none" and "DM-drug" separately, the *sum* of the rates (total incidence of T2 diabetes) and the rate-ratio (essentially the odds of starting on pharmacological treatment given diagnosis) are more likely of interest. Incidentally, modeling of the two rates by age-period-cohort models [1] immediately yields age-period-cohort models for the sum and the ratio too. We shall return to this in more detail in the modelling chapter.

Mortality rates can in the same vein be modeled separately or jointly for the two diabetes states, and again the sum or the ratio of the rates can be derived from the separate models. Likewise with the ratio relative to the mortality rates of the non-diabetic part of the population ("Well").

1.3 Data structure and construction

The structure of the analysis data sets is basically one table of person years and types of events for each of the three transient states (the white boxes in figure 1.1). The tables have different classifying sets; the "Well" dataset is classified by sex, age and calendar time; the "DM-none" by sex, age, calendar time and duration of DM and the "DM-drug" dataset by sex, age, calendar time, duration of DM and duration of drug treatment.

Data is constructed by tabulating events of DM and death, and the person-years spent in each state. The person-years in the "Well" state is obtained by taking the person-years in the entire Danish population and subtracting the person-years among those in the two DM states.

In the tabulations, age and calendar time is grouped in 1-year intervals, and time since diagnosis of DM, respectively drug initiation in 6-month intervals, the latter giving up to 40 different classes; table 1.1 gives an overview of the classifiers and outcome variables in the datasets.

Table 1.1: *Classifiers and outcome counters in the analysis data sets*

		Dataset (transient state)			
		Well	DM-none	DM-drug	Levels
Events:	DM-none	•			
	DM-drug	•	•		
	Dead	•	•	•	
Classifiers:	sex	•	•	•	2
	age	•	•	•	100
	cal.time	•	•	•	20
	DM dur.		•	•	40
	drug dur.			•	40
No. cells(1000s)		4	160	≈ 3000	

For the analysis of rates out of "Well" we will have $2 \times 100 \times 20 = 4,000$ entries, for the analysis of rates out of "DM-none" we will have $2 \times 100 \times 20 \times 40 = 160,000$ entries and for the analysis of rates out of "DM-drug" we will have (at most)
$2 \times 100 \times 20 \times 40 \times 40 = 6,400,000$. The latter is a gross exaggeration, since we will not have all 1600 possible combinations of DM duration and drug duration in the dataset, notably the latter is never larger than the former, so we will certainly have less than 3,200,000 records in that dataset.

Analysis of the rates is straight-forward using the number of events as response in Poisson-models with the log-person-years as offset and the effects of the time-scales as smooth parametric curves (natural splines).

1.4 SAS code enumerating follow-up

Here we devise the SAS-code to create for the three outlined datsets for analysis of incidence and mortality.

Note that we strat by renaming the variables from the mother dataset to names more mnemonically pertinent to the task at hand, and then make the Lexis split of data, followed by the relevant tabulation.

```
NOTE: Copyright (c) 2002-2012 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.4 (TS1M3)
      Licensed to FORSKNING 2, Site 50800723.
NOTE: This session is executing on the X64_ES08R2  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_ES08R2 WIN 6.1.7601 Service Pack 1 Server

NOTE: SAS initialization used:
      real time           15.38 seconds
      cpu time            1.76 seconds

1     libname her "../data" ;
NOTE: Libref HER was successfully assigned as follows:
      Engine:        V9
      Physical Name: E:\workdata\705093\BXC\data
2     options nosource2 ;
3     %inc "../../sas/Lexis.sas" ;
160
161   data her.base ;
162     set her.personbase_bxc ( rename = ( DOBirth = doBth
163                                         DoDeath = doDth
164                                  DoEndStudykorr = doEnd
165                                          DoRMPS = doMed
166                                   first_DM_korr = doDM ) ) ;
167     doIni = "01JAN1995"d ; * Convenience constant ;
168     format doEnd doMed doDM ddmmyy10. ;
169   run ;

NOTE: There were 420488 observations read from the data set HER.PERSONBASE_BXC.
NOTE: The data set HER.BASE has 420488 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           1.04 seconds
      cpu time            0.35 seconds


170
171   ************************************************************************
172   * Note that the Lexis macro uses the variables entry, exit and fail
173   * if they are not specified in the call of the macro. enttry and exit
174   * are date variables and fail the event indicator, 0 being the
175   * censoring code by default ;
176
177   ************************************************************************
178   * Set up data for enumerating PY and events of different types from
179   * the DMnone state - fishy data displayed and fixed ;
```

```
180  data DMnone ;
181    set her.base ;
182    if ( doDM > .z and ( doDM < doMed or doMed le .z ) ) ;
183    entry = max( doIni, doDM  ) ;
184     exit = min( doEnd, doMed, doDth ) ;
185   * fail is 2 if the exit is to medication, and 1 if it is to death ;
186     fail = ( doMed > .z and doMed = exit ) * 2 +
187            ( doDth > .z and doDth = exit ) ;
188    if fail gt 2 then do ;
189       put "What:" fail= dodm= doMed= doEnd= doDth= ;
190       fail = 1 ;
191       end ;
192  run ;

What:fail=3 doDM=11/04/2003 doMed=24/04/2003 doEnd=24/04/2003 doDth=24/04/2003
What:fail=3 doDM=17/05/2010 doMed=05/07/2010 doEnd=05/07/2010 doDth=05/07/2010
What:fail=3 doDM=11/02/2013 doMed=21/02/2013 doEnd=21/02/2013 doDth=21/02/2013
What:fail=3 doDM=26/04/1978 doMed=14/02/1995 doEnd=14/02/1995 doDth=14/02/1995
What:fail=3 doDM=06/02/2002 doMed=25/03/2010 doEnd=25/03/2010 doDth=25/03/2010
What:fail=3 doDM=13/02/2008 doMed=06/03/2008 doEnd=06/03/2008 doDth=06/03/2008
What:fail=3 doDM=12/04/2006 doMed=20/04/2006 doEnd=20/04/2006 doDth=20/04/2006
What:fail=3 doDM=27/09/2005 doMed=24/10/2005 doEnd=24/10/2005 doDth=24/10/2005
NOTE: There were 420488 observations read from the data set HER.BASE.
NOTE: The data set WORK.DMNONE has 216298 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.79 seconds
      cpu time            0.29 seconds


193
194  %Lexis( data = DMnone,
195             out = byP,
196          breaks = 0 to 125 by 1,
197          origin = '01JAN1900'd,
198           scale = 365.25,
199            left = P ) ;

NOTE: There were 216298 observations read from the data set WORK.DMNONE.
NOTE: The data set WORK.DISCRD has 0 observations and 14 variables.
NOTE: The data set WORK.BYP has 811158 observations and 14 variables.
NOTE: DATA statement used (Total process time):
      real time           2.62 seconds
      cpu time            2.62 seconds


200
201  data byP ; set byP ; P = P+1900 ; run ;

NOTE: There were 811158 observations read from the data set WORK.BYP.
NOTE: The data set WORK.BYP has 811158 observations and 14 variables.
NOTE: DATA statement used (Total process time):
      real time           0.60 seconds
      cpu time            0.60 seconds


202
203  %Lexis( data = byP,
204             out = byAP,
205          breaks = 0 to 125 by 1,
206          origin = doBth,
207           scale = 365.25,
208            left = A ) ;
```

```
NOTE: There were 811158 observations read from the data set WORK.BYP.
NOTE: The data set WORK.DISCRD has 0 observations and 15 variables.
NOTE: The data set WORK.BYAP has 1354864 observations and 15 variables.
NOTE: DATA statement used (Total process time):
      real time           9.11 seconds
      cpu time            9.11 seconds


209
210  %Lexis( data = byAP,
211           out = byAPD,
212         breaks = 0 to 125 by 0.5,
213         origin = doDM,
214          scale = 365.25,
215           left = dur,
216           risk = Y ) ;

NOTE: There were 1354864 observations read from the data set WORK.BYAP.
NOTE: The data set WORK.DISCRD has 0 observations and 17 variables.
NOTE: The data set WORK.BYAPD has 2387227 observations and 17 variables.
NOTE: DATA statement used (Total process time):
      real time           27.12 seconds
      cpu time            27.12 seconds


217
218  data byAPD ;
219    set byAPD ;
220    DMd = ( fail = 2 ) ;
221    Dth = ( fail = 1 ) ;
222  run ;

NOTE: There were 2387227 observations read from the data set WORK.BYAPD.
NOTE: The data set WORK.BYAPD has 2387227 observations and 19 variables.
NOTE: DATA statement used (Total process time):
      real time           2.30 seconds
      cpu time            2.30 seconds


223
224  proc summary  data=byAPD  nway ;
225    class sex A P dur ;
226    var Y DMd Dth ;
227    output out = DMnone ( keep = sex A P dur Y DMd Dth )
228           sum = ;
229  run ;

NOTE: There were 2387227 observations read from the data set WORK.BYAPD.
NOTE: The data set WORK.DMNONE has 107501 observations and 7 variables.
NOTE: PROCEDURE SUMMARY used (Total process time):
      real time           3.04 seconds
      cpu time            6.72 seconds


230
231  data her.DMnone ;
232    set DMnone ;
233    label A="Age" P="Date" dur="DM duration" ;
234  run ;

NOTE: There were 107501 observations read from the data set WORK.DMNONE.
```

```
NOTE: The data set HER.DMNONE has 107501 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.21 seconds
      cpu time            0.07 seconds


235
236  ********************************************************************
237  * Set up data for enumerating PY and events of different types from
238  * the DMdrug state ;
239  data DMdrug ;
240    set her.base ;
241    if ( doMed > .z ) ;
242    entry = doMed ;
243     exit = min( doEnd, doDth ) ;
244     fail = ( doDth > .z and doDth = exit ) ;
245  run ;

NOTE: There were 420488 observations read from the data set HER.BASE.
NOTE: The data set WORK.DMDRUG has 363570 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           0.81 seconds
      cpu time            0.29 seconds


246
247  %Lexis( data = DMdrug,
248             out = byP,
249          breaks = 0 to 125 by 1,
250          origin = "01JAN1900"d,
251           scale = 365.25,
252            left = P ) ;

NOTE: There were 363570 observations read from the data set WORK.DMDRUG.
NOTE: The data set WORK.DISCRD has 0 observations and 14 variables.
NOTE: The data set WORK.BYP has 2769658 observations and 14 variables.
NOTE: DATA statement used (Total process time):
      real time           4.86 seconds
      cpu time            4.86 seconds


253
254  data byP ; set byP ; P = P+1900 ; run ;

NOTE: There were 2769658 observations read from the data set WORK.BYP.
NOTE: The data set WORK.BYP has 2769658 observations and 14 variables.
NOTE: DATA statement used (Total process time):
      real time           2.18 seconds
      cpu time            2.18 seconds


255
256  %Lexis( data = byP,
257             out = byAP,
258          breaks = 0 to 125 by 1,
259          origin = doBth,
260           scale = 365.25,
261            left = A ) ;

NOTE: There were 2769658 observations read from the data set WORK.BYP.
NOTE: The data set WORK.DISCRD has 0 observations and 15 variables.
NOTE: The data set WORK.BYAP has 5305662 observations and 15 variables.
```

```
NOTE: DATA statement used (Total process time):
      real time           23.32 seconds
      cpu time            23.30 seconds


262
263  %Lexis( data = byAP,
264           out = byAPD,
265        breaks = 0 to 125 by 0.5,
266        origin = doDM,
267         scale = 365.25,
268          left = dur ) ;

NOTE: There were 5305662 observations read from the data set WORK.BYAP.
NOTE: The data set WORK.DISCRD has 0 observations and 16 variables.
NOTE: The data set WORK.BYAPD has 10268718 observations and 16 variables.
NOTE: DATA statement used (Total process time):
      real time           1:16.88
      cpu time            1:16.70


269
270  %Lexis( data = byAPD,
271           out = byAPDd,
272        breaks = 0 to 125 by 0.5,
273        origin = doMed,
274         scale = 365.25,
275          left = ddur,
276          risk = Y ) ;

NOTE: There were 10268718 observations read from the data set WORK.BYAPD.
NOTE: The data set WORK.DISCRD has 0 observations and 18 variables.
NOTE: The data set WORK.BYAPDD has 12778047 observations and 18 variables.
NOTE: DATA statement used (Total process time):
      real time           2:40.72
      cpu time            2:40.07


277
278  proc summary  data = byAPDd ( rename = ( fail=Dth ) )  nway ;
279    class sex A P dur ddur ;
280    var Y Dth ;
281    output out = DMdrug ( keep = sex A P dur ddur Y Dth )
282           sum = ;
283  run ;

NOTE: There were 12778047 observations read from the data set WORK.BYAPDD.
NOTE: The data set WORK.DMDRUG has 1023047 observations and 7 variables.
NOTE: PROCEDURE SUMMARY used (Total process time):
      real time           13.18 seconds
      cpu time            31.85 seconds


284
285  data her.DMdrug ;
286    set DMdrug ;
287    label A="Age" P="Date" dur="DM duration" ddur="Drug duration" ;
288  run ;

NOTE: There were 1023047 observations read from the data set WORK.DMDRUG.
NOTE: The data set HER.DMDRUG has 1023047 observations and 7 variables.
NOTE: DATA statement used (Total process time):
```

```
      real time           1.32 seconds
      cpu time            0.59 seconds


289
290   ********************************************************************
291   * Enumerate the total deaths and risk time in DMnone and DMdrug for
292   * subtraction from the total Danish population data ;
293   data FU ;
294     set her.DMnone ( keep = sex A P Y Dth )
295         her.DMdrug ( keep = sex A P Y Dth ) ;
296   run ;

NOTE: There were 107501 observations read from the data set HER.DMNONE.
NOTE: There were 1023047 observations read from the data set HER.DMDRUG.
NOTE: The data set WORK.FU has 1130548 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           2.27 seconds
      cpu time            0.31 seconds


297
298   proc summary  data = FU  nway ;
299     class sex A P ;
300     var Y Dth ;
301     output out = YDth ( keep = sex A P Y Dth )
302            sum = ;
303   run ;

NOTE: There were 1130548 observations read from the data set WORK.FU.
NOTE: The data set WORK.YDTH has 3152 observations and 5 variables.
NOTE: PROCEDURE SUMMARY used (Total process time):
      real time           0.90 seconds
      cpu time            1.81 seconds


304
305   ********************************************************************
306   * Set up data for enumerating events of different types from
307   * the Well state ;
308   data well ;
309     set her.base ;
310     DM  = ( doDM  > doIni & doDM < doMed ) ;
311     DMd = ( doMed > doIni & doDM = doMed ) ;
312       A = floor( ( doDM - doBth ) / 365.25 ) ;
313       P = floor( ( doDM         ) / 365.25 ) + 1960 ;
314   run ;

NOTE: There were 420488 observations read from the data set HER.BASE.
NOTE: The data set WORK.WELL has 420488 observations and 12 variables.
NOTE: DATA statement used (Total process time):
      real time           1.10 seconds
      cpu time            0.40 seconds


315
316   proc summary  data = well  nway ;
317     class sex A P ;
318     var DM DMd ;
319     output out = DMnd ( keep = sex A P DM DMd )
320            sum = ;
321   run ;
```

```
NOTE: There were 420488 observations read from the data set WORK.WELL.
NOTE: The data set WORK.DMND has 5153 observations and 5 variables.
NOTE: PROCEDURE SUMMARY used (Total process time):
      real time           0.38 seconds
      cpu time            0.84 seconds


322
323  ********************************************************************
324  * Merge the follow-up events in Well and the person-years and deaths and
325  * PY in the two DM states for subsequent subtraction from total
326  * population Y and D ;
327  data her.Well ;
328    merge YDth /* variables  Y and Dth */
329          DMnd /* variables DM and DMd */ ;
330    by sex A P ;
331  run ;

NOTE: There were 3152 observations read from the data set WORK.YDTH.
NOTE: There were 5153 observations read from the data set WORK.DMND.
NOTE: The data set HER.WELL has 5365 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.21 seconds
      cpu time            0.03 seconds


332
333  ********************************************************************
334  * Convert the relevant datasets to Xport datasets ;
335  libname xDMn xport "..\data\DMnone.xpt" ;
NOTE: Libref XDMN was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: E:\workdata\705093\BXC\data\DMnone.xpt
336  libname xDMd xport "..\data\DMdrug.xpt" ;
NOTE: Libref XDMD was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: E:\workdata\705093\BXC\data\DMdrug.xpt
337  libname xWll xport "..\data\Well.xpt"   ;
NOTE: Libref XWLL was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: E:\workdata\705093\BXC\data\Well.xpt
338  proc copy in=her out=xDMn memtype=data ; select DMnone ; run ;

NOTE: Copying HER.DMNONE to XDMN.DMNONE (memtype=DATA).
NOTE: There were 107501 observations read from the data set HER.DMNONE.
NOTE: The data set XDMN.DMNONE has 107501 observations and 7 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           0.43 seconds
      cpu time            0.15 seconds


339  proc copy in=her out=xDMd memtype=data ; select DMdrug ; run ;

NOTE: Copying HER.DMDRUG to XDMD.DMDRUG (memtype=DATA).
NOTE: There were 1023047 observations read from the data set HER.DMDRUG.
NOTE: The data set XDMD.DMDRUG has 1023047 observations and 7 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           2.44 seconds
      cpu time            1.23 seconds
```

```
340  proc copy in=her out=xWll memtype=data ; select Well   ; run ;

NOTE: Copying HER.WELL to XWLL.WELL (memtype=DATA).
NOTE: There were 5365 observations read from the data set HER.WELL.
NOTE: The data set XWLL.WELL has 5365 observations and 7 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           0.04 seconds
      cpu time            0.01 seconds
```

1.5 Population, deaths and person-years in DK

Here we read data set of population size and no. of deaths extracted from the data bank of Statistics Denmark. The aim is to construct a dataset of deaths and PY for the entire DK population classiefied by sex, age and period (1-year classes) as well as a dataset of population size classified by sex, age and date (in one year classes/equidistances). These will be used as the necessary population backgroudn for the analysis of incidence/mortality and prevalence.

We need the `N2Y` function from `Epi` and the `na.locf` from the `zoo` package:

```
> library(Epi)
> library(zoo)
```

1.5.1 Population data

The population data has a line with the year ahead of data for each year, which we must carry forward and subsequently delete:

```
> pop <- read.csv("/home/bendix/sdc/demodb/raw/pop-1980-2016.csv")
> head( pop )
     P  A     M     F
1 1980 NA    NA    NA
2   NA  0 30347 28886
3   NA  1 31791 30373
4   NA  2 31780 30379
5   NA  3 33736 31868
6   NA  4 36663 35477
> pop$P <- na.locf(pop$P)
> pop <- subset(pop,!is.na(A))
> head(pop)
     P A     M     F
2 1980 0 30347 28886
3 1980 1 31791 30373
4 1980 2 31780 30379
5 1980 3 33736 31868
6 1980 4 36663 35477
7 1980 5 36275 34876
> summary(pop)
       P             A               M                F
 Min.   :1980   Min.   :  0.0   Min.   :    0.0   Min.   :    0
 1st Qu.:1989   1st Qu.: 31.0   1st Qu.:  569.2   1st Qu.: 1860
 Median :1998   Median : 62.5   Median :25607.5   Median :26420
 Mean   :1998   Mean   : 62.5   Mean   :20857.5   Mean   :21346
 3rd Qu.:2007   3rd Qu.: 94.0   3rd Qu.:36199.5   3rd Qu.:35149
 Max.   :2016   Max.   :125.0   Max.   :46208.0   Max.   :44295
```

We now have the population *size* i 1-year classes at 1 January each year, so we write these presence data to a file to be used as basis for the analysis of prevalences of DM:

```
> N.dk <- subset( rbind( data.frame( sex="M", pop[,c("A","P")], N=pop$M ),
+                        data.frame( sex="F", pop[,c("A","P")], N=pop$F ) ),
+                 A<99 )
> head( N.dk )
  sex A    P     N
2   M 0 1980 30347
3   M 1 1980 31791
4   M 2 1980 31780
5   M 3 1980 33736
6   M 4 1980 36663
7   M 5 1980 36275
> write.csv( subset( N.dk, P>1994 & P<2016 ), file="../data/nDK.csv", row.names=FALSE )
```

1.5.2 Person-years

Using the presence data we use `N2Y` to compute PY in Lexis triangles; however deaths do not come in triangles, so we collapse to A-sets (classified only by age and period) — separately for men and women.

```
> Y.m <- transform( N2Y( A=A, P=P, N=M, data=pop ), A=floor(A), P=floor(P) )
> Y.f <- transform( N2Y( A=A, P=P, N=F, data=pop ), A=floor(A), P=floor(P) )
> Y.m <- aggregate( Y.m$Y, Y.m[,c("A","P")], FUN=sum )
> Y.f <- aggregate( Y.f$Y, Y.f[,c("A","P")], FUN=sum )
```

We then stack the data for the two sexes, and restrict to ages 0–98 as this is the range in which the mortality data is available:

```
> Y <- subset( rbind( data.frame( sex="M", Y.m ),
+                     data.frame( sex="F", Y.f ) ),
+              A<99 )
> names(Y)[4] <- "Y"
> str(Y)
'data.frame':        7128 obs. of  4 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  0 1 2 3 4 5 6 7 8 9 ...
 $ P  : num  1980 1980 1980 1980 1980 1980 1980 1980 1980 1980 ...
 $ Y  : num  29777 31096 31796 32756 35209 ...
> summary(Y)
 sex            A               P               Y
 M:3564   Min.   : 0   Min.   :1980   Min.   :   63.83
 F:3564   1st Qu.:24   1st Qu.:1989   1st Qu.:19359.75
          Median :49   Median :1998   Median :30858.92
          Mean   :49   Mean   :1998   Mean   :26837.99
          3rd Qu.:74   3rd Qu.:2006   3rd Qu.:36623.12
          Max.   :98   Max.   :2015   Max.   :45388.50
```

1.5.3 Death counts

Data are in the same format as the population data, so we do the same exercise as before using `na.locf`:

```
> dth <- read.csv("/home/bendix/sdc/demodb/raw/dth-1980-2015.csv")
> head(dth)
```

```
     P  A   M   F
1 1980 NA  NA  NA
2   NA  0 274 210
3   NA  1  26  20
4   NA  2  21  12
5   NA  3  19   7
6   NA  4   7  12
```

```
> dth$P <- na.locf(dth$P)
> dth <- subset(dth,!is.na(A))
> head(dth)
```

```
     P A   M   F
2 1980 0 274 210
3 1980 1  26  20
4 1980 2  21  12
5 1980 3  19   7
6 1980 4   7  12
7 1980 5   9   7
```

The deaths are only given till age 98, plus the summary category 99+, which we exclude:

```
> D <- subset( rbind( data.frame( sex="M", dth[,c("A","P")], D=dth$M ),
+                     data.frame( sex="F", dth[,c("A","P")], D=dth$F ) ),
+              A<99 )
> str( D )
'data.frame':        7128 obs. of  4 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : int  0 1 2 3 4 5 6 7 8 9 ...
 $ P  : int  1980 1980 1980 1980 1980 1980 1980 1980 1980 1980 ...
 $ D  : int  274 26 21 19 7 9 12 6 14 10 ...
```

1.5.4 Merging deaths and PY

With the PY and the deaths classified in the same way, we can now merge them to the dataset which is needed as basis for the mortality analyses:

```
> M.dk <- merge( Y, D )
> str( M.dk )
'data.frame':        7128 obs. of  5 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  1980 1981 1982 1983 1984 ...
 $ Y  : num  28422 26976 25836 25296 25108 ...
 $ D  : int  210 181 177 163 156 182 204 177 201 204 ...
```

```
> head( M.dk )
```

```
  sex A    P       Y   D
1   F 0 1980 28422.5 210
2   F 0 1981 26976.0 181
3   F 0 1982 25835.5 177
4   F 0 1983 25296.5 163
5   F 0 1984 25108.5 156
6   F 0 1985 25844.0 182
```

```
> summary( M.dk )
```

```
 sex            A             P             Y                  D
 M:3564   Min.   : 0   Min.   :1980   Min.   :   63.83   Min.   :   0.0
 F:3564   1st Qu.:24   1st Qu.:1989   1st Qu.:19359.75   1st Qu.:  20.0
          Median :49   Median :1998   Median :30858.92   Median : 123.0
          Mean   :49   Mean   :1998   Mean   :26837.99   Mean   : 287.1
          3rd Qu.:74   3rd Qu.:2006   3rd Qu.:36623.12   3rd Qu.: 498.0
          Max.   :98   Max.   :2015   Max.   :45388.50   Max.   :1276.0
```

```
> write.csv( subset( M.dk, P>1994 & P<2015 ), file="../data/mDK.csv", row.names=FALSE )
```

We can test that we actually can get a readable dataset out of this:

```
> xDK <- read.csv( file="../data/mDK.csv" )
> head( xDK )
  sex A    P       Y   D
1   F 0 1995 34072.0 153
2   F 0 1996 33494.5 157
3   F 0 1997 32903.0 163
4   F 0 1998 32499.5 148
5   F 0 1999 32263.5 112
6   F 0 2000 32534.0 149
> summary( xDK )
 sex            A            P             Y                  D
 F:1980   Min.   : 0   Min.   :1995   Min.   :   86.83   Min.   :   0.0
 M:1980   1st Qu.:24   1st Qu.:2000   1st Qu.:19271.75   1st Qu.:  16.0
          Median :49   Median :2004   Median :32625.08   Median : 122.0
          Mean   :49   Mean   :2004   Mean   :27412.83   Mean   : 282.3
          3rd Qu.:74   3rd Qu.:2009   3rd Qu.:36703.83   3rd Qu.: 494.0
          Max.   :98   Max.   :2014   Max.   :45388.50   Max.   :1276.0
```

This is the dataset uploaded to Statistics Denmark to from the basis for incidence and mortality for the non-diabetic part of the population.

Chapter 2

Incidence of DM

This chapter describes the modeling of the rates of DM and the ratio of these.

2.1 Theory

Suppose we have separate age-period-cohort models for the two types of diabetes incidence:

$$\log\big(\lambda_i(a,p)\big) = f_i(a) + g_i(p) + h_i(p-a), \quad i = n, d$$

where n refers to DM-none (diabetes without drug treatment) and d refers to DM-drug (diabetes with drug treatment). We can fit both models separately and show effects using a cohort-flat parametrization for example, using a suitable reference date for the period effect, 1 January 2010, say. The log-RR between the two incidences is merely

$$\log(\mathrm{RR}_{d \text{ vs } n}) = \big(f_d(a) - f_n(a)\big) + \big(g_d(p) - g_n(p)\big) + \big(h_d(p-a) - h_n(p-a)\big)$$

which again is an age-period-cohort model, but now for the RR between drug and non-drug incidence of DM. If the cohort restriction and the reference dates for the period effects are identical the RR will obey the same constraints as well.

The RR will tell how the odds of initiation on drugs (given diagnosis of DM) varies by age and period (and how this may be modified by age).

Incidentally, we can of course also without problems *add* the two sets of rates, but since rates are additive on the rate-scale and *not* on the log-rate scale, the implicit model fitted for the sum of the rates will not be an age-period-cohort model.

However, the separate models for the two types of T2D diagnosis can be fitted as one model with an interaction term, using a stacked dataset for the two transitions (so a dataset where the person-years are replicated but with different versions of the response - one being DM-none, the other DM-drug). In that interaction model it could be tested whether the cohort effects were identical, whether the period effects were identical and whether the age-effects were proportional. These tests basically correspond to an investigation of how the components of the rate-ratios compare to 1.

2.2 Follow-up data

2.2.1 The Well state

First load the relevant follow-up data and adjust the person-years and deaths from the well state, but first we need a few packages:

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

First the population data with the person-years and deaths in the entire population:

```
> popy <- read.csv(file="../data/mDK.csv")
> head(popy)
  sex A    P       Y   D
1   F 0 1995 34072.0 153
2   F 0 1996 33494.5 157
3   F 0 1997 32903.0 163
4   F 0 1998 32499.5 148
5   F 0 1999 32263.5 112
6   F 0 2000 32534.0 149
> str(popy)
'data.frame':        3960 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ P  : int  1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 ...
 $ Y  : num  34072 33495 32903 32500 32264 ...
 $ D  : int  153 157 163 148 112 149 156 125 121 130 ...
> subset( popy, A<2 & P==1999 )
     sex A    P        Y   D
5      F 0 1999 32263.50 112
65     F 1 1999 32695.83   9
1985   M 0 1999 33990.00 170
2045   M 1 1999 34559.50  12
> popy <- subset( popy, A>29 )
> table( popy$A )
30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
90 91 92 93 94 95 96 97 98
40 40 40 40 40 40 40 40 40
> table( popy$P )
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
 138  138  138  138  138  138  138  138  138  138  138  138  138  138  138  138  138  138
2013 2014
 138  138
```

The number of transitions from Well to DM states as well as the person-years and deaths from these states:

```
> well <- read.xport(file="../data/Well.xpt")
> names(well) <- c("sex","A","P","Yd","Dd","DM","DMd")
> well$sex <- factor( well$sex, labels=c("M","F") )
> well <- subset( well, P>1994 & A<99 )
> head(well)
   sex  A    P        Yd Dd DM DMd
29   M 30 1995  9.283368  0 10   8
30   M 30 1996 10.414784  0  7  10
31   M 30 1997  8.388775  1 11   5
32   M 30 1998  8.234771  0 11   8
33   M 30 1999  9.779603  1  7   9
34   M 30 2000  8.357974  0 12   4
```

```
> str(well)
'data.frame':        2760 obs. of  7 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  30 30 30 30 30 30 30 30 30 30 ...
 $ P  : num  1995 1996 1997 1998 1999 ...
 $ Yd : num  9.28 10.41 8.39 8.23 9.78 ...
 $ Dd : num  0 0 1 0 1 0 1 1 0 0 ...
 $ DM : num  10 7 11 11 7 12 11 21 11 9 ...
 $ DMd: num  8 10 5 8 9 4 4 2 8 10 ...
> summary(well)
 sex            A               P               Yd                  Dd
 M:1380   Min.   :30   Min.   :1995   Min.   :   2.932   Min.   :  0.00
 F:1380   1st Qu.:47   1st Qu.:2000   1st Qu.: 352.025   1st Qu.:  7.00
          Median :64   Median :2004   Median : 954.428   Median : 39.00
          Mean   :64   Mean   :2004   Mean   :1117.124   Mean   : 60.46
          3rd Qu.:81   3rd Qu.:2009   3rd Qu.:1603.229   3rd Qu.:110.00
          Max.   :98   Max.   :2014   Max.   :5291.743   Max.   :227.00

       DM              DMd
 Min.   :  0.00   Min.   :  0.00
 1st Qu.: 11.00   1st Qu.: 25.00
 Median : 28.00   Median : 65.00
 Mean   : 34.21   Mean   : 74.56
 3rd Qu.: 52.00   3rd Qu.:101.00
 Max.   :149.00   Max.   :441.00
 NA's   :22       NA's   :22
> table( well$A )
30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89
40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 40
90 91 92 93 94 95 96 97 98
40 40 40 40 40 40 40 40 40
> table( well$P )
1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012
 138  138  138  138  138  138  138  138  138  138  138  138  138  138  138  138  138  138
2013 2014
 138  138
```

We now merge the observation dataset to the population dataset and insert the relevant 0s etc.:

```
> well <- merge( popy, well, all.x=TRUE )
> well[is.na(well)] <- 0
> well <- transform( well, Y = Y - Yd,
+                          D = D - Dd,
+                          A = A + 0.5,
+                          P = P + 0.5 )
> save( well, file="../data/well.Rda" )
```

Now the variable `Y` has the person-years in the state Well, and the exit counts are in `D` (deaths), `DM` (diabetes, no drugs) and `DMd` (diabetes, drug initiated).

2.2.2 DM without drug treatment

There are two possible exits from the DM-none state, to DM-drug and to Death; here we groom the tabulated dataset and save it as an R analysis dataset:

```
> DMn <- read.xport(file="../data/DMnone.xpt")
> head(DMn)
```

```
  SEX  A    P DUR           Y DMD DTH
1   1 30 1994 0.0 0.016427105   0   0
2   1 30 1994 0.5 0.006160164   0   0
3   1 30 1995 0.0 4.938056126   7   0
4   1 30 1995 0.5 0.625256674   2   0
5   1 30 1996 0.0 2.371321013   1   0
6   1 30 1996 0.5 0.601300479   0   0
```

```
> names(DMn) <- c("sex","A","P","dur","Y","DMd","Dth")
> DMn$sex <- factor( DMn$sex, labels=c("M","F") )
> DMn <- transform( subset( DMn, P>1994 & A<99 ),
+                   A = A + 0.5,
+                   P = P + 0.5,
+                 dur = dur + 0.25 )
> head(DMn)
```

```
  sex    A      P  dur         Y DMd Dth
3   M 30.5 1995.5 0.25 4.9380561   7   0
4   M 30.5 1995.5 0.75 0.6252567   2   0
5   M 30.5 1996.5 0.25 2.3713210   1   0
6   M 30.5 1996.5 0.75 0.6013005   0   0
7   M 30.5 1997.5 0.25 2.8052704   7   0
8   M 30.5 1997.5 0.75 0.2474333   0   0
```

```
> str(DMn)
```

```
'data.frame':        101531 obs. of  7 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 ...
 $ P  : num  1996 1996 1996 1996 1998 ...
 $ dur: num  0.25 0.75 0.25 0.75 0.25 0.75 0.25 0.75 0.25 0.75 ...
 $ Y  : num  4.938 0.625 2.371 0.601 2.805 ...
 $ DMd: num  7 2 1 0 7 0 4 1 2 0 ...
 $ Dth: num  0 0 0 0 0 0 0 0 1 0 ...
```

```
> summary(DMn)
```

```
 sex             A               P             dur              Y
 M:47973   Min.   :30.50   Min.   :1996   Min.   : 0.25   Min.   : 0.00068
 F:53558   1st Qu.:55.50   1st Qu.:2000   1st Qu.: 4.75   1st Qu.: 0.64613
           Median :68.50   Median :2006   Median :10.25   Median : 2.30322
           Mean   :68.07   Mean   :2006   Mean   :11.21   Mean   : 5.35331
           3rd Qu.:81.50   3rd Qu.:2010   3rd Qu.:16.25   3rd Qu.: 7.04466
           Max.   :98.50   Max.   :2014   Max.   :44.25   Max.   :70.26728
      DMd              Dth
 Min.   :  0.000   Min.   : 0.0000
 1st Qu.:  0.000   1st Qu.: 0.0000
 Median :  0.000   Median : 0.0000
 Mean   :  1.568   Mean   : 0.3158
 3rd Qu.:  1.000   3rd Qu.: 0.0000
 Max.   :256.000   Max.   :16.0000
```

```
> table( DMn$A )
```

```
30.5 31.5 32.5 33.5 34.5 35.5 36.5 37.5 38.5 39.5 40.5 41.5 42.5 43.5 44.5 45.5 46.5 47.5
  79  158  239  317  397  474  553  634  710  789  867  939 1013 1087 1159 1231 1305 1371
48.5 49.5 50.5 51.5 52.5 53.5 54.5 55.5 56.5 57.5 58.5 59.5 60.5 61.5 62.5 63.5 64.5 65.5
1443 1515 1592 1656 1716 1776 1819 1859 1882 1924 1947 1966 2001 2005 2015 2023 2034 2011
66.5 67.5 68.5 69.5 70.5 71.5 72.5 73.5 74.5 75.5 76.5 77.5 78.5 79.5 80.5 81.5 82.5 83.5
2002 1983 1977 1961 1953 1927 1921 1907 1895 1872 1870 1870 1856 1837 1824 1799 1797 1784
84.5 85.5 86.5 87.5 88.5 89.5 90.5 91.5 92.5 93.5 94.5 95.5 96.5 97.5 98.5
1747 1743 1729 1688 1645 1620 1569 1517 1427 1337 1251 1127 1008  877  705
```

```
> table( DMn$P )
```

```
1995.5 1996.5 1997.5 1998.5 1999.5 2000.5 2001.5 2002.5 2003.5 2004.5 2005.5 2006.5
  4753   4331   4423   4487   4550   4639   4705   4780   4857   4905   5016   5124
2007.5 2008.5 2009.5 2010.5 2011.5 2012.5 2013.5 2014.5
  5230   5356   5458   5568   5669   5804   5912   5964
```

```
> table( DMn$dur )
```

```
 0.25  0.75  1.25  1.75  2.25  2.75  3.25  3.75  4.25  4.75  5.25  5.75  6.25  6.75  7.25
 2729  2697  2691  2684  2640  2635  2596  2591  2553  2547  2509  2501  2465  2456  2408
 7.75  8.25  8.75  9.25  9.75 10.25 10.75 11.25 11.75 12.25 12.75 13.25 13.75 14.25 14.75
 2403  2358  2349  2300  2280  2224  2212  2145  2149  2070  2074  2007  1989  1924  1913
15.25 15.75 16.25 16.75 17.25 17.75 18.25 18.75 19.25 19.75 20.25 20.75 21.25 21.75 22.25
 1850  1818  1739  1717  1637  1622  1527  1473  1361  1298  1185  1122  1021   984   870
22.75 23.25 23.75 24.25 24.75 25.25 25.75 26.25 26.75 27.25 27.75 28.25 28.75 29.25 29.75
  832   741   698   620   588   513   476   444   414   386   369   331   317   283   277
30.25 30.75 31.25 31.75 32.25 32.75 33.25 33.75 34.25 34.75 35.25 35.75 36.25 36.75 37.25
  242   234   201   195   161   155   127   119    87    87    60    64    38    41    22
37.75 38.25 38.75 39.25 39.75 40.25 40.75 41.25 41.75 42.25 42.75 43.25 43.75 44.25
   23     7     9     7     9     6     6     4     4     4     2     2     2     1
> save( DMn, file="../data/DMn.Rda" )
```

2.2.3 DM with drug treatment

There is only one possible exit from the DM-drug state, to Death; here we groom the tabulated dataset and save it as an R analysis dataset:

```
> DMd <- read.xport(file="../data/DMdrug.xpt")
> head(DMd)
  SEX  A    P DUR DDUR         Y DTH
1   1 30 1995 0.0  0.0 2.9284736   0
2   1 30 1995 0.5  0.0 0.6132786   0
3   1 30 1995 0.5  0.5 0.1783025   0
4   1 30 1996 0.0  0.0 5.1512663   0
5   1 30 1996 0.5  0.0 0.2813142   0
6   1 30 1996 0.5  0.5 2.0095825   0
> names(DMd) <- c("sex","A","P","dur","ddur","Y","Dth")
> DMd$sex <- factor( DMd$sex, labels=c("M","F") )
> DMd <- transform( subset( DMd, P>1994 & A<99 ),
+                   A = A + 0.5,
+                   P = P + 0.5,
+                 dur = dur + 0.25,
+                ddur = ddur + 0.25 )
> head(DMd)
  sex    A      P  dur ddur         Y Dth
1   M 30.5 1995.5 0.25 0.25 2.9284736   0
2   M 30.5 1995.5 0.75 0.25 0.6132786   0
3   M 30.5 1995.5 0.75 0.75 0.1783025   0
4   M 30.5 1996.5 0.25 0.25 5.1512663   0
5   M 30.5 1996.5 0.75 0.25 0.2813142   0
6   M 30.5 1996.5 0.75 0.75 2.0095825   0
> str(DMd)
'data.frame':        1019319 obs. of  7 variables:
 $ sex : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A   : num  30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 ...
 $ P   : num  1996 1996 1996 1996 1996 ...
 $ dur : num  0.25 0.75 0.75 0.25 0.75 0.75 0.25 0.75 0.75 0.25 ...
 $ ddur: num  0.25 0.25 0.75 0.25 0.25 0.75 0.25 0.25 0.75 0.25 ...
 $ Y   : num  2.928 0.613 0.178 5.151 0.281 ...
 $ Dth : num  0 0 0 0 0 0 1 0 0 0 ...
> summary(DMd)
 sex              A               P             dur             ddur
 M:505004   Min.   :30.5   Min.   :1996   Min.   : 0.25   Min.   : 0.25
 F:514315   1st Qu.:56.5   1st Qu.:2004   1st Qu.: 7.75   1st Qu.: 2.25
            Median :68.5   Median :2008   Median :12.25   Median : 5.75
            Mean   :67.7   Mean   :2008   Mean   :13.48   Mean   : 6.48
            3rd Qu.:78.5   3rd Qu.:2012   3rd Qu.:17.75   3rd Qu.: 9.75
            Max.   :98.5   Max.   :2014   Max.   :54.75   Max.   :19.75
       Y                 Dth
 Min.   :  0.0000   Min.   : 0.0000
```

```
 1st Qu.:  0.1961   1st Qu.: 0.0000
 Median :  0.4822   Median : 0.0000
 Mean   :  2.4916   Mean   : 0.1323
 3rd Qu.:  1.2762   3rd Qu.: 0.0000
 Max.   :231.8460   Max.   :17.0000
> table( DMd$A )
 30.5  31.5  32.5  33.5  34.5  35.5  36.5  37.5  38.5  39.5  40.5  41.5  42.5  43.5  44.5
  119   348   669  1099  1629  2214  2884  3665  4486  5355  6275  7221  8171  9202 10249
 45.5  46.5  47.5  48.5  49.5  50.5  51.5  52.5  53.5  54.5  55.5  56.5  57.5  58.5  59.5
11290 12262 13222 14162 15032 15853 16685 17516 18248 18923 19628 20268 20818 21344 21865
 60.5  61.5  62.5  63.5  64.5  65.5  66.5  67.5  68.5  69.5  70.5  71.5  72.5  73.5  74.5
22349 22881 23299 23623 23917 24131 24337 24396 24351 24271 24112 23874 23565 23392 23315
 75.5  76.5  77.5  78.5  79.5  80.5  81.5  82.5  83.5  84.5  85.5  86.5  87.5  88.5  89.5
23127 22863 22529 22038 21561 21089 20553 19950 19275 18381 17426 16409 15277 14055 12652
 90.5  91.5  92.5  93.5  94.5  95.5  96.5  97.5  98.5
11323  9868  8413  7036  5726  4652  3703  2803  2095
> table( DMd$P )
1995.5 1996.5 1997.5 1998.5 1999.5 2000.5 2001.5 2002.5 2003.5 2004.5 2005.5 2006.5
  9592  16921  21404  25680  29954  34220  38398  42487  46515  50714  54735  58623
2007.5 2008.5 2009.5 2010.5 2011.5 2012.5 2013.5 2014.5
 62227  65851  69372  72704  76098  79222  81820  82782
> table( DMd$dur )
 0.25  0.75  1.25  1.75  2.25  2.75  3.25  3.75  4.25  4.75  5.25  5.75  6.25  6.75  7.25
 2747  5376  7578  9807 11750 13733 15346 17162 18608 20201 21462 22795 23790 24900 25699
 7.75  8.25  8.75  9.25  9.75 10.25 10.75 11.25 11.75 12.25 12.75 13.25 13.75 14.25 14.75
26595 27092 27831 28050 28462 28321 28359 27953 27797 27115 26732 25877 25445 24527 23903
15.25 15.75 16.25 16.75 17.25 17.75 18.25 18.75 19.25 19.75 20.25 20.75 21.25 21.75 22.25
22775 22084 21058 20350 19308 18568 17593 16753 15734 14757 13607 12709 11739 10952 10153
22.75 23.25 23.75 24.25 24.75 25.25 25.75 26.25 26.75 27.25 27.75 28.25 28.75 29.25 29.75
 9494  8771  8168  7610  6980  6345  5939  5649  5207  4929  4542  4292  3975  3816  3449
30.25 30.75 31.25 31.75 32.25 32.75 33.25 33.75 34.25 34.75 35.25 35.75 36.25 36.75 37.25
 3225  2947  2782  2522  2411  2215  2070  1900  1784  1582  1393  1272  1173  1053   960
37.75 38.25 38.75 39.25 39.75 40.25 40.75 41.25 41.75 42.25 42.75 43.25 43.75 44.25 44.75
  842   754   665   634   559   525   445   426   355   347   275   259   213   213   169
45.25 45.75 46.25 46.75 47.25 47.75 48.25 48.75 49.25 49.75 50.25 50.75 51.25 51.75 52.25
  142   118   116   100    87    73    63    54    53    44    40    30    27    24    16
52.75 53.25 53.75 54.25 54.75
   16    10     8     5     4
> table( DMd$ddur )
 0.25  0.75  1.25  1.75  2.25  2.75  3.25  3.75  4.25  4.75  5.25  5.75  6.25  6.75  7.25
53906 55163 50338 51616 46768 47904 43164 44193 39517 40651 36157 37331 33000 34088 29900
 7.75  8.25  8.75  9.25  9.75 10.25 10.75 11.25 11.75 12.25 12.75 13.25 13.75 14.25 14.75
31055 26956 28151 24145 25321 21490 22670 18920 20104 16497 17656 14243 15411 12051 13182
15.25 15.75 16.25 16.75 17.25 17.75 18.25 18.75 19.25 19.75
10041 11119  8129  9112  6284  7268  4601  5597  2931  2689
> save( DMd, file="../data/DMd.Rda" )
```

We have now the three analysis datsets in R-format.

2.3 APC models for incidence of DM

We now have all three datasets for the necessary analysis of the incidence and mortality rates. To this end we load the `well` dataset and model both the total and the separte incidence rates by APC-models.

```
> library( Epi )
> clear()
> load( file="../data/well.Rda" )
> lls()
  name mode class      size
1 well list data.frame 2760 9
```

```
> head( well )
  sex    A      P        Y  D        Yd Dd DM DMd
1   F 30.5 1995.5 41439.44 16 22.888433  0 17  16
2   F 30.5 1996.5 42979.33 14 20.002738  0 14   8
3   F 30.5 1997.5 42323.70 21 14.463381  0 16   6
4   F 30.5 1998.5 39372.82 18 11.849418  0 12  10
5   F 30.5 1999.5 37259.98 24 11.015743  0  9   9
6   F 30.5 2000.5 36882.30 15  9.700205  0  9   9
```

First we define the knots to be used for the effects

```
> a.kn <- with( well, quantile( rep(  A,D), probs=(1:5-0.5)/5 ) )
> p.kn <- with( well, quantile( rep(P  ,D), probs=(1:3-0.5)/3 ) )
> c.kn <- with( well, quantile( rep(P-A,D), probs=(1:4-0.5)/4 ) )
```

Then we can fit all the models; separately for men and women and separately for persons on and not on dug at initiation.

```
> Mtot  <- with( subset( well, sex=="M" ),
+               apc.fit( A=A, P=P, D=DM+DMd, Y=Y/1000,
+                        ref.p=2010, parm="APC", model="ns",
+                        npar=list(A=a.kn,P=p.kn,C=c.kn) ) )
[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                    1375     9597.5
Age-drift              1374     7129.8  1  2467.70 < 2.2e-16 ***
Age-Cohort             1372     7066.5  2    63.30 1.798e-14 ***
Age-Period-Cohort      1371     6711.6  1   354.84 < 2.2e-16 ***
Age-Period             1373     6809.9 -2   -98.28 < 2.2e-16 ***
Age-drift              1374     7129.8 -1  -319.85 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> round( Mtot$Drift, 2 )
    exp(Est.) 2.5% 97.5%
APC      1.02 1.02  1.02
A-d      1.02 1.02  1.02
> Ftot  <- with( subset( well, sex=="F" ),
+               apc.fit( A=A, P=P, D=DM+DMd, Y=Y/1000,
+                        ref.p=2010, parm="APC", model="ns",
+                        npar=list(A=a.kn,P=p.kn,C=c.kn) ) )
[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                    1375    11165.3
Age-drift              1374     8748.4  1  2416.88 < 2.2e-16 ***
Age-Cohort             1372     8507.9  2   240.51 < 2.2e-16 ***
Age-Period-Cohort      1371     7851.0  1   656.83 < 2.2e-16 ***
Age-Period             1373     8253.4 -2  -402.39 < 2.2e-16 ***
Age-drift              1374     8748.4 -1  -494.95 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> round( Ftot$Drift, 2 )
    exp(Est.) 2.5% 97.5%
APC      1.03 1.03  1.03
A-d      1.02 1.02  1.02
> Mnone <- with( subset( well, sex=="M" ),
+               apc.fit( A=A, P=P, D=DM, Y=Y/1000,
+                        ref.p=2010, parm="APC", model="ns",
+                        npar=list(A=a.kn,P=p.kn,C=c.kn) ) )
```

```
[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                    1375     9483.4
Age-drift              1374     6684.6  1   2798.9 < 2.2e-16 ***
Age-Cohort             1372     6510.6  2    174.0 < 2.2e-16 ***
Age-Period-Cohort      1371     3453.7  1   3056.9 < 2.2e-16 ***
Age-Period             1373     3489.4 -2    -35.8 1.725e-08 ***
Age-drift              1374     6684.6 -1  -3195.1 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> round( Mnone$Drift, 2 )
    exp(Est.) 2.5% 97.5%
APC      0.96 0.96  0.96
A-d      0.96 0.96  0.96
> Fnone <- with( subset( well, sex=="F" ),
+               apc.fit( A=A, P=P, D=DM, Y=Y/1000,
+                        ref.p=2010, parm="APC", model="ns",
+                        npar=list(A=a.kn,P=p.kn,C=c.kn) ) )
[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                    1375     8688.2
Age-drift              1374     5792.0  1  2896.15  < 2e-16 ***
Age-Cohort             1372     5690.5  2   101.50  < 2e-16 ***
Age-Period-Cohort      1371     2854.3  1  2836.26  < 2e-16 ***
Age-Period             1373     2860.8 -2    -6.54  0.03794 *
Age-drift              1374     5792.0 -1 -2931.22  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> round( Fnone$Drift, 2 )
    exp(Est.) 2.5% 97.5%
APC      0.96 0.96  0.96
A-d      0.95 0.95  0.96
> Mdrug <- with( subset( well, sex=="M" ),
+               apc.fit( A=A, P=P, D=DMd, Y=Y/1000,
+                        ref.p=2010, parm="APC", model="ns",
+                        npar=list(A=a.kn,P=p.kn,C=c.kn) ) )
[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                    1375    16224.4
Age-drift              1374     6670.4  1   9554.1 < 2.2e-16 ***
Age-Cohort             1372     6241.6  2    428.7 < 2.2e-16 ***
Age-Period-Cohort      1371     6241.2  1      0.4  0.525242
Age-Period             1373     6656.2 -2   -414.9 < 2.2e-16 ***
Age-drift              1374     6670.4 -1    -14.2  0.000166 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> round( Mdrug$Drift, 2 )
    exp(Est.) 2.5% 97.5%
APC      1.05 1.05  1.06
A-d      1.05 1.05  1.05
> Fdrug <- with( subset( well, sex=="F" ),
+               apc.fit( A=A, P=P, D=DMd, Y=Y/1000,
+                        ref.p=2010, parm="APC", model="ns",
+                        npar=list(A=a.kn,P=p.kn,C=c.kn) ) )
```

```
[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                    1375    18794.1
Age-drift              1374     9858.7  1   8935.4 < 2.2e-16 ***
Age-Cohort             1372     9093.2  2    765.6 < 2.2e-16 ***
Age-Period-Cohort      1371     8968.7  1    124.4 < 2.2e-16 ***
Age-Period             1373     9831.7 -2   -863.0 < 2.2e-16 ***
Age-drift              1374     9858.7 -1    -27.0 2.018e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> round( Fdrug$Drift, 2 )
    exp(Est.) 2.5% 97.5%
APC      1.06 1.06  1.06
A-d      1.06 1.06  1.06
```

Although formally extremely significnat is is clear that the non-linear component of the cohort effects are not as pronounced as the non-linear effects of the period effects.

2.3.1 Parametrizations of the APC-models

Having fitted these models we can now plot the estimated effects (even if it formally makes little sense):

```
> pla <- function( mm=Mtot, ff=Ftot, r.txt = "DM incidence per 1000 PY", mf=1 ) {
+ apc.frame( a.lab = seq(30,100,20),
+             a.tic = seq(30,100,10),
+            cp.lab = seq(1900,2020,20),
+            cp.tic = seq(1900,2020,10),
+             r.lab = c(5,10,20,50,100,200)/10*mf,
+             r.tic = c(5:10/10,2:10,20)*mf,
+             r.txt = r.txt,
+            rr.txt = "",
+            rr.ref = 5*mf,
+               gap = 8 )
+ apc.lines( mm, col="blue" )
+ apc.lines( ff, col="red" ) }
> pla()


> par( mfrow=c(1,2) )
> pla( Mnone, Fnone, "DM-none incidence per 1000 PY", mf=1/5 )
> pla( Mdrug, Fdrug, "DM-drug incidence per 1000 PY" )
```

With the separately fitted models we can also show how the RR between the two types of incidence behave:

```
> names( Mnone )
[1] "Type"  "Model" "Age"   "Per"   "Coh"   "Drift" "Ref"   "Anova" "Knots"
> str( Mnone$Age )
 num [1:69, 1:4] 30.5 31.5 32.5 33.5 34.5 35.5 36.5 37.5 38.5 39.5 ...
 - attr(*, "dimnames")=List of 2
  ..$ : NULL
  ..$ : chr [1:4] "Age" "Rate" "2.5%" "97.5%"
> iratio <-
+ function( R1, R0 )
+ {
+ # Small function to calculate RR with CIs from independent rates with
+ # CIs; R1 and R0 are assumed 4-column matrices with time, rate, lower and
+ # upper confidence limits repectively. Assuming they are independent
```

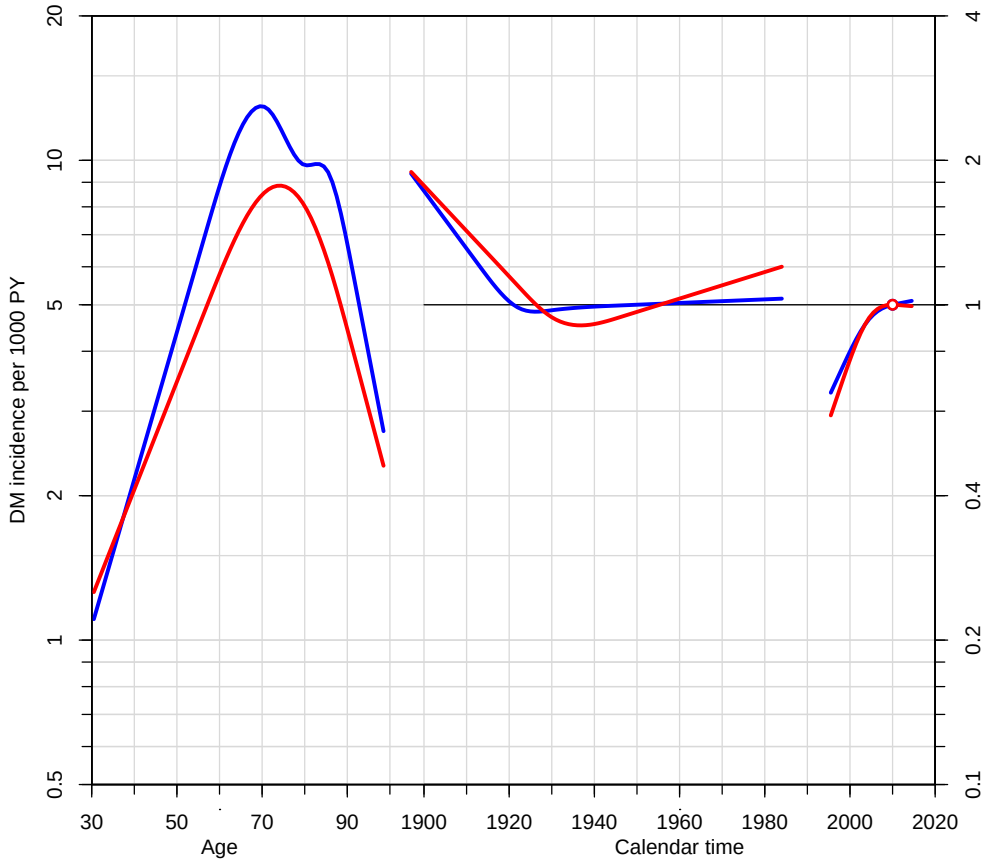

Figure 2.1: *APC-model for the total incidence of DM, parametrization constrained cohort effect to 0 slope, 0 level and a reference date as 2010-01-01. Blue:men, red:women.*

```
+ # the function returns the RR (1 vs 0) with confidence interval
+ l1 <- log(R1[,-1])
+ s1 <- (l1[,3]-l1[,2])/(1.96*2)
+ l0 <- log(R0[,-1])
+ s0 <- (l0[,3]-l0[,2])/(1.96*2)
+ lr <- cbind( R1[,1],
+              exp( cbind(l1[,1]-l0[,1],sqrt(s1^2+s0^2)) %*% ci.mat() ) )
+ }
> Mage <- iratio( Mdrug$Age, Mnone$Age )
> Mper <- iratio( Mdrug$Per, Mnone$Per )
> Mcoh <- iratio( Mdrug$Coh, Mnone$Coh )
> Fage <- iratio( Fdrug$Age, Fnone$Age )
> Fper <- iratio( Fdrug$Per, Fnone$Per )
> Fcoh <- iratio( Fdrug$Coh, Fnone$Coh )
> head( Mage )
          Estimate     2.5%    97.5%
[1,] 30.5 1.497703 1.432113 1.566296
[2,] 31.5 1.541424 1.476447 1.609260
[3,] 32.5 1.586421 1.522135 1.653422
[4,] 33.5 1.632732 1.569214 1.698820
[5,] 34.5 1.680395 1.617725 1.745492
[6,] 35.5 1.729449 1.667705 1.793478
```

We now have all the ratios and we can the plot them in an APC-frame:

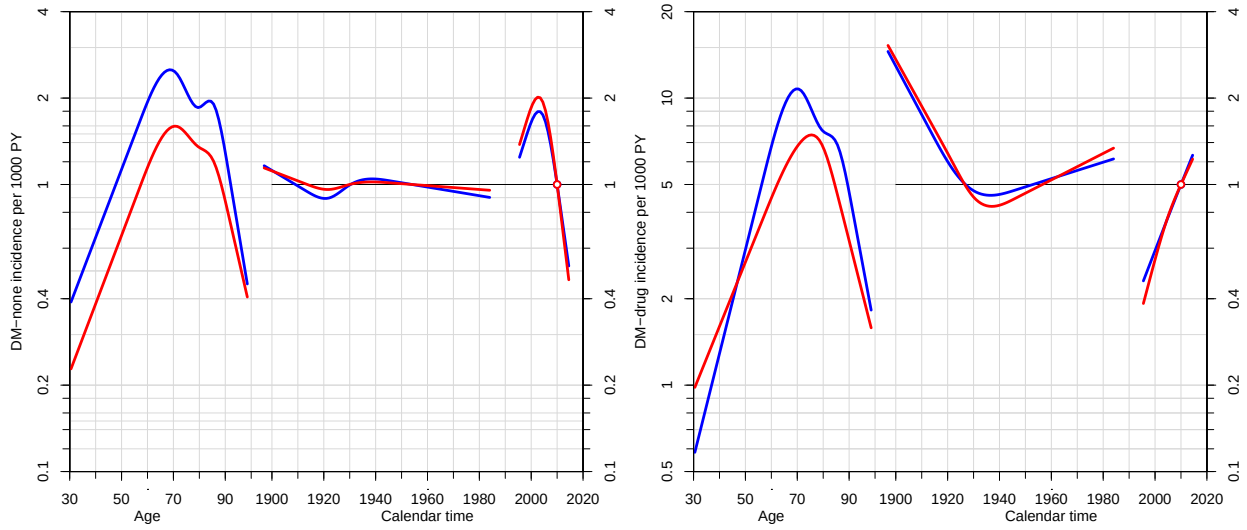

Figure 2.2: *APC-models for incidences of DM without (left) and DM with drug-treatment (right) at diagnosis. Note that the rate-scale is different for the two sets of incidences. The parametrization constrained cohort effect to 0 slope, 0 level and a reference date as 2010-01-01. Blue:men, red:women.*

```
> apc.frame( a.lab = seq(30,100,20),
+            a.tic = seq(30,100,10),
+           cp.lab = seq(1900,2020,20),
+           cp.tic = seq(1900,2020,10),
+            r.lab = c(0.1,0.2,0.5,1,2,5,10),
+            r.tic = c(1:9/10,1:10),
+            r.txt = "Rate ratio, Drug vs. none",
+           rr.txt = "",
+           rr.ref = 1,
+              gap = 8 )
>    matlines( Mage[,1], Mage[,-1], type="l",
+              lty=1, col="blue", lwd=c(3,1,1) )
>    matlines( Fage[,1], Fage[,-1], type="l",
+              lty=1, col="red", lwd=c(3,1,1) )
> pc.matlines( Mper[,1], Mper[,-1], type="l",
+              lty=1, col="blue", lwd=c(3,1,1) )
> pc.matlines( Fper[,1], Fper[,-1], type="l",
+              lty=1, col="red", lwd=c(3,1,1) )
> pc.matlines( Mcoh[,1], Mcoh[,-1], type="l",
+              lty=1, col="blue", lwd=c(3,1,1) )
> pc.matlines( Fcoh[,1], Fcoh[,-1], type="l",
+              lty=1, col="red", lwd=c(3,1,1) )
> pc.points( 2010, 1, pch=16 )
```

2.3.2 Predicted incidences and odds

It may be simpler to attach meaning to fitted rates than to components of the parameter vectors, particularly when multiplw time scales are involved.

Hence we plot the overall predicted values, but in order to extract these sufficiently detailed we re-fit the models in terms of A and P:

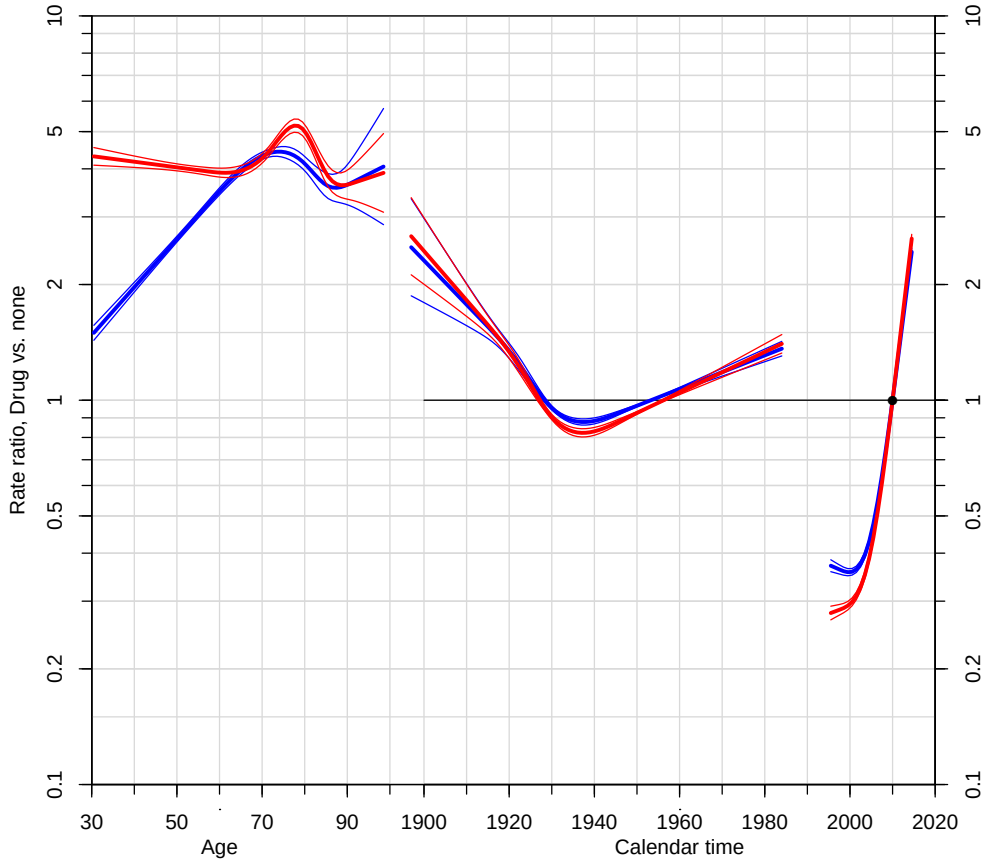

Figure 2.3: *APC-model for the RR of drug to non-drug initiated incidences of DM. The parametrization constrained cohort effect to 0 slope, 0 level and a reference date as 2010-01-01. Blue:men, red:women.*

```
> m.tot <- glm( DM+DMd ~ Ns(  A,knots=a.kn) +
+                        Ns(P  ,knots=p.kn) +
+                        Ns(P-A,knots=c.kn),
+               offset = log(Y),
+               family = poisson,
+               data = subset( well, sex=="M" ) )
> m.drug <- update( m.tot, DMd ~ . )
> m.none <- update( m.tot, DM  ~ . )
> f.tot  <- update( m.tot, data = subset( well, sex=="F" ) )
> f.drug <- update( f.tot, DMd ~ . )
> f.none <- update( f.tot, DM  ~ . )
```

Then we set up an array to hold the predicted values with CIs so we can plot them easier:

```
> a.pt <- seq(30,99,1/4)
> p.pt <- 1995:2015
> nd <- data.frame
> predinc <- NArray( list( P = p.pt,
+                          A = a.pt,
+                        sex = levels(well$sex),
+                       type = c("Total","None","Drug","D vs N"),
+                       what = c("est","lo","hi") ) )
> nd <- data.frame( A=a.pt, P=1995, Y=1000 )
```

```
> for( pp in p.pt )
+    {
+    nd$P <- pp
+ predinc[paste(pp),,"M","Total",] <- ci.pred( m.tot , newdata=nd )
+ predinc[paste(pp),,"F","Total",] <- ci.pred( f.tot , newdata=nd )
+ predinc[paste(pp),,"M","None" ,] <- ci.pred( m.none, newdata=nd )
+ predinc[paste(pp),,"F","None" ,] <- ci.pred( f.none, newdata=nd )
+ predinc[paste(pp),,"M","Drug" ,] <- ci.pred( m.drug, newdata=nd )
+ predinc[paste(pp),,"F","Drug" ,] <- ci.pred( f.drug, newdata=nd )
+    }
```

We can then calculate the predicted odds of drug initiation:

```
> ldr <- log( predinc[,,,"Drug",1] )
> sdr <- log( predinc[,,,"Drug",3] / predinc[,,,"Drug",2] )/(1.96*2)
> lno <- log( predinc[,,,"None",1] )
> sno <- log( predinc[,,,"None",3] / predinc[,,,"None",2] )/(1.96*2)
> lRR <- ldr - lno
> sRR <- sqrt(sdr^2+sno^2)
> predinc[,,,"D vs N",1] <- exp(lRR)
> predinc[,,,"D vs N",2] <- exp(lRR-1.96*sRR)
> predinc[,,,"D vs N",3] <- exp(lRR+1.96*sRR)
```

The predicted incidence rates for men at mid-year 1996, 1998,. . . ,2014 are shown in figure 2.4

```
> par( mfrow=c(1,2) )
> for( sx in c("M","F") )
+    {
+ matplot( a.pt, t(predinc[0:10*2+1,,sx,"Total",1]),
+          type="l", lwd=c(3,1), lty=1, col=gray(1:11/15), log="y",
+          xlab="Age at follow-up", ylim=c(1,15),
+          ylab="Incidence rate of DM (per 1000 PY)" )
+    }
```

If we in parallel look at the odds of starting on drugs given a diagnosis we get a picture of

```
> par( mfrow=c(1,2) )
> for( sx in c("M","F") )
+ {
+ matplot( a.pt, t(predinc[0:10*2+1,,sx,"D vs N",1]),
+          type="l", lwd=c(3,1), lty=1, col=gray(1:11/15), log="y",
+          xlab="Age at follow-up", ylim=c(0.5,20),
+          ylab=paste(sx,": Odds of starting on drugs",sep="") )
+ text( 40, 20, sx, cex=1.5, font=2 )
+ }


> par( mfrow=c(1,2) )
> for( sx in c("M","F") )
+ {
+ matplot( a.pt, cbind( t(predinc[0:5*4+1,,sx,"D vs N",1]),
+                       t(predinc[0:5*4+1,,sx,"D vs N",2]),
+                       t(predinc[0:5*4+1,,sx,"D vs N",3]) ),
+          type="l", lwd=rep(c(3,1),c(6,12)), lty=1, col=gray(1:6/7), log="y",
+          xlab="Age at follow-up", ylim=c(0.5,20),
+          ylab="Odds of starting on drugs" )
+ text( 40, 20, sx, cex=1.5, font=2 )
+ }
```

We see a clear increase by age for men, less so for women. For both sexes, the secularincrease is marked.

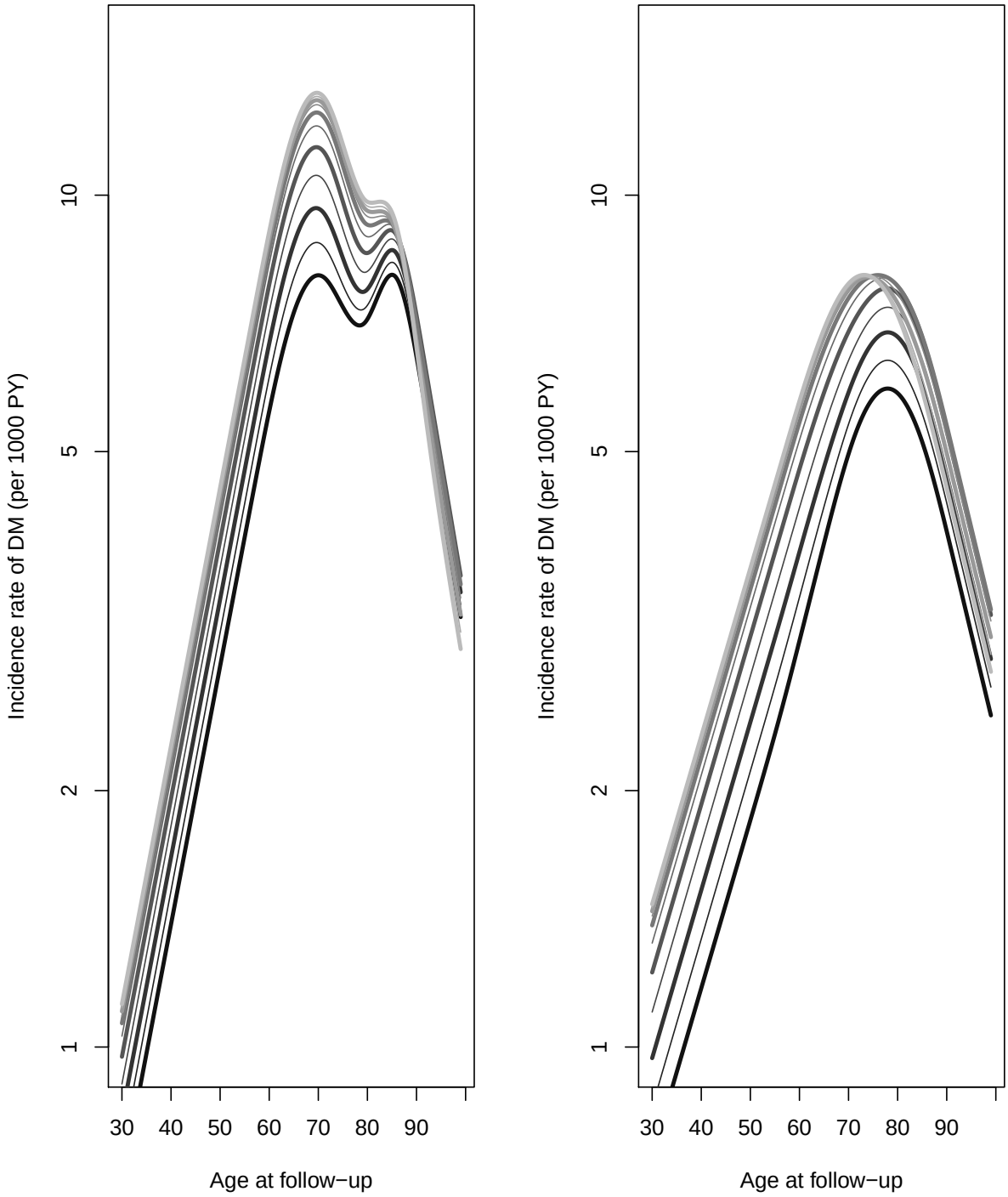

Figure 2.4: *Cross-sectional incidence rates of all diabetes based on the revised criteria, for the dates 1 January 1995, 1997, . . . 2015, the earliest plotted with the darkest colours, men in left panel women in right panel.*

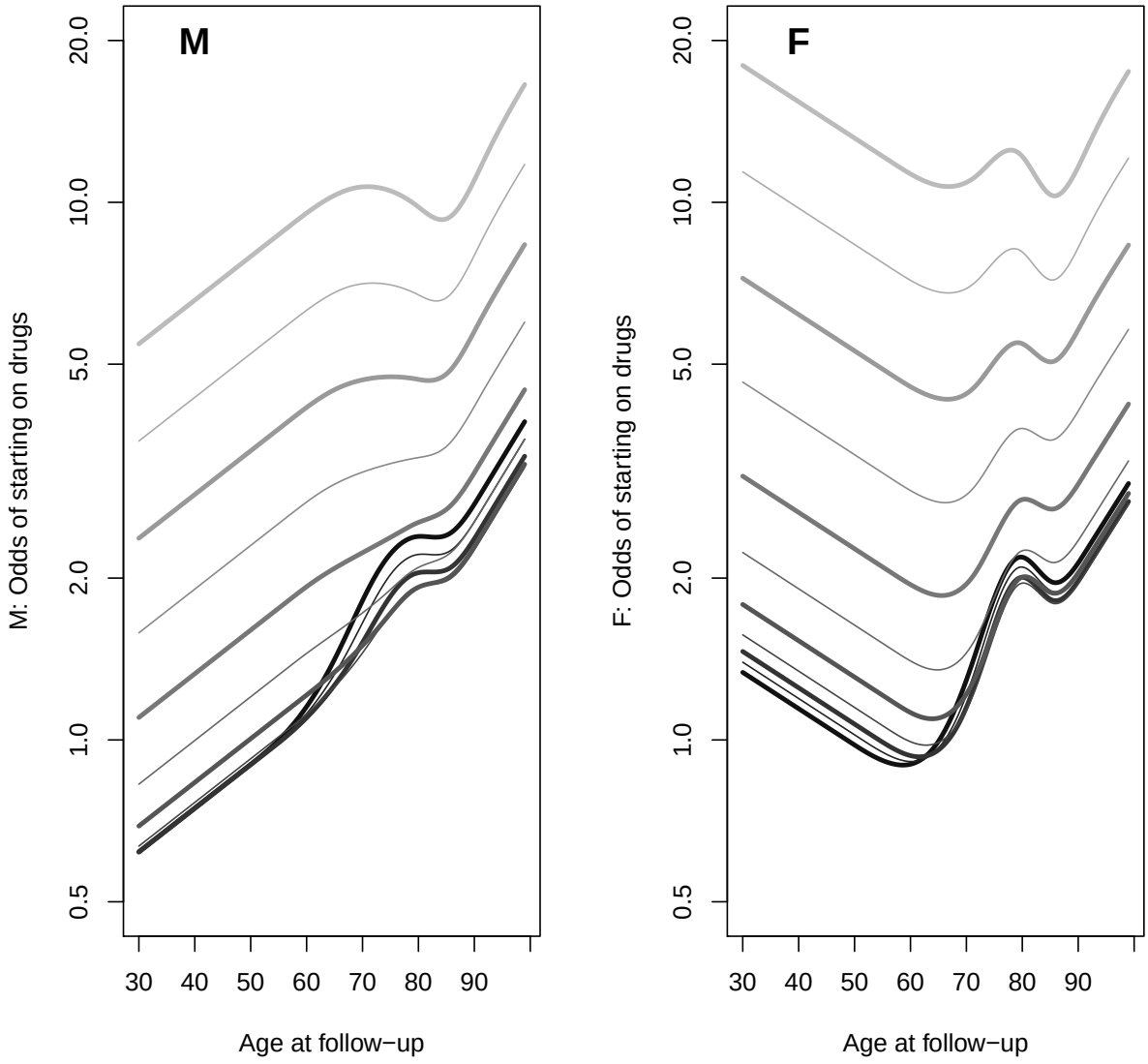

Figure 2.5: *Cross-sectional odds of starting drugs for the dates 1 January 1995, 1997, . . . 2015, the oldest plotted with the darkest colours, every other date witha thin line.*

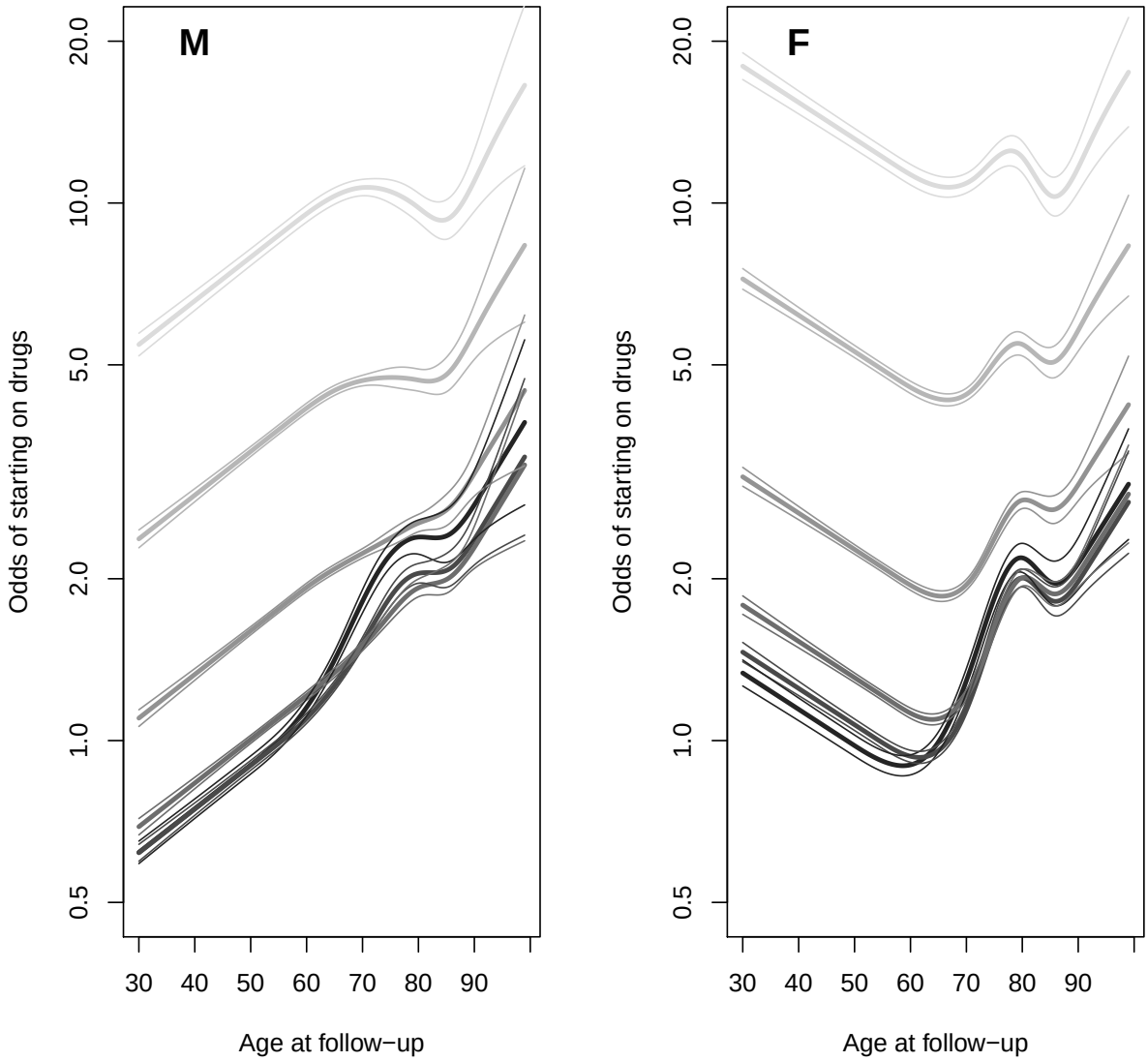

Figure 2.6: *Cross-sectional odds of starting drugs for the dates 1 January 1995, 1999, 2003, 2007, 2011, 2015, with confidence intervals (thin lines); the oldest plotted with the darkest colours, men in left panel women in right panel.*

Chapter 3

Mortality and years of life lost

3.1 APC models for mortality

We want to model the mortality among diabetes patients and among persons without diabetes in order to asses the years of life lost and how these have developed over the study period.

The basic building blocks are mortality rates much in the same format as that set up for the incidence rates, however further classified by diabetes duration so that we can assess how mortality rates among diabetes patients depend on diabetes duration and medication duration.

However, calculations of life-years lost requires knowledge of mortality rates till the end of life for every one; in practice age 100 or close. And by that token for person diagnosed in age 60, say, how diabetes duration up to 40 years influence mortality. Which is not available in data, so in the initial calculations we shall ignore duration of diabetes.

```
> library( Epi )
> sessionInfo()
R version 3.2.3 (2015-12-10)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows Server 2008 R2 x64 (build 7601) Service Pack 1

locale:
[1] LC_COLLATE=Danish_Denmark.1252  LC_CTYPE=Danish_Denmark.1252
[3] LC_MONETARY=Danish_Denmark.1252 LC_NUMERIC=C
[5] LC_TIME=Danish_Denmark.1252

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

other attached packages:
[1] Epi_2.0

loaded via a namespace (and not attached):
 [1] cmprsk_2.2-7    compiler_3.2.3  MASS_7.3-45     plyr_1.8.3      tools_3.2.3
 [6] parallel_3.2.3  survival_2.38-3 etm_0.6-2       Rcpp_0.12.3     splines_3.2.3
[11] grid_3.2.3      lattice_0.20-33
> clear()
> load( file="../data/well.Rda" )
> load( file="../data/DMn.Rda" )
> load( file="../data/DMd.Rda" )
> lls()
  name mode class      size
1 DMd  list data.frame 1019319 7
```

```
2 DMn  list data.frame 101531 7
3 well list data.frame 2760 9
> head( well )
  sex    A      P         Y  D        Yd Dd DM DMd
1   F 30.5 1995.5 41439.44 16 22.888433  0 17  16
2   F 30.5 1996.5 42979.33 14 20.002738  0 14   8
3   F 30.5 1997.5 42323.70 21 14.463381  0 16   6
4   F 30.5 1998.5 39372.82 18 11.849418  0 12  10
5   F 30.5 1999.5 37259.98 24 11.015743  0  9   9
6   F 30.5 2000.5 36882.30 15  9.700205  0  9   9
> head( DMn )
  sex    A      P  dur         Y DMd Dth
3   M 30.5 1995.5 0.25 4.9380561   7   0
4   M 30.5 1995.5 0.75 0.6252567   2   0
5   M 30.5 1996.5 0.25 2.3713210   1   0
6   M 30.5 1996.5 0.75 0.6013005   0   0
7   M 30.5 1997.5 0.25 2.8052704   7   0
8   M 30.5 1997.5 0.75 0.2474333   0   0
> head( DMd )
  sex    A      P  dur ddur         Y Dth
1   M 30.5 1995.5 0.25 0.25 2.9284736   0
2   M 30.5 1995.5 0.75 0.25 0.6132786   0
3   M 30.5 1995.5 0.75 0.75 0.1783025   0
4   M 30.5 1996.5 0.25 0.25 5.1512663   0
5   M 30.5 1996.5 0.75 0.25 0.2813142   0
6   M 30.5 1996.5 0.75 0.75 2.0095825   0
```

Since we will model mortality jointly for the the two types of diabetes patients we stack the two datasets, exclude the duration variables and rename the death counter to the same as in the well dataset:

```
> DMall <- rbind( DMn[,c("sex","A","P","Y","Dth")],
+                 DMd[,c("sex","A","P","Y","Dth")] )
> names( DMall )[5] <- "D"
> str( DMall )
'data.frame':        1120850 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 30.5 ...
 $ P  : num  1996 1996 1996 1996 1998 ...
 $ Y  : num  4.938 0.625 2.371 0.601 2.805 ...
 $ D  : num  0 0 0 0 0 0 0 0 1 0 ...
> head( DMall )
  sex    A      P         Y D
3   M 30.5 1995.5 4.9380561 0
4   M 30.5 1995.5 0.6252567 0
5   M 30.5 1996.5 2.3713210 0
6   M 30.5 1996.5 0.6013005 0
7   M 30.5 1997.5 2.8052704 0
8   M 30.5 1997.5 0.2474333 0
```

We fit models for the mortality among the non-diabetic part of the population and for the diabetic part (ignoring drug status) separately, but we first need knots for the three effects:

```
> ( a.kn <- seq(35,90,,6) )
[1] 35 46 57 68 79 90
> ( p.kn <- with( well, quantile( rep(P  ,Dd), probs=(1:3-0.5)/3 ) ) )
16.66667%       50% 83.33333%
   1999.5    2006.5    2012.5
> ( c.kn <- with( well, quantile( rep(P-A,Dd), probs=(1:5-0.5)/5 ) ) )
 10%  30%  50%  70%  90%
1914 1921 1928 1935 1946
```

With datasets and knots for spline effects in place we can fit the models:

```
> mw <- glm( D ~ Ns(  A,knots=a.kn) +
+                Ns(P  ,knots=p.kn) +
+                Ns(P-A,knots=c.kn),
+            offset = log(Y),
+            family = poisson,
+              data = subset( well, sex=="M" ) )
> fw <- update( mw,
+              data = subset( well, sex=="F" ) )

> md <- update( mw,
+            data = subset( DMall, sex=="M" ) )
> system.time(
+ fd <- update( mw,
+            data = subset( DMall, sex=="F" ) ) )
```

With these models in place we can now extract the estimated mortality rates and put the in an array (not that we are not calculating confidence limits in this instance). It is not really the mortality rates we compute here, is the *cumulative* mortality rates over intervals of length `intl`:

```
> intl <- 1/10 # Length of the intervals we use for integration
> a.pt <- seq(30,100,intl)
> p.pt <- 1995:2015
> Mort <- NArray( list( A = a.pt,
+                       P = p.pt,
+                     sex = levels(well$sex),
+                      tp = c("Well","DM") ) )
> nd <- data.frame( A=a.pt+intl/2, P=1995, Y=intl )
> for( pp in p.pt )
+     {
+     nd$P <- pp
+ Mort[,paste(pp),"M","Well"] <- ci.pred( mw, newdata=nd )[,1]
+ Mort[,paste(pp),"F","Well"] <- ci.pred( fw, newdata=nd )[,1]
+ Mort[,paste(pp),"M","DM"  ] <- ci.pred( md, newdata=nd )[,1]
+ Mort[,paste(pp),"F","DM"  ] <- ci.pred( fd, newdata=nd )[,1]
+     }
```

We can the compute the expected survival (conditional on survival till age 30) — note that since we used `Y` equal to the interval length, we just need the cumulative sum here.

```
> Surv <- exp( -apply( Mort, 2:4, cumsum ) )
> str( Surv )
 num [1:701, 1:21, 1:2, 1:2] 1 1 1 1 1 ...
 - attr(*, "dimnames")=List of 4
  ..$ A  : chr [1:701] "30" "30.1" "30.2" "30.3" ...
  ..$ P  : chr [1:21] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "F" "M"
  ..$ tp : chr [1:2] "Well" "DM"
```

The years of life lost to diabetes at a given age of diagnosis is the difference between the areas under the (conditional) survival curves for non-DM and DM persons. Note that this in a sense may be a bit exaggerating the life loss, as this is really a comparison with a non-DM population that is immune to DM, because the non-DM survival function does not take occurrence of DM and the ensuing increased mortality into account.

In order to compute the expected residual lifetime from any point in age we need the successive integrals from each age and till the end, but these are basically just the cumulative sums from the end, so for convenience we need a function that does this, `musmuc`. In order to get the integral right we of course multiply with `intl`, and since we are interested in the the integral of the `conditional` survival functions from different ages, we also must divide by the survival probabilities — conveniently available in `Surv` too:

```
> musmuc <- function(x) rev(cumsum(rev(x)))
> Erl <- apply( Surv, 2:4, musmuc ) * intl / Surv
```

Now we have the expected residual life times for DM and non-DM persons we can just subtract and we have the expected life time lost to diabetes by sex and calendar time:

```
> round( ftable( Erl[c(1,101,201),,,], row.vars=2 ), 1 )
     A     30                40                50
     sex    F         M       F         M       F         M
     tp  Well   DM Well   DM Well   DM Well   DM Well   DM Well   DM
P
1995     49.3 36.7 44.7 30.8 39.6 27.8 35.3 23.5 30.4 20.2 26.3 17.1
1996     49.5 37.2 45.0 31.3 39.8 28.2 35.6 24.0 30.6 20.6 26.6 17.5
1997     49.7 37.8 45.3 31.9 40.0 28.7 35.8 24.5 30.7 21.1 26.8 17.9
1998     49.9 38.3 45.5 32.4 40.2 29.2 36.1 25.0 30.9 21.5 27.1 18.3
1999     50.1 38.8 45.8 33.0 40.4 29.7 36.3 25.5 31.1 21.9 27.3 18.7
2000     50.3 39.4 46.1 33.6 40.6 30.2 36.6 26.0 31.3 22.3 27.6 19.1
2001     50.5 39.9 46.4 34.1 40.8 30.7 36.9 26.5 31.4 22.8 27.8 19.5
2002     50.7 40.4 46.6 34.7 41.0 31.3 37.1 27.0 31.6 23.2 28.1 19.9
2003     50.9 41.0 46.9 35.3 41.2 31.8 37.4 27.5 31.8 23.7 28.3 20.3
2004     51.1 41.5 47.2 35.8 41.4 32.3 37.7 28.0 32.0 24.1 28.6 20.8
2005     51.3 42.1 47.5 36.4 41.6 32.8 37.9 28.5 32.2 24.5 28.8 21.2
2006     51.6 42.6 47.7 37.0 41.8 33.3 38.2 29.0 32.4 25.0 29.1 21.6
2007     51.8 43.1 48.0 37.5 42.0 33.8 38.5 29.5 32.6 25.4 29.3 22.0
2008     52.0 43.6 48.3 38.1 42.3 34.3 38.7 30.0 32.8 25.9 29.6 22.5
2009     52.3 44.2 48.6 38.7 42.5 34.8 39.0 30.5 33.1 26.3 29.8 22.9
2010     52.5 44.7 48.8 39.2 42.7 35.3 39.3 31.0 33.3 26.8 30.1 23.3
2011     52.8 45.2 49.1 39.8 43.0 35.8 39.5 31.5 33.5 27.2 30.3 23.8
2012     53.0 45.7 49.4 40.4 43.2 36.3 39.8 32.0 33.7 27.7 30.5 24.2
2013     53.3 46.3 49.6 40.9 43.5 36.8 40.0 32.5 34.0 28.1 30.8 24.7
2014     53.5 46.8 49.9 41.5 43.7 37.3 40.3 33.0 34.2 28.6 31.0 25.1
2015     53.8 47.3 50.1 42.1 44.0 37.8 40.5 33.6 34.4 29.0 31.3 25.6
> Yll <- Erl[,,"Well"]-Erl[,,"DM"]
> round( ftable( Yll[c(1,101,201,301,401),,2:1], row.vars=2 ), 1 )
     A     30        40        50        60        70
     sex    M    F    M    F    M    F    M    F    M    F
P
1995     14.0 12.6 11.8 11.9  9.2 10.2  6.6  7.7  4.1  5.1
1996     13.7 12.2 11.5 11.6  9.1  9.9  6.5  7.5  4.1  5.0
1997     13.4 11.9 11.3 11.3  8.9  9.7  6.4  7.3  4.0  4.9
1998     13.1 11.6 11.1 11.0  8.8  9.4  6.3  7.2  4.0  4.8
1999     12.8 11.2 10.9 10.6  8.6  9.2  6.2  7.0  3.9  4.7
2000     12.5 10.9 10.6 10.3  8.5  8.9  6.2  6.8  3.9  4.6
2001     12.2 10.6 10.4 10.0  8.3  8.7  6.1  6.6  3.8  4.4
2002     11.9 10.2 10.2  9.7  8.1  8.4  6.0  6.5  3.8  4.3
2003     11.6  9.9  9.9  9.4  8.0  8.2  5.9  6.3  3.7  4.2
2004     11.4  9.6  9.7  9.1  7.8  7.9  5.8  6.1  3.7  4.1
2005     11.1  9.3  9.5  8.8  7.6  7.7  5.7  5.9  3.6  4.0
2006     10.8  9.0  9.2  8.5  7.5  7.4  5.6  5.7  3.6  3.9
2007     10.5  8.7  9.0  8.3  7.3  7.2  5.4  5.6  3.5  3.8
2008     10.2  8.4  8.7  8.0  7.1  7.0  5.3  5.4  3.5  3.7
2009      9.9  8.1  8.5  7.7  6.9  6.7  5.2  5.2  3.4  3.6
2010      9.6  7.8  8.3  7.5  6.7  6.5  5.1  5.1  3.3  3.4
2011      9.3  7.6  8.0  7.2  6.5  6.3  4.9  4.9  3.3  3.3
2012      9.0  7.3  7.7  6.9  6.3  6.1  4.8  4.7  3.2  3.2
2013      8.7  7.0  7.5  6.7  6.1  5.9  4.6  4.6  3.1  3.1
2014      8.4  6.8  7.2  6.4  5.9  5.6  4.5  4.4  3.0  3.0
2015      8.1  6.5  7.0  6.2  5.7  5.4  4.3  4.2  2.9  2.9
```

Thus we see that the YLL for man has been decreasing more than for women

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> matplot( a.pt, Yll[,,"M"],
+          type="l", lty=1, col="blue", lwd=2:1, yaxs="i",
+          ylim=c(0,15), xlab="Age at DM diagnosis",
```

```
+          ylab="Years of life lost to DM" )
> abline( v=50, h=1:15, col=gray(0.8) )
> matplot( a.pt, Yll[,,"F"],
+          type="l", lty=1, col="red", lwd=2:1, yaxs="i",
+          ylim=c(0,15), xlab="Age at DM diagnosis",
+          ylab="Years of life lost to DM" )
> abline( v=50, h=1:15, col=gray(0.8) )
```

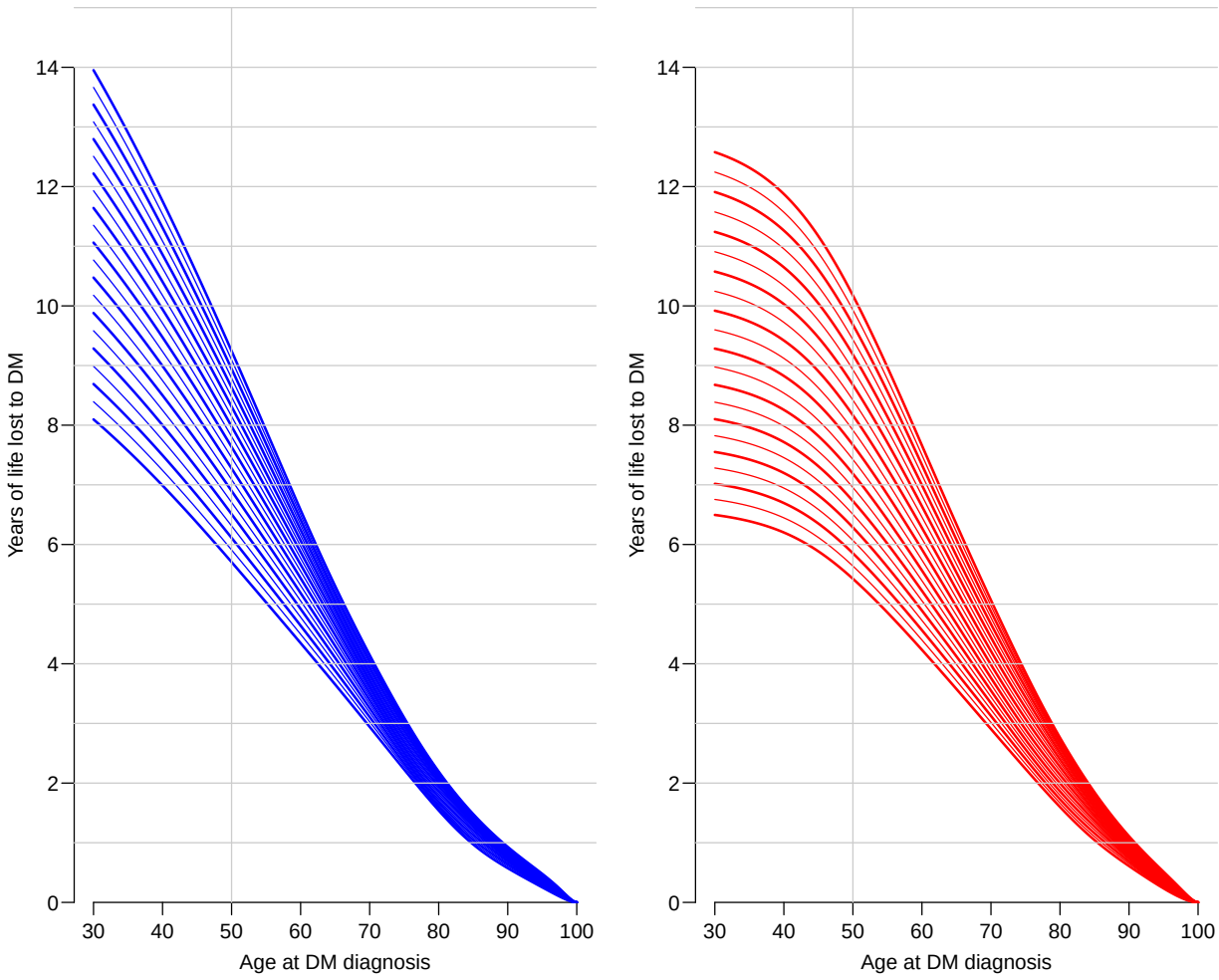

Figure 3.1: *Years of life lost to DM, using cross-section mortality rates for the dates 1 January 1995 through 2015, and assuming immunity to DM in the non-DM population. Blue is men, red is women.*

```
> par( mfrow=c(1,1), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> matplot( a.pt, cbind(Yll[,,"M"],NA,Yll[,,"F"]),
+          type="l", lty=1, lwd=2:1, yaxs="i",
+          col=rep(c("blue","red"),each=dim(Yll)[2]+1),
+          ylim=c(0,15), xlab="Age at DM diagnosis",
+          ylab="Years of life lost to DM" )
> abline( v=50, h=1:15, col=gray(0.8) )
```

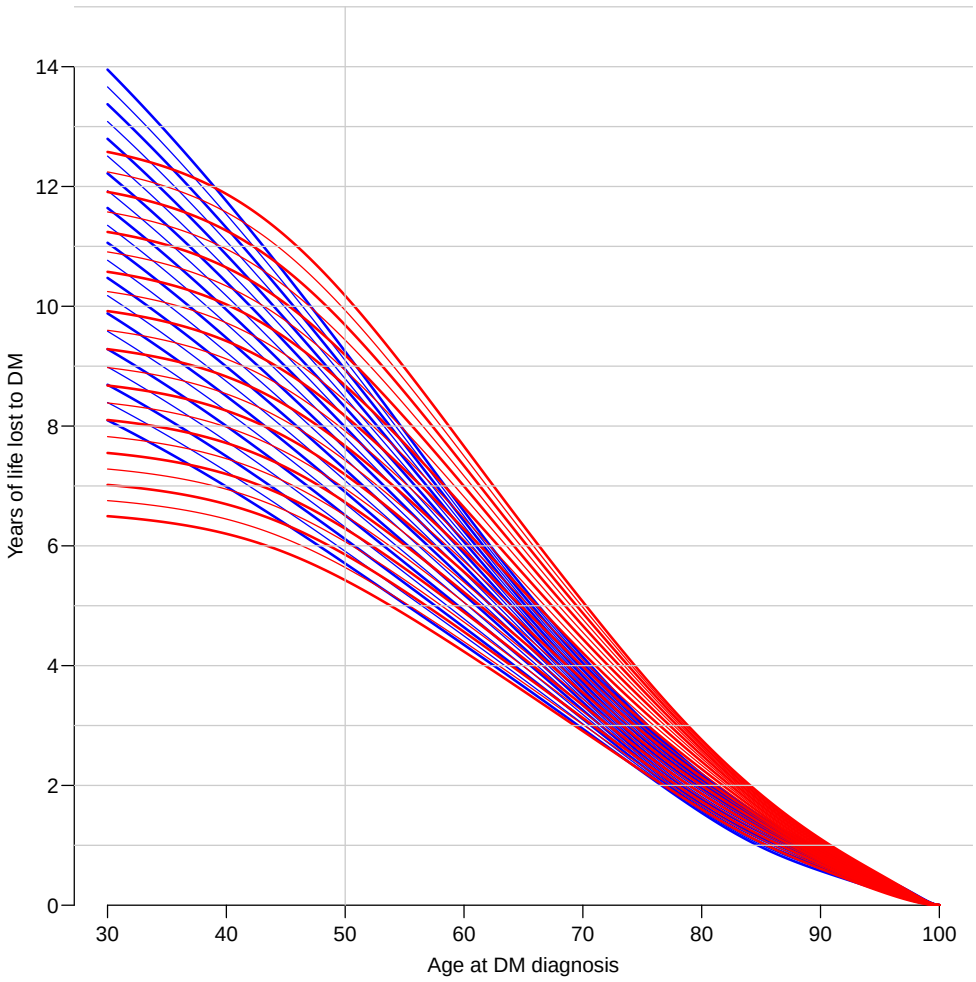

Figure 3.2: *Years of life lost to DM, using cross-section mortality rates for the dates 1 January 1995 through 2015, and assuming immunity to DM in the non-DM population. Blue is men, red is women.*

Chapter 4

Prevalence

We have two sets of prevalences to describe: Total DM and drug-treated DM.

4.1 Constructing prevalence data

The following SAS-program outouts a file of prevalent cases by sex, age and date - the dates being 1 January 1995–2015.

```
NOTE: Copyright (c) 2002-2012 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.4 (TS1M3)
      Licensed to FORSKNING 1, Site 50800722.
NOTE: This session is executing on the X64_ES08R2  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_ES08R2 WIN 6.1.7601 Service Pack 1 Server

NOTE: SAS initialization used:
      real time           16.44 seconds
      cpu time            3.27 seconds

1    libname her "../data" ;
NOTE: Libref HER was successfully assigned as follows:
      Engine:        V9
      Physical Name: E:\workdata\705093\BXC\data
2
3    * Ouput one record for every peron passing a 1st January alive with DM
4    * Note that DM is anyone with DM - both with and witout drugs ;
5    data prev ;
6      set her.personbase_bxc ( rename = ( DOBirth = doBth
7                                          DoDeath = doDth
8                                   DoEndStudykorr = doEnd
9                                          DoRMPS = doMed
10                                  first_DM_korr = doDM ) ) ;
11     do P = 1995 to 2015 by 1 ;
12        A = floor( (mdy(1,1,P)-doBth)/365.25 );
13       DM = ( doDM <mdy(1,1,P) and (doDth>mdy(12,31,P-1) or doDth=.) ) ;
14      DMd = ( doMed<mdy(1,1,P) and (doDth>mdy(12,31,P-1) or doDth=.) ) ;
```

```
15       if DM then output ;
16     end ;
17   run ;

NOTE: There were 420488 observations read from the data set HER.PERSONBASE_BXC.
NOTE: The data set WORK.PREV has 3265796 observations and 11 variables.
NOTE: DATA statement used (Total process time):
      real time           28.40 seconds
      cpu time            15.10 seconds


18
19   * Summarise the number of DM and DMd by sex, age and date ;
20   proc summary  data = prev  nway ;
21     class sex A P ;
22     var DM DMd ;
23     output out = her.prev ( keep = sex A P DM DMd )
24            sum = ;
25   run ;

NOTE: There were 3265796 observations read from the data set WORK.PREV.
NOTE: The data set HER.PREV has 3124 observations and 5 variables.
NOTE: PROCEDURE SUMMARY used (Total process time):
      real time           2.09 seconds
      cpu time            5.92 seconds


26
27   libname xPrv xport "..\data\Prev.xpt"   ;
NOTE: Libref XPRV was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: E:\workdata\705093\BXC\data\Prev.xpt
28   proc copy in=her out=xPrv memtype=data ; select prev ; run ;

NOTE: Copying HER.PREV to XPRV.PREV (memtype=DATA).
NOTE: There were 3124 observations read from the data set HER.PREV.
NOTE: The data set XPRV.PREV has 3124 observations and 5 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           0.10 seconds
      cpu time            0.09 seconds
```

4.2 Analysis of prevalences

If we want to describe the prevalence of non-drug treated DM we must subtract the two numbers.

References

[1] B Carstensen. Age-Period-Cohort models for the Lexis diagram. *Statistics in Medicine*, 26(15):3018–3045, July 2007.