A Danish Diabetes Register

SDC
January 2017
http://bendixcarstensen.com/DMreg)
Draft version 7
Confidential!

Compiled Tuesday 24th January, 2017, 15:44
from: /home/bendix/sdc/DMreg/NewReg/NRx.tex

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
<bcar0029@regionh.dk> <b@bxc.dk>
http://BendixCarstensen.com

Contents

1 Background **1**
- 1.1 NDR criteria . . . 1
- 1.2 RUKS . . . 1
- 1.3 A new register . . . 2
  - 1.3.1 Type of diabetes . . . 3
  - 1.3.2 GDM . . . 3
  - 1.3.3 PCOS . . . 3
  - 1.3.4 Summary . . . 4
  - 1.3.5 Differences to RUKS . . . 4

2 Data acquisition **5**
- 2.1 Diabetes patients . . . 5
- 2.2 Population data . . . 5
- 2.3 Follow-up . . . 5
  - 2.3.1 Analysis dataset . . . 5
- 2.4 Prevalences . . . 6
- 2.5 Datasets of follow-up and prevalence . . . 6
  - 2.5.1 Rate data . . . 7
  - 2.5.2 Prevalence data . . . 8
  - 2.5.3 Crude tables for teaching . . . 9

3 Prevalence and ocurrence rates **10**
- 3.1 Prerequisites . . . 10
- 3.2 Diabetes prevalences . . . 11
  - 3.2.1 Trends in prevalence . . . 13
  - 3.2.2 Prevalence age-period interaction . . . 16
- 3.3 DM incidence . . . 19
  - 3.3.1 Age-Period-Cohort modelling . . . 19
  - 3.3.2 Time-trends in rates . . . 20
  - 3.3.3 Summary of the APC modelling . . . 21
  - 3.3.4 Prediction by age for different cohorts . . . 21
  - 3.3.5 Prediction by age across periods . . . 23
  - 3.3.6 Saving the fitted models . . . 24
- 3.4 Mortality . . . 28
  - 3.4.1 Mortality in non-diabetics . . . 28
  - 3.4.2 Mortality among DM patients . . . 30
- 3.5 Summary of APC models for incidence and mortality . . . 35

3.5.1 Time-trends in mortality rates . . . 35

4 Prediction of rates **38**
4.1 Analysis date . . . 38
4.1.1 Datasets for rate modeling . . . 41
4.2 Models for incidence and mortality rates . . . 41
4.3 Incidence rates . . . 43
4.3.1 Incidence rate predictions . . . 44
4.3.2 A damping extrapolation . . . 44
4.4 Mortality rates . . . 47
4.4.1 Diabetes patients . . . 47
4.4.2 Persons without diabetes . . . 49
4.5 Average trends . . . 49
4.6 Trends in estimated rates . . . 50
4.7 Extrapolation of rates . . . 51

5 Components of prevalence **54**
5.1 Transition probabilities . . . 55
5.2 Prediction of the observed prevalences . . . 57
5.2.1 Checking the prediction . . . 59
5.3 How much is attributable to what? . . . 63
5.4 Number of diabetes patients in Denmark . . . 68
5.5 Timetrend in the components . . . 72

6 Predicting prevalence of diabetes **76**
6.1 Predicted rates . . . 76
6.2 Transition probabilities . . . 77
6.3 Prediction of the observed prevalences . . . 79
6.4 The actual numbers of diabetes patients in Denmark . . . 81
6.5 Population forecast from DST . . . 82
6.6 Timetrend in prevalent number of DM patients . . . 85
References . . . 94

7 SAS programs **95**
7.1 Rationale and overview . . . 95
7.2 Program execution . . . 96
7.3 Program documentation . . . 96
7.3.1 `optslibs.sas` . . . 97
7.3.2 `xgdm.sas` . . . 97
7.4 `00-base` . . . 98
7.4.1 `00-base.lst` . . . 98
7.5 `01-npr` . . . 98
7.5.1 `01-npr.lst` . . . 105
7.6 `02-dvdd` . . . 114
7.6.1 `02-dvdd.lst` . . . 118
7.7 `03-ndr` . . . 123
7.7.1 `03-ndr.lst` . . . 125
7.8 `04-rmps` . . . 126

7.8.1 04-rmps.lst . . . 131
7.9 05-diab . . . 138
7.9.1 05-diab.lst . . . 141
7.10 06-define . . . 143
7.10.1 06-define.lst . . . 147
7.11 06a-comp . . . 161
7.11.1 06a-comp.lst . . . 162
7.12 Tabulation of follow-up and prevalence . . . 165
7.13 08-mkFU . . . 165
7.13.1 08-mkFU.lst . . . 169
7.14 09-mkPr . . . 171
7.14.1 09-mkPr.lst . . . 173

Chapter 1

Background

The maintenance of the National Diabetes Register (NDR) has been discontinued by the Health Data Authority (Sundhedsdatastyrelsen). It is being replaced by the Register of Selected Chronic Diseases (RUKS) which however does not encompass recisely the same persons.

1.1 NDR criteria

The "old" NDR, established 2006, covering the period from 1995 (in terms of incidence) was based on the following criteria [2]:

`lpr:` recording of diabetes as diagnosis in the NPR

`fodt:` use of the service "foot-therapy for diabetes patients" in the National Health Services Register (NHSR).

`bl5i1:` the date of the 5^{th} blood glucose measurement within a period of one year in the NHSR.

`bl2i5:` two measurements of blood glucose per year in 5 consecutive years. The date is defined as the 2^{nd} blood glucose measurement within the 5^{th} period of one year.

`oad:` date of 2^{nd} purchase of OAD as recorded in the Register of Medicines Products Statistics (RMPS).

`ins:` date of 2^{nd} purchase of insulin as recorded in the RMPS.

The inclusion date was the earliest of the dates where any of these 6 criteria were met.

It has been pointed out the the two blood-glucose (purely *procedural*) criteria included many persons that were unlikely to be diabetic patients, notably women only being *tested* for gestational diabetes (GDM) [3].

1.2 RUKS

The alleged replacement of the NDR is the Register of Selected Chronic Diseases (Register over Udvalgte Kroniske Sygdomme, RUKS). Among the 8 diseases selected for the register, are T1 diabetes and T2 diabetes.

From the register description, the RUKS seems to be a register of repeated prevalences, in the sense that the register has a separate version as of each 1 January (?).

- Type 2 DM:
  - Persons recorded with ICD10 code E11 in NPR, as the latest diagnosis.
    Persons are not included on the basis of a single NPR contact with code E11, at least one more contact (E10 or E11) or purchase of OAD or insulin is required.
  - Persons who have purchased OADs (A10B from the RMPS), and at least two purchases of either A10A (insulins) or A10B (other antidiabetic drugs).
  - Women who have a diagnosis of PCOS or have only purchased metformin (and no other OADs or insulin) and have purchased either clomifen (G03GB02) or estrogen (G03HB) are *excluded*.
  - Persons who have had no diabetes recordings in NPR or RMPS during the last 10 years are *excluded*.
  - The term "latest" for the NPR criterion and the exclusion referring to "last 10 years" seems to indicate that the register is defined relative to a particular creation date for the register, although this is not explicitly stated.
- Type 1 diabetes:
  - Persons recorded with ICD10 code E10 in NPR, as the latest diagnosis.
    Persons are not included on the basis of NPR contacts with code E11, at least one purchase of insulins is required
  - Persons who have purchased insulins (A10A from the RMPS), and at least two purchases of either A10A (insulins) or A10B (OADs).
  - Persons already classified as T2 above are *excluded*.
  - Women with a diagnosis of GDM (ICD10 024.4) and only have purchased anti diabetic medication in a window from 280 days before the first till 280 days after the last recording of GDM are excluded.

1.3 A new register

The following is an attempt to reconstruct the NDR, using (almost) the same criteria as in the original NDR, with an additional effort to define persons as either T1 or T2.

The basic content of the register is one record per person with sex, type of diabetes and dates of birth, DM and death. Additionally, the register will have the dates for meeting each of the defining criteria (name of the date in the register):

`doOAD` Date of first recorded purchase of OAD (A10B).

`doIns` Date of first recorded purchase of insulin (A10A).

`doNPR` Date of the first recorded contact date with a diagnosis of diabetes in the NPR.

`doDVD` Recorded date of diagnosis in the Danish adult diabetes register (DADD). If none recorde the date of reporting is used.

`doDiab` Recorded date of eye-screening in the Danish eye-screening database for diabetes patients.

1.3.1 Type of diabetes

The classification of patients as T1 or T2 based on register date only is not accurate, and the approach chosen here is therefore to identify persons with T1D with reasonably high specificity, and classify the rest as T2D. Thus T2D will be equivalent to "cannot be classified as T1D with reasonable certainty", and hence the classification should be used cautiously; the persons recorded as T1D are very likely to be T1D, but there is some under-reporting. Consequently, some T1D patients that are erroneously classified as T2D, but the precise size of this problem is unknown.

The practical implementation of the type classification is:

- use the DADD recordings of diabetes type (T1,T2,Other,Unkn) and classify persons as T1 resp. T2 if more than half of the recordings are T1 resp. T2. The rest are left unclassified.

- use the NPR to classify persons as T1 resp. T2 if if more than half of the recordings are T1 (E10) resp. T2 (E11). The other codes (E12-E13) are ignored, and thus some persons are left unclassified.

- The classification from DADD is always used. If a person is not classified or does not appear in DADD, the classification from NPR us used.

- Finally, persons that have purchased OAD before age 15 or insulin before age 30 are classified as T1 (except if classified as T2 in DADD); all other are classified as T2.

1.3.2 GDM

If a person is recorded with a diagnosis of GDM in the NPR, the person cannot enter the register on any criterion during the next 365 days. To account for registration delays the window starts 30 days prior to the date of GDM.

GDM diagnoses that are less than 200 days from the previous one are disregarded; so from the earliest GDM for a given person, no GDM diagnosis in the next 200 days is counted, from the next after this, another window of 200 days is used etc.

1.3.3 PCOS

If a person is recorded with PCOS in the NPR, this person cannot be included on the basis of metformin purchase in the period from the earliest PCOS diagnosis until the person's 45^{th} birthday.

Moreover, if a person's only drug dispensations are metformin between age 20 and 45, the person is considered a possible PCOS cases and these dispensations are not counted. Dispensations after age 45 for women with a PCOS diagnosis are considered as diabetes medication and lead to in inclusion in the register at the first date of purchase after the 45th birthday.

1.3.4 Summary

The register we attempt to build is a register where persons are captured by a set of criteria and then kept in the register with this inclusion date. In principle we might use the health registers to define en exit date as well (for example 10 years), but we have chosen to let prospective useres define this type of dates themselves.

1.3.5 Differences to RUKS

The proposed approach differs from RUKS in the following points:

- Persons may be included in our register even if only one purchase of OAD/Insulin is recorded.
- Only NPR diagnosis of PCOS is used, and persons deemed to suffer from PCOS can actually be included at a time after age 40 based in metformin purchase alone.

  It is not entirely clear whether persons who meet the criteria for PCOS before age 40 and only meet other criteria after age 40 are included in RUKS with a date of inclusion equal to the first purchase of metformin.
- RUKS maintains a window no inclusion of 280 days *before* date of GDM till 280 days after. Our approach only use a window of 30 days before (to account for registration delays) and of 365 days after the date of GDM.
- The RUKS approach to definition T1/T2 is based on recordings in the NPR and classify persons as T1/T2 according to the most frequently occurring of E10 and E11, whereas our approach only classify persons if more than half of the recordings are E10 resp E11, also taking other codes into account. It seems that RUKS do not include the code E12–E14 as diabetes.
- It appears as if RUKS is a prevalence register, where persons are included with reference to a given date. RUKS seems not to be available as a research register — which is quite logical for a register that have different persons recorded for different reference dates.

Chapter 2

Data acquisition

2.1 Diabetes patients

We have reconstructed the Danish Diabetes Register based on information from the National Patient Register, the Prescription Register (officially: Register of Medicines Products Statistics) and the Danish Adult Diabtes Database (DADD) and the eye examination databas (Diabase). The register contains id, date of birth, date of DM and date of death (among other things).

In the register we have defined T1 diabetes as those persons we are reasonably certain to be T1 patients namely those under 15 at first purchase of any anti-diabetic drugs, those under 30 at first purchase of insulin, and those recorded as T1 patients in the DADD or in NPR (the latter excluding those recorded as T2 in DADD).

2.2 Population data

We have had access to records for the entire population containing date of birth, death, emigration and immigration. From this we constructed a dataset with one record per period spent in Denmark; a new record is started whenever a person enters the study population (by birth, immigration or crossing 1995-01-01 alive). The follow-up represented by this record is terminated by emigration death or the end of the study period, currently 2015-12-31.

2.3 Follow-up

We used the Diabetes Register to subdivide the population records of follow-up by state of follow up in "noDM", "T1" and "T2". Also the register information was used to count the number of DM-events; that is diagnoses of T1 resp. T2 and deaths.

2.3.1 Analysis dataset

The constructed dataset or follow-up records was further subdivided by current age and calendar time in 1-year intervals, and the resulting dataset was tabulated by sex, state, age and calendar time of follow-up and date of birth; the latter three in 1-year categories. Each entry in the tabulation contains the person-years at risk, the number of deaths and the

number of diagnoses of T1 and T2, respectively. Obviously, the number of diagnoses of T1 and T2 in the states "T1" and "T2" is 0.

This dataset allows us to model incidence rates of T1 and T2 diabetes as well as mortality rates separately for the three states. In this report we will pool the two types of diabetes and thus base predictions etc. on incidence rates of any type of DM and mortality rates among persons.

2.4 Prevalences

Further when we make projections we shall also need the prevalence of DM at different dates (1 January each year 1996–2016). These numbers were also constructed from the follow-up dataset, by simply extracting those that were alive at the dates and classify these a being either "noDM", "T1" or "T2".

The construction of the follow-up and prevalence data from the register is documented in the SAS-programs `08-mkFU` (p. 165 ff.) and `09-mkPr` (p. 171 ff.).

2.5 Datasets of follow-up and prevalence

First load the relevant package:

```
> options( width=95 )
> library( Epi )
> library( splines )
> library( haven )
> print( sessionInfo(), l=F )
R version 3.3.2 (2016-10-31)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows Server 2012 R2 x64 (build 9600)

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

other attached packages:
[1] haven_1.0.0 Epi_2.7

loaded via a namespace (and not attached):
 [1] cmprsk_2.2-7      MASS_7.3-45       assertthat_0.1    Matrix_1.2-7.1    plyr_1.8.4
 [6] parallel_3.3.2    tools_3.3.2       survival_2.40-1   tibble_1.2        etm_0.6-2
[11] Rcpp_0.12.8       grid_3.3.2        numDeriv_2016.8-1 lattice_0.20-34
```

We first load the follow-up dataset and then the prevalence dataset:

```
> FU  <- read_sas( "../data/futab.sas7bdat" )
> str( FU )
Classes 'tbl_df', 'tbl' and 'data.frame':       25723 obs. of  10 variables:
 $ sex  : num  1 1 1 1 1 1 1 1 1 1 ...
 $ state: chr  "T1" "T1" "T1" "T1" ...
 $ A    : atomic  0 0 0 0 0 0 0 0 0 0 ...
  ..- attr(*, "label")= chr "Age at FU"
 $ P    : atomic  1995 1996 1997 1997 1998 ...
  ..- attr(*, "label")= chr "Date of FU"
 $ C    : num  1995 1995 1996 1997 1997 ...
 $ D    : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
 $ T1    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ T2    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ Y     : num  0.001563 0.000476 0.00017 0.000461 0.000517 ...
 $ U     : num  0 1 1 0 1 0 1 0 1 0 ...
 - attr(*, "label")= chr "FUTAB"
```

```
> prv <- read_sas( "../data/prv.sas7bdat" )
> str( prv )
```

```
Classes 'tbl_df', 'tbl' and 'data.frame':       13545 obs. of  5 variables:
 $ pdat : num  1995 1995 1995 1995 1995 ...
 $ state: chr  "T1" "T1" "T1" "T1" ...
 $ sex  : num  1 1 1 1 1 1 1 1 1 1 ...
 $ age  : num  1 2 3 4 5 6 7 8 9 10 ...
 $ n    : num  3 8 7 17 22 28 40 29 33 42 ...
 - attr(*, "label")= chr "PRV"
```

2.5.1 Rate data

We want a dataframe with number of DM cases (`X`) and deaths among DM patients (`D.DM`) and non-DM perons (`D.nD`) and and person-years among DM patients (`Y.DM`) and non-DM persons (`Y.nD`), classified by sex, age and date.

However, the data from the `futab` dataset is clearly not usable for analysis of rates till 1996 (see the output tabel from `08-mkFU` (page 165), so we restrict the follow-up data to the period after 1996-01-01.

```
> DM <- subset( FU, P>1995 & state %in% c("T1","T2") )[,c("sex","A","P","U","D","Y")]
> DM <- aggregate( DM[,c("D","Y")], by=DM[,c("sex","A","P","U")], FUN=sum )
> names( DM )[5:6] <- c("D.DM","Y.DM")
> head( DM )
```

```
  sex A    P U D.DM        Y.DM
1   2 0 1996 0    0 0.001073922
2   1 1 1996 0    0 0.002034223
3   2 1 1996 0    0 0.003585216
4   1 2 1996 0    0 0.006972621
5   2 2 1996 0    0 0.004438741
6   1 3 1996 0    0 0.004505818
```

```
> nD <- subset( FU, P>1995 & state %in% c("Well") )[,c("sex","A","P","U","T1","T2","D","Y")]
> nD$X <- nD$T1 + nD$T2
> nD <- aggregate( nD[,c("X","D","Y")], by=nD[,c("sex","A","P","U")], FUN=sum )
> names( nD )[6:7] <- c("D.nD","Y.nD")
> head( nD )
```

```
  sex A    P U X D.nD     Y.nD
1   1 0 1996 0 0    0 17.56979
2   2 0 1996 0 5    0 16.38617
3   1 1 1996 0 4   15 18.34906
4   2 1 1996 0 6   10 17.48155
5   1 2 1996 0 7    2 18.06545
6   2 2 1996 0 1    2 17.26346
```

```
> TT <- merge( nD, DM, all=TRUE )
> TT[is.na(TT)] <- 0
> TT <- transform( TT, A = A + (1+U)/3,
+                      P = P + (2-U)/3,
+                    sex = factor( sex, labels=c("M","F") ) )
> head( TT )
```

```
  sex         A        P U X D.nD     Y.nD D.DM         Y.DM
1   M 0.3333333 1996.667 0 0    0 17.56979    0 0.0000000000
2   M 0.6666667 1996.333 1 1   28 17.88386    0 0.0004757016
3   M 0.3333333 1997.667 0 1    1 17.72566    0 0.0004613279
4   M 0.6666667 1997.333 1 1   18 17.32057    0 0.0001704312
5   M 0.3333333 1998.667 0 2    2 17.11636    0 0.0009267625
6   M 0.6666667 1998.333 1 2   18 17.09326    0 0.0007015743
```

```
> summary( TT )
```

```
 sex            A                P               U               X              D.nD
 M:4000   Min.   : 0.3333   Min.   :1996   Min.   :0.0   Min.   :  0.0   Min.   :  0
 F:4000   1st Qu.:25.1667   1st Qu.:2001   1st Qu.:0.0   1st Qu.:  6.0   1st Qu.:  7
          Median :50.0000   Median :2006   Median :0.5   Median : 21.0   Median : 51
          Mean   :50.0000   Mean   :2006   Mean   :0.5   Mean   : 41.2   Mean   :119
          3rd Qu.:74.8333   3rd Qu.:2011   3rd Qu.:1.0   3rd Qu.: 68.0   3rd Qu.:205
          Max.   :99.6667   Max.   :2016   Max.   :1.0   Max.   :280.0   Max.   :590
      Y.nD               D.DM             Y.DM
 Min.   : 0.02235   Min.   :  0.00   Min.   :0.00000
 1st Qu.: 8.50906   1st Qu.:  0.00   1st Qu.:0.06703
 Median :16.05813   Median :  5.00   Median :0.24374
 Mean   :13.22582   Mean   : 18.92   Mean   :0.41975
 3rd Qu.:17.90357   3rd Qu.: 33.00   3rd Qu.:0.63896
 Max.   :23.34236   Max.   :116.00   Max.   :2.70543
```

The numbers may look a bit odd but it is mainly the mortality among 0-year old that appear in the upper triangle, which indeed *is* a bit strange. Finally we save the data for analysis:

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

2.5.2 Prevalence data

We need a version of the prevalence dataset in 1-year classes for analysis of prevalence:

```
> P1tab <- aggregate( data.frame(   N = prv$n ),
+                     data.frame( typ = Relevel( prv$state, list( DM=1:2, Pop=3 ) ),
+                                 sex = factor(prv$sex,labels=c("M","F")),
+                                   A = floor(prv$age),
+                                   P = prv$pdat ),
+                     FUN=sum )
WARNING:  prv$state has been converted to a factor with levels:
 T1 T2 Well
```

```
> str( P1tab )
```

```
'data.frame':        9345 obs. of  5 variables:
 $ typ: Factor w/ 2 levels "DM","Pop": 2 1 2 1 2 1 2 1 2 1 ...
 $ sex: Factor w/ 2 levels "M","F": 1 2 2 1 1 2 2 1 1 2 ...
 $ A  : num  0 0 0 1 1 1 1 2 2 2 ...
 $ P  : num  1995 1995 1995 1995 1995 ...
 $ N  : num  35821 1 34257 3 34787 ...
```

```
> PD <- subset( P1tab, typ=="DM"  )[,-1]
> PN <- subset( P1tab, typ=="Pop" )[,-1]
> names( PD )[grep("N",names(PD))] <- "X"
> Ptab <- merge( PD, PN, all=TRUE )
> Ptab[is.na(Ptab)] <- 0
> Ptab$N <- Ptab$X + Ptab$N
> head( Ptab )
```

```
  sex A    P X     N
1   M 0 1995 0 35821
2   M 0 1996 2 36258
3   M 0 1997 0 34920
4   M 0 1998 1 34936
5   M 0 1999 2 34151
6   M 0 2000 2 34035
> save( Ptab, file="../data/Ptab.Rda" )
```

2.5.3 Crude tables for teaching

```
> FUtab <- aggregate( TT[,5:9], data.frame( sex=TT$sex, A=floor(TT$A/5)*5, P=floor(TT$P) ),
> str( FUtab )
'data.frame':        800 obs. of  8 variables:
 $ sex : Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ A   : num  0 0 5 5 10 10 15 15 20 20 ...
 $ P   : num  1996 1996 1996 1996 1996 ...
 $ X   : num  27 27 29 33 43 47 44 42 61 65 ...
 $ D.nD: num  72 61 27 18 30 23 95 40 166 51 ...
 $ Y.nD: num  177 168 159 151 141 ...
 $ D.DM: num  0 0 0 0 0 0 0 0 1 0 ...
 $ Y.DM: num  0.051 0.045 0.179 0.132 0.267 ...
> head( FUtab )
  sex  A    P  X D.nD     Y.nD D.DM       Y.DM
1   M  0 1996 27   72 177.1678    0 0.05100890
2   F  0 1996 27   61 167.8859    0 0.04497604
3   M  5 1996 29   27 158.9996    0 0.17851403
4   F  5 1996 33   18 150.6390    0 0.13218891
5   M 10 1996 43   30 140.8958    0 0.26694867
6   F 10 1996 47   23 134.9699    0 0.25601985
> Ptab <- aggregate( data.frame(   N = prv$n ),
+                    data.frame( typ = Relevel( prv$state, list( DM=1:2, Pop=3 ) ),
+                                sex = factor(prv$sex,labels=c("M","F")),
+                                  A = floor(prv$age/5)*5,
+                                  P = prv$pdat ),
+                    FUN=sum )
WARNING:  prv$state has been converted to a factor with levels:
 T1 T2 Well
> str( Ptab )
'data.frame':        1925 obs. of  5 variables:
 $ typ: Factor w/ 2 levels "DM","Pop": 1 2 1 2 1 2 1 2 1 2 ...
 $ sex: Factor w/ 2 levels "M","F": 1 1 2 2 1 1 2 2 1 1 ...
 $ A  : num  0 0 0 0 5 5 5 5 10 10 ...
 $ P  : num  1995 1995 1995 1995 1995 ...
 $ N  : num  41 172220 37 163686 163 ...
> head( Ptab )
  typ sex A    P      N
1  DM   M 0 1995     41
2 Pop   M 0 1995 172220
3  DM   F 0 1995     37
4 Pop   F 0 1995 163686
5  DM   M 5 1995    163
6 Pop   M 5 1995 150446
> save( FUtab, Ptab, file="DM5x1.Rda" )
```

Chapter 3

Prevalence and ocurrence rates

3.1 Prerequisites

We will make models for the rates as observed in small intervals of age and calendar time, so we start by specifying the interval length, 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 midpoints of the age-classes (`a.pt` and the time-points `t.pt` — note that we shall make predictions of rates all the way to 2030 — this will mainly be with the purpose of predicting the total number of DM persons in the population:

```
> int <- 1/12
> a.pt <- seq(0,100,int)[-1] - int/2
> t.pt <- seq(1996,2040,int)
> p.pt <- t.pt[-1] - int/2
```

We shall model all the rates by age-period-cohort models separately for men and women, both on a multiplicative and an additive scale. As a sensitivity analysis we will also model the rates only by an age-period model.

We will use natural splines to model the effects of age, period and cohort, and for all analyses we will use the same `number` of knots for these three effects, but of course place them differently based on the location of information, *i.e.* the events:

```
> nk.a <- 8
> nk.p <- 6
> nk.c <- 8
```

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, 1 )
 [1] 0.1000000 0.1888889 0.2777778 0.3666667 0.4555556 0.5444444 0.6333333 0.7222222
 [9] 0.8111111 0.9000000
> 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, 4 )
```

```
 [1] 0.0250000 0.1305556 0.2361111 0.3416667 0.4472222 0.5527778 0.6583333 0.7638889
 [9] 0.8694444 0.9750000
```

So this function generates `nk`, equidistant points in the interval (0,1) where the outer points are $1/(2 \times$ `nk`$)$ from the end. The second parameter is for modifying the multiplier 2, so that the outer intervals are $1/($`nk` $\times$ `bd`:

```
> save( qn, int, a.pt, t.pt, p.pt, nk.a, nk.p, nk.c , file="../data/inits.Rda" )
```

3.2 Diabetes prevalences

Finally, in order to get the machinery working, we need the observed prevalences and population size at the starting point, that is at 1.1.1996. These are available from the same tabulation of the diabetes register as before — note that the `N` in the `Ptab` is the number of persons *without* diabetes, and that the remaining code relies on `N` being the `total` number of persons:

```
> load( file="../data/Ptab.Rda" )
> load( file="../data/inits.Rda" )
> pr <- subset( Ptab, P>1995 )
> summary( pr )
 sex            A                P              X                N
 M:2276   Min.   :  0.00   Min.   :1996   Min.   :   0.0   Min.   :    1
 F:2304   1st Qu.: 27.00   1st Qu.:2001   1st Qu.:  85.0   1st Qu.:12503
          Median : 54.00   Median :2006   Median : 388.0   Median :31840
          Mean   : 54.06   Mean   :2006   Mean   : 774.3   Mean   :25061
          3rd Qu.: 81.00   3rd Qu.:2011   3rd Qu.:1198.5   3rd Qu.:36366
          Max.   :117.00   Max.   :2016   Max.   :5348.0   Max.   :46309
```

```
> head( pr )
  sex A    P X     N
2   M 0 1996 2 36258
3   M 0 1997 0 34920
4   M 0 1998 1 34936
5   M 0 1999 2 34151
6   M 0 2000 2 34035
7   M 0 2001 1 34432
```

These are empirical prevalences (`X`—no. of cases of DM, `N`—population size) for each of the 21 dates 1.1.1996 – 1.1.2016 in 1-year intervals, but to get the machinery running we will need the number of diabetes cases in age intervals of length `int`.

So we model the prevalences as of 1 January each of the years 1996—2016, as a smooth function of age, and use the predicted prevalences to produce the prevalence of diabetes in each of the smaller age-classes that we use for the simulation. We use a log-link binomial model with a smooth spline with 12 knots:

```
> ( kp.a <- c( 15, with( pr, quantile( rep(A,X), qn(12) ) ) ) )
          4.166667%     12.5% 20.83333% 29.16667%     37.5% 45.83333% 54.16667%     62.5%
       15        30        43        50        55        59        62        65        68
70.83333% 79.16667%     87.5% 95.83333%
       71        75        79        86
```

```
> pr.fit <- NArray( list( sex = c("M","F"),
+                           A = a.pt,
+                           P = sort(unique(pr$P)) ) )
> str( pr.fit )
 logi [1:2, 1:1200, 1:21] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ sex: chr [1:2] "M" "F"
  ..$ A  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667
  ..$ P  : chr [1:21] "1996" "1997" "1998" "1999" ...
> prod( dim(pr.fit) )
[1] 50400
```

Note that the array of the fitted prevalences has 1200 ages — we used the single-year *observations* of prevalence to predict the prevalences at a set of narrowly spaced ages.

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.mod <- glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ),
+                family = binomial(link="log"),
+                  data = subset( pr, sex==sx & P==as.numeric(dt) ) )
+ pr.fit[sx,,dt] <- predict( pr.mod,
+                            newdata = data.frame( A = a.pt ),
+                            type = "response" )
+    }
> tt <- pr.fit[,c(1:2,NA,floor(dim(pr.fit)[2]*3/4)+1:5),]
> dimnames( tt )[[2]] <- round( as.numeric( dimnames( tt )[[2]] ), 2 )
> str( tt )
 num [1:2, 1:8, 1:21] 0.000626 0.000579 0.00063 0.000583 NA ...
 - attr(*, "dimnames")=List of 3
  ..$ sex: chr [1:2] "M" "F"
  ..$ A  : chr [1:8] "0.04" "0.12" NA "75.04" ...
  ..$ P  : chr [1:21] "1996" "1997" "1998" "1999" ...
> round( ftable( tt[,,-(5:10)], row.vars=2:1 )*100, 1 )
          P 1996 1997 1998 1999 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016
A     sex
0.04  M      0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
      F      0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
0.12  M      0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
      F      0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
NA    M       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
      F       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
75.04 M      5.7  6.1  6.5  6.8 10.6 11.3 12.0 12.9 13.7 14.6 15.5 16.0 16.2 16.4 16.6
      F      4.9  5.2  5.4  5.6  8.2  8.7  9.1  9.6 10.0 10.4 11.0 11.5 11.6 11.7 11.8
75.12 M      5.8  6.1  6.5  6.8 10.6 11.3 12.0 12.9 13.7 14.6 15.5 16.0 16.2 16.4 16.6
      F      4.9  5.2  5.4  5.6  8.2  8.7  9.1  9.7 10.0 10.4 11.1 11.5 11.7 11.7 11.9
75.21 M      5.8  6.1  6.5  6.9 10.6 11.3 12.0 12.9 13.7 14.6 15.5 16.0 16.2 16.4 16.6
      F      4.9  5.2  5.4  5.6  8.2  8.7  9.1  9.7 10.1 10.5 11.1 11.5 11.7 11.7 11.9
75.29 M      5.8  6.1  6.5  6.9 10.6 11.3 12.0 12.9 13.7 14.6 15.5 16.1 16.3 16.4 16.7
      F      5.0  5.2  5.5  5.6  8.2  8.7  9.1  9.7 10.1 10.5 11.1 11.6 11.7 11.8 11.9
75.38 M      5.8  6.1  6.5  6.9 10.6 11.3 12.0 12.9 13.7 14.6 15.5 16.1 16.3 16.4 16.7
      F      5.0  5.2  5.5  5.6  8.2  8.7  9.2  9.7 10.1 10.5 11.1 11.6 11.7 11.8 12.0
```

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

```
> plp <- function(grid=FALSE){
+ 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(0,20), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+           type="l", lty=1, col="blue", lwd=c(1,2) )
+ if( grid ) abline(h=0:22,v=2:9*10,col=gray(0.9))
+ matlines( a.pt, pr.fit["M",,]*100,
+            type="l", lty=1, col="blue", lwd=c(1,2) )
+ text( 39, 19.5, "Men", adj=1, col="blue", cex=1.2 )
+ a89 <- grep( "89.", dimnames(pr.fit)[[2]] )[1]
+ a80 <- grep( "80.", dimnames(pr.fit)[[2]] )[1]
+ text( 89, pr.fit["M",a89,"1996"]* 99, "1996", col="blue", adj=c(1,1) )
+ text( 80, pr.fit["M",a80,"2016"]*101, "2016", col="blue", adj=c(0,0) )
+ axis( side=1 )
+ axis( side=2 )
+ axis( side=2, at=0:20, labels=NA )
+ matplot( a.pt, pr.fit["F",,]*100,
+           ylim=c(0,20), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+           type="l", lty=1, col="red", lwd=c(1,2) )
+ if( grid ) abline(h=0:22,v=2:9*10,col=gray(0.9))
+ matlines( a.pt, pr.fit["F",,]*100,
+            type="l", lty=1, col="red", lwd=c(1,2) )
+ text( 39, 19.5, "Women", adj=1, col="red", cex=1.2 )
+ text( 89, pr.fit["F",a89,"1996"]* 99, "1996", col="red", adj=c(1,1) )
+ text( 80, pr.fit["F",a80,"2016"]*101, "2016", 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 )
+ }
> plp()
```

```
> plp(grid=TRUE)
```

For the calculations we shall only use the estimated prevalences as of 1.1.1996 as starting point for the simulation:

```
> save( pr, pr.fit, file="../data/prFit.Rda" )
```

3.2.1 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:

```
> p.pr <- as.numeric( dimnames( pr.fit )[["P"]] )
> 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 ) )
```

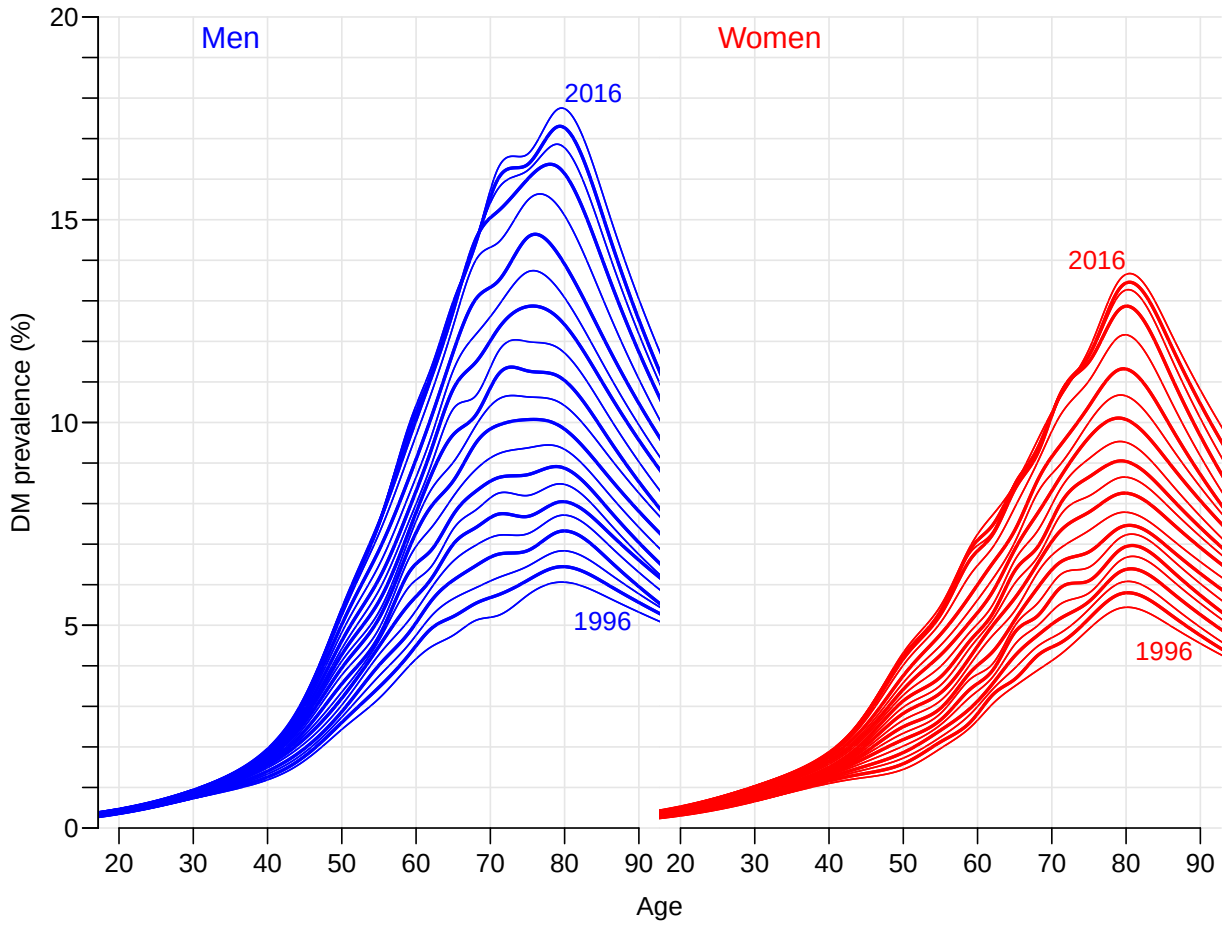

Figure 3.1: *Smoothed age-specific prevalences for the years 1996–2016. Blue is men, red is women.*

```
+ pr.chg[sx,] <- ( ci.exp( lmod, subset="P" ) - 1 ) * 100
+ pr.lfit[sx,,] <- predict( lmod,
+                           newdata = data.frame( A=rep(a.pt,     length(p.pr)),
+                                                 P=rep(p.pr,each=length(a.pt)) ),
+                              type = "response" )
+    }
```

This model is of course a simplification of the model above, which has a totally unspecified 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 )
> a89 <- grep( "89.", dimnames(pr.fit)[[2]] )[1]
> a80 <- grep( "80.", dimnames(pr.fit)[[2]] )[1]
```

```
> text( 89, pr.fit["M",a89,"1996"]* 99, "1996", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M",a80,"2016"]*101, "2016", 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",a89,"1996"]* 99, "1996", col="red", adj=c(1,1) )
> text( 80, pr.fit["F",a80,"2016"]*101, "2016", 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 )
```

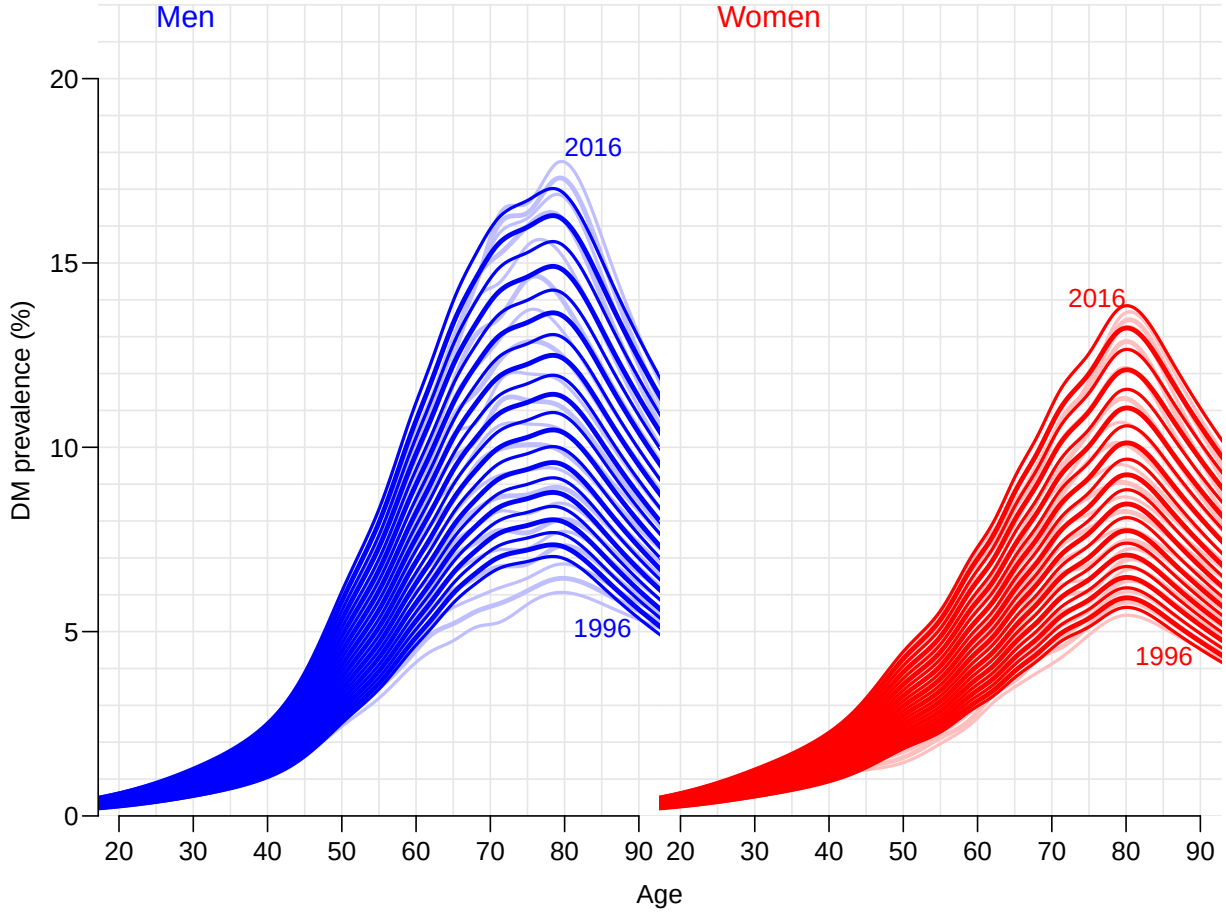

Figure 3.2: *Smoothed age-specific prevalences for the 17-year period 1995–2015 using a model with constant annual relative change in prevalences (full color) compared to the smoothing of the single years (pale color). Blue is men, red is women.*

From figure 3.2 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 2015. So the

overall estimate of some 6% in relative annual increase of prevalences over the 20-year period 1995–2015, is *not* providing an adequate summary:

```
> round( pr.chg, 2 )
  % chg/y   lo   hi
M    4.51 4.49 4.54
F    4.57 4.55 4.60
```

3.2.2 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  41  55  64  71  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 ),
+                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.pr)),
+                                                 P=rep(p.pr,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 )
> 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) )
```

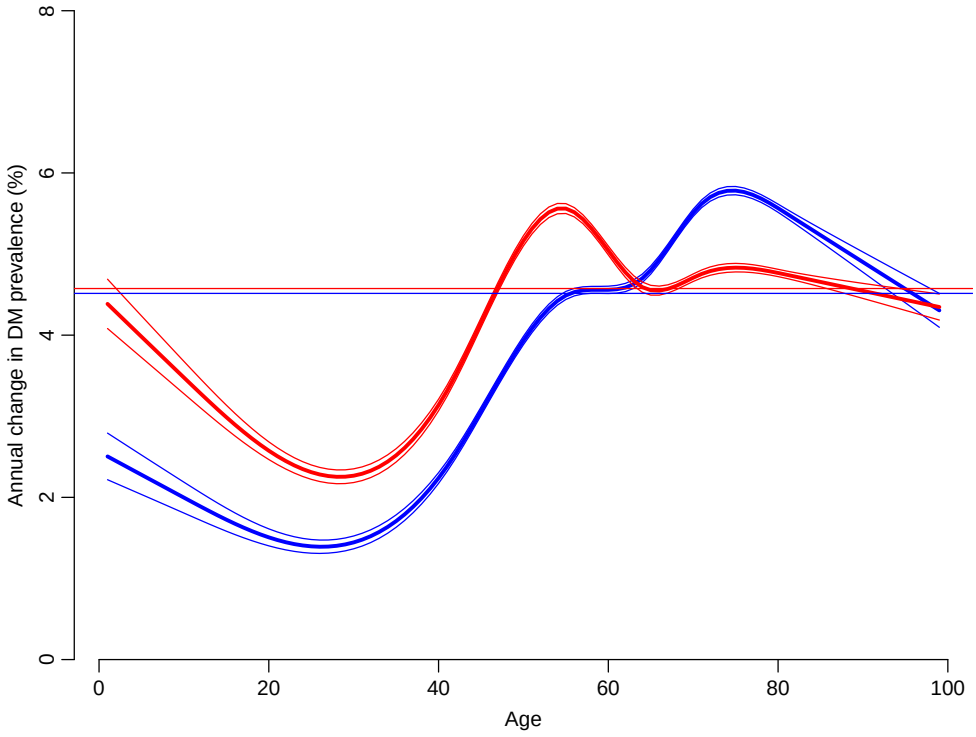

Figure 3.3: *The estimated change in prevalence in different ages, separately for men (blue) and women (red). The horizontal lines indicate the estimate from the naïve model with constant change for all ages.*

```
> 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 )
> a89 <- grep( "89.", dimnames(pr.fit)[[2]] )[1]
> a80 <- grep( "80.", dimnames(pr.fit)[[2]] )[1]
> text( 89, pr.fit["M",a89,"1996"]* 99, "1996", col="blue", adj=c(1,1) )
> text( 80, pr.fit["M",a80,"2016"]*101, "2016", 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",a89,"1996"]* 99, "1996", col="red", adj=c(1,1) )
> text( 80, pr.fit["F",a80,"2016"]*101, "2016", 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 3.4 is seen that the model captures the actual pattern much better than the simple model with an annual change common across ages.

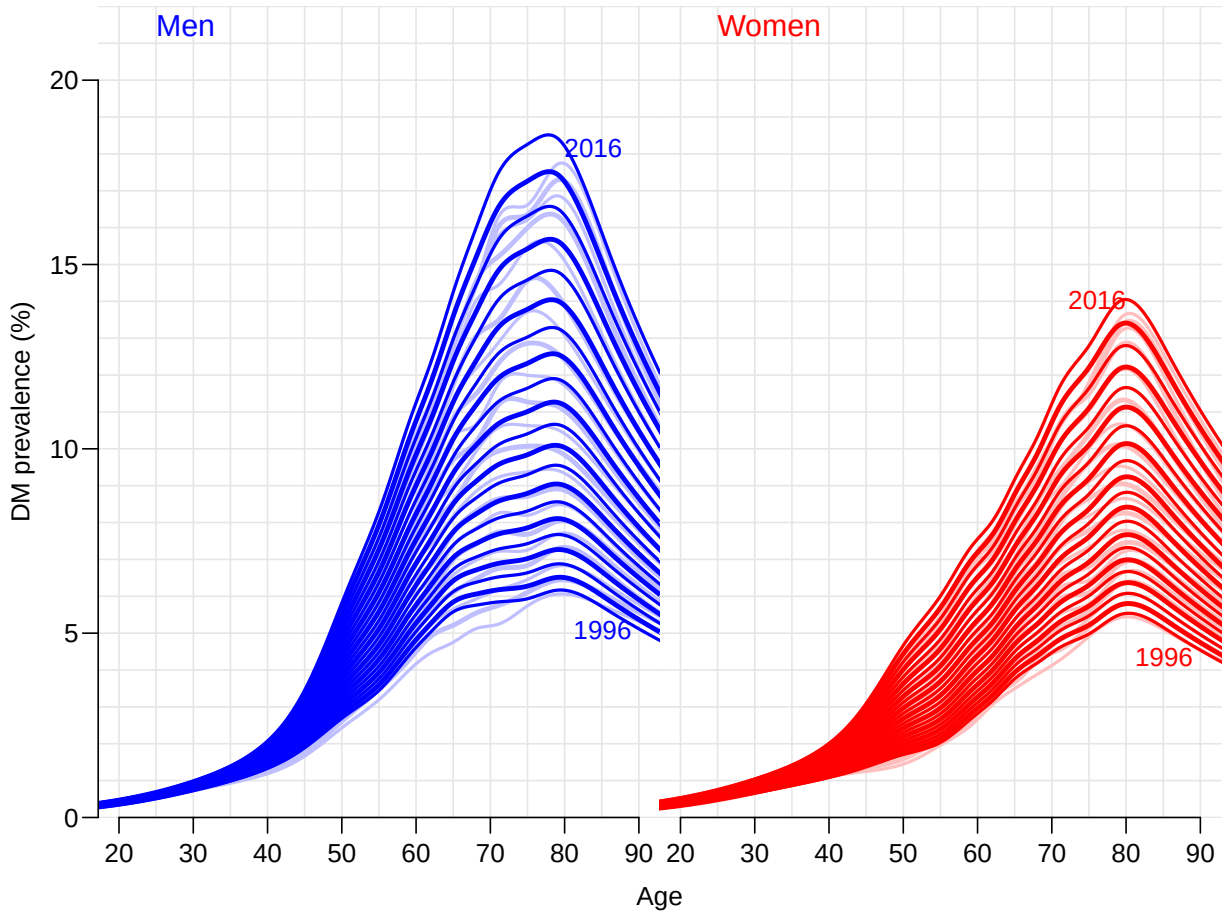

Figure 3.4: *Smoothed age-specific prevalences for the 20-year period 1996–2016 using a model with age-specific constant annual relative change in prevalences (full color) compared to the smoothing of the single years (pale color). Blue is men, red is women.*

3.3 DM incidence

In this chapter we use the constructed follow-up dataset for the entire population:

```
> options( width=95 )
> library( Epi )
> library( splines )
> load( file="../data/TT.Rda" )
```

3.3.1 Age-Period-Cohort modelling

We will use `X` and `Y.nD` as response variables in the analysis of diabetes incidence rates; we fit separate APC-models for total diabetes incidence men and women and plot the estimates together:

```
> DD <- TT[,c("sex","A","P","X","Y.nD")]
> names(DD)[4:5] <- c("D","Y")
> ( A.kn <- quantile( rep(DD$A     ,DD$D), probs=(1:8-0.5)/8 ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
35.33333 47.66667 54.33333 59.66667 64.33333 68.66667 74.33333 82.66667
> ( P.kn <- quantile( rep(DD$P     ,DD$D), probs=(1:6-0.5)/6 ) )
8.333333%       25% 41.66667% 58.33333%       75% 91.66667%
 1998.333  2002.333  2005.667  2009.333  2011.667  2014.333
> ( C.kn <- quantile( rep(DD$P-DD$A,DD$D), probs=(1:8-0.5)/8 ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1922.333 1931.667 1938.333 1943.333 1947.667 1953.333 1960.333 1972.667
> acpM <- apc.fit( subset(DD,sex=="M"), ref.c=1950, parm="ACP",
+                                       npar=list(A=A.kn,P=P.kn,C=C.kn) )
[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                    3992      16129
Age-drift              3991      13747  1  2382.46 < 2.2e-16 ***
Age-Cohort             3985      13683  6    63.27 9.741e-12 ***
Age-Period-Cohort      3981      11179  4  2504.16 < 2.2e-16 ***
Age-Period             3987      11259 -6   -79.66 4.207e-15 ***
Age-drift              3991      13747 -4 -2487.77 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> acpF <- apc.fit( subset(DD,sex=="F"), ref.c=1950, parm="ACP",
+                                       npar=list(A=A.kn,P=P.kn,C=C.kn) )
[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                    3992      16229
Age-drift              3991      13964  1  2265.17 < 2.2e-16 ***
Age-Cohort             3985      13678  6   285.53 < 2.2e-16 ***
Age-Period-Cohort      3981      11291  4  2386.63 < 2.2e-16 ***
Age-Period             3987      11743 -6  -451.46 < 2.2e-16 ***
Age-drift              3991      13964 -4 -2220.70 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> apcM <- apc.fit( subset(DD,sex=="M"), ref.p=2000, parm="APC",
+                                       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                    3992      16129
Age-drift              3991      13747  1  2382.46 < 2.2e-16 ***
Age-Cohort             3985      13683  6    63.27 9.741e-12 ***
Age-Period-Cohort      3981      11179  4  2504.16 < 2.2e-16 ***
Age-Period             3987      11259 -6   -79.66 4.207e-15 ***
Age-drift              3991      13747 -4 -2487.77 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> apcF <- apc.fit( subset(DD,sex=="F"), ref.p=2000, parm="APC",
+                                       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                    3992      16229
Age-drift              3991      13964  1  2265.17 < 2.2e-16 ***
Age-Cohort             3985      13678  6   285.53 < 2.2e-16 ***
Age-Period-Cohort      3981      11291  4  2386.63 < 2.2e-16 ***
Age-Period             3987      11743 -6  -451.46 < 2.2e-16 ***
Age-drift              3991      13964 -4 -2220.70 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we stick to the period-major parametrization as in figure 3.5, we are essentially referring to cross-sectional rates, and they seem to have a peak around age 80. However since there 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.

3.3.2 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 (D-weights) 2.15 2.06 2.24  2.64 2.54 2.74
A-d             2.00 1.92 2.09  2.20 2.10 2.29
```

Thus we see that the average annual trend in rates negigible for men and only slightly more pronounced for women.

3.3.3 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(1900,2020,20), cp.tic=seq(1900,2020,10),
+            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=15, 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(1900,2020,20), cp.tic=seq(1900,2020,10),
+            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=15, 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 3.5 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.

3.3.4 Prediction by age for different 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,
```

[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:

```
+                 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 ),
+              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$Model ) - fitted( Mapc ) )
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-5.116e-13 -1.299e-14  0.000e+00  5.030e-14  8.171e-14  1.023e-12
> summary( fitted( apcF$Model ) - fitted( Fapc ) )
      Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
-4.121e-13  0.000e+00  4.086e-14  4.674e-14  8.171e-14  6.537e-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<2016 & 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':        320 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 1998 1998 1998 ...
 $ gc : num  1910 1910 1910 1910 1910 1910 1910 1910 1910 1910 ...
 $ D  : num  363 645 322 477 266 495 242 382 175 326 ...
 $ Y  : num  45.8 95.1 39.2 84.8 33.2 ...
```

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 )
```

3.3.5 Prediction by age across periods

Alternatively we could show the rates in different ages as a function of calendar time; we shall do this for ages 20, 30,...,90, note that we use the natural spline property of linearity to boldly predict rates beyond 2016-01-01.

```
> prf <- data.frame( A = rep( 2:8*10, each=31 ),
+                    P = rep( 1995:2025, 7 ),
+                    Y = 1 )[outer(c(NA,1:31),0:6*31,"+")[-1],]
```

Then we can show the predicted rates:

```
> plp <-
+ function( mod, clr, pr, ya="l" )
+     {
+ matplot( prf$P, ci.pred( mod, prf ),
+          type="l", lty=c("solid","22","66","11","33")[c(1,1,1,1,3,3,4,5,5)],
+          lwd=c(3,1,1), col=clr,
+          log="y", xaxt="n", xlab="", ylab="", ylim=c(0.2,20), yaxt=ya )
+ abline( v = 2016, col="gray" )
+     }
> par( mfcol=c(3,2), mar=c(0,0,0,0), oma=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plp( Mapc, "blue", prf ) ; text( 1995, 20, "APC", adj=c(0,1) )
> plp( Map , "blue", prf ) ; text( 1995, 20, "AP" , adj=c(0,1) )
> plp( Mac , "blue", prf ) ; text( 1995, 20, "AC" , adj=c(0,1) )
> axis( side=1 )
> plp( Fapc, "red", prf, "n" )
> plp( Fap , "red", prf, "n" )
> plp( Fac , "red", prf, "n" )
> axis( side=1 )
> mtext( "DM incidence rate per 1000 PY", side=2, outer=TRUE, line=2, las=0 )
> mtext( "Date of FU", side=1, outer=TRUE, line=2 )
```

3.3.6 Saving the fitted models

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

```
> save( DD, file="../data/incdata.Rda" )
> save( acpM, apcM, acpF, apcF,
+       Mapc, Mac, Map, Fapc, Fac, Fap, file="../data/inc.Rda" )
```

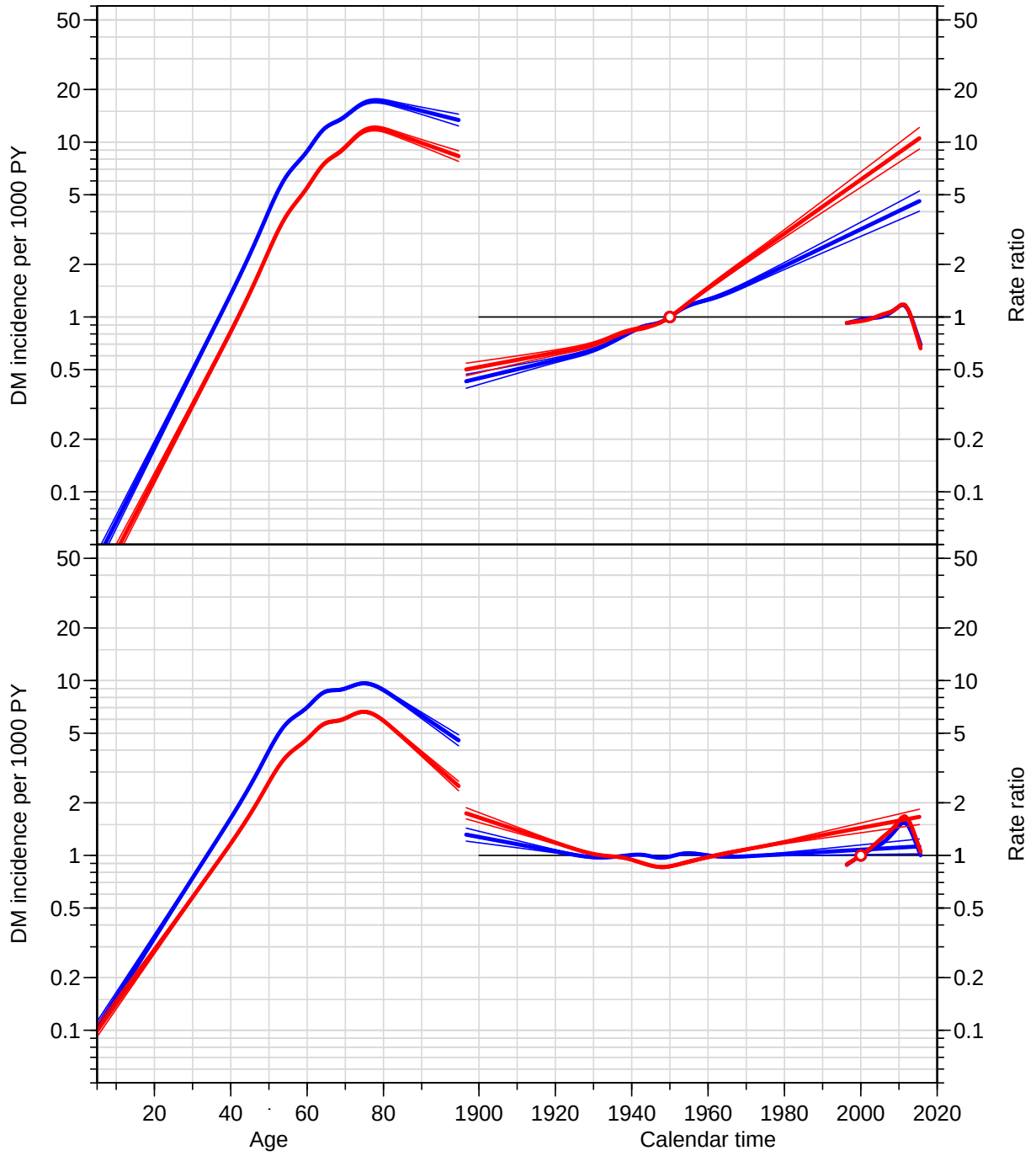

Figure 3.5: *Age-Period-Cohort models for DM incidence among men (blue) and women (red), using the same scaling in the two plots. The top panel is the parametrization with horizontal period effect and cohort reference 1950, bottom panel is the parametrization with horizontal cohort effect and period reference 2000.*

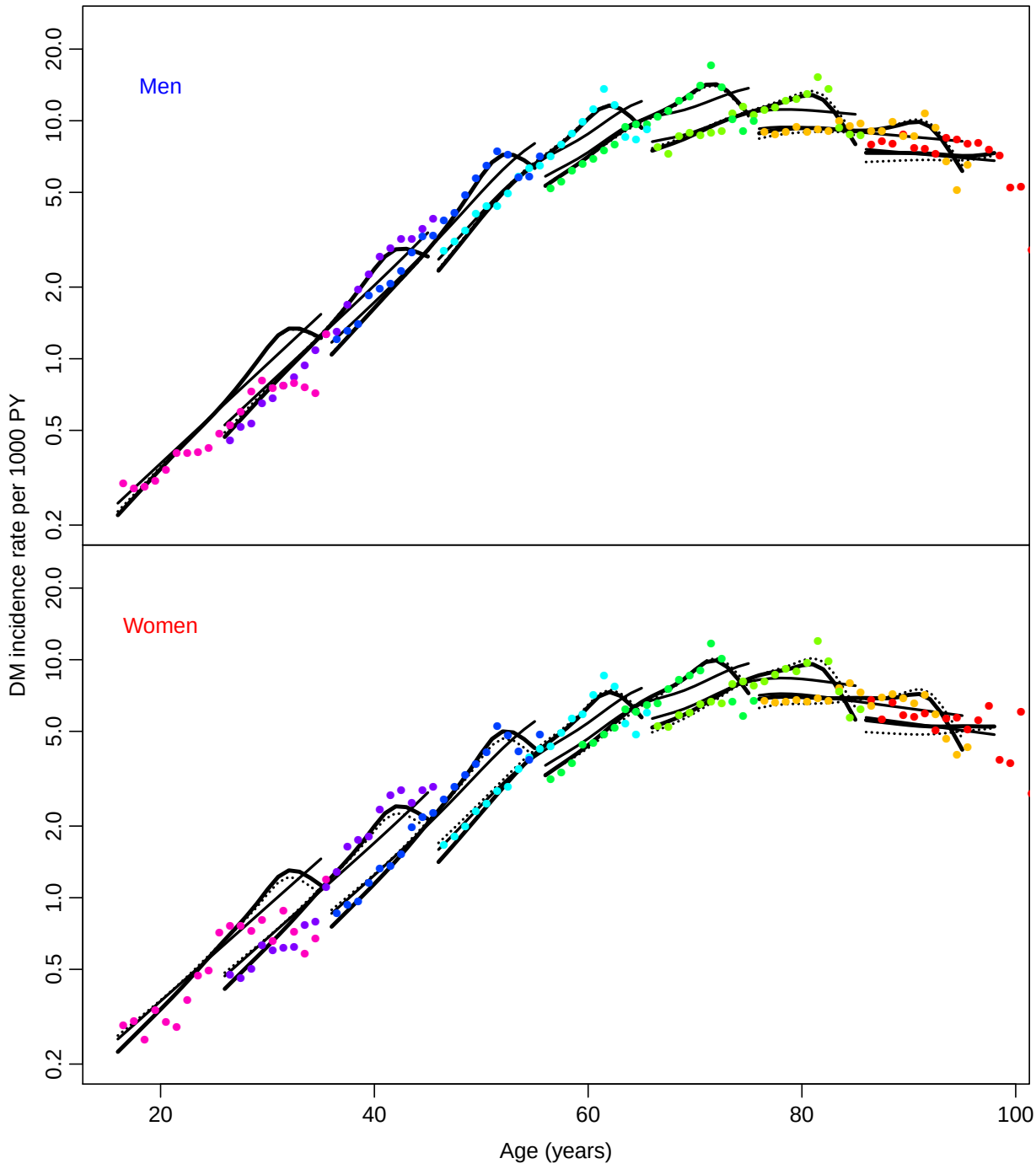

Figure 3.6: *Fitted age-specific incidence rates for the cohorts 1910,. . . ,1980: Full thick line: APC-model, broken line: AP model and full thin line: AC-model. Empirical age-specific rates from C-sets for 1-year period and 10-year cohorts are given as colored dots, colored separately for each cohort.*

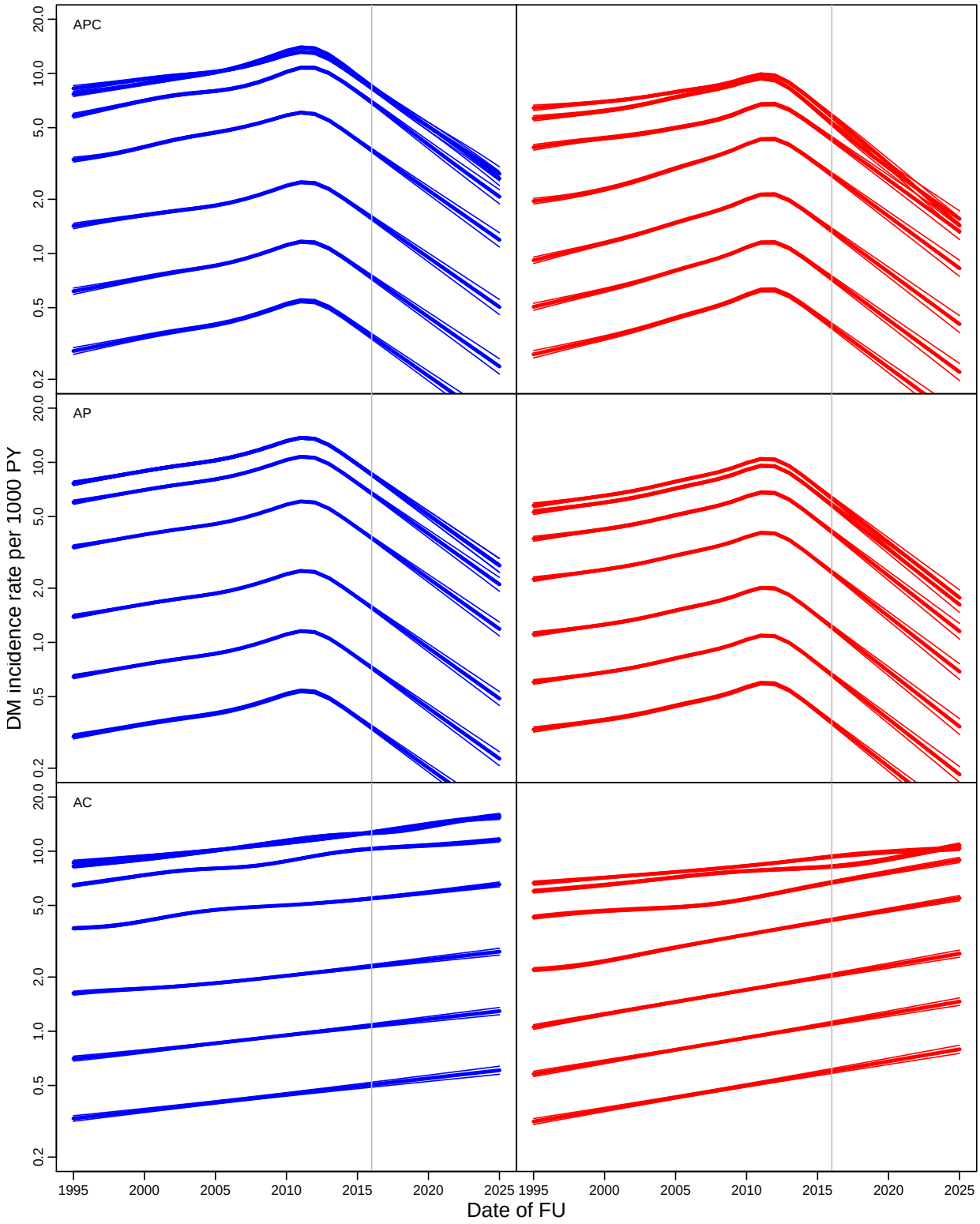

Figure 3.7: *Fitted age-specific incidence rates for ages 20, 30, . . . , 80. Blue: man, red: women, model from top: APC, AP, AC.*

3.4 Mortality

3.4.1 Mortality in non-diabetics

We are going to use `Y.nD` and `Y.nD` as response variables in the analysis of mortality rates; we fit separate models for men and women and plot the estimates together:

```
> nD <- TT[,c("sex","A","P","D.nD","Y.nD")]
> names( nD )[4:5] <- c("D","Y")
> ( A.nD.kn <- c( 10, 20,
+              quantile( rep(nD$A     ,nD$D), probs=(1:8-0.5)/8 ) ) )
                    6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
10.00000 20.00000 50.66667 64.33333 71.66667 77.33333 81.33333 85.33333 88.66667 93.66667
> ( P.nD.kn <- quantile( rep(nD$P     ,nD$D), probs=(1:6-0.5)/6 ) )
8.333333%       25% 41.66667% 58.33333%       75% 91.66667%
 1997.333  2000.667  2003.667  2006.667  2010.333  2014.333
> ( C.nD.kn <- quantile( rep(nD$P-nD$A,nD$D), probs=(1:8-0.5)/8 ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1909.667 1915.667 1920.333 1924.333 1928.667 1934.333 1942.667 1955.667
> nDacpM <- apc.fit( subset(nD,sex=="M"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=A.nD.kn,P=P.nD.kn,C=C.nD.kn) )
[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                    3990    18794.7
Age-drift              3989     7763.3  1  11031.4 < 2.2e-16 ***
Age-Cohort             3983     6581.4  6   1181.9 < 2.2e-16 ***
Age-Period-Cohort      3979     6568.2  4     13.3   0.01011 *
Age-Period             3985     7709.3 -6  -1141.2 < 2.2e-16 ***
Age-drift              3989     7763.3 -4    -54.0 5.313e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> nDacpF <- apc.fit( subset(nD,sex=="F"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=A.nD.kn,P=P.nD.kn,C=C.nD.kn) )
[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                    3990    13636.1
Age-drift              3989     7014.9  1   6621.3 < 2.2e-16 ***
Age-Cohort             3983     5469.6  6   1545.3 < 2.2e-16 ***
Age-Period-Cohort      3979     5444.9  4     24.7 5.784e-05 ***
Age-Period             3985     6905.4 -6  -1460.5 < 2.2e-16 ***
Age-drift              3989     7014.9 -4   -109.5 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 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" )
```

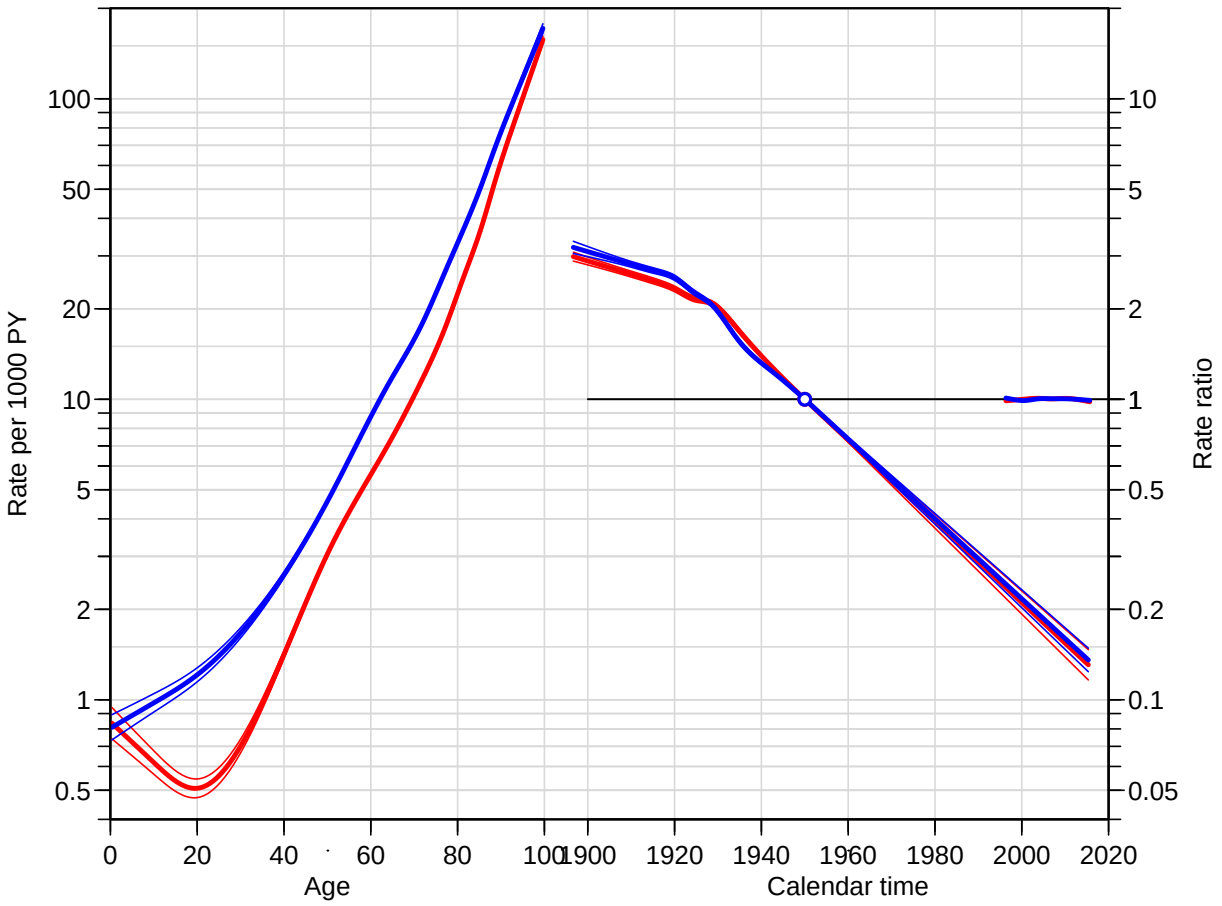

Figure 3.8: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), cohort effects constrained to be 1 at 1950, period slope to be 0. Blue: Men; red: Women.*

We also fit using the period-major parametrization:

```
> nDapcM <- apc.fit( subset(nD,sex=="M"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=A.nD.kn,P=P.nD.kn,C=C.nD.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                    3990    18794.7
Age-drift              3989     7763.3  1  11031.4 < 2.2e-16 ***
Age-Cohort             3983     6581.4  6   1181.9 < 2.2e-16 ***
Age-Period-Cohort      3979     6568.2  4     13.3   0.01011 *
Age-Period             3985     7709.3 -6  -1141.2 < 2.2e-16 ***
Age-drift              3989     7763.3 -4    -54.0 5.313e-11 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> nDapcF <- apc.fit( subset(nD,sex=="F"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=A.nD.kn,P=P.nD.kn,C=C.nD.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                    3990    13636.1
Age-drift              3989     7014.9  1   6621.3 < 2.2e-16 ***
Age-Cohort             3983     5469.6  6   1545.3 < 2.2e-16 ***
Age-Period-Cohort      3979     5444.9  4     24.7 5.784e-05 ***
Age-Period             3985     6905.4 -6  -1460.5 < 2.2e-16 ***
Age-drift              3989     7014.9 -4   -109.5 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> 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" )
```

3.4.2 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 <- TT[,c("sex","A","P","D.DM","Y.DM")]
> names( DM )[4:5] <- c("D","Y")
> DM <- subset( DM, Y>0 )
```

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

```
> ( A.DM.kn <- quantile( rep(DM$A     ,DM$D), probs=(1:8-0.5)/8 ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
56.33333 66.33333 71.66667 75.66667 79.66667 83.33333 86.66667 91.66667
```

```
> ( P.DM.kn <- quantile( rep(DM$P     ,DM$D), probs=(1:6-0.5)/6 ) )
```

```
8.333333%       25% 41.66667% 58.33333%       75% 91.66667%
 1998.333  2002.333  2005.667  2009.333  2011.667  2014.667
```

```
> ( C.DM.kn <- quantile( rep(DM$P-DM$A,DM$D), probs=(1:8-0.5)/8 ) )
```

```
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1912.667 1919.333 1923.333 1927.333 1931.667 1936.333 1942.333 1951.667
```

```
> DMacpM <- apc.fit( subset(DM,sex=="M"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=A.DM.kn,P=P.DM.kn,C=C.DM.kn) )
```

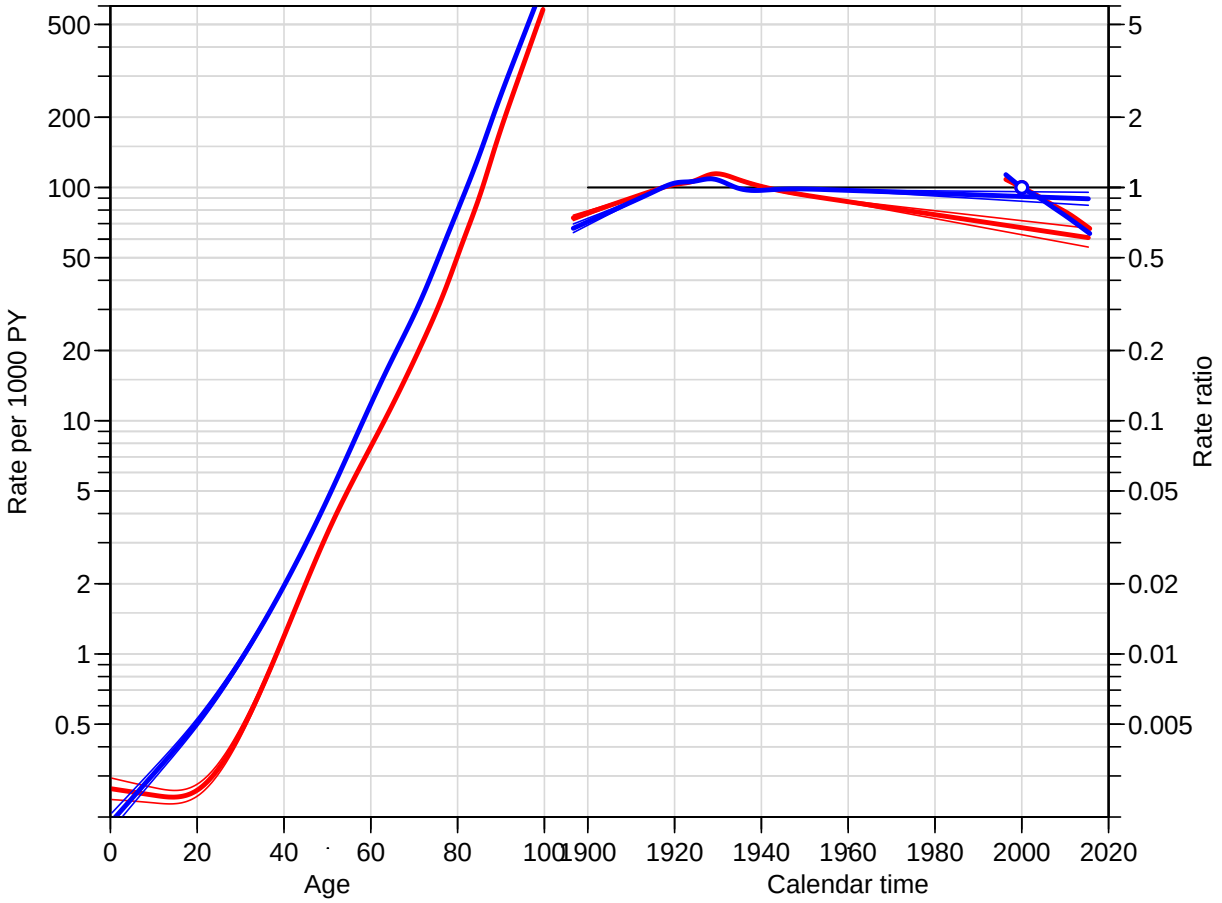

Figure 3.9: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

```
[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                    3984     7139.2
Age-drift              3983     3626.6  1   3512.6  < 2e-16 ***
Age-Cohort             3977     3463.0  6    163.6  < 2e-16 ***
Age-Period-Cohort      3973     3455.1  4      8.0  0.09198 .
Age-Period             3979     3617.6 -6   -162.5  < 2e-16 ***
Age-drift              3983     3626.6 -4     -9.1  0.05967 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> DMacpF <- apc.fit( subset(DM,sex=="F"),
+                    ref.c=1950,
+                    parm="ACP",
+                    npar=list(A=A.DM.kn,P=P.DM.kn,C=C.DM.kn) )

[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                     3977     5621.6
Age-drift               3976     3499.6  1  2122.00 < 2.2e-16 ***
Age-Cohort              3970     3311.6  6   188.08 < 2.2e-16 ***
Age-Period-Cohort       3966     3296.7  4    14.82  0.005091 **
Age-Period              3972     3485.2 -6  -188.44 < 2.2e-16 ***
Age-drift               3976     3499.6 -4   -14.46  0.005974 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> 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" )
```

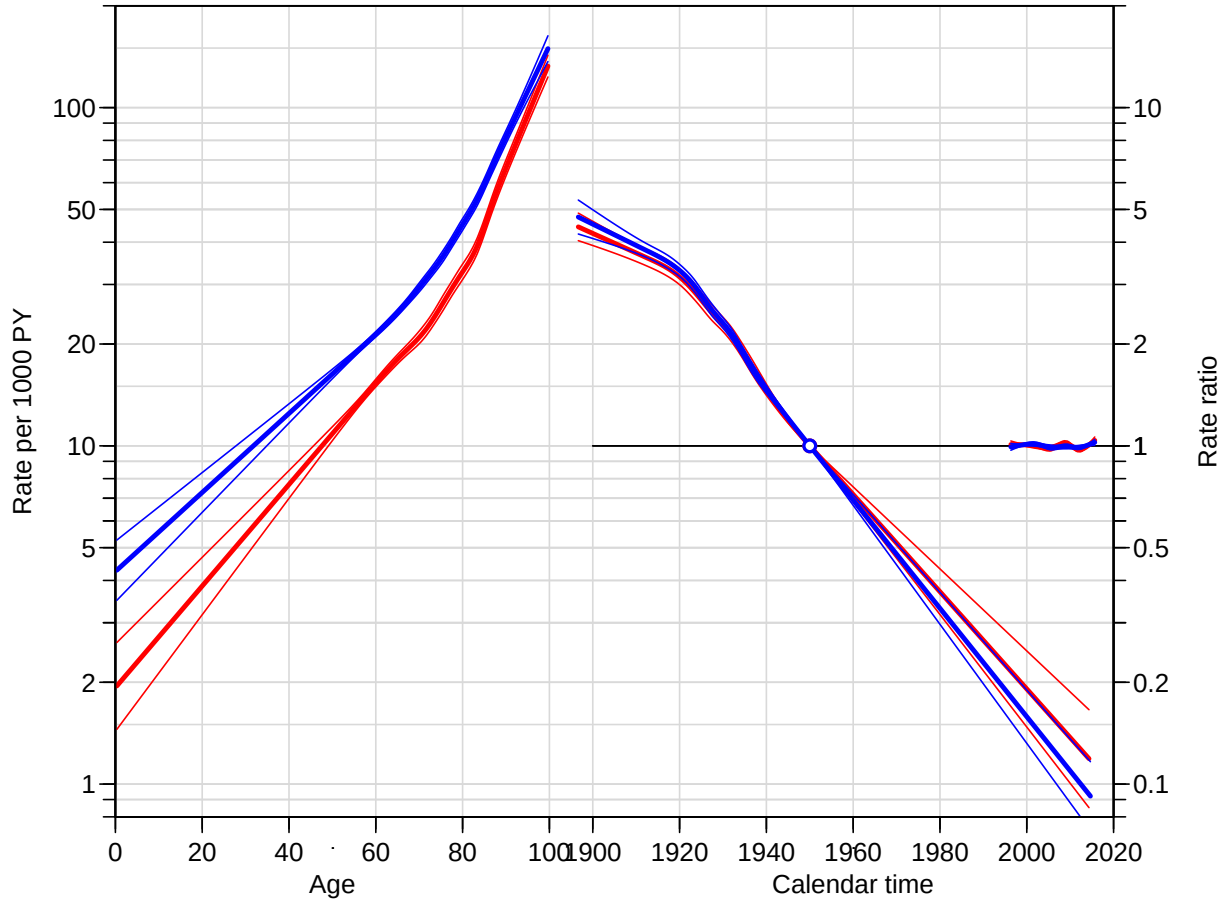

Figure 3.10: *Estimates from an APC-model for mortality among DM patients in Denmark 1995–2011 (original definition), cohort constrained to be 1 at 1950, period slope to be 0. Blue: Men, red: Women.*

We also fit using the period-major parametrization:

```
> DMapcM <- apc.fit( subset(DM,sex=="M"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=A.DM.kn,P=P.DM.kn,C=C.DM.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                    3984     7139.2
Age-drift              3983     3626.6  1   3512.6  < 2e-16 ***
Age-Cohort             3977     3463.0  6    163.6  < 2e-16 ***
Age-Period-Cohort      3973     3455.1  4      8.0  0.09198 .
Age-Period             3979     3617.6 -6   -162.5  < 2e-16 ***
Age-drift              3983     3626.6 -4     -9.1  0.05967 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> DMapcF <- apc.fit( subset(DM,sex=="F"),
+                    ref.p=2000,
+                    parm="APC",
+                    npar=list(A=A.DM.kn,P=P.DM.kn,C=C.DM.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                    3977     5621.6
Age-drift              3976     3499.6  1  2122.00 < 2.2e-16 ***
Age-Cohort             3970     3311.6  6   188.08 < 2.2e-16 ***
Age-Period-Cohort      3966     3296.7  4    14.82  0.005091 **
Age-Period             3972     3485.2 -6  -188.44 < 2.2e-16 ***
Age-drift              3976     3499.6 -4   -14.46  0.005974 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> 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       100
```

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

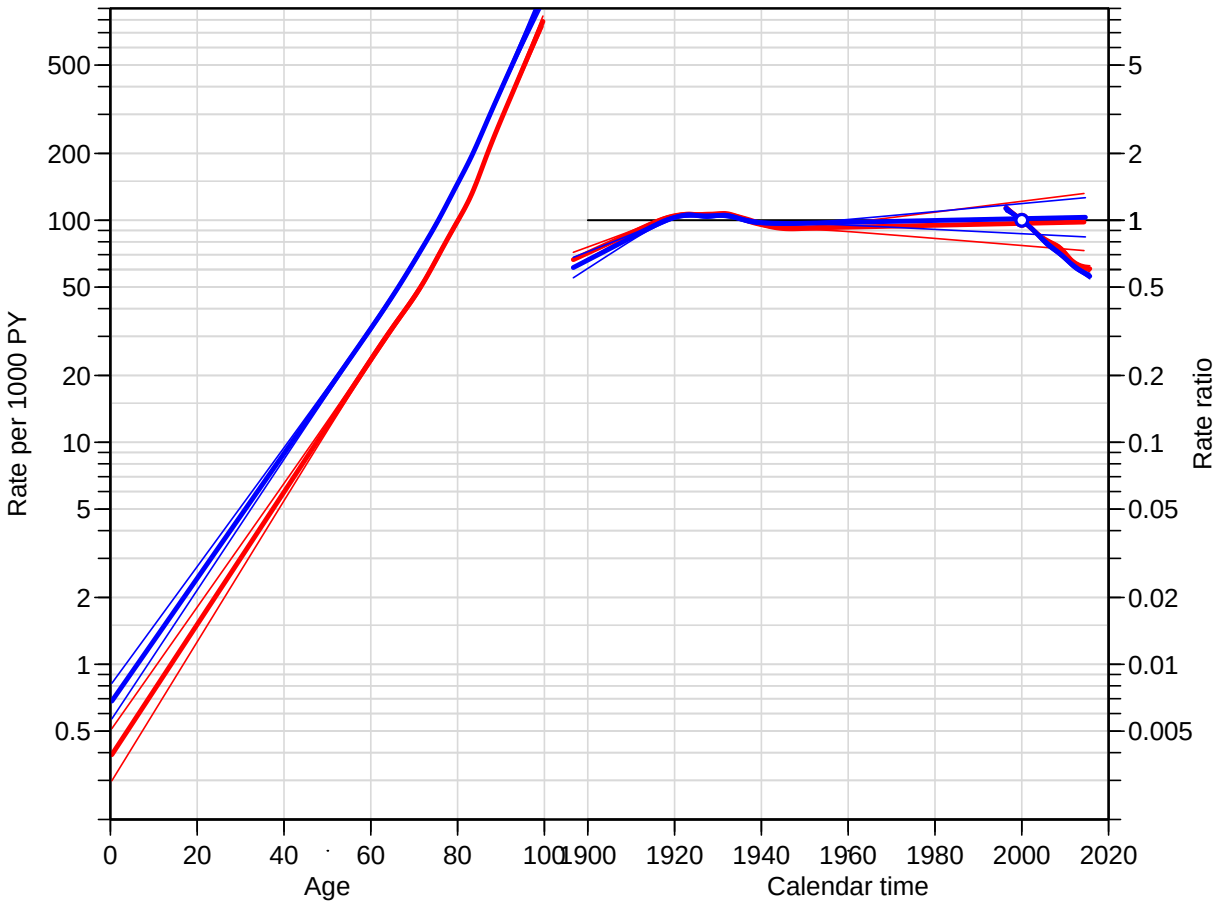

Figure 3.11: *Estimates from an APC-model for mortality among non-diabetic individuals in Denmark 1995–2011 (original definition of DM), period constrained to be 1 at 2000, cohort slope to be 0. Blue: Men, red: Women.*

3.5 Summary of APC models for incidence and 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 base, clearly there is no epidemiologically significant period-effect.

```
> load( file="../data/inc.Rda" )
> par( mfrow=c(3,2), mar=c(0,0,0,0), oma=c(3,4,2,4), mgp=c(3,1,0)/1.6, las=1 )
> frm <- function( side, wh ) {
+ apc.frame( a.lab=seq(0,80,20), a.tic=c(0,seq(10,100,10)),
+           cp.lab=seq(1900,2020,20), cp.tic=seq(1900,2020,10),
+            r.lab=c(outer(c(1,2,5),10^(-1:1)),100), tic.fac=2,
+            r.tic=c(outer(2:9,10^(-1:1)),100,150,200), rr.ref=10,
+            gap=15, r.txt=wh, side=side ) }
> frm( 2, "DM incidence per 1000 PY" )
> lines( acpM, col="blue", ci=TRUE )
> lines( acpF, col="red" , ci=TRUE )
> frm( 4 )
> lines( apcM, col="blue", ci=TRUE )
> lines( apcF, col="red" , ci=TRUE )
> frm( 2, "non-DM mortality per 1000 PY" )
> lines( nDacpM, col="blue", ci=TRUE )
> lines( nDacpF, col="red" , ci=TRUE )
> frm( 4 )
> lines( nDapcM, col="blue", ci=TRUE )
> lines( nDapcF, col="red" , ci=TRUE )
> frm( 1:2, "DM mortality per 1000 PY" )
> lines( DMacpM, col="blue", ci=TRUE )
> lines( DMacpF, col="red" , ci=TRUE )
> frm( c(1,4) )
> lines( DMapcM, col="blue", ci=TRUE )
> lines( DMapcF, col="red" , ci=TRUE )
> mtext( "Cohort major", side=3, line=0.5, at=0.25, outer=TRUE )
> mtext( "Period major", side=3, line=0.5, at=0.75, outer=TRUE )
```

3.5.1 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 (D-weights)                  A-d
                  exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
who    sex
non-DM M              -2.9 -2.9  -2.8      -2.6 -2.7  -2.6
       F              -2.4 -2.5  -2.4      -2.0 -2.0  -1.9
DM     M              -3.7 -3.8  -3.6      -3.6 -3.7  -3.5
       F              -3.4 -3.5  -3.2      -3.1 -3.2  -2.9
```

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 patients than among persons without diabetes.

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

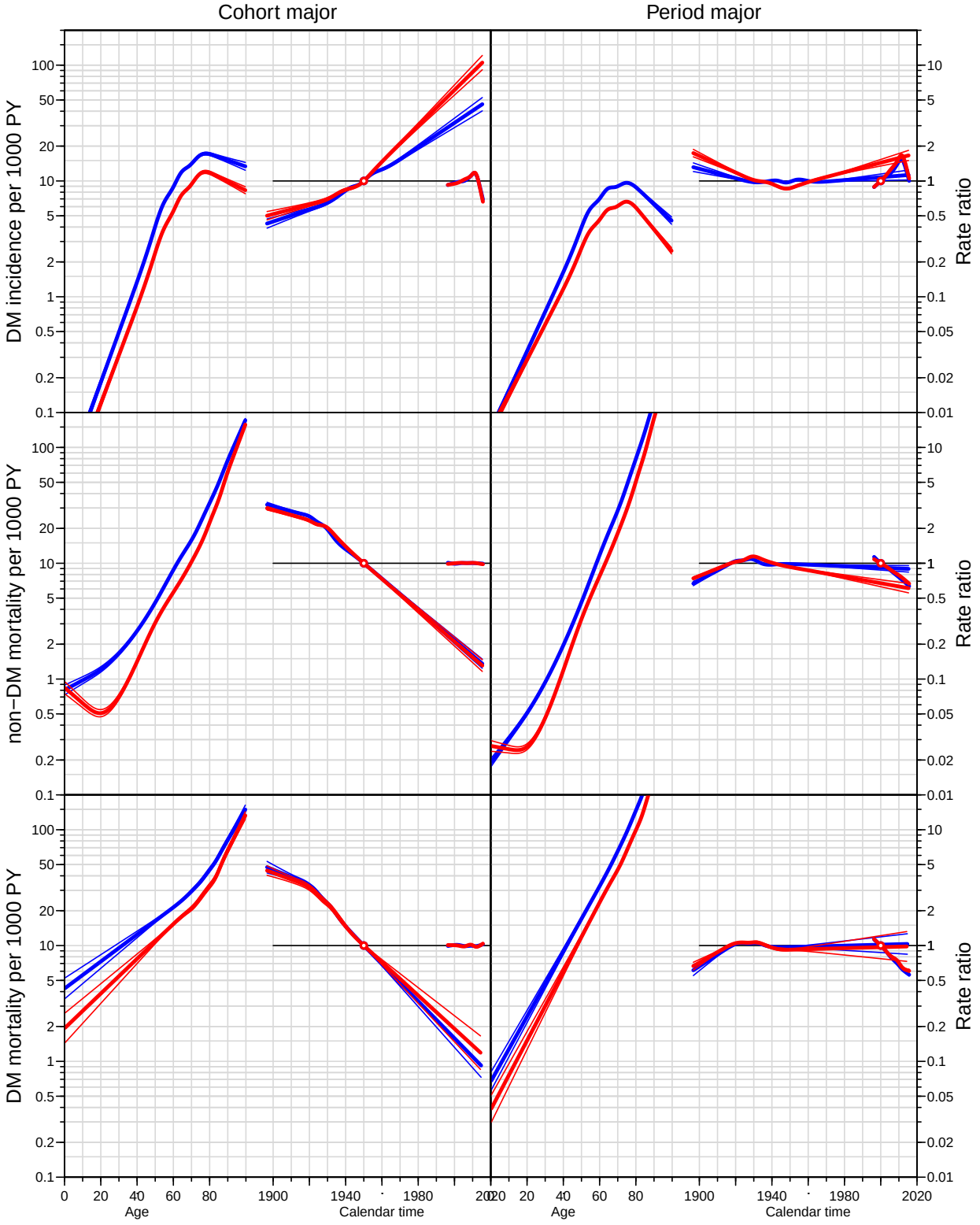

Figure 3.12: *Age-Period-Cohort models for mortality among men (blue) and women (red). Top panel is the mortality among non-diabetics and the lower panel is the mortality among diabetes patients.*

Chapter 4

Prediction of rates

Despite the previous section's modeling of incidence and mortlity rates by APC-models we re-fit the models for prediction purposes.

4.1 Analysis date

We model the incidence and mortality rates from a tabulation of the diabetes register made in conjunction with the annual update of the register. The analytical units are the Lexis triangles, classified by age, period and cohort ($\blacktriangledown$ and $\blacktriangle$):

```
> load( '../data/inits.Rda' )
> load( '../data/TT.Rda' )
> # convert to person years from person-millenia as used in the SAS-programs
> TT[,c("Y.nD","Y.DM")] <- TT[,c("Y.nD","Y.DM")]*1000
> TT <- subset( TT, P > 1996 )
> str( TT )
'data.frame':        8000 obs. of  9 variables:
 $ sex : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A   : num  0.333 0.667 0.333 0.667 0.333 ...
 $ P   : num  1997 1996 1998 1997 1999 ...
 $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
 $ X   : num  0 1 1 1 2 2 2 2 1 3 ...
 $ D.nD: num  0 28 1 18 2 18 0 11 0 21 ...
 $ Y.nD: num  17570 17884 17726 17321 17116 ...
 $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...
 $ Y.DM: num  0 0.476 0.461 0.17 0.927 ...
> head( TT )
  sex         A        P U X D.nD     Y.nD D.DM      Y.DM
1   M 0.3333333 1996.667 0 0    0 17569.79    0 0.0000000
2   M 0.6666667 1996.333 1 1   28 17883.86    0 0.4757016
3   M 0.3333333 1997.667 0 1    1 17725.66    0 0.4613279
4   M 0.6666667 1997.333 1 1   18 17320.57    0 0.1704312
5   M 0.3333333 1998.667 0 2    2 17116.36    0 0.9267625
6   M 0.6666667 1998.333 1 2   18 17093.26    0 0.7015743
> summary( TT )
 sex            A                P              U             X              D.nD
 M:4000   Min.   : 0.3333   Min.   :1996   Min.   :0.0   Min.   :  0.0   Min.   :  0
 F:4000   1st Qu.:25.1667   1st Qu.:2001   1st Qu.:0.0   1st Qu.:  6.0   1st Qu.:  7
          Median :50.0000   Median :2006   Median :0.5   Median : 21.0   Median : 51
          Mean   :50.0000   Mean   :2006   Mean   :0.5   Mean   : 41.2   Mean   :119
```

```
          3rd Qu.:74.8333   3rd Qu.:2011   3rd Qu.:1.0   3rd Qu.: 68.0   3rd Qu.:205
          Max.   :99.6667   Max.   :2016   Max.   :1.0   Max.   :280.0   Max.   :590
      Y.nD               D.DM             Y.DM
 Min.   :   22.35   Min.   :  0.00   Min.   :   0.00
 1st Qu.: 8509.06   1st Qu.:  0.00   1st Qu.:  67.03
 Median :16058.13   Median :  5.00   Median : 243.74
 Mean   :13225.82   Mean   : 18.92   Mean   : 419.75
 3rd Qu.:17903.57   3rd Qu.: 33.00   3rd Qu.: 638.96
 Max.   :23342.36   Max.   :116.00   Max.   :2705.43
> table( TT$Y.DM==0,TT$Y.nD==0 )
       FALSE
  FALSE  7977
  TRUE     23
```

We see from the tabulation that we truly have data in Lexis-triangles:

```
> with( subset( TT, A<5 & P<1999 ),
+       print( table( Age=round(A,2),
+                     Per=round(P,2) ),
+              zero.print="." ) )
      Per
Age    1996.33 1996.67 1997.33 1997.67 1998.33 1998.67
  0.33       .       2       .       2       .       2
  0.67       2       .       2       .       2       .
  1.33       .       2       .       2       .       2
  1.67       2       .       2       .       2       .
  2.33       .       2       .       2       .       2
  2.67       2       .       2       .       2       .
  3.33       .       2       .       2       .       2
  3.67       2       .       2       .       2       .
  4.33       .       2       .       2       .       2
  4.67       2       .       2       .       2       .
```

A brief overview of the number of events and PY:

```
> tt <- xtabs( cbind(X,D.nD,D.DM,Y.nD,Y.DM) ~ sex + floor(P),
+              data = TT )
> round( ftable( addmargins(tt,1:2), row.vars=1:2 ) )
                      X      D.nD      D.DM      Y.nD      Y.DM
sex floor(P)
M   1996           6602     26894      2958   2551813     47789
    1997           6528     26205      3069   2559636     51223
    1998           6991     25607      3124   2565493     54920
    1999           7468     25235      3415   2570306     58725
    2000           7414     24463      3406   2575532     62791
    2001           7491     24655      3504   2582024     66845
    2002           7720     24359      3747   2587849     70815
    2003           8785     24141      3872   2591456     75133
    2004           9061     23316      3871   2593652     80158
    2005           8864     22600      3947   2597064     85083
    2006           9228     22863      4124   2602379     90126
    2007           9772     22646      4242   2611223     95272
    2008          10628     22224      4280   2624040    101316
    2009          11246     21934      4763   2635748    107637
    2010          12188     21689      4788   2644351    114590
    2011          13907     20833      4928   2651824    122766
    2012          12145     20419      5141   2659128    130855
    2013           9267     20523      5290   2670097    136249
```

```
    2014          8867     19938      5470   2688000    139869
    2015          9912     20296      5532   2689703    142459
    Sum         184084    460840     83471  52251318   1834623
F   1996          5413     27411      2661   2620753     43025
    1997          5277     27274      2728   2628018     45520
    1998          5501     26407      2765   2633905     48237
    1999          5811     27243      2873   2638757     50976
    2000          5862     26260      2971   2644189     53927
    2001          6012     26533      3045   2650333     56871
    2002          6010     26784      3110   2655795     59848
    2003          6939     25872      3253   2658828     62949
    2004          7057     24718      3124   2661464     66811
    2005          7196     24459      3323   2664904     70570
    2006          7276     24431      3424   2669731     74431
    2007          8007     24679      3566   2676849     78479
    2008          8481     23711      3484   2687706     83325
    2009          8610     23783      3710   2699204     88124
    2010          9365     23162      3884   2708667     93337
    2011         11220     22341      3802   2716290     99759
    2012          9804     22023      3837   2722084    106525
    2013          7309     21752      3981   2731402    111039
    2014          6648     21080      4063   2745803    113931
    2015          7681     21276      4271   2740586    115710
    Sum         145479    491199     67875  53555267   1523395
Sum 1996         12015     54305      5619   5172566     90814
    1997         11805     53479      5797   5187655     96743
    1998         12492     52014      5889   5199397    103158
    1999         13279     52478      6288   5209063    109701
    2000         13276     50723      6377   5219720    116717
    2001         13503     51188      6549   5232356    123717
    2002         13730     51143      6857   5243644    130663
    2003         15724     50013      7125   5250284    138082
    2004         16118     48034      6995   5255116    146969
    2005         16060     47059      7270   5261968    155653
    2006         16504     47294      7548   5272110    164557
    2007         17779     47325      7808   5288072    173752
    2008         19109     45935      7764   5311746    184641
    2009         19856     45717      8473   5334951    195761
    2010         21553     44851      8672   5353018    207927
    2011         25127     43174      8730   5368114    222524
    2012         21949     42442      8978   5381212    237380
    2013         16576     42275      9271   5401500    247288
    2014         15515     41018      9533   5433802    253800
    2015         17593     41572      9803   5430289    258170
    Sum         329563    952039    151346 105806585   3358018
```

```
> tt <- xtabs( cbind(X,D.nD,Y.nD,D.DM,Y.DM) ~ sex + gP,
+              data = transform( subset( TT, P > 1996 ),
+                                gP = factor( (P>=2001)+(P>=2006)+(P>=2011),
+                                             labels=c("1996-2000",
+                                                      "2001-2005",
+                                                      "2006-2010",
+                                                      "2011-2015") ) ) )
> round( ftable( addmargins(tt,1:2), row.vars=1:2 ) )
```

```
                     X      D.nD      Y.nD      D.DM      Y.DM
sex gP
M   1996-2000    35003    128404  12822780     15972    275448
    2001-2005    41921    119071  12952045     18941    378035
    2006-2010    53062    111356  13117741     22197    508942
```

```
    2011-2015      54098    102009  13358752     26361    672198
    Sum           184084    460840  52251318     83471   1834623
F   1996-2000      27864    134595  13165621     13998    241685
    2001-2005      33214    128366  13291324     15855    317049
    2006-2010      41739    119766  13442157     18068    417696
    2011-2015      42662    108472  13656165     19954    546964
    Sum           145479    491199  53555267     67875   1523395
Sum 1996-2000      62867    262999  25988401     29970    517133
    2001-2005      75135    247437  26243369     34796    695084
    2006-2010      94801    231122  26559898     40265    926638
    2011-2015      96760    210481  27014917     46315   1219162
    Sum           329563    952039 105806585    151346   3358018
```

We set up arrays to hold the predicted incidence and mortality rates from the different models, separately for the two sexes:

```
> Lambda <- Mu.W <- Mu.DM <- NArray( list( a = a.pt,
+                                          p = p.pt,
+                                        sex = c("M","F"),
+                                        mod = c("ap","apc","att","i20","i25","i30") ) )
> str( Lambda )
 logi [1:1200, 1:528, 1:2, 1:6] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ a  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667
  ..$ p  : chr [1:528] "1996.04166666667" "1996.125" "1996.20833333333" "1996.29166666667"
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:6] "ap" "apc" "att" "i20" ...
> prod( dim(Lambda) )
[1] 7603200
```

Note that we have a level on dimension 4, `att`, designed to hold predictions where we attenuate the linear trajectories from the period and cohort terms.

4.1.1 Datasets for rate modeling

First we construct simple datasets for APC and Lee-Carter analysis separately for the three types of transitions we are going to consider.

We can explore the entire set of models for men and women separately using the `apc.LCa` function:

```
> incdat <- TT[,c("A","P",    "X","Y.nD","sex")]
> mnDdat <- TT[,c("A","P","D.nD","Y.nD","sex")]
> mDMdat <- TT[,c("A","P","D.DM","Y.DM","sex")]
> names( incdat )[3:4] <-
+ names( mnDdat )[3:4] <-
+ names( mDMdat )[3:4] <-c("D","Y")
> mDMdat <- subset( mDMdat, Y>0 )
```

4.2 Models for incidence and mortality rates

There are two immediate alternatives to the age-period-cohort models on the log-scale; age-period-cohort models on the additive scale and Lee-Carter models. These may provide a better fit to data (some of the expanded Lee-Carter models necessarily do).

```
> minc <- apc.LCa( subset(incdat,sex=="M"), eps=1e-4 )
> finc <- apc.LCa( subset(incdat,sex=="F"), eps=1e-4 )
> mmnD <- apc.LCa( subset(mnDdat,sex=="M"), eps=1e-4 )
> fmnD <- apc.LCa( subset(mnDdat,sex=="F"), eps=1e-4 )
> mmDM <- apc.LCa( subset(mDMdat,sex=="M"), eps=1e-4 )
> fmDM <- apc.LCa( subset(mDMdat,sex=="F"), eps=1e-4 )
> save( incdat, mnDdat, mDMdat,
+       minc, finc, mmnD, fmnD, mmDM, fmDM,
+       file = "imdat.Rda" )
```

Once we have fitted all models for all transitions we can show their relative merits in terms of deviance, and assess to what extent the APC models would suffice or whether substantial improvements could be obtained by using a Lee-Carter model:

```
> load( file = "imdat.Rda" )
> par( mfcol=c(3,2) )
> show.apc.LCa( minc, col.txt="blue" ); text( 10, 90, "DM\n incidence", cex=2 )
> show.apc.LCa( mmDM, col.txt="blue" ); text( 10, 90, "DM\n mortality", cex=2 )
> show.apc.LCa( mmnD, col.txt="blue" ); text( 10, 90, "non-DM\n mortality", cex=2 )
> show.apc.LCa( finc, col.txt="red"  )
> show.apc.LCa( fmDM, col.txt="red"  )
> show.apc.LCa( fmnD, col.txt="red"  )
```

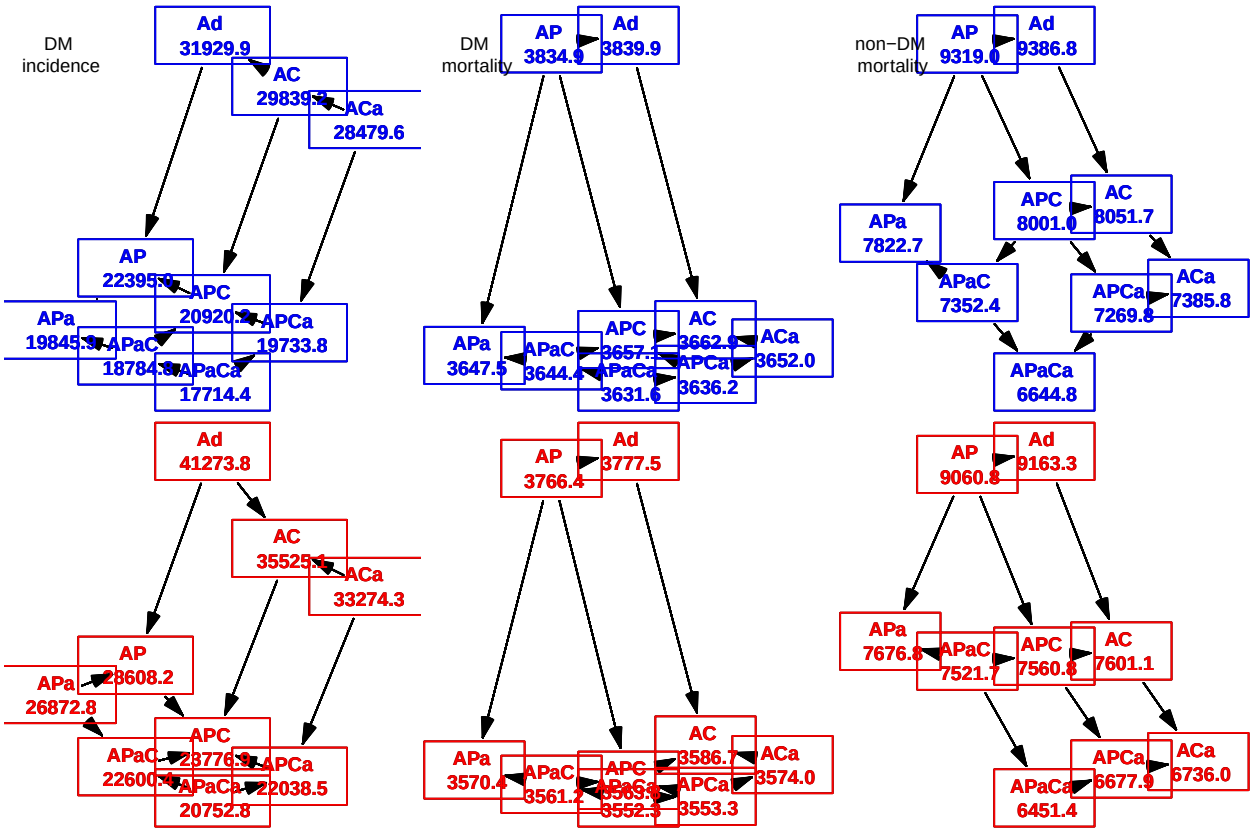

Figure 4.1: *Relative fit of APC and Lee-Carter models for incidence and mortality rates for men and women respectively.*

From the relative model fits shown in figure 4.1 we see that for diabetes incidence and mortality the APC is quite well fitting relative to the classical Lee-Carter model (APa),

but that the Lee-Carter type models offer some improvement relative to the APC model for non-DM mortality but only if cohort-age effect is included.

Thus, we shall proceed with APC-models for the incidence and mortality rates.

4.3 Incidence rates

Based on this we can now derive the location of the knots for this model (first loading the prerequisites):

```
> load( file="../data/inits.Rda" )
> ( ki.a <- with( incdat, quantile( rep(  A,D), qn(nk.a) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
35.33333 47.66667 54.33333 59.66667 64.33333 68.66667 74.33333 82.66667
> ( ki.p <- with( incdat, quantile( rep(P  ,D), qn(nk.p) ) ) )
8.333333%       25% 41.66667% 58.33333%       75% 91.66667%
 1998.333  2002.333  2005.667  2009.333  2011.667  2014.333
> ( ki.c <- with( incdat, quantile( rep(P-A,D), qn(nk.c) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1922.333 1931.667 1938.333 1943.333 1947.667 1953.333 1960.333 1972.667
```

The model we set up is an age-period-cohort model with these three terms in it. As we are only going to use the model for predictions we need not bother about parametrization issues, so it is not an issue that the model we fit is formally over-parametrized. However we want to extract the average trend from the APC-model, so we also fit the model with the parametrization that allows us to extract the trend.

```
> m.inc.ap <- glm( D ~ Ns(A,knots=ki.a) + Ns(P,knots=ki.p),
+                      offset = log( Y ),
+                      family = poisson,
+                        data = subset(incdat,sex=="M") )
> m.inc.aPC <- update( m.inc.ap, . ~ . - Ns(P  ,kn=ki.p) + I(P) +
+                              detrend( Ns(P  ,kn=ki.p), P  , D ) +
+                              detrend( Ns(P-A,kn=ki.c), P-A, D ) )
> m.inc.apc <- update( m.inc.ap, . ~ . + Ns(P-A,kn=ki.c) )
> c( m.inc.apc$deviance, m.inc.aPC$deviance )
[1] 11179.21 11179.21
> f.inc.ap  <- update( m.inc.ap , data = subset(incdat,sex=="F") )
> f.inc.apc <- update( m.inc.apc, data = subset(incdat,sex=="F") )
> f.inc.aPC <- update( m.inc.aPC, data = subset(incdat,sex=="F") )
```

The average annual trends in incidence from the multiplicative models:

```
> inc.chg <- rbind( ci.exp(m.inc.aPC,subset="I\\(P")-1,
+                   ci.exp(f.inc.aPC,subset="I\\(P")-1 )*100
> rownames( inc.chg ) <- c("DM incidence change     Men",
+                          "                       Women")
> round( inc.chg, 1 )
                          exp(Est.) 2.5% 97.5%
DM incidence change     Men       2.2  2.1   2.2
                      Women       2.6  2.5   2.7
```

The average increase is higher for women than for men, but the period effect is massively non-linear, so these summary figures are not really informative, see the comparative figure with the mortality rates, **??**

4.3.1 Incidence rate predictions

Finally we need the predicted incidence rates at a grid of points suitable for the calculations of predicted prevalences. We make the predictions for all combinations of `a.pt` and `p.pt`.

However, 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`,

```
> nd <- data.frame( A = rep(a.pt,     length(p.pt)),
+                   P = rep(p.pt,each=length(a.pt)),
+                   Y = int )
> head( nd )
           A        P          Y
1 0.04166667 1996.042 0.08333333
2 0.12500000 1996.042 0.08333333
3 0.20833333 1996.042 0.08333333
4 0.29166667 1996.042 0.08333333
5 0.37500000 1996.042 0.08333333
6 0.45833333 1996.042 0.08333333
> dim( nd )
[1] 633600      3
```

Note that the prediction data frame was set up with age varying fastest, and the `Lambda` array with age before period, so that the column-major storage of arrays conforms with the predictions form `nd`:

```
> Lambda[,,"M","ap" ] <- predict.glm( m.inc.ap , type="response", newdata=nd )
> Lambda[,,"F","ap" ] <- predict.glm( f.inc.ap , type="response", newdata=nd )
> Lambda[,,"M","apc"] <- predict.glm( m.inc.apc, type="response", newdata=nd )
> Lambda[,,"F","apc"] <- predict.glm( f.inc.apc, type="response", newdata=nd )
```

Thus we have the incidence rates that we need for two sexes, and for two different modeling approaches.

4.3.2 A damping extrapolation

Instead of using the naive extrapolation we may dampen the trend derived from the naive application of the natural splines. I the vein of the recommendation by Sasieni [**?**] we use an attenuation of the trend beyond 2016 for period and beyond 2000 for cohort. The attenuation is traditionally set to 0.92 per year, but we shall be more conservative.

The arithmetic goes as follows: Suppose the slope of the period or cohort effect is β, but that it would be an exaggeration to continue the period effect indefinitely at a slope of β, so we choose a *damping* factor, d, say, such that the slope of the effect at tie t is not β, but rather βd^t. Thus for the effect $f(t)$ we have that:

$$f'(t) = \beta d^t \quad \Leftrightarrow \quad f(t) = k + \beta d^t / \log(d) \quad \Rightarrow \quad f(0) = k + \beta / \log(d)$$

Now, if the original linear extrapolation we have from the natural spline parametrization has the form:

$$f(t) = \alpha + \beta t$$

then at the prediction start (for convenience of notation, $t = 0$):

$$f(0) = \alpha = k + \beta/\log(d) \quad \Leftrightarrow \quad k = \alpha - \beta/\log(d)$$

These formulae are implemented in the function `damp`, which assumes that the predicted effects at times `t` are in `fval`, and that the effects are linear beyond `t0`, but that we attenuate the effect with the damping factor `damp`, in cancer epidemiology by default often set to 0.92, but we shall be a bit more conservative and set it to 0.88 (which also is a figure taken out of thin air):

```
> damp <-
+ function( fval, t, t0, dfac=0.88 )
+     {
+ # where are the ealiest an latest prediction point after t0
+ wh <- match( range( t[t>t0] ), t )
+ # slope of predicted curve - we assume linearity
+ beta <- diff( fval[wh] ) / diff( t[wh] )
+ # the attenuated curve - revision beyond first point after t0:
+ kons <- fval[wh[1]] - beta/log(dfac)
+ ifelse( t>t[wh[1]], kons + beta * dfac^(t-t[wh[1]]) / log( dfac ), fval )
+     }
> t <- seq( 0,30,,100 )
> f <- 2.5 + 0.4 * t
> t0 <- 16
> matplot( t, cbind( f, damp(f,t,5) ), lty=1, lwd=c(8,2), type="l" )
> abline( v= 5 )
```

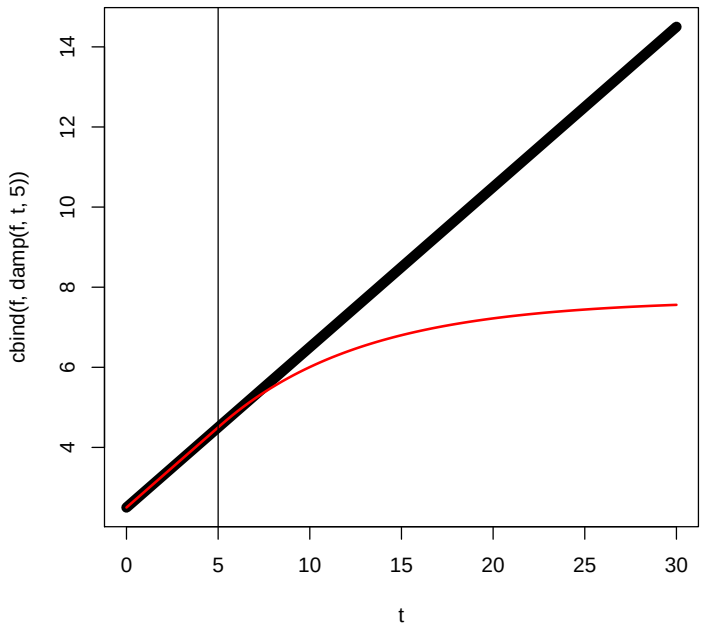

Figure 4.2: *Illustration of the* `damp` *function for attenuation of linear effects.*

We can now employ this damping function on the terms-predictions from the apc-models, and show how it pans out in practice for the period and cohort effects. We adjust the

predictions for the period and cohort terms and use these to reconstruct the predicted (cumulative) rates, note in particular that the intercept is not in the terms components, it must be retrieved from the attribute `const` in the `predict.glm` object:

```
> predatt.apc <-
+ function( obj, nd, dfac=0.88, p.end=2016, c.end=2000 )
+    {
+ # predicted terms --- linear beyond outer knots
+ pc <- po <- predict.glm( obj, type="terms", newdata=nd )
+ # replace with the attenuated curves
+ pc[,2] <- damp( pc[,2], nd$P     , p.end, dfac=dfac )
+ pc[,3] <- damp( pc[,3], nd$P-nd$A, c.end, dfac=dfac )
+ # return predicted rates as well as original and revised terms predictions
+ list( pr.resp  = exp( apply( pc, 1, sum ) + attr(pc,"const") + log(nd$Y) ),
+       pr.terms = pc,
+       pr.org   = po )
+    }
```

We have observed that the incidence rates show a decreasing tendency over the last few years of observation, hence we may want not only to investigate a scenario where rates are kept constant, but also one where we increase the period effect by 2% per year — this is only going to be used for the incidence rates as a sensitivity analysis, and together with the

```
> predinc.apc <-
+ function( obj, nd, dfac=0.025, p.end=2016, c.end=2000 )
+    {
+ # predicted terms --- linear beyond outer knots
+ pc <- po <- predict.glm( obj, type="terms", newdata=nd )
+ # replace with the revised curves
+ mxp <- match( max( nd$P[nd$P<p.end] ), nd$P )
+ pc[,2] <- ifelse( nd$P>p.end, pc[mxp,2] + pmax(0,(nd$P-p.end)*dfac), pc[,2] )
+ # return predicted rates as well as original and revised terms predictions
+ list( pr.resp  = exp( apply( pc, 1, sum ) + attr(pc,"const") + log(nd$Y) ),
+       pr.terms = pc,
+       pr.org   = po )
+    }
```

We can then illustrate the two variants for the terms

```
> zz <- predatt.apc( m.inc.apc, nd )
> ww <- predinc.apc( m.inc.apc, nd )
> str( zz )
List of 3
 $ pr.resp : Named num [1:633600] 5.50e-06 5.54e-06 5.57e-06 5.61e-06 5.64e-06 ...
  ..- attr(*, "names")= chr [1:633600] "1" "2" "3" "4" ...
 $ pr.terms: num [1:633600, 1:3] -3.28 -3.27 -3.27 -3.26 -3.25 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:633600] "1" "2" "3" "4" ...
  .. ..$ : chr [1:3] "Ns(A, knots = ki.a)" "Ns(P, knots = ki.p)" "Ns(P - A, kn = ki.c)"
  ..- attr(*, "constant")= num -6.09
 $ pr.org  : num [1:633600, 1:3] -3.28 -3.27 -3.27 -3.26 -3.25 ...
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:633600] "1" "2" "3" "4" ...
  .. ..$ : chr [1:3] "Ns(A, knots = ki.a)" "Ns(P, knots = ki.p)" "Ns(P - A, kn = ki.c)"
  ..- attr(*, "constant")= num -6.09
> str( nd )
```

```
'data.frame':        633600 obs. of  3 variables:
 $ A: num  0.0417 0.125 0.2083 0.2917 0.375 ...
 $ P: num  1996 1996 1996 1996 1996 ...
 $ Y: num  0.0833 0.0833 0.0833 0.0833 0.0833 ...
> par( mfrow=c(1,3) )
> oA <- order( nd$A )
> oP <- order( nd$P )
> oC <- order( nd$P-nd$A )
> matplot( nd$A[oA], cbind( zz$pr.org[oA,1], zz$pr.terms[oA,1] ),
+          type="l", lty=1, lwd=c(5,1), col=gray(c(6,1)/9) )
> matplot( nd$P[oP], cbind( zz$pr.org[oP,2], zz$pr.terms[oP,2], ww$pr.terms[oP,2] ),
+          type="l", lty=1, lwd=c(5,1,1), col=c(gray(c(6,1)/9),"red") )
> abline( v=2016 )
> matplot( (nd$P-nd$A)[oC], cbind( zz$pr.org[oC,3], zz$pr.terms[oC,3] ),
+          type="l", lty=1, lwd=c(5,1), col=gray(c(6,1)/9) )
> abline( v=2000 )
```

With this machinery in place we are now able to add the predicted (cumulative) rates to the `Lambda` array:

```
> Lambda[,,"M","att"] <- predatt.apc( m.inc.apc, nd, p.end=2016, c.end=2000 )$pr.resp
> Lambda[,,"F","att"] <- predatt.apc( f.inc.apc, nd, p.end=2016, c.end=2000 )$pr.resp
> Lambda[,,"M","i20"] <- predinc.apc( m.inc.apc, nd, dfac=0.020, p.end=2016, c.end=2000 )$p
> Lambda[,,"F","i20"] <- predinc.apc( f.inc.apc, nd, dfac=0.020, p.end=2016, c.end=2000 )$p
> Lambda[,,"M","i25"] <- predinc.apc( m.inc.apc, nd, dfac=0.025, p.end=2016, c.end=2000 )$p
> Lambda[,,"F","i25"] <- predinc.apc( f.inc.apc, nd, dfac=0.025, p.end=2016, c.end=2000 )$p
> Lambda[,,"M","i30"] <- predinc.apc( m.inc.apc, nd, dfac=0.030, p.end=2016, c.end=2000 )$p
> Lambda[,,"F","i30"] <- predinc.apc( f.inc.apc, nd, dfac=0.030, p.end=2016, c.end=2000 )$p
```

4.4 Mortality rates

4.4.1 Diabetes patients

First we fix the position of knots by age, period and cohort as we did for the incidence rates, and then we fit the same set of models, and make the same set of predictions, and put in a similarly defined array:

```
> ( kmd.a <- with( mDMdat, quantile( rep(A  ,D), qn(nk.a) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
56.33333 66.33333 71.66667 75.66667 79.66667 83.33333 86.66667 91.66667
> ( kmd.p <- with( mDMdat, quantile( rep(P  ,D), qn(nk.p) ) ) )
8.333333%       25% 41.66667% 58.33333%       75% 91.66667%
 1998.333  2002.333  2005.667  2009.333  2011.667  2014.667
> ( kmd.c <- with( mDMdat, quantile( rep(P-A,D), qn(nk.c) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1912.667 1919.333 1923.333 1927.333 1931.667 1936.333 1942.333 1951.667
> m.md.ap <- glm( D ~ Ns(A,knots=kmd.a) + Ns(P,knots=kmd.p),
+                         offset = log(Y),
+                         family = poisson,
+                           data = subset( mDMdat, sex=="M" ) )
> m.md.aPC <- update( m.md.ap, . ~ . - Ns(P  ,kn=kmd.p) + I(P) +
+                             detrend( Ns(P  ,kn=kmd.p), P  , D ) +
+                             detrend( Ns(P-A,kn=kmd.c), P-A, D ) )
```

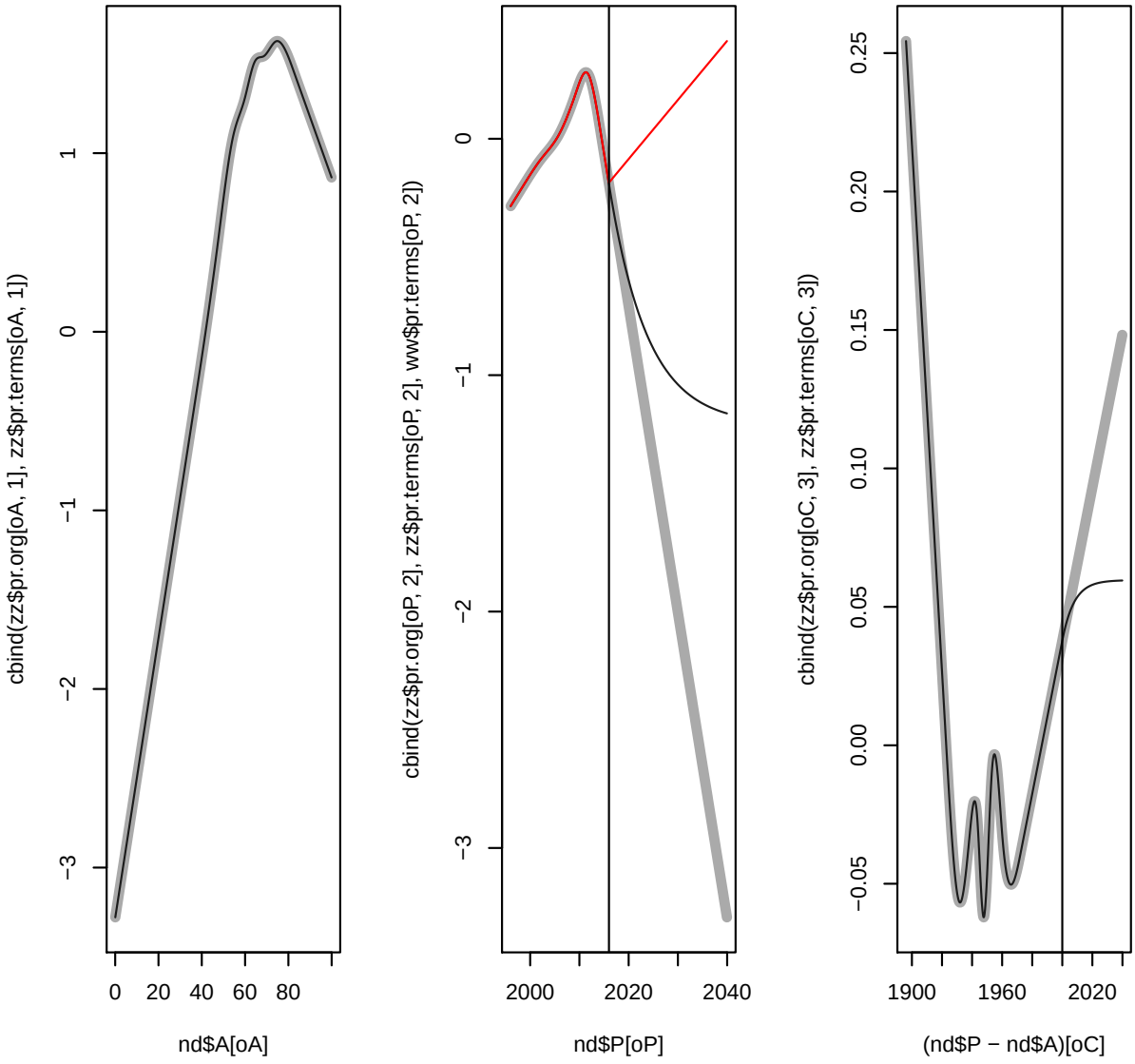

Figure 4.3: *Demonstration of the attenuation of the period and cohort terms from an APC-model. The attenuation effect used is the magic number 0.88 per year; that is, the slope of the effects decreases by a factor 0.88 per year, corresponding to a 47% decrease at 5 years and 72% decrease at 10 years in the slope of the effects.*
The red line is an annual increase i perido effect of 2.5% intended as a sensitivity analysis tool.

```
> m.md.apc <- update( m.md.ap, . ~ . + Ns(P-A,kn=kmd.c) )
> f.md.ap  <- update( m.md.ap , data = subset( mDMdat, sex=="F" ) )
> f.md.apc <- update( m.md.apc, data = subset( mDMdat, sex=="F" ) )
> f.md.aPC <- update( m.md.aPC, data = subset( mDMdat, sex=="F" ) )
> Mu.DM[,,"M","ap" ] <- predict.glm( m.md.ap , type="response", newdata=nd )
> Mu.DM[,,"F","ap" ] <- predict.glm( f.md.ap , type="response", newdata=nd )
> Mu.DM[,,"M","apc"] <- predict.glm( m.md.apc, type="response", newdata=nd )
> Mu.DM[,,"F","apc"] <- predict.glm( f.md.apc, type="response", newdata=nd )
```

```
> Mu.DM[,,"M","att"] <- predatt.apc( m.md.apc, nd )$pr.resp
> Mu.DM[,,"F","att"] <- predatt.apc( f.md.apc, nd )$pr.resp
> Mu.DM[,,"M","i20"] <- predinc.apc( m.md.apc, nd, dfac=0.020 )$pr.resp
> Mu.DM[,,"F","i20"] <- predinc.apc( f.md.apc, nd, dfac=0.020 )$pr.resp
> Mu.DM[,,"M","i25"] <- predinc.apc( m.md.apc, nd, dfac=0.025 )$pr.resp
> Mu.DM[,,"F","i25"] <- predinc.apc( f.md.apc, nd, dfac=0.025 )$pr.resp
> Mu.DM[,,"M","i30"] <- predinc.apc( m.md.apc, nd, dfac=0.030 )$pr.resp
> Mu.DM[,,"F","i30"] <- predinc.apc( f.md.apc, nd, dfac=0.030 )$pr.resp
```

4.4.2 Persons without diabetes

The mortality in the population without diabetes is modeled in exactly the same way, except we put in knots early in age:

```
> ( kmw.a <- with( mnDdat, c( 5, 15,
+                          quantile( rep(A  ,D), qn(nk.a) ) ) ) )
                     6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
 5.00000 15.00000 50.66667 64.33333 71.66667 77.33333 81.33333 85.33333 88.66667 93.66667
> ( kmw.p <- with( mnDdat, quantile( rep(P  ,D), qn(nk.p) ) ) )
8.333333%       25% 41.66667% 58.33333%       75% 91.66667%
 1997.333  2000.667  2003.667  2006.667  2010.333  2014.333
> ( kmw.c <- with( mnDdat, quantile( rep(P-A,D), qn(nk.c) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1909.667 1915.667 1920.333 1924.333 1928.667 1934.333 1942.667 1955.667
> m.mw.ap <- glm( D ~ Ns(A,knots=kmw.a) + Ns(P,knots=kmw.p),
+                     offset = log(Y),
+                     family = poisson,
+                       data = subset( mnDdat, sex=="M" ) )
> m.mw.aPC <- update( m.mw.ap, . ~ . - Ns(P  ,kn=kmw.p) + I(P) +
+                             detrend( Ns(P  ,kn=kmw.p), P  , D ) +
+                             detrend( Ns(P-A,kn=kmw.c), P-A, D ) )
> m.mw.apc <- update( m.mw.ap, . ~ . + Ns(P-A,kn=kmw.c) )
> f.mw.ap  <- update( m.mw.ap , data = subset( mnDdat, sex=="F" ) )
> f.mw.apc <- update( m.mw.apc, data = subset( mnDdat, sex=="F" ) )
> f.mw.aPC <- update( m.mw.aPC, data = subset( mnDdat, sex=="F" ) )
> Mu.W[,,"M","ap" ] <- predict.glm( m.mw.ap , type="response", newdata=nd )
> Mu.W[,,"F","ap" ] <- predict.glm( f.mw.ap , type="response", newdata=nd )
> Mu.W[,,"M","apc"] <- predict.glm( m.mw.apc, type="response", newdata=nd )
> Mu.W[,,"F","apc"] <- predict.glm( f.mw.apc, type="response", newdata=nd )
> Mu.W[,,"M","att"] <- predatt.apc( m.mw.apc, nd )$pr.resp
> Mu.W[,,"F","att"] <- predatt.apc( f.mw.apc, nd )$pr.resp
> Mu.W[,,"M","i20"] <- predinc.apc( m.mw.apc, nd, dfac=0.020 )$pr.resp
> Mu.W[,,"F","i20"] <- predinc.apc( f.mw.apc, nd, dfac=0.020 )$pr.resp
> Mu.W[,,"M","i25"] <- predinc.apc( m.mw.apc, nd, dfac=0.025 )$pr.resp
> Mu.W[,,"F","i25"] <- predinc.apc( f.mw.apc, nd, dfac=0.025 )$pr.resp
> Mu.W[,,"M","i30"] <- predinc.apc( m.mw.apc, nd, dfac=0.030 )$pr.resp
> Mu.W[,,"F","i30"] <- predinc.apc( f.mw.apc, nd, dfac=0.030 )$pr.resp
```

4.5 Average trends

The average annual trends in all of the rates can now be summarized:

```
> mort.chg <- rbind( ci.exp(m.md.aPC,subset="I\\(P")-1,
+                    ci.exp(f.md.aPC,subset="I\\(P")-1,
+                    ci.exp(m.mw.aPC,subset="I\\(P")-1,
+                    ci.exp(f.mw.aPC,subset="I\\(P")-1 )*100
> rownames( mort.chg ) <- c("Mortality change,   DM: Men",
+                           "                        Women",
+                           "Mortality change, Well: Men",
+                           "                        Women")
> round( rbind( inc.chg, mort.chg ), 1 )
                            exp(Est.) 2.5% 97.5%
DM incidence change     Men       2.2  2.1   2.2
                        Women     2.6  2.5   2.7
Mortality change,   DM: Men      -3.7 -3.8  -3.6
                        Women    -3.4 -3.5  -3.2
Mortality change, Well: Men      -2.9 -2.9  -2.8
                        Women    -2.4 -2.5  -2.4
```

Thus it appears that the incidence rates of diabetes are increasing by some 2.2% per year for women but 3.5% per year for men, while mortality rates are decreasing 3.6% per your for persons with diabetes, but only 2.5–2.9% per year for perons without.

For convenience of calculations and for subsequent use, we save the estimated rates and other quantities of interest:

```
> save( Lambda, Mu.W, Mu.DM, a.pt, p.pt, t.pt, int, qn, file="../data/rateEsts.Rda" )
> load( file="../data/rateEsts.Rda" )
```

4.6 Trends in estimated rates

In order to show how the models predict the time trends in diabetes incidence and mortality, we make a graphical display of the estimated rates in ages 20,. . . ,90 versus calendar time:

A brief overview of the mortality and incidence rates over time:

```
> pts <- as.numeric( dimnames(Lambda)[[2]] )[1:240]
> ( dimnames(Lambda)[[1]][agr <- seq(240,1080,120)] )
[1] "19.9583333333333" "29.9583333333333" "39.9583333333333" "49.9583333333333"
[5] "59.9583333333333" "69.9583333333333" "79.9583333333333" "89.9583333333333"
> par( mfrow=c(3,2), mar=c(1,1,1,1), oma=c(3,3,0,0), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> yticks <- outer( 1:9, 10^(-2:1) )[1:30]
> rpl <- function( Lambda, sx, yl ){
+ matplot( pts, t(Lambda[agr,1:240,sx,"apc"])*1000,
+          log="y", yaxt="n", ylim=c(0.01,30), ylab="", xaxt="n",
+          lty=1, lwd=5, type="l",
+          col=gray(4:11/13) )#if(sx=="M") "blue" else "red" )
+ mtext( side=2, yl, line=2.5, las=0 )
+ }
> rpl( Lambda, "M", "DM incidence per 1000 PY" )
>                           axis( side=2 ) ; axis( side=2, at=yticks, labels=NA )
> text( 1996, 30, "Men", adj=c(0,1), cex=1.5 )
> rpl( Lambda, "F", "" )
> text( 1996, 30, "Women", adj=c(0,1), cex=1.5 )
> rpl(  Mu.W , "M", "Population mortality per 1000 PY" )
>                           axis( side=2 ) ; axis( side=2, at=yticks, labels=NA )
> rpl(  Mu.W , "F", "" )
```

```
> rpl(  Mu.DM, "M", "DM mortality per 1000 PY" )
>                          axis( side=2 ) ; axis( side=2, at=yticks, labels=NA )
>                          axis( side=1 ) ; axis( side=1, at=1996:2016, labels=NA )
> rpl(  Mu.DM, "F", "" ) ; axis( side=1 ) ; axis( side=1, at=1996:2016, labels=NA )
> mtext( "Date of follow-up", side=1, line=2, outer=TRUE )
```

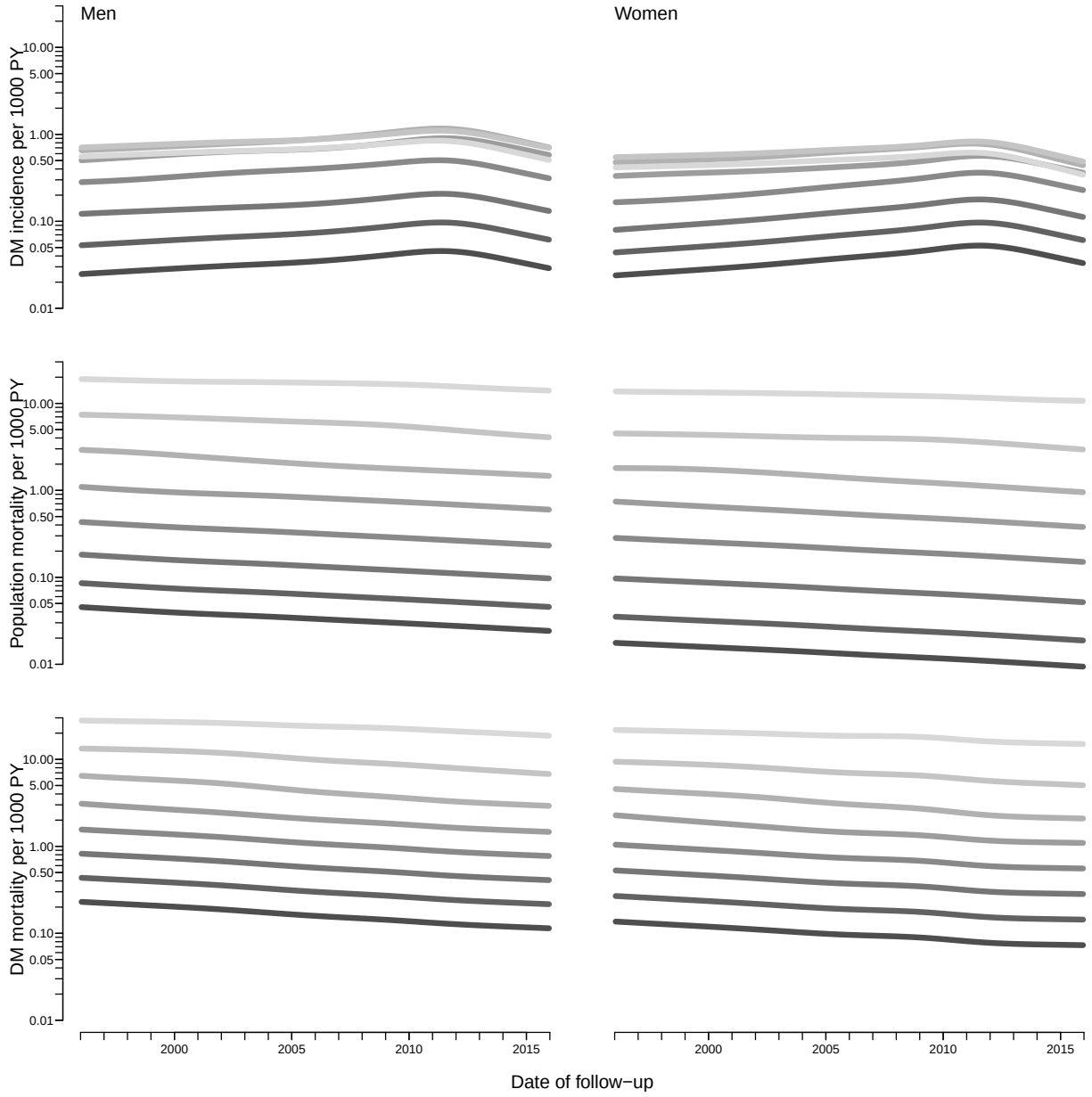

Figure 4.4: *Trends in incidence and mortality rates for ages 20 (darkest), 30, . . . , 90 (lightest), as estimated from the age-period-cohort models.*

4.7 Extrapolation of rates

It is possible to extrapolate the rates beyond the observed dates by simply extending the linear part of the natural splines; in fact this is already done in the rate-objects `Lambda`,

`Mu.W` and `Mu.DM`. However, as seen in figure 4.5 the predicted decline in diabetes rates is presumably way too dramatic.

```
> pts <- as.numeric( dimnames(Lambda)[[2]] )
> ( dimnames(Lambda)[[1]][agr <- seq(240,1080,120)] )
[1] "19.9583333333333" "29.9583333333333" "39.9583333333333" "49.9583333333333"
[5] "59.9583333333333" "69.9583333333333" "79.9583333333333" "89.9583333333333"
> par( mfrow=c(3,2), mar=c(1,1,1,1), oma=c(3,3,0,0), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> yticks <- outer( 1:9, 10^(-2:1) )[1:30]
> rpl <- function( Lambda, sx, yl, inc=FALSE ){
+ matplot( pts, t(Lambda[agr,,sx,"apc"])*1000,
+          log="y", yaxt="n", ylim=c(0.01,30), ylab="", xaxt="n",
+          lty=1, lwd=5, type="l", col=gray(4:11/13) )
+ matlines( pts, t(Lambda[agr,,sx,"att"])*1000,
+           lty=1, lwd=1, type="l", col="black" )
+ if( inc )
+ matlines( pts, t(Lambda[agr,,sx,"i25"])*1000,
+           lty=1, lwd=1, type="l", col="red" )
+ abline( v=2016, lty=3, col=gray(0.6) )
+ mtext( side=2, yl, line=2.5, las=0 )
+ }
> rpl( Lambda, "M", "DM incidence per 1000 PY", inc=TRUE )
>                          axis( side=2 ) ; axis( side=2, at=yticks, labels=NA, tcl=-0.3 )
> text( 1996, 30, "Men", adj=c(0,1), cex=1.5 )
> rpl( Lambda, "F", "", inc=TRUE )
> text( 1996, 30, "Women", adj=c(0,1), cex=1.5 )
> rpl(  Mu.W , "M", "Population mortality per 1000 PY" )
>                          axis( side=2 ) ; axis( side=2, at=yticks, labels=NA, tcl=-0.3 )
> rpl(  Mu.W , "F", "" )
> rpl(  Mu.DM, "M", "DM mortality per 1000 PY" )
>                          axis( side=2 ) ; axis( side=2, at=yticks, labels=NA, tcl=-0.3 )
>                          axis( side=1 ) ; axis( side=1, at=1996:2030, labels=NA, tcl=-0.3
> rpl(  Mu.DM, "F", "" ) ; axis( side=1 ) ; axis( side=1, at=1996:2030, labels=NA, tcl=-0.3
> mtext( "Date of follow-up", side=1, line=2, outer=TRUE )
```

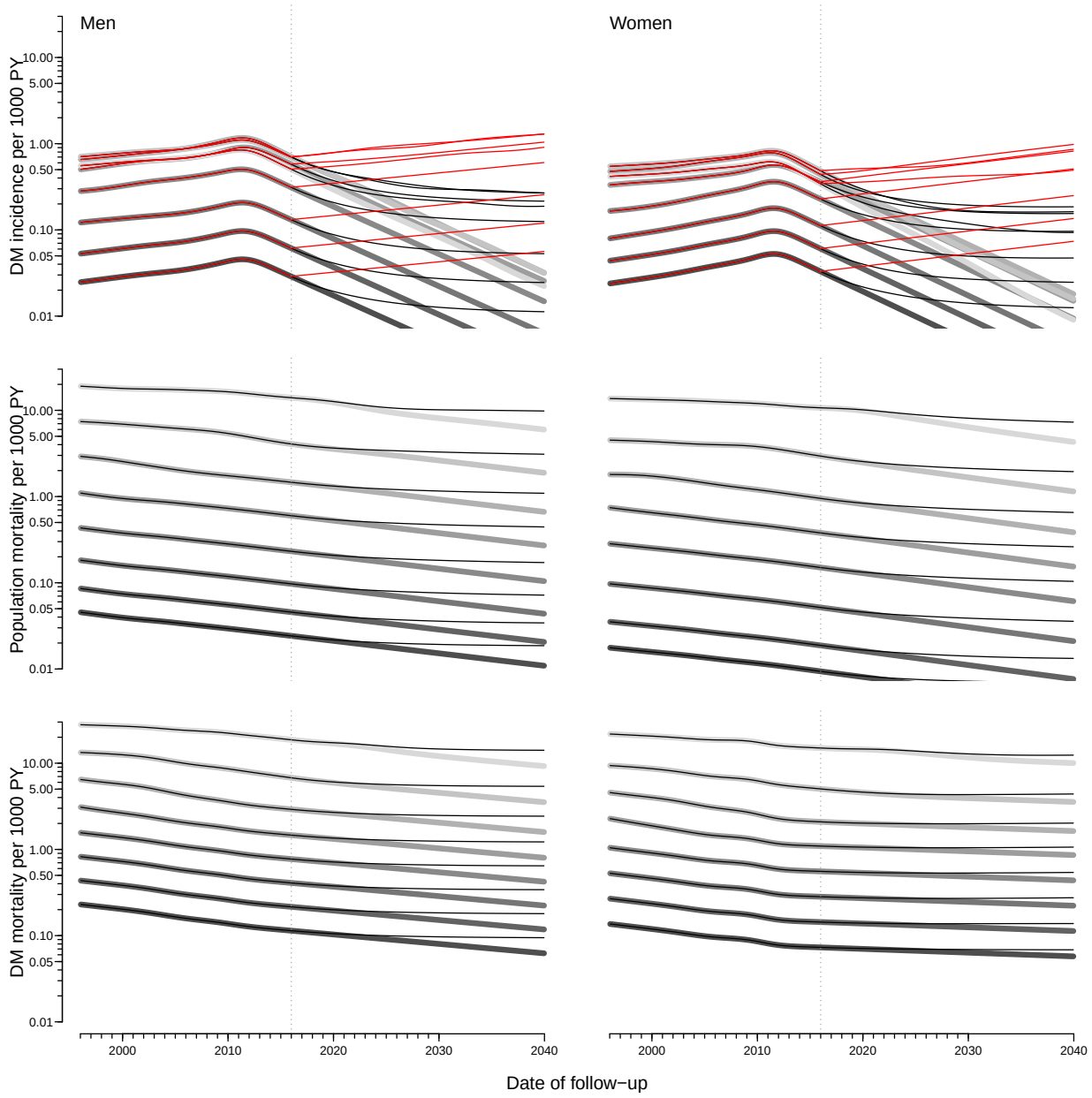

Figure 4.5: *Trends in incidence and mortality rates for ages 20 (darkest), 30, . . . , 90 (lightest), as estimated from the age-period-cohort models and predicted by naive extrapolation of the natural splines. The vertical dotted lines indicate the end of available data, and the thin overlaid lines represent the rate predictions based on attenuated period- and cohort-effects. The red lines are predictions assuming an annual increase in period effects of 2.5%/year — only used for the incidence rates.*

Chapter 5

Components of prevalence

The purpose of this chapter is to use the estimated transition rates to predict the prevalences at later times. This is in itself not an interesting endeavour, 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 an interval length as chosen.

```
> load( file="../data/rateEsts.Rda")
> lls()
  name   mode     class    size
1 a.pt   numeric  numeric  1200
2 int    numeric  numeric  1
3 Lambda numeric  array    1200 528 2 6
4 Mu.DM  numeric  array    1200 528 2 6
5 Mu.W   numeric  array    1200 528 2 6
6 p.pt   numeric  numeric  528
7 qn     function function 1
8 t.pt   numeric  numeric  529
```

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

`pr.fit` — array of empirical age-specific prevalences at 1.1.1996–1.1.2016, smoothed by natural splines.

`TR` — array of transition probabilities between states Well and DM and Death. Transition probabilities are computed under the 4 different scenarios combining mortality and incidence rates either as they actually developed 1996–2016 or assuming they were constant at the 1996 level. These refer to intervals of length `int` and are therefore labeled on the period dimension by the midpoint of these, a total of 20/`int`.

`prv` — array of predicted prevalences based on the initial prevalences at 1.1.1996 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 1996. Likewise, the period dimension is expanded by one relative to that in `TR` (replacing `p.pt` (period points)

with `t.pt` (time points)), since this refer to points in time and not time intervals (periods).

`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.

5.1 Transition probabilities

In order to get the predicted *number* of persons 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.

Note that the structures `Lambda`, `Mu.W` and `Mu.DM`, contain predicted rates up to 2040 under different scenarios, that we do not need:

```
> str( Lambda )
 num [1:1200, 1:528, 1:2, 1:6] 5.62e-06 5.65e-06 5.69e-06 5.72e-06 5.76e-06 ...
 - attr(*, "dimnames")=List of 4
  ..$ a  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667"
  ..$ p  : chr [1:528] "1996.04166666667" "1996.125" "1996.20833333333" "1996.29166666667"
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:6] "ap" "apc" "att" "i20" ...
```

. . . we only need the dates till 2016, and the models `ap` and `apc`:

```
> # 2nd dimension of rates is all the way to 2040, only need to 2016
> dimnames(Lambda)[[2]][250+1:4]
[1] "2016.875"         "2016.95833333333" "2017.04166666667" "2017.125"
> Lambda <- Lambda[,1:252,,1:2]
> Mu.W   <- Mu.W  [,1:252,,1:2]
> Mu.DM  <- Mu.DM [,1:252,,1:2]
> states <- c("Well","DM")
> TR <- NArray( c( dimnames(Lambda),
+                  list( from = states,
+                          to = states,
+                       scene = c("obs","m-fix","i-fix","all-f" ) ) ) )
> str( TR )
 logi [1:1200, 1:252, 1:2, 1:2, 1:2, 1:2, 1:4] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 7
  ..$ a    : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667"
  ..$ p    : chr [1:252] "1996.04166666667" "1996.125" "1996.20833333333" "1996.29166666667"
  ..$ sex  : chr [1:2] "M" "F"
  ..$ mod  : chr [1:2] "ap" "apc"
  ..$ from : chr [1:2] "Well" "DM"
  ..$ to   : chr [1:2] "Well" "DM"
  ..$ scene: chr [1:4] "obs" "m-fix" "i-fix" "all-f"
```

The situation where both the mortality rates and incidence rates are fixed at the 1996 level is trivial, because transition probabilities in that case only depend on age and not on period.

In order to fill `TR`, we need the cumulative incidences over intervals of length `int`. But these were exactly the ones we predicted in the previous sections by setting the person years equal to `int` in the data frame supplied to the `newdata` argument.

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 0.083 year):

```
> TR[,,,,"Well","Well","obs"] <- exp(-Lambda-Mu.W)
> TR[,,,,"Well","DM"  ,"obs"] <-      Lambda
> TR[,,,,"DM"  ,"Well","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; we only compute how many of the persons alive that end up being alive at the next timepoint

When we fix the mortality or incidence at the 1996 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) for either incidence, mortality or both:

```
> ( np <- dim(Lambda)[2] )
  p
252
```

```
> TR[,,,,"Well","Well","m-fix"] <- exp(-Lambda-Mu.W[,rep(1,np),,])
> TR[,,,,"Well","DM"  ,"m-fix"] <-      Lambda
> TR[,,,,"DM"  ,"Well","m-fix"] <- 0
> TR[,,,,"DM"  ,"DM"  ,"m-fix"] <- exp(      -Mu.DM[,rep(1,np),,])
```

```
> TR[,,,,"Well","Well","i-fix"] <- exp(-Lambda[,rep(1,np),,]-Mu.W)
> TR[,,,,"Well","DM"  ,"i-fix"] <-      Lambda[,rep(1,np),,]
> TR[,,,,"DM"  ,"Well","i-fix"] <- 0
> TR[,,,,"DM"  ,"DM"  ,"i-fix"] <- exp(-Mu.DM)
```

```
> TR[,,,,"Well","Well","all-f"] <- exp(-Lambda[,rep(1,np),,]-Mu.W[,rep(1,np),,])
> TR[,,,,"Well","DM"  ,"all-f"] <-      Lambda[,rep(1,np),,]
> TR[,,,,"DM"  ,"Well","all-f"] <- 0
> TR[,,,,"DM"  ,"DM"  ,"all-f"] <- exp( -Mu.DM[,rep(1,np),,])
```

We have now collected the transition probabilities between "Well" and "DM" as well as the probabilities of remaining in each of these, all referring to a duration of `int`, a total of about 30 mil. numbers:

```
> str( TR )
 num [1:1200, 1:252, 1:2, 1:2, 1:2, 1:2, 1:4] 1 1 1 1 1 ...
 - attr(*, "dimnames")=List of 7
  ..$ a    : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.2916666666666
  ..$ p    : chr [1:252] "1996.04166666667" "1996.125" "1996.20833333333" "1996.29166666667
  ..$ sex  : chr [1:2] "M" "F"
  ..$ mod  : chr [1:2] "ap" "apc"
  ..$ from : chr [1:2] "Well" "DM"
  ..$ to   : chr [1:2] "Well" "DM"
  ..$ scene: chr [1:4] "obs" "m-fix" "i-fix" "all-f"
```

```
> prod( dim(TR) )
```

```
[1] 19353600
```

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

5.2 Prediction of the observed prevalences

Note that we do not need to predict the population size; we can get away with only predicting the prevalences as fractions. When we multiply the fraction of persons in states (Well,DM) with the transition matrix, we get fraction of the persons in the previous state that are in states (Well,DM), which does not sum to 1 (because of the ones dying), so we must rescale to prevalence age in each step.

First we set up an array to hold the predicted prevalences under different scenarios. 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 1996, 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 1.1.1996).

```
> dpr <- c( dimnames(Lambda)[1:4],
+           list( c(dimnames(TR)[["scene"]],"mort","inc","const") ) )
> names( dpr )[c(2,5)] <- c("t","what")
> t.pt[250+1:4]
[1] 2016.833 2016.917 2017.000 2017.083

> dpr[["t"]] <- t.pt[1:253]
> prv <- NArray(dpr)
> str( prv )
 logi [1:1200, 1:253, 1:2, 1:2, 1:7] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.29166666666666
  ..$ t   : chr [1:253] "1996" "1996.08333333333" "1996.16666666667" "1996.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...

> prod( dim(prv) )
[1] 8500800
```

Then we make a loop that updates the prevalences at 1.1.1996 to those at subsequent times, but first we must initialize the prevalences as modeled at 1996-01-01, as well as the prevalences at ages 0 (the new-born) that we set to 0; note that we exploit the column major storage of arrays to get `prv` filled with identical values across the two last dimensions (model and scenario).

```
> load( file="../data/prFit.Rda" )
> str( pr.fit ); str( prv )
 num [1:2, 1:1200, 1:21] 0.000626 0.000579 0.00063 0.000583 0.000635 ...
 - attr(*, "dimnames")=List of 3
  ..$ sex: chr [1:2] "M" "F"
  ..$ A  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667
  ..$ P  : chr [1:21] "1996" "1997" "1998" "1999" ...
 logi [1:1200, 1:253, 1:2, 1:2, 1:7] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.29166666666666
  ..$ t   : chr [1:253] "1996" "1996.08333333333" "1996.16666666667" "1996.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

```
> # Smoothed prevalences at 1.1.1996 - the starting values
> for( sx in c("M","F") ) prv[,1,sx,,] <- pr.fit[sx,,1]
> # Prevalences at age 0
> prv[1,,"M",,] <- 0
> prv[1,,"F",,] <- 0
> tt <- prv[c(1:2,floor(dim(prv)[1]/1.5)+1:2),1:2,,,]
> dimnames( tt )[[1]] <- round( as.numeric(dimnames( tt )[[1]]), 3 )
> dimnames( tt )[[2]] <- round( as.numeric(dimnames( tt )[[2]]), 3 )
> round( ftable( tt, row.vars=4:1 )*100, 1 )
                        what obs m-fix i-fix all-f mort inc const
mod sex t        a
ap  M   1996     0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125       0.1   0.1   0.1   0.1  0.1 0.1   0.1
                 66.708      5.0   5.0   5.0   5.0  5.0 5.0   5.0
                 66.792      5.0   5.0   5.0   5.0  5.0 5.0   5.0
         1996.083 0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125        NA    NA    NA    NA   NA  NA    NA
                 66.708       NA    NA    NA    NA   NA  NA    NA
                 66.792       NA    NA    NA    NA   NA  NA    NA
    F   1996     0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125       0.1   0.1   0.1   0.1  0.1 0.1   0.1
                 66.708      3.7   3.7   3.7   3.7  3.7 3.7   3.7
                 66.792      3.7   3.7   3.7   3.7  3.7 3.7   3.7
         1996.083 0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125        NA    NA    NA    NA   NA  NA    NA
                 66.708       NA    NA    NA    NA   NA  NA    NA
                 66.792       NA    NA    NA    NA   NA  NA    NA
apc M   1996     0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125       0.1   0.1   0.1   0.1  0.1 0.1   0.1
                 66.708      5.0   5.0   5.0   5.0  5.0 5.0   5.0
                 66.792      5.0   5.0   5.0   5.0  5.0 5.0   5.0
         1996.083 0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125        NA    NA    NA    NA   NA  NA    NA
                 66.708       NA    NA    NA    NA   NA  NA    NA
                 66.792       NA    NA    NA    NA   NA  NA    NA
    F   1996     0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125       0.1   0.1   0.1   0.1  0.1 0.1   0.1
                 66.708      3.7   3.7   3.7   3.7  3.7 3.7   3.7
                 66.792      3.7   3.7   3.7   3.7  3.7 3.7   3.7
         1996.083 0.042       0.0   0.0   0.0   0.0  0.0 0.0   0.0
                 0.125        NA    NA    NA    NA   NA  NA    NA
                 66.708       NA    NA    NA    NA   NA  NA    NA
                 66.792       NA    NA    NA    NA   NA  NA    NA
```

So now we have checked that we have put initial values correctly into `prv`, basically at the period edge (at 1996) and the age edge (at 0). Then we can compute the predicted prevalences under the different scenarios. We take the fraction of the population in age class `ia` at time `ip` that end up as diabetes patients at time `ip+1` (and hence in age class `ia+1`), and divide by the fraction of all that remain alive, which is the diabetes patients (survivors and new), *plus* those who survive free of diabetes:

```
> system.time(
+ for( ip in 1:(dim(prv)[2]-1) )
+ for( ia in 1:(dim(prv)[1]-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,,,"Well","DM"  ,] ) /
```

```
+   (    prv[ia,ip,,1:4]  * TR[ia,ip,,"DM"  ,"DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"Well","DM"  ,]
+    +(1-prv[ia,ip,,1:4]) * TR[ia,ip,,"Well","Well",] )
+            )
   user  system elapsed
  12.53    0.02   12.55
```

Note that the reason that the last dimension, `scene`, is explicitly mentioned in the array `prv` is because the length of this dimension is 7, but the corresponding in `TR` only 4 — recall 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 1996.

Moreover, note that the code above is particularly simple because we only need to compute the prevalence at the next date and age. If we had had a more elaborate model with, say complications states, the calculations in the loop would have been a matrix-multiplication updating the state-distribution, but this simplification would have been at the expense of another three loop-levels, namely over the the three last dimensions of the `prv` array — the matrix machinery only operates on 2-dimensonal structures, matrices. (Well, maybe some `apply` construction could be set up).

We can then show a few of the predicted prevalences in (

```
> round( prv[1:4,1:4,1,1,1,drop=F]*100, 3 )
, , sex = M, mod = ap, what = obs

                    t
a                    1996 1996.08333333333 1996.16666666667 1996.25
  0.0416666666666667 0.000            0.000             0.000   0.000
  0.125              0.063            0.001             0.001   0.001
  0.208333333333333  0.063            0.064             0.001   0.001
  0.291666666666667  0.064            0.064             0.064   0.002
```

```
> save( a.pt, prv, file="../data/prv.Rda" )
> load(            file="../data/prv.Rda" )
```

5.2.1 Checking the prediction

With this initial prediction in place we can now check whether we have made a reasonable approximation to the observed prevalences at 1.1.2016.

In the array `prv` are all the prevalences as predicted from the prevalence in 1996 using the estimated incidences and mortalities; predicted at intervals of `inc` whereas we have the smoothed empirical prevalences in 1996, 2001, 2006, 2011 and 2016 in the array `pr.fit`:

Thus we have the predicted age-specific prevalences for men in say 2000 in `prv[,"2000","M","apc","obs"]`, and the smoothed empirical in `pr.fit[,"M","2000"]`. We now plot these in the same plot:

```
> nd <- c( grep("2001",dimnames(prv)[[2]])[1],
+          grep("2006",dimnames(prv)[[2]])[1],
+          grep("2011",dimnames(prv)[[2]])[1],
+          grep("2016",dimnames(prv)[[2]])[1] )
> ( wh <- dimnames(prv)[[2]][nd] )
[1] "2001" "2006" "2011" "2016"
```

```
> 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, pr.fit["M",,wh]*100,
+          xlim=c(10,95), ylim=c(0,18), yaxs="i",
+          xlab="Age", ylab="Prevalence (%)",
+          type="l", col="blue", lty=1, lwd=2 )
> axis( side=1, at=1:9*10, labels=NA )
> matlines( a.pt, prv[,wh,"M","apc","obs"]*100,
+          type="l", col="blue", lty="32", lwd=4 )
> matplot( a.pt, pr.fit["F",,wh]*100,
+          xlim=c(10,95), ylim=c(0,18), yaxs="i",
+          xlab="Age", ylab="", yaxt="n",
+          type="l", col="red", lty=1, lwd=2 )
> matlines( a.pt, prv[,wh,"F","apc","obs"]*100,
+           type="l", col="red", lty="32", lwd=3 )
> axis( side=1, at=1:9*10, labels=NA )
> mtext( "Prevalence of DM (%)", side=2, line=2, las=0, outer=TRUE )
> mtext( "Age", side=1, line=2, las=0, outer=TRUE )


> 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[,wh,"M","apc","obs"],
+                      prv[,wh,"M","ap" ,"obs"],
+                      pr.fit["M",,wh])*100,
+          xlim=c(10,95), ylim=c(0,18), yaxs="i",
+          xlab="Age", ylab="Prevalence (%)",
+          type="l", col="black", lty=rep(c(0,1),c(8,4)), lwd=2 )
> matlines( a.pt, cbind(prv[,wh,"M","ap" ,"obs"],
+                       prv[,wh,"M","apc","obs"])*100,
+           type="l", col="blue", lty=rep(c("12","42"),c(4,4)), lwd=4 )
> axis( side=1, at=1:9*10, labels=NA )
> matplot( a.pt, cbind(prv[,wh,"F","apc","obs"],
+                      prv[,wh,"F","ap" ,"obs"],
+                      pr.fit["F",,wh])*100,
+          xlim=c(10,95), ylim=c(0,18), yaxs="i",
+          xlab="Age", ylab="", yaxt="n",
+          type="l", col="black", lty=rep(c(0,1),c(8,4)), lwd=2 )
> matlines( a.pt, cbind(prv[,wh,"F","ap" ,"obs"],
+                       prv[,wh,"F","apc","obs"])*100,
+           type="l", col="red", lty=rep(c("12","42"),c(4,4)), lwd=4 )
> axis( side=1, at=1:9*10, labels=NA )
> mtext( "Prevalence of DM (%)", side=2, line=2, las=0, outer=TRUE )
> mtext( "Age", side=1, line=2, las=0, outer=TRUE )
```

Since the APC-models for rates clearly provide a better fit (figure 5.1), we shall use these in the reporting of the different scenarios.

We now compare the predicted prevalences under the four scenarios at 1.1.2016:

```
> str( prv )
 num [1:1200, 1:253, 1:2, 1:2, 1:7] 0 0.00063 0.000635 0.000639 0.000644 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.29166666666666
  ..$ t   : chr [1:253] "1996" "1996.08333333333" "1996.16666666667" "1996.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
> dimnames(prv)[["t"]][np <- 241]
```

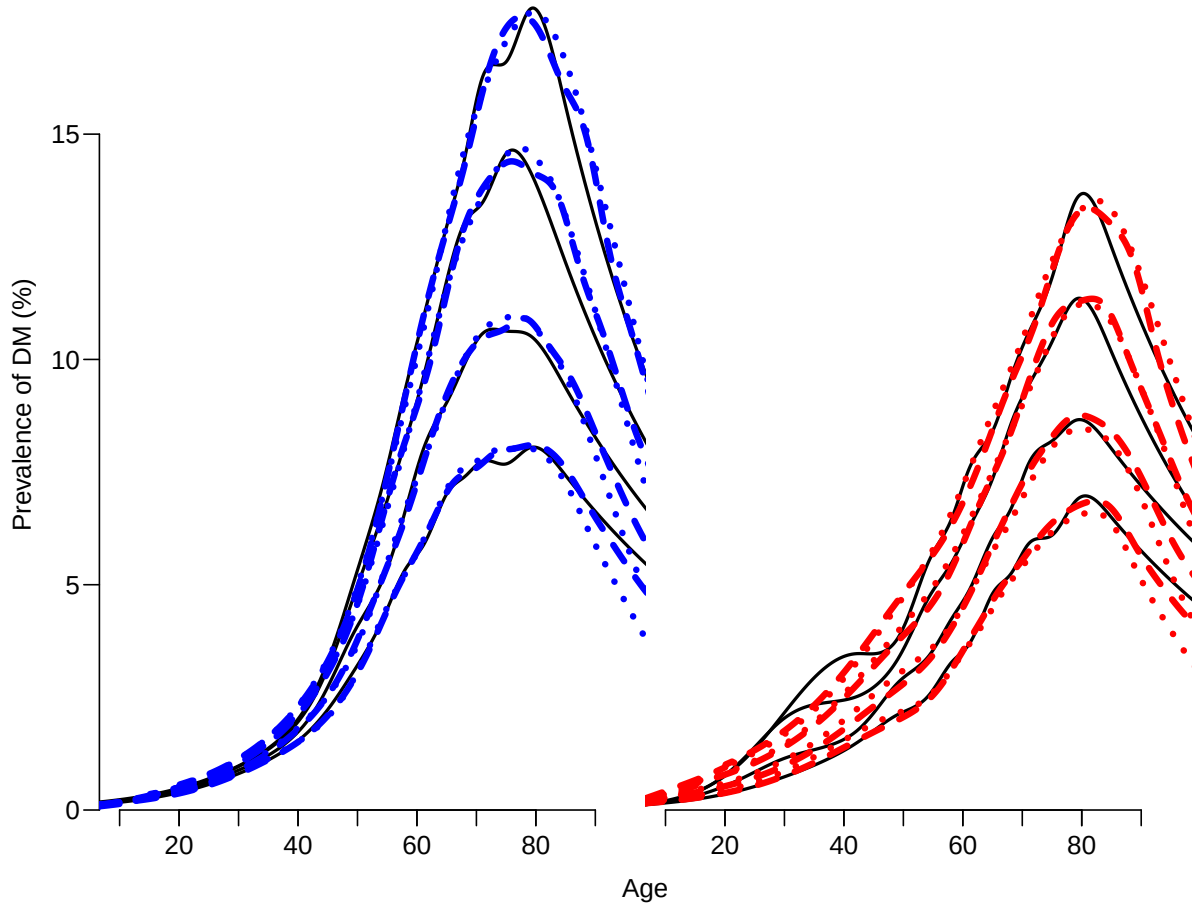

Figure 5.1: *Plot of observed (full lines) and predicted prevalences in 2001, using simple age-period-models (dotted lines) or age-period-cohort models (broken lines). Clearly the broken lines gives the better approximation to the smoothed empirical rates (black lines).*

```
[1] "2016"
> prv[floor(dim(prv)[1]/1.5)+1:5,np,"M","apc",]*100
                  what
a                      obs    m-fix    i-fix    all-f mort inc const
  66.7083333333333 13.74514 12.52098 10.45399 9.454924   NA  NA    NA
  66.7916666666667 13.78295 12.54914 10.49134 9.483485   NA  NA    NA
  66.875           13.82081 12.57731 10.52854 9.511871   NA  NA    NA
  66.9583333333333 13.85873 12.60549 10.56559 9.540077   NA  NA    NA
  67.0416666666667 13.89672 12.63369 10.60249 9.568100   NA  NA    NA
> 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" )
> lpl <- function( sx, cl )
+ matplot( a.pt, cbind(prv[,np,sx,"apc",c("obs","m-fix","i-fix","all-f")],
+                      prv[, 1,sx,"apc",1  ])*100,
+          xlim=c(20,90), ylim=c(0,18), yaxs="i", yaxt="n",
+          xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=c(1,2,1,2,1),
+          col=cl,   lwd=c(5,5,3,3,1) )
> lpl( "M", "blue" )
> axis( side=2 )
> axis( side=2, at=0:18, labels=NA, tcl=-0.3)
```

```
> lpl( "F", "red" )
> dimnames(prv)[[5]]
[1] "obs"   "m-fix" "i-fix" "all-f" "mort"  "inc"   "const"
```

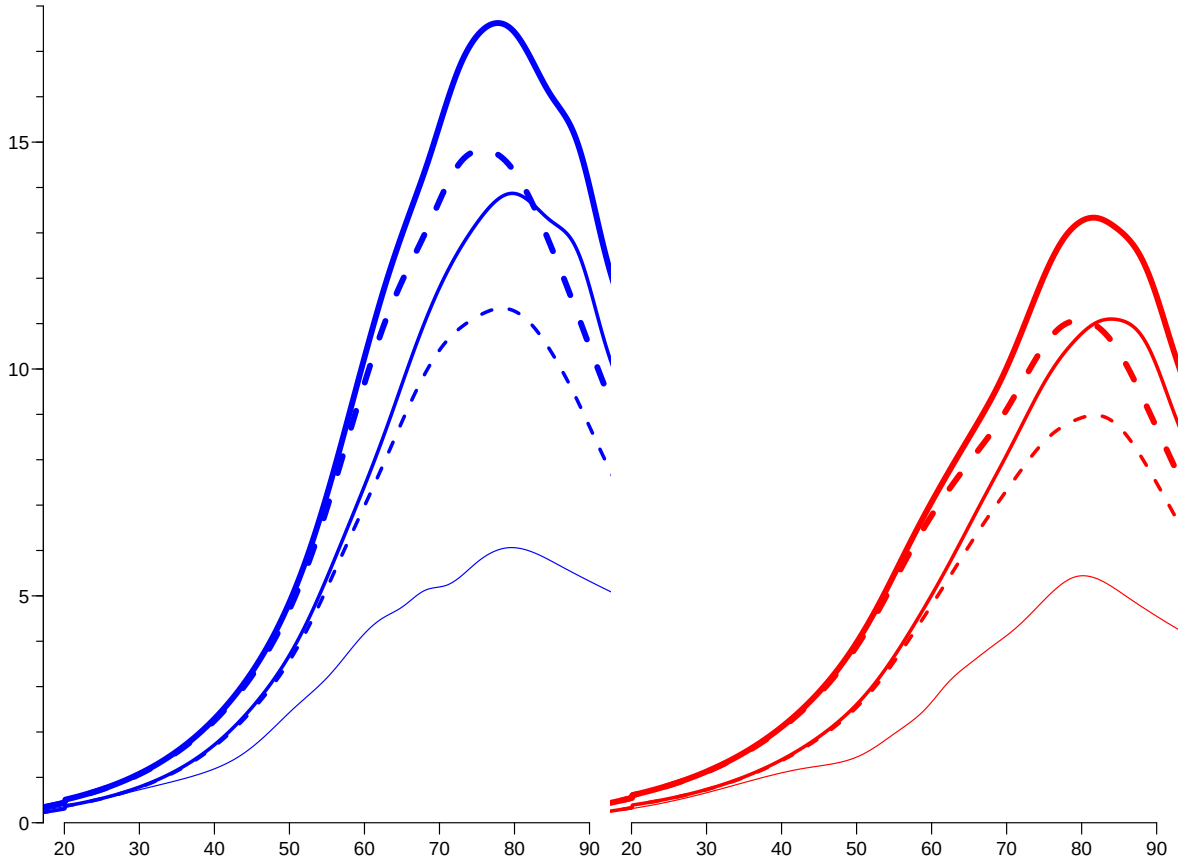

Figure 5.2: *The predicted prevalences under different scenarios:*
Full lines: Mortality rates evolve as observed.
Broken lines: Mortality rates remain as 1996.
Thick lines: Incidence rates evolve as observed.
Thin lines: Incidence rates remain as in 1996.
The very thin lines lowest in the two displays are the observed age-specific prevalences in 1996.

```
> scen <- c("Mort changes, Inc changes",
+           "Mort 1996, Inc changes",
+           "Mort changes, Inc 1996",
+           "Mort 1996, Inc 1996")
> c.a <- dimnames(prv)[[1]][floor(dim(prv)[1]/1.5)]
> n.a <- as.numeric( c.a )
> hts <- prv[c.a,np,"M","apc",1:4]*100
> cau.exp <-
+ function( wh=1:4, fill=FALSE )
+ {
+ pdf( paste( "comp-DMpr-", paste(wh,collapse=""), if( fill ) "F",
+             ".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[,np,"M","apc",],prv[,1,"M","apc",1])*100,
+           xlim=c(20,90), ylim=c(0,18), xlab="Age",
+           ylab="Prevalence (%)", yaxs="i",
+           type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
+ axis( side=2, at=0:18, labels=NA, tcl=-0.3 )
+ mtext( "Prevalence (%)", side=2, line=2.5, outer=T, las=0 )
+ matlines( a.pt, prv[,np,"M","apc",]*100,
+           type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="blue" )
+ matlines( a.pt, prv[,1,"M","apc",]*100,
+            type="l", lty=1, lwd=1, col="blue" )
+ 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[,np,"M","apc",wh[1]],
+                   rev(prv[,np,"M","apc",wh[2]]))*100,
+                     col=rgb(0,0,1,0.3), border="transparent" )
+ matplot( a.pt, cbind(prv[,np,"F","apc",],prv[,1,"F","apc",1])*100,
+           xlim=c(20,90), ylim=c(0,18), xlab="Age", yaxt="n", yaxs="i",
+           type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
+ matlines( a.pt, prv[,np,"F","apc",]*100,
+           type="l", lty=rep(c("11","22"),2), lwd=c(4,4,2,2)+1, col="red" )
+ matlines( a.pt, prv[,1,"F","apc",]*100, type="l", lty=1, lwd=1, col="red" )
+ if( fill ) polygon( c(a.pt,rev(a.pt)),
+                     c(prv[,np,"F","apc",wh[1]],
+                   rev(prv[,np,"F","apc",wh[2]]))*100,
+                     col=rgb(1,0,0,0.3), border="transparent" )
+ dev.off()
+ }
> cau.exp(1:4)

pdf
  2

> 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 5.2 shows the predicted prevalences under 4 different scenarios compared to the observed prevalences as of 1.1.1996.

5.3 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 )[5]
$what
[1] "obs"   "m-fix" "i-fix" "all-f" "mort"  "inc"   "const"
> par( mfrow=c(1,2), mar=c(3,0,0,0), oma=c(0,4,1,1),
+      mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matplot( a.pt, cbind( prv[,np,"M","apc","obs"  ]-
+                       prv[,np,"M","apc","m-fix"],
+                       prv[,np,"M","apc","i-fix"]-
+                       prv[,np,"M","apc","all-f"] )*100,
+          xlim=c(20,90), ylim=c(0,5), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=1, lwd=c(4,2)+1, col="blue", yaxs="i" )
> axis( side=2, at=0:14/2, labels=NA, tcl=-0.3 )
> mtext( "Prevalence difference (%)", side=2, line=2.5, outer=T, las=0 )
> matlines(a.pt, cbind( prv[,np,"M","apc","obs"  ]-
+                       prv[,np,"M","apc","i-fix"],
+                       prv[,np,"M","apc","m-fix"]-
+                       prv[,np,"M","apc","all-f"] )*100,
+          type="l", lty="22", lwd=c(4,2)+1, col="blue" )
> matplot( a.pt, cbind( prv[,np,"F","apc","obs"  ]-
+                       prv[,np,"F","apc","m-fix"],
+                       prv[,np,"F","apc","i-fix"]-
+                       prv[,np,"F","apc","all-f"] )*100,
+          xlim=c(20,90), ylim=c(0,5), xlab="Age", yaxt="n", yaxs="i",
+          type="l", lty=1, lwd=c(4,2)+1, col="red" )
> matlines(a.pt, cbind( prv[,np,"F","apc","obs"  ]-
+                       prv[,np,"F","apc","i-fix"],
+                       prv[,np,"F","apc","m-fix"]-
+                       prv[,np,"F","apc","all-f"] )*100,
+          type="l", lty="22", lwd=c(4,2)+1, col="red" )
```

From figure 5.3 we see that the two different ways of computing the contribution give pretty much the same results — the differences never exceed some 0.5%. Therefore, if we want to attribute fractions of the prevalence in 2010 to decreasing mortality and increasing incidence respectively, 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 1996 ("`all-f`"). This is obtained by taking the average of the two curves in each scenario.

The thin lines at the bottom of figure 5.2 represent the prevalence at 1.1.1996, so it is pretty clear that the incidence an mortality rates as observed by 1996 did not provide for at steady state.

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

1. the "inherited" prevalences from 1996.
2. the prevalence attributable to rates of mortality and incidence as of 1996 — the "epidemiological disequilibrium" as of 1996.
3. the prevalence attributable to the *increase* in the incidence rates.
4. the prevalence attributable to the *decrease* in the mortality rates.

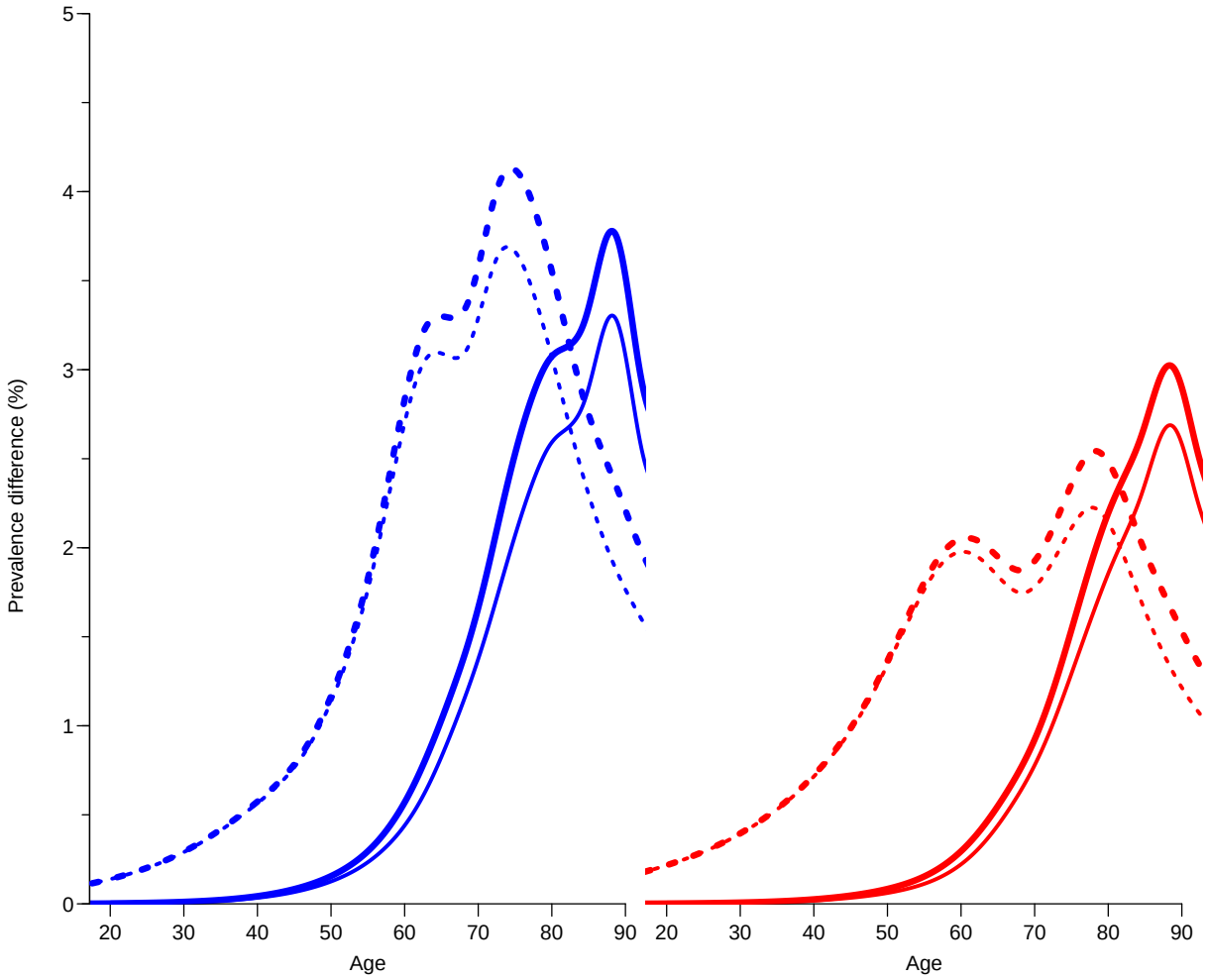

Figure 5.3: *Suggested contributions to age-specific prevalences from increasing mortalities over the period 1996–2016; the thick lines are obtained by subtracting the prediction based on fixing one rate from the one using the observed rates; thin lines based on subtracting the prediction based on fixing both rates from that where only 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 we now fill out the remaining 3 dimensions of `prv`:

```
> prv[,,,,"mort" ] <- ( prv[,,,,"obs"  ]-prv[,,,,"m-fix"] +
+                       prv[,,,,"i-fix"]-prv[,,,,"all-f"] ) / 2
> prv[,,,,"inc"  ] <- ( prv[,,,,"obs"  ]-prv[,,,,"i-fix"] +
+                       prv[,,,,"m-fix"]-prv[,,,,"all-f"] ) / 2
> prv[,,,,"const"] <-   prv[,,,,"all-f"]-prv[,rep(1,dim(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 15 dates 1996,...,2010.

First we define a function to make the component plots, and then we can plot the resulting development for men and women, for convenience we also put the latter in a function.

```
> 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=col[i-1])
+ 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 )
+ }
> one.comp <-
+ function( sex, clr )
+ {
+ par( mfrow=c(3,6), mar=c(0,0,0,0), oma=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
+ for(j in 1:18)
+ {
+ yn <- round(seq(1,dim(prv)[2],,18))
+ poly.parts( a.pt,
+             cbind(prv[,1    ,sex,"apc","obs"],
+                   prv[,yn[j],sex,"apc",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,18),
+             txt=dimnames(prv)[[2]][yn[j]] )
+ abline(h=0)
+ if( j==1 ) text( rep(25,3), c(13,15,17)+0.5,
+                  c("Imbalance","Incidence","Mortality"),
+                  col=clr[2:4], font=2, adj=0, cex=1.2 )
+ if( j %in% c(1,7,13) ) axis( side=2 )
+ if( j %in% 13:18 ) axis( side=1 )
+ mtext( "Age", side=1, outer=TRUE, line=1.5, font=1, las=0 )
+ mtext( "Prevalence of DM", side=2, outer=TRUE, line=1.5, font=1, las=0 )
+ }
+ }

> 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" )
> clr <- rgb(c(3,2,1.5,0)/3,c(3,2,1.5,0)/3,1)
> poly.parts( a.pt, cbind(prv[,1 ,"M","apc","obs"],
+                         prv[,np,"M","apc",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,18) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> axis( side=2 )
> axis( side=2, at=1:18, labels=NA, tcl=-0.3 )
> text( rep(25,3), 17:19-1.5,
+       c("Imbalance","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=1.0 )
> clr <- rgb(1,c(3,2,1.5,0)/3,c(3,2,1.5,0)/3)
> poly.parts( a.pt, cbind(prv[,1 ,"F","apc","obs"],
+                         prv[,np,"F","apc",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,18) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> text( rep(25,3), 17:19-1.5,
+       c("Imbalance","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=1.0 )
> mtext( "Age", side=1, outer=TRUE, line=0.8, font=1, las=0 )
> mtext( "Prevalence of DM (%)", side=2, outer=TRUE, line=2, font=1, las=0 )
```

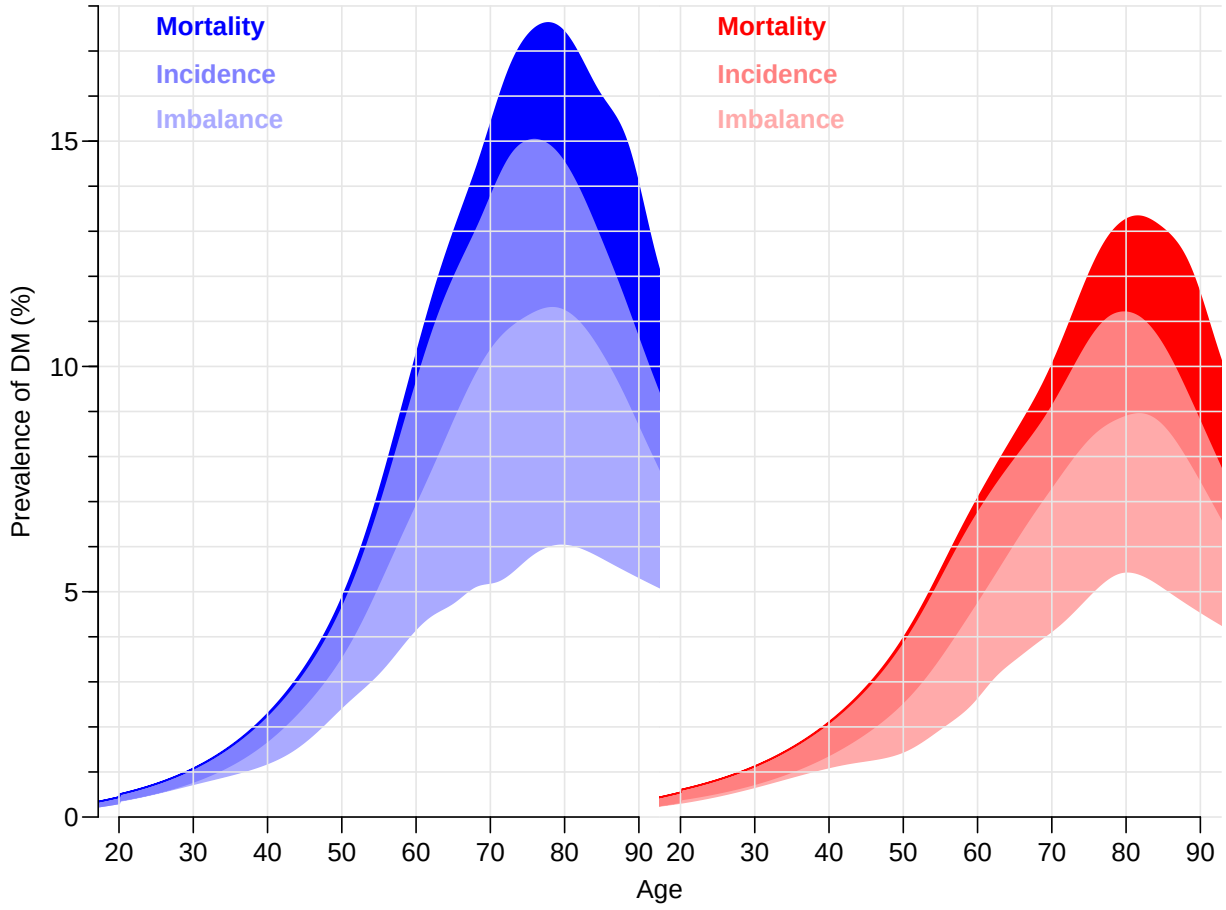

Figure 5.4: *Changes in predicted age-specific prevalences of DM in Denmark 2016 among men (blue) and women (red), partitioned by the contribution from rates as they were in 1996 ("Imbalance"), increases in incidence and decrease in mortality, respectively.*

5.4 Number 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 1996, 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` (but not extent) as `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). However `prn` will have 100 age-classes rather than 1200 (100/`int`), and only 18 dates: `prv`.

```
> dn <- dimnames(prv)
> dn[[1]] <- 0:99
> dn[[2]] <- 1996:2016
> dn[[5]] <- dn[[5]][c(5:7,1)]
> prn <- NArray( dn )
> length(prv) ; str( prv ) ; table( prv<0 )
[1] 8500800
 num [1:1200, 1:253, 1:2, 1:2, 1:7] 0 0.00063 0.000635 0.000639 0.000644 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.29166666666666
  ..$ t   : chr [1:253] "1996" "1996.08333333333" "1996.16666666667" "1996.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
  FALSE    TRUE
8321273  179527
> length(prn) ; str( prn )
[1] 33600
 logi [1:100, 1:21, 1:2, 1:2, 1:4] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:100] "0" "1" "2" "3" ...
  ..$ t   : chr [1:21] "1996" "1997" "1998" "1999" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:4] "mort" "inc" "const" "obs"
```

In order to fill in the numbers we use the estimated age-specific prevalences at 1st January each year, that is at the dates 1996-01-01,...,2016-01-01 in the entries along the `t`-dimension of `prv`. Moreover we want the prevalences for a 1 year age class rather than age-classes of length `int`. So we take the average prevalences from `prv` over each one-year age-interval. The vectors `wh.a` and `wh.p` will hold the number of the age and period classes from `prv` which have the desired prevalences (as proportions) that we will use for multiplication with the population figures:

```
> comp <- c("mort","inc","const")
> str( prn[,,,,comp ] )
 logi [1:100, 1:21, 1:2, 1:2, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:100] "0" "1" "2" "3" ...
  ..$ t   : chr [1:21] "1996" "1997" "1998" "1999" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:3] "mort" "inc" "const"
> # Find the dates in the predicted prevalences prv that matches the
> # dates in prn where empirical rates are available.
> prv.p <- as.numeric( dimnames(prv)[["t"]] )
> prn.p <- as.numeric( dimnames(prn)[["t"]] )
> wh.p <- match( prn.p, prv.p )
> if( any(is.na(wh.p)) ) # Need to find approximate dates if they do not match
+ for( ip in 1:length(prn.p) )
+    {
+    dd <- abs( prn.p[ip]-prv.p )
+    wh.p[ip] <- (1:length(dd))[dd==min(dd)]
+    }
> wh.p
 [1]   1  13  25  37  49  61  73  85  97 109 121 133 145 157 169 181 193 205 217 229 241
> prv <- pmax( prv, 0 )
> # Ages in the two arrays
> prv.a <- as.numeric( dimnames(prv)[["a"]] )
> prn.a <- as.numeric( dimnames(prn)[["a"]] )
> for( ip in 1:length(wh.p) )
+ for( ia in 1:length(prn.a) )
+    {
+ wh.a <- which( prn.a[ia]==floor(prv.a) )
+ prn[ia,ip,,,comp ] <- apply( prv[wh.a,wh.p[ip],,, comp], 2:4, mean )
+ prn[ia,ip,,,"obs"] <- apply( prv[wh.a,       1,,,"obs"], 2:3, mean )
+    }
```

Now `prn` contains the estimated prevalences (as fractions) for 100 age classes and the 21 dates. We need to multiply these prevalences by the population figures at these times. The population figures are in `pr`:

```
> head( pr )
  sex A    P X     N
2   M 0 1996 2 36258
3   M 0 1997 0 34920
4   M 0 1998 1 34936
5   M 0 1999 2 34151
6   M 0 2000 2 34035
7   M 0 2001 1 34432
> pop <- xtabs( N ~ A + P + sex, data=pr )[1:100,,]
> dmp <- xtabs( X ~ A + P + sex, data=pr )[1:100,,]
> str( pop )
 table [1:100, 1:21, 1:2] 36258 36077 35003 35344 33830 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:21] "1996" "1997" "1998" "1999" ...
  ..$ sex: chr [1:2] "M" "F"
> str( dmp )
```

```
 table [1:100, 1:21, 1:2] 2 8 10 18 14 21 33 38 52 39 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:21] "1996" "1997" "1998" "1999" ...
  ..$ sex: chr [1:2] "M" "F"
> str( prn )
 num [1:100, 1:21, 1:2, 1:2, 1:4] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:100] "0" "1" "2" "3" ...
  ..$ t   : chr [1:21] "1996" "1997" "1998" "1999" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:4] "mort" "inc" "const" "obs"
> for( i in dimnames(prn)[[4]] )
+ for( j in dimnames(prn)[[5]] )
+ prn[,,,i,j] <- prn[,,,i,j] * pop
```

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(=0) 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 <- "2016"
> oo <- c("mort","inc","const","obs")
> 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="transparent",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) )
+ axis( side=2, at=seq(0,100,5), labels=NA, tcl=-0.3 )
+ mtext( pp, at=-lim, adj=1.4, cex=1.3, font=1 )
+ 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( "comp-obs-film.pdf", width=8, height=6 )
> for( pp in paste(1996:2016) ) draw.dmp(pp)
> dev.off()
```

```
null device
          1
> for( pp in paste(1996:2016) )
+ {
+ pdf( paste("comp-obs-", pp, ".pdf", sep=""), width=8, height=6 )
+ draw.dmp(pp)
+ dev.off()
+ }
```

Now we can also draw a population pyramid of the *predicted* number of DM patients using colors that range from very light to full:

```
> shd <- c(0.0, 1.1, 2.0, 2.8) / 3
> een <- rep(1,4)
> clr <- rgb( c(een,rev(shd)),
+             c(shd,rev(shd)),
+             c(shd,    een ) )
> clr
[1] "#FF0000" "#FF5E5E" "#FFAAAA" "#FFEEEE" "#EEEEFF" "#AAAAFF" "#5E5EFF" "#0000FF"
> # 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
> oo <- c("mort","inc","const","obs")
> lim <- 6
> 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","apc",       ], 1, sum),
+                                  prn[,pp,"M","apc",    oo ],
+                                  prn[,pp,"F","apc",rev(oo)] ) )/ 1000,
+          horiz=TRUE, col=clr[c(1,8:2)], border=rep("transparent",8),
+          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) )
+ axis( side=2, at=seq(0,100,5), labels=NA, tcl=-0.3 )
+ tt <- addmargins( apply( prn[,pp,,"apc",],2:3, sum ), 2 )
+ nn <- tt / tt[,5] * 100
+ ppos <- seq(1,5.9,,5)-0.1
+ npos <- -rev(ppos)
+ mtext( pp, at=-lim, adj=1.8, line=2, cex=1.2, font=1 )
+ mtext( c(lg<- c("Mort","Inc","Imbal","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( "comp-DMpr-film.pdf", width=9, height=6 )
> for( pp in paste(1996:2016) ) draw.pyr(pp)
> dev.off()
null device
          1
> for( pp in paste(1996:2016) )
+ {
+ pdf( paste("comp-DMpr-", pp, ".pdf", sep=""), width=8, height=6 )
+ draw.pyr(pp)
+ dev.off()
+ }
```

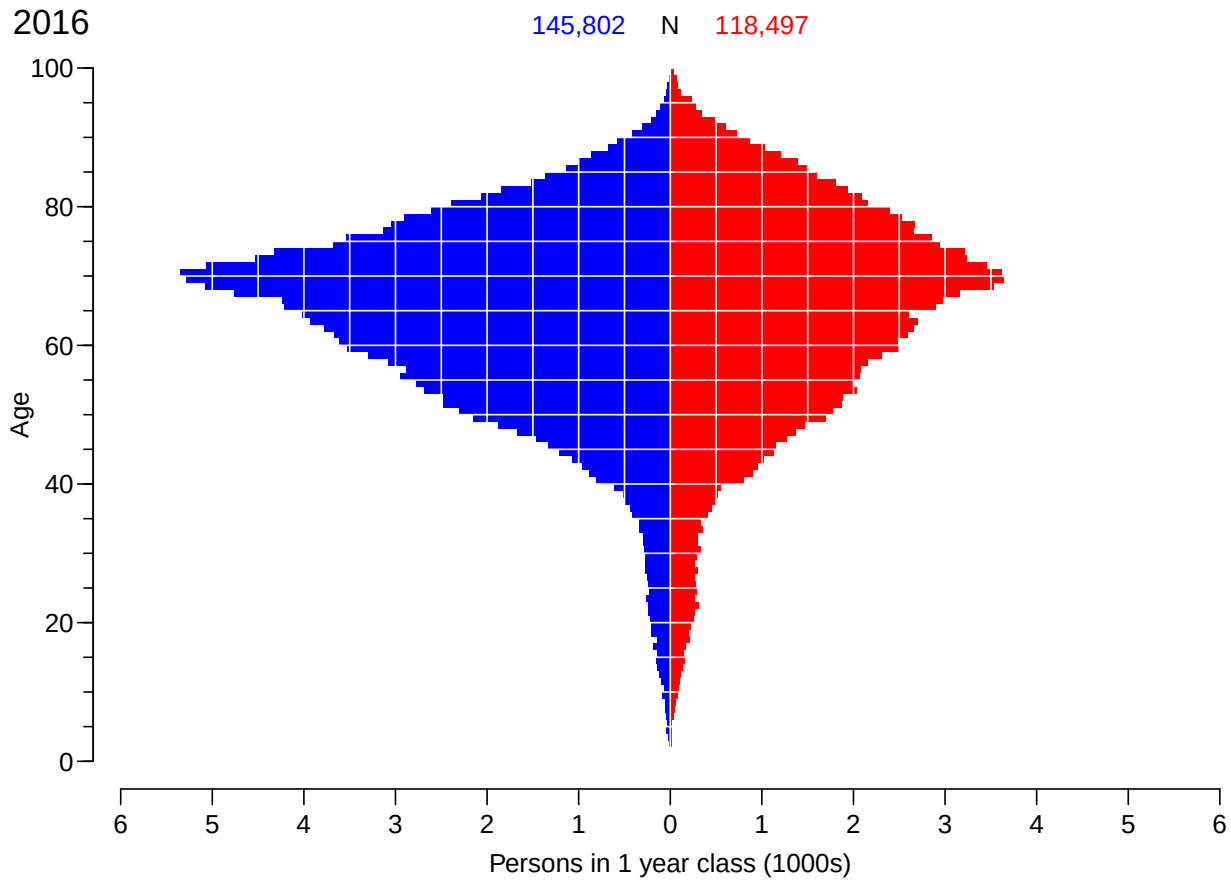

Figure 5.5: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2016.*

5.5 Timetrend in the components

It is of course also of interest to see how large a fraction of the DM-patients in various ages that can be attributed to the different components at different times.

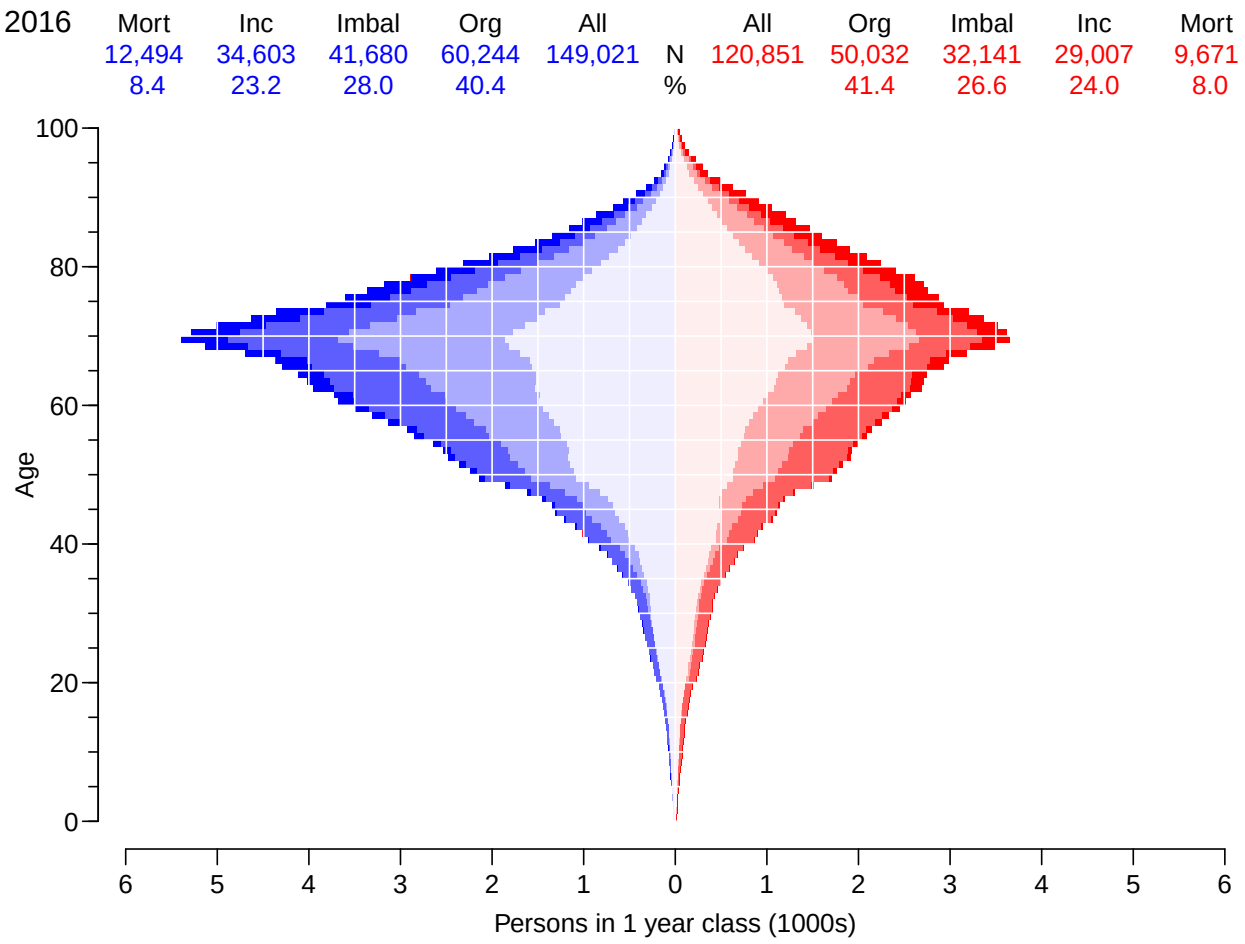

Figure 5.6: *Age-distribution of the predicted no. of diabetes cases in Denmark as of 1.1.2016, subdivided by the components of disease prevalence: Mort: decrease in mortality, Inc: increase in incidence, Imbal: constant rates from 1996 (non-steady-state imbalance between incidence and mortality in 1996, Org: age-specific prevalence in 1996.*

```
> str( prv )
 num [1:1200, 1:253, 1:2, 1:2, 1:7] 0 0.00063 0.000635 0.000639 0.000644 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.29166666666666
  ..$ t   : chr [1:253] "1996" "1996.08333333333" "1996.16666666667" "1996.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "ap" "apc"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
> dimnames( prv )[[5]]
[1] "obs"   "m-fix" "i-fix" "all-f" "mort"  "inc"   "const"
```

We extract the prevalences at ages 60, 70 and 80, and compute the fractions of all DM patients attributable to each component:

```
> aloc <- match( 6:8*10, floor(as.numeric(dimnames(prv)[[1]])) )
> ptrend <- ( prv[aloc,,,"apc",-(2:4)] + prv[aloc-1,,,"apc",-(2:4)] )/2
> str( ptrend )
 num [1:3, 1:253, 1:2, 1:4] 0.0417 0.052 0.0606 0.0419 0.0524 ...
 - attr(*, "dimnames")=List of 4
```

```
  ..$ a   : chr [1:3] "60.0416666666667" "70.0416666666667" "80.0416666666667"
  ..$ t   : chr [1:253] "1996" "1996.08333333333" "1996.16666666667" "1996.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ what: chr [1:4] "obs" "mort" "inc" "const"
> # Fraction of all DM at each age
> ptrend[,,,] <- ptrend[,,,]/ptrend[,,,rep("obs",4)]
> # Reduce to fraction attributable to org. prevalence
> ptrend[,,,"obs"] <- ptrend[,,,"obs"  ] -
+                     ptrend[,,,"mort" ] -
+                     ptrend[,,,"inc"  ] -
+                     ptrend[,,,"const"]
```

We can now plot the stacked fractions of the components for ages 60, 70 and 80 separately for men and women:

```
> p.pt <- as.numeric(dimnames(ptrend)[[2]])
> par( mfcol=c(3,2), mar=c(0,0,0,2), oma=c(3,4,1,0), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> for( sx in c("M","F") )
+ for( ag in 1:3 )
+ {
+ plot( NA, xlim=c(1996,2016), ylim=0:1*100,
+           xaxs="i", yaxs="i", xaxt="n", yaxt="n",xlab="",ylab="" )
+ #if( sx=="M" ){
+ #axis( side=2 )
+ #axis( side=2, at=1:9*10, labels=NA ) }
+ if( ag==3 ){
+ axis( side=1 )
+ axis( side=1, at=1:20+1996, labels=NA, tcl=-0.3 ) }
+ polygon( c(p.pt,rev(p.pt)), c( ptrend[ag,,sx,"mort"] *100,
+                               rev(ptrend[ag,,sx,"mort"])*  0),
+          col=clr[if(sx=="M") 8 else 1], border="transparent" )
+ polygon( c(p.pt,rev(p.pt)), c(ptrend[ag,,sx,"mort"],
+                              rev(ptrend[ag,,sx,"mort"]+
+                                  ptrend[ag,,sx,"inc" ]))*100,
+          col=clr[if(sx=="M") 7 else 2], border="transparent" )
+ polygon( c(p.pt,rev(p.pt)), c(ptrend[ag,,sx,"mort"]+
+                                  ptrend[ag,,sx,"inc" ],
+                              rev(ptrend[ag,,sx,"mort"]+
+                                  ptrend[ag,,sx,"inc" ]+
+                                  ptrend[ag,,sx,"const"]))*100,
+          col=clr[if(sx=="M") 6 else 3], border="transparent" )
+ abline( v=seq(2000,2015,5), h=1:9*10, col="white" )
+ axis( side=4, at=1:7*10, tcl=-0.5 )
+ axis( side=4, at=0:14*5, tcl=-0.4, labels=NA )
+ axis( side=4, at=0:70  , tcl=-0.2, labels=NA )
+ text( 1997, 75, paste( if(sx=="F") "Women\n" else "Men\n","age",(6:8*10)[ag]),
+       cex=1.2, font=2, adj=0 )
+ box(col=gray(0.7),bty="o")
+ }
> mtext( side=1, "Date", line=2, outer=TRUE, cex=0.67 )
> mtext( side=2, "Prevalence component (%)", line=2, outer=TRUE, cex=0.67, las=0 )
```

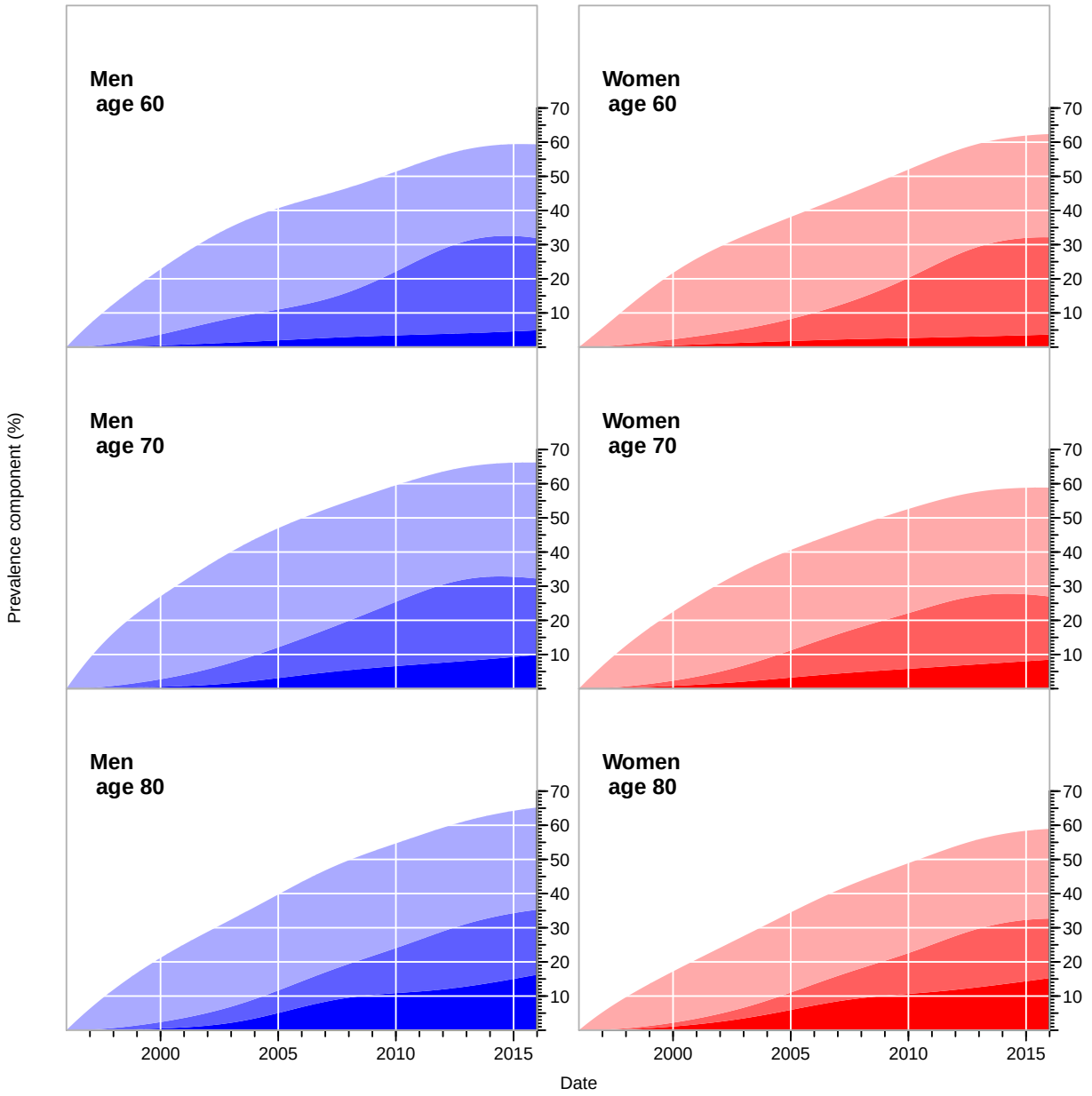

Figure 5.7: *Fraction of the prevalent cases at different times attributable to a) declining mortality (bottom, full color), b) increasing incidence (middle, pale color) and c) prevalence/mortality imbalance at 1996 (top, weak color). The white area above the curves corresponds to the fraction of the cases that would have been around if incidence and mortality rates had remained as in 1996.*

Chapter 6

Predicting prevalence of diabetes

In the previous chapter we predicted the number of prevalent cases, or more precisely the age-specific prevalences of DM over the period 1996 – 2016 under different scenarios.

In this chapter we shall essentially repeat this for the period 2016–2030, using three different scenarios:

- Use the naively predicted rates from the APC-model with natural splines — the "`apc`" component if the rate-arrays.

  This will give a prediction of the numbers which is the least credible.

- Use the attenuated rates — the "`att`" component of the rate-arrays.

```
> library( Epi )
> library( splines )
> clear()
```

6.1 Predicted rates

In the previous chapter we did predictions of the incidence and mortality rates, and used prevalences at 1996-01-01 as starting point to predict prevalences at 2016-01-01.

Here we shall do the same as in the exercise assessing the components of prevalence, except that we will start with the observed (smoothed) age-specific prevalences at 2016-01-01 and then use the three different scenarios laid out above to predict the prevalences each year till 2030.

First we load the estimated / predicted rates

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

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

`pr.fit` — array of predicted age-specific prevalences at 1995-01-01 to 2016-01-01, smoothed by natural splines. This is actually derived in the section on prevalence:

```
> load( file="../data/prFit.Rda" )
> str( pr.fit )
```

```
 num [1:2, 1:1200, 1:21] 0.000626 0.000579 0.00063 0.000583 0.000635 ...
 - attr(*, "dimnames")=List of 3
  ..$ sex: chr [1:2] "M" "F"
  ..$ A  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.2916666666
  ..$ P  : chr [1:21] "1996" "1997" "1998" "1999" ...
```

`TR` — array of transition probabilities between states Well and DM and Death. Transition probabilities are computed under the 3 different scenarios combining mortality and incidence rates either as they actually developed 1996–2015. These refer to intervals of length `int` and are therefore labeled on the period dimension by the midpoint of these, a total of 15/`int`.

`prv` — array of predicted prevalences based on the initial prevalences at 2016-01-01 and the transition probabilities as put in `TR`. The scenario dimension refers to the 3 scenarios: "lin", "att" and "fix". Moreover, 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.

6.2 Transition probabilities

In order to get the predicted *number* of persons 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. But we will only need the rates from 2016 and onward, so we restrict the arrays with the rates to this period:

```
> dimnames( Lambda )[[2]][240:241]
[1] "2015.95833333333" "2016.04166666667"

> rLambda <- Lambda[,-(1:240),,]
> rMu.W   <-  Mu.W [,-(1:240),,]
> rMu.DM  <-  Mu.DM[,-(1:240),,]
> states <- c("Well","DM")
> TR <- NArray( c( dimnames( rLambda ),
+                  list( from = states,
+                        to = states ) ) )
> dimnames( TR )[[4]][1] <- "fix"
> str( TR )

 logi [1:1200, 1:288, 1:2, 1:6, 1:2, 1:2] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 6
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.2916666666666
  ..$ p   : chr [1:288] "2016.04166666667" "2016.125" "2016.20833333333" "2016.2916666667"
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:6] "fix" "apc" "att" "i20" ...
  ..$ from: chr [1:2] "Well" "DM"
  ..$ to  : chr [1:2] "Well" "DM"
```

The situation where both the mortality rates and incidence rates are fixed at the 2016 level ("`fix`") is trivial, because transition probabilities in that case only depend on age and not on period.

In order to fill `TR`, we need the cumulative incidences over intervals of length `int`. But these were exactly the ones we predicted in the previous sections by setting the person years equal to `int` in the data frame supplied to the `newdata` argument.

So we can now compute the one-`int`-step transition matrices for every combination of `a.pt` and `p.pt`, in steps of `int` (in this case 0.0833 year).

```
> TR[,,,,"Well","Well"] <-     exp(-rLambda-rMu.W)
> TR[,,,,"Well","DM"  ] <- 1 - exp(-rLambda)
> TR[,,,,"DM"  ,"Well"] <- 0
> TR[,,,,"DM"  ,"DM"  ] <-     exp(-rMu.DM)
```

Note that we here fill in the transition probabilities from the age-period model in the `fix` slot if the 4th dimension of `TR`, but we overwrite this with the rates as of 2016:

```
> TR[,,,"fix","Well","Well"] <-     exp(-rLambda[,rep(1,dim(TR)[2]),,"apc"]-rMu.W[,rep(1,di
> TR[,,,"fix","Well","DM"  ] <- 1 - exp(-rLambda[,rep(1,dim(TR)[2]),,"apc"])
> TR[,,,"fix","DM"  ,"Well"] <- 0
> TR[,,,"fix","DM"  ,"DM"  ] <-     exp(- rMu.DM[,rep(1,dim(TR)[2]),,"apc"])
```

Likewise we fill in the entries referring to the three scenarios of increasing incidence rates — note that it is only for the incidence rates we impose an increase in rates:

```
> TR[,,,"i20","Well","Well"] <-     exp(-rLambda[,,,"i20"]-rMu.W[,,,"att"])
> TR[,,,"i20","Well","DM"  ] <- 1 - exp(-rLambda[,,,"i20"])
> TR[,,,"i20","DM"  ,"Well"] <- 0
> TR[,,,"i20","DM"  ,"DM"  ] <-     exp(- rMu.DM[,,,"att"])
> TR[,,,"i25","Well","Well"] <-     exp(-rLambda[,,,"i25"]-rMu.W[,,,"att"])
> TR[,,,"i25","Well","DM"  ] <- 1 - exp(-rLambda[,,,"i25"])
> TR[,,,"i25","DM"  ,"Well"] <- 0
> TR[,,,"i25","DM"  ,"DM"  ] <-     exp(- rMu.DM[,,,"att"])
> TR[,,,"i30","Well","Well"] <-     exp(-rLambda[,,,"i30"]-rMu.W[,,,"att"])
> TR[,,,"i30","Well","DM"  ] <- 1 - exp(-rLambda[,,,"i30"])
> TR[,,,"i30","DM"  ,"Well"] <- 0
> TR[,,,"i30","DM"  ,"DM"  ] <-     exp(- rMu.DM[,,,"att"])
```

Finally, 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; we only compute how many of the persons alive that end up being alive at the next time point

We have now collected the transition probabilities between "Well" and "DM" as well as the probabilities of remaining in each of these, all referring to a duration of `int`, a total of almost 5 mil. numbers:

```
> str( TR )
 num [1:1200, 1:288, 1:2, 1:6, 1:2, 1:2] 1 1 1 1 1 ...
 - attr(*, "dimnames")=List of 6
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.29166666666666
  ..$ p   : chr [1:288] "2016.04166666667" "2016.125" "2016.20833333333" "2016.29166666667"
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:6] "fix" "apc" "att" "i20" ...
  ..$ from: chr [1:2] "Well" "DM"
  ..$ to  : chr [1:2] "Well" "DM"
> prod( dim(TR) )
[1] 16588800
> save( TR, file="../data/TRf.Rda" )
```

6.3 Prediction of the observed prevalences

Note that we do not need to predict the population size; we can get away with only predicting the prevalences as fractions. When we multiply the fraction of persons in states (Well,DM) with the transition matrix, we get fraction of the persons in the previous state that are in states (Well,DM), which does not sum to 1 (because of the ones dying), so we must rescale to prevalence age in each step.

First we set up an array to hold the predicted prevalences under different scenarios. 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 2016-01-01).

```
> dpr <- dimnames(TR)[1:4]
> names( dpr )[2] <- "t"
> dpr[["t"]] <- t.pt[-(1:240)]
> prv <- NArray( dpr )
> str( prv )
 logi [1:1200, 1:289, 1:2, 1:6] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ a  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667
  ..$ t  : chr [1:289] "2016" "2016.08333333333" "2016.16666666667" "2016.25" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:6] "fix" "apc" "att" "i20" ...
> prod( dim(prv) )
[1] 4161600
```

Thus we must make a loop that updates the prevalences at 2016-01-01 to those at subsequent times, but first we must initialize the prevalences as modeled on 2016-01-01, as well as the prevalences at ages 0 (the new-born) that we set to 0; note that we exploit the column major storage of arrays to get `prv` filled with identical values across the two last dimensions (model and scenario).

```
> load( file="../data/prFit.Rda" )
> str( pr.fit )
 num [1:2, 1:1200, 1:21] 0.000626 0.000579 0.00063 0.000583 0.000635 ...
 - attr(*, "dimnames")=List of 3
  ..$ sex: chr [1:2] "M" "F"
  ..$ A  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667
  ..$ P  : chr [1:21] "1996" "1997" "1998" "1999" ...
> # Smoothed prevalences at 2016-01-01 - the starting values
> for( sx in c("M","F") ) prv[,"2016",sx,] <- pr.fit[sx,,"2016"]
> # Prevalences at age 0
> prv[1,,,] <- 0
> tt <- prv[c(1:2,floor(dim(prv)[1]/1.5)+1:3),1:3,,]
> dimnames( tt )[[1]] <- round( as.numeric(dimnames( tt )[[1]]), 2 )
> dimnames( tt )[[2]] <- round( as.numeric(dimnames( tt )[[2]]), 2 )
> round( ftable( tt, col.vars=3:4 )*100, 1 )
             sex    M                            F
             mod  fix  apc  att  i20  i25  i30  fix  apc  att  i20  i25  i30
a    t
0.04 2016        0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
```

```
      2016.08      0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
      2016.17      0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.12  2016         0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
      2016.08       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
      2016.17       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
66.71 2016        13.8 13.8 13.8 13.8 13.8 13.8  9.0  9.0  9.0  9.0  9.0  9.0
      2016.08       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
      2016.17       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
66.79 2016        13.8 13.8 13.8 13.8 13.8 13.8  9.1  9.1  9.1  9.1  9.1  9.1
      2016.08       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
      2016.17       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
66.88 2016        13.9 13.9 13.9 13.9 13.9 13.9  9.1  9.1  9.1  9.1  9.1  9.1
      2016.08       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
      2016.17       NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
```

So now it is checked that we have put the initial values correctly into `prv`. Then we can compute the predicted prevalences under the different scenarios. We take the fraction of the population in age class `ia` at time `ip` that end up as diabetes patients at time `ip+1` (and hence in age class `ia+1`), and divide by the fraction of all that remain alive, which is the diabetes patients, *plus* those who survive free of diabetes:

```
> system.time(
+ for( ip in 1:(dim(prv)[2]-1) )
+ for( ia in 1:(dim(prv)[1]-1) )
+ prv[ia+1,ip+1,,] <-
+   (    prv[ia,ip,,]  * TR[ia,ip,,"DM"  ,"DM"  ]
+    +(1-prv[ia,ip,,]) * TR[ia,ip,,"Well","DM"  ] ) /
+   (    prv[ia,ip,,]  * TR[ia,ip,,"DM"  ,"DM"  ]
+    +(1-prv[ia,ip,,]) * TR[ia,ip,,"Well","DM"  ]
+    +(1-prv[ia,ip,,]) * TR[ia,ip,,"Well","Well"] )
+             )
   user  system elapsed
  12.94    0.00   12.94
```

Note that the code above is particularly simple because we only need to compute the prevalence at the next date and age. If we had had a more elaborate model with, say complications states, the calculations in the loop would have been a matrix-multiplication updating the state-distribution, but this simplification would have been at the expense of another three loop-levels, namely over the the three last dimensions of the `prv` array.

We can then show a few of the predicted prevalences in (

```
> round( prv[1:4,1:2,1,1,drop=F]*100, 3 )
, , sex = M, mod = fix

                    t
a                     2016 2016.08333333333
  0.0416666666666667 0.000             0.000
  0.125              0.121             0.001
  0.208333333333333  0.121             0.121
  0.291666666666667  0.122             0.122

> save( a.pt, prv, file="../data/prv-comp.Rda" )
> load(            file="../data/prv-comp.Rda" )
```

6.4 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.

To show the number of patients we set up an array `prn` with `structure` (but not extent) as `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). However `prn` will have 100 age-classes rather than 1200 (100/`int`), and only 15 dates (2016–2030): `prv`.

```
> dn <- dimnames(prv)
> dn[[1]] <- 0:99
> dn[[2]] <- 2016:2030
> prn <- NArray( dn )
> table( (prv>0) + (prv>=0) )
      1       2
   3468 4158132
> str( prv )
 num [1:1200, 1:289, 1:2, 1:6] 0 0.00121 0.00121 0.00122 0.00123 ...
 - attr(*, "dimnames")=List of 4
  ..$ a  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667
  ..$ t  : chr [1:289] "2016" "2016.08333333333" "2016.16666666667" "2016.25" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:6] "fix" "apc" "att" "i20" ...
> str( prn )
 logi [1:100, 1:15, 1:2, 1:6] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ a  : chr [1:100] "0" "1" "2" "3" ...
  ..$ t  : chr [1:15] "2016" "2017" "2018" "2019" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:6] "fix" "apc" "att" "i20" ...
```

In order to fill in the numbers we use the estimates age-specific prevalences at 1st January each year, that is at the dates 2016-01-01,...,2030-01-01 in the entries along the `t`-dimension of `prv`. Moreover we want the prevalences for a 1 year age class rather than age-classes of length `int`. So we take the average prevalences from `prv` over each one-year age-interval. The vectors `wh.a` and `wh.p` will hold the number of the age and period classes from `prv` which have the desired prevalences (as proportions) that we will use for multiplication with the population figures:

```
> # Find the dates in the predicted prevalences prv that matches the
> # dates in prn where empirical rates are available.
> prv.p <- as.numeric( dimnames(prv)[["t"]] )
> prn.p <- as.numeric( dimnames(prn)[["t"]] )
>  wh.p <- match( prn.p, prv.p )
> if( any(is.na(wh.p)) ) # Need to find approximate dates if they do not match
+ for( ip in 1:length(prn.p) )
+    {
+    dd <- abs( prn.p[ip]-prv.p )
+    wh.p[ip] <- (1:length(dd))[dd==min(dd)]
+    }
> wh.p
```

```
 [1]   1  13  25  37  49  61  73  85  97 109 121 133 145 157 169
> prv <- pmax( prv, 0 )
> # Ages in the two arrays
> prv.a <- as.numeric( dimnames(prv)[["a"]] )
> prn.a <- as.numeric( dimnames(prn)[["a"]] )
> for( ip in 1:length(wh.p) )
+ for( ia in 1:length(prn.a) )
+    {
+ wh.a <- which( prn.a[ia]==floor(prv.a) )
+ prn[ia,ip,,] <- apply( prv[wh.a,wh.p[ip],,], 2:3, mean )
+    }
> range( prv )
[1] 0.000000 0.275464
> range( prn )
[1] 6.336452e-06 2.246479e-01
```

Now `prn` contains the prevalences (as fractions) for 100 age classes and the 15 dates. We need to multiply these prevalences by the population figures at these times.

6.5 Population forecast from DST

One prerequisite for the prediction is that we have some knowledge about the future population size of Denmark, so we have acquired this from the data bank of Statistics Denmark:

```
> bef <- read.csv2( "../data/bef2040.csv", header=TRUE )
> bef$sex <- ifelse( bef$sex==" ", NA, bef$sex )
> for( j in 1:3 )
+ for( i in 2:nrow(bef) )
+     if( is.na(bef[i,j]) ) bef[i,j] <- bef[i-1,j]
> table( bef$sex )
   2    3
3277 3277
> bef$sex <- factor( bef$sex, levels=3:2, labels=c("M","F") )
> bef <- subset( bef, !is.na(N) )
> addmargins( xtabs( N ~ P + sex, data = bef ), 2 )
      sex
P            M       F     Sum
  2016 2837887 2869364 5707251
  2017 2863749 2891251 5755000
  2018 2887236 2911306 5798542
  2019 2906801 2928091 5834892
  2020 2923339 2942471 5865810
  2021 2938021 2955444 5893465
  2022 2951822 2967966 5919788
  2023 2965270 2980437 5945707
  2024 2978491 2992896 5971387
  2025 2991402 3005238 5996640
  2026 3003925 3017277 6021202
  2027 3016021 3028944 6044965
  2028 3027617 3040154 6067771
  2029 3038690 3050838 6089528
  2030 3049243 3060972 6110215
  2031 3059248 3070527 6129775
```

```
  2032 3068731 3079506 6148237
  2033 3077703 3087895 6165598
  2034 3086217 3095728 6181945
  2035 3094266 3102972 6197238
  2036 3101872 3109674 6211546
  2037 3109101 3115831 6224932
  2038 3115961 3121489 6237450
  2039 3122496 3126666 6249162
  2040 3128698 3131429 6260127
```

```
> str( bef )
```

```
'data.frame':        6300 obs. of  4 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : int  0 0 0 0 0 0 0 0 0 0 ...
 $ P  : int  2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 ...
 $ N  : int  30049 30897 31660 32444 33259 34197 35259 36232 36895 37240 ...
```

We need the population figures in an array of the same shape as (some of the dimensions of) `prv`

```
> pop <- xtabs( N ~ A + P + sex,
+               data = subset( bef, A<100 & P<2031 ) )
> str( pop )
```

```
 int [1:100, 1:15, 1:2] 30049 29750 29459 30834 31038 33643 33495 34918 34180 34688 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:15] "2016" "2017" "2018" "2019" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = N ~ A + P + sex, data = subset(bef, A < 100 &
```

```
> str( prn )
```

```
 num [1:100, 1:15, 1:2, 1:6] 0.00114 0.00133 0.00142 0.00152 0.00162 ...
 - attr(*, "dimnames")=List of 4
  ..$ a  : chr [1:100] "0" "1" "2" "3" ...
  ..$ t  : chr [1:15] "2016" "2017" "2018" "2019" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:6] "fix" "apc" "att" "i20" ...
```

```
> dmp <- prn
> for( im in dimnames(dmp)[[4]] ) dmp[,,,im] <- prn[,,,im] * pop
> save( dmp, file="dmp.Rda" )
> load( file="dmp.Rda" )
```

First we draw simple population pyramids of the age-distribution of the diabetes patients in Denmark, as predicted under different scenarios:

```
> # 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(=0) 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 <- "2016"
> lim <- 6
> clr <- c("red","blue")
> draw.dmp <-
+ function(pp,wh)
+ {
+ par( mar=c(3,3,3,0), mgp=c(3,1,0)/1.6, las=1 )
```

```
+ barplot( height=t( cbind( -dmp[,pp,"M",wh],
+                            dmp[,pp,"M",wh],
+                            dmp[,pp,"F",wh] ) )/ 1000,
+          horiz=TRUE, col=clr,
+          border="transparent",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=1, at=seq(-lim,lim,0.5), labels=NA, tcl=-0.3 )
+ axis( side=2, at=seq(0,100,20) )
+ axis( side=2, at=seq(0,100,5), labels=NA, tcl=-0.3 )
+ mtext( pp, at=-lim, adj=1.2, cex=1.0, font=2 )
+ mtext( formatC(sum(dmp[,pp,"M",wh]),0,format="f",big.mark=","),
+        at=-1, col="blue", line=0, cex=0.99, adj=1 )
+ mtext( formatC(sum(dmp[,pp,"F",wh]),0,format="f",big.mark=","),
+        at= 1, col="red" , line=0, cex=0.99, adj=0 )
+ mtext( "N", at=0, line=0, cex=0.99 )
+ }
> pdf( "pred-inc-film.pdf", width=8, height=6 )
> for( pp in paste(2016:2030) ) draw.dmp(pp,"fix")
> dev.off()
null device
          1
> pdf( "pred-apc-film.pdf", width=8, height=6 )
> for( pp in paste(2016:2030) ) draw.dmp(pp,"apc")
> dev.off()
null device
          1
> pdf( "pred-att-film.pdf", width=8, height=6 )
> for( pp in paste(2016:2030) ) draw.dmp(pp,"att")
> dev.off()
null device
          1
> for( pp in paste(2016:2030) )
+ {
+ pdf( paste("pred-att-", pp, ".pdf", sep=""), width=8, height=6 )
+ draw.dmp(pp,"att")
+ dev.off()
+ }

> par( mfrow=c(2,2), mar=c(3,3,0,0), oma=c(0,0,2,0), mgp=c(3,1,0)/1.6, las=1 )
> draw.dmp("2016","att")
> draw.dmp("2020","att")
> draw.dmp("2025","att")
> draw.dmp("2030","att")
> mtext( "Incidence rate decrease attennuates from 2016", side=3, line=0, outer=TRUE )

> par( mfrow=c(2,2), mar=c(3,3,0,0), oma=c(0,0,2,0), mgp=c(3,1,0)/1.6, las=1 )
> draw.dmp("2016","fix")
> draw.dmp("2020","fix")
> draw.dmp("2025","fix")
> draw.dmp("2030","fix")
> mtext( "Incidence and mortality rates constant from 2016", side=3, line=0, outer=TRUE )
```

```
> par( mfrow=c(2,2), mar=c(3,3,0,0), oma=c(0,0,2,0), mgp=c(3,1,0)/1.6, las=1 )
> draw.dmp("2016","apc")
> draw.dmp("2020","apc")
> draw.dmp("2025","apc")
> draw.dmp("2030","apc")
> mtext( "Naive linear prediction from 2016", side=3, line=0, outer=TRUE )


> par( mfrow=c(2,2), mar=c(3,3,0,0), oma=c(0,0,2,0), mgp=c(3,1,0)/1.6, las=1 )
> draw.dmp("2016","i20")
> draw.dmp("2020","i20")
> draw.dmp("2025","i20")
> draw.dmp("2030","i20")
> mtext( "Incidence rates increase 2.0%/y from 2016, mortality decrease", side=3, line=0, o


> par( mfrow=c(2,2), mar=c(3,3,0,0), oma=c(0,0,2,0), mgp=c(3,1,0)/1.6, las=1 )
> draw.dmp("2016","i25")
> draw.dmp("2020","i25")
> draw.dmp("2025","i25")
> draw.dmp("2030","i25")
> mtext( "Incidence rates increase 2.5%/y from 2016, mortality decrease", side=3, line=0, o


> par( mfrow=c(2,2), mar=c(3,3,0,0), oma=c(0,0,2,0), mgp=c(3,1,0)/1.6, las=1 )
> draw.dmp("2016","i30")
> draw.dmp("2020","i30")
> draw.dmp("2025","i30")
> draw.dmp("2030","i30")
> mtext( "Incidence rates increase 3.0%/y from 2016, mortality decrease", side=3, line=0, o
```

6.6 Timetrend in prevalent number of DM patients

We would like to see the overall change in the number of diabetes patients, as recorded in the structure `dmp`

```
> DMall <- dmp[,,"M",] + dmp[,,"F",]
> DMcum <- apply( DMall, 2:3, cumsum )
> DMcum <- DMcum[c(1,1:100),,]
> DMcum[1,,] <- 0
> DMcum <- DMcum/1000
> str( DMcum )
 num [1:101, 1:15, 1:6] 0 0.0663 0.1427 0.2242 0.3151 ...
 - attr(*, "dimnames")=List of 3
  ..$ a  : chr [1:101] "0" "0" "1" "2" ...
  ..$ t  : chr [1:15] "2016" "2017" "2018" "2019" ...
  ..$ mod: chr [1:6] "fix" "apc" "att" "i20" ...
```

Finally we can plot the predicted numbers from the different scenarios:

```
> range( DMcum )
[1]   0.000 369.521
```

```
> ryr <- c(2016:2030,2030:2016)
> leg <- c("All rates fixed at 2016 level",
+          "Linear projection from 2016",
+          "Attenuated linear projection",
+          "DM incidence increasing 2.0%/y",
+          "DM incidence increasing 2.5%/y",
+          "DM incidence increasing 3.0%/y")
> par( mfrow=c(3,2), mar=c(2,1,0,3), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> for( j in c(2,3,1,4:6) )
+    {
+ plot( NA,
+       xlim=c(2015.5,2030), xlab="", xaxt="n", xaxs="i",
+       ylim=c(0,420), yaxs="i", yaxt="n", ylab="" )
+ axis( side=4, at=0:4*100 )
+ axis( side=4, at=seq(0,4,1/4)*100, labels=NA, tcl=-0.3 )
+ axis( side=1, at=2015+1:3*5 )
+ axis( side=1, at=2016:2030, labels=NA, tcl=-0.3 )
+ for( i in 1:10 ) polygon( ryr, c( DMcum[1+(i-1)*10,,j],
+                              rev( DMcum[1+ i   *10,,j] ) ),
+                           col=gray( (17-i)/18 ), border=gray(0.8) ) #"transparent" )
+ abline( h=seq(50,400,50), v=c(2020,2025), col=gray(1), lty="16" )
+ for( i in seq(55,85,10) ) text( 2029, DMcum[paste(i),"2029",j],
+                                 paste( i-5,"-",i+4,sep="") )
+ text( 2016, 400, paste( dimnames(DMcum)[[3]][j], ": ", leg[j], sep="" ), adj=c(0,1) )
+    }
```

From figure 6.7 it appears that it is the decreasing incidence rates of diabetes that carries the major differences of more than 100,000 patients in 2030. The decrease in the number of incident cases is very recent; during the period 2012–2014 there was a drop and a very slight pick-up during 2015. Tabulations in the SAS-programs `08-mkFU` (p. **??**) and `09-mkPr` (p. **??**), show that there is no particular break in data where the decrease occur.

Thus the prediction of the number of future patients is crucially dependent on the tiny amount of information available about future diabetes incidence rates in the rather odd behaviour of the rates in 2012–2015.

However, even the quite brutal assumption of a pick up of increasing DM incidence rates by 2.5% per year will not bring the predicted number of patients over 400,000 in 2030. So to say that the number of diabetes patients is less than this in 2030 seems to be a fairly safe bet.

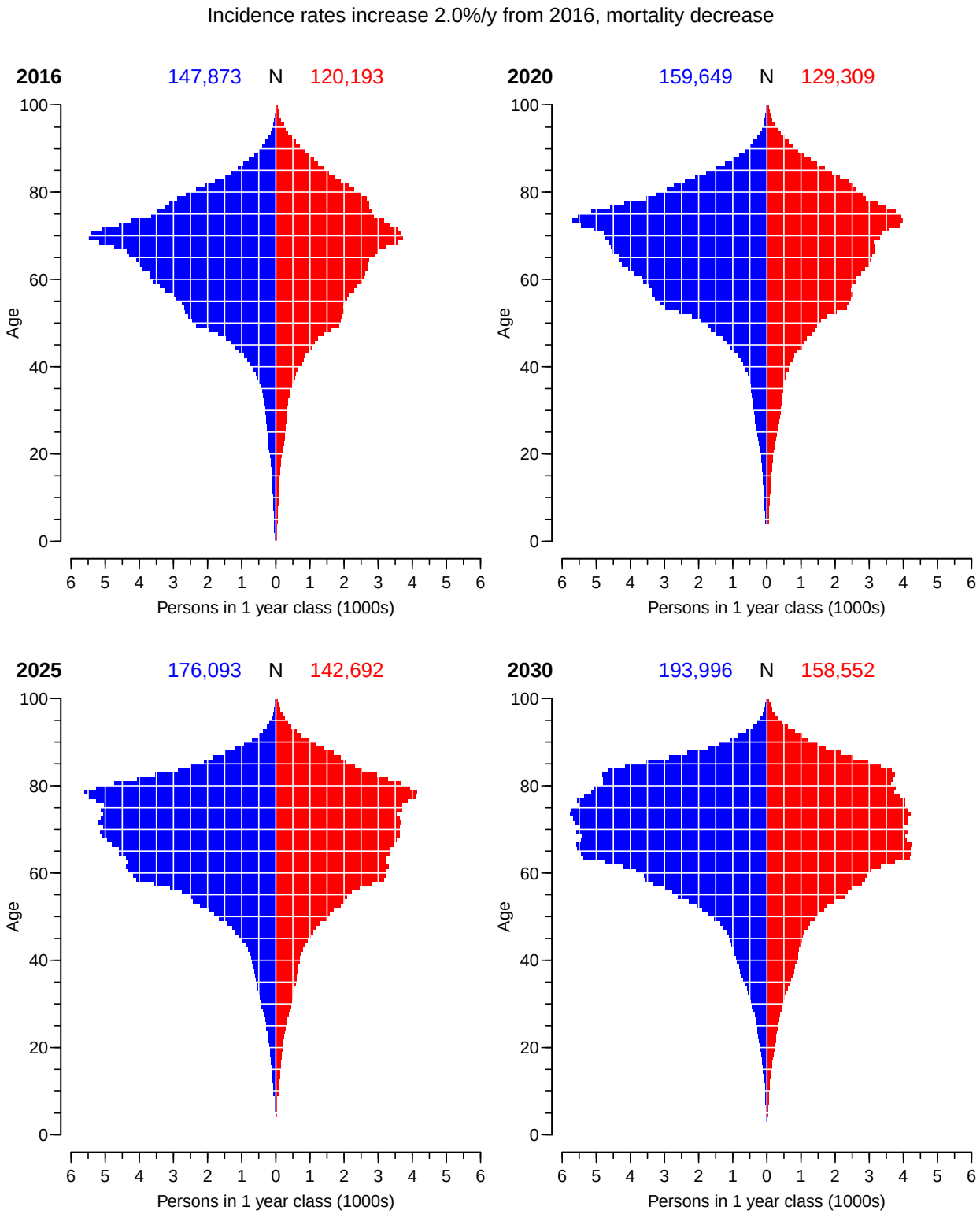

Figure 6.1: *Empirical age-distribution of the diabetes cases in Denmark based on a unchanged mortality and increasing incidence rates (2.0%/y) from 2016.*

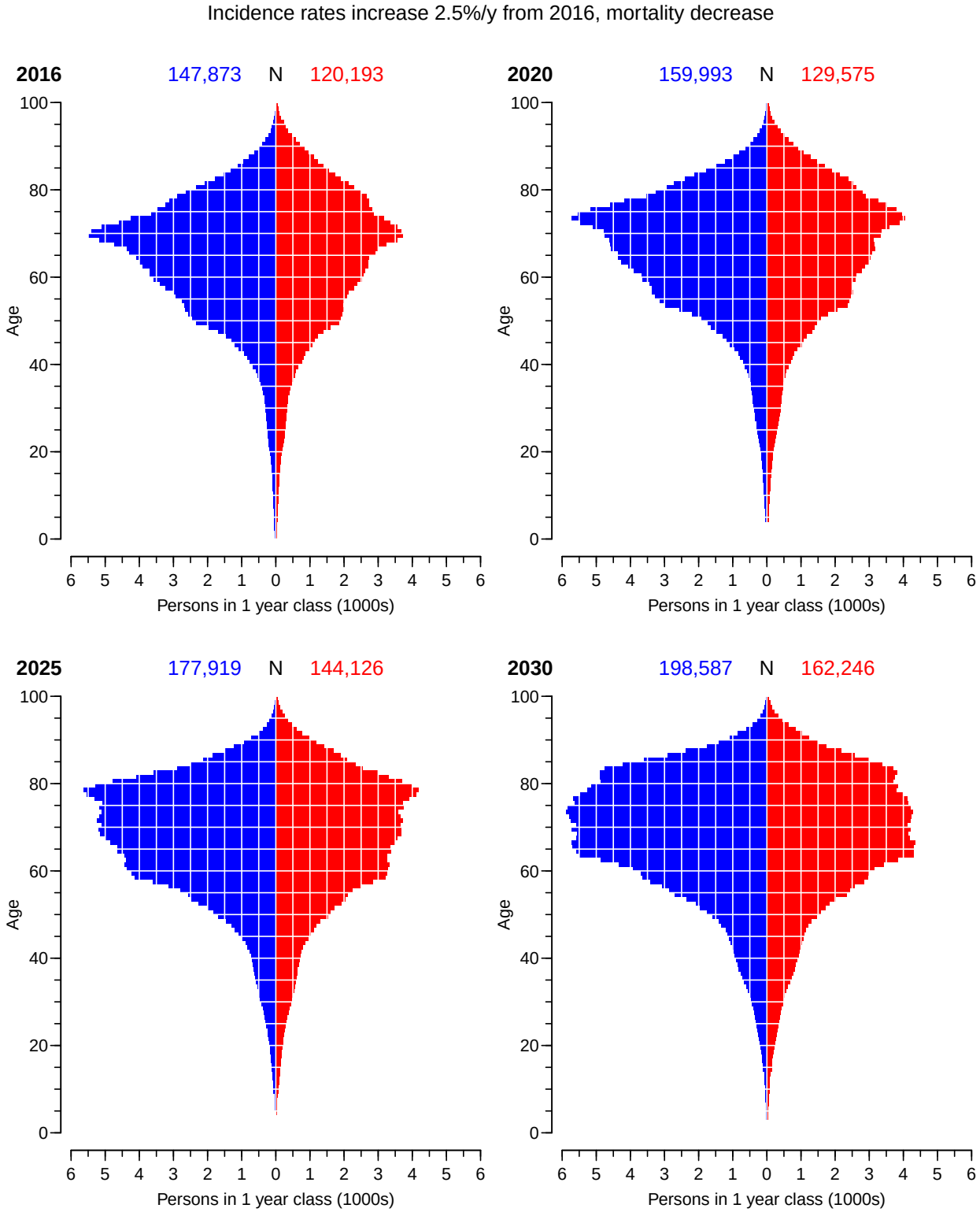

Figure 6.2: *Empirical age-distribution of the diabetes cases in Denmark based on a unchanged mortality and increasing incidence rates (2.5%/y) from 2016.*

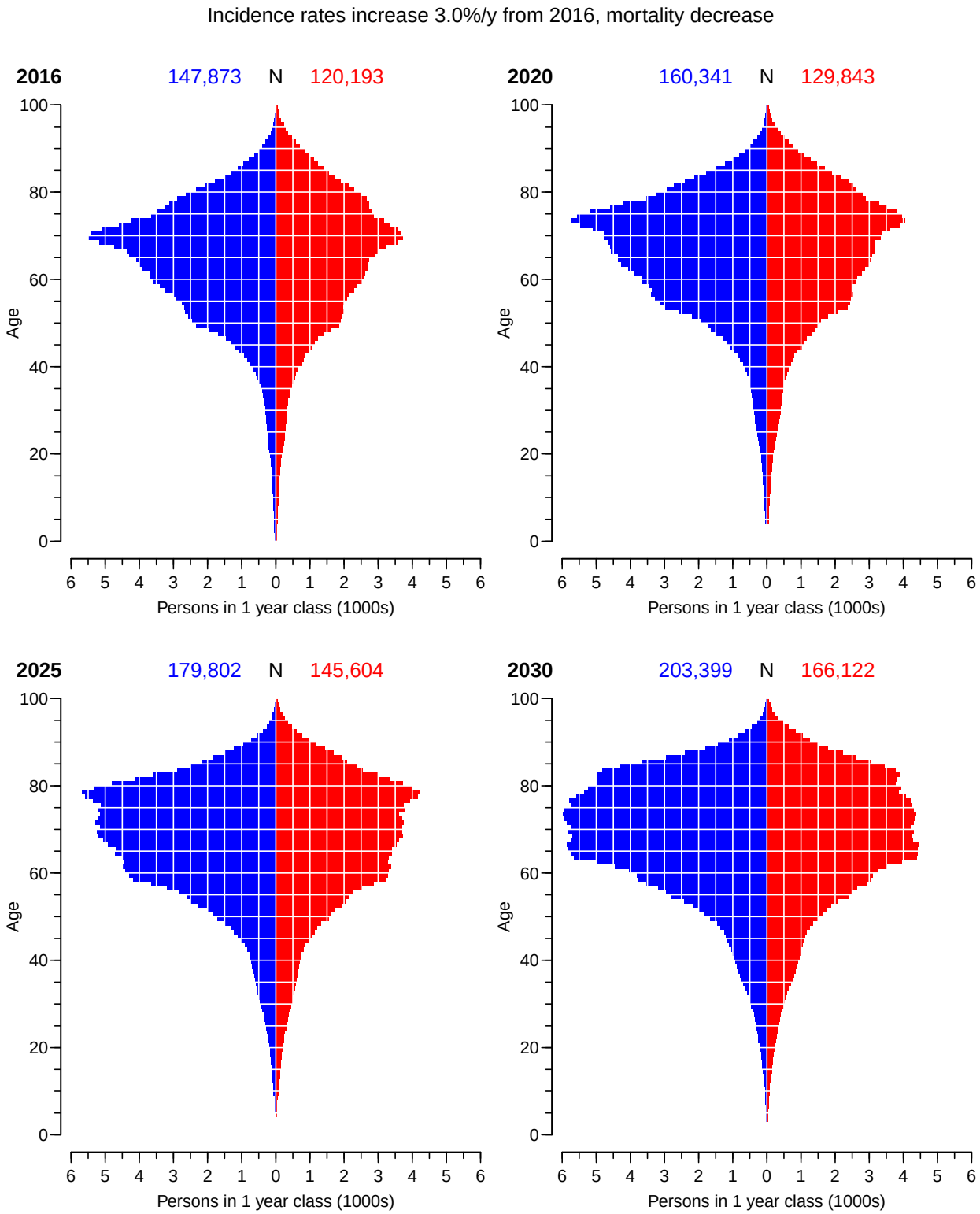

Figure 6.3: *Empirical age-distribution of the diabetes cases in Denmark based on a unchanged mortality and increasing incidence rates (3.0%/y) from 2016.*

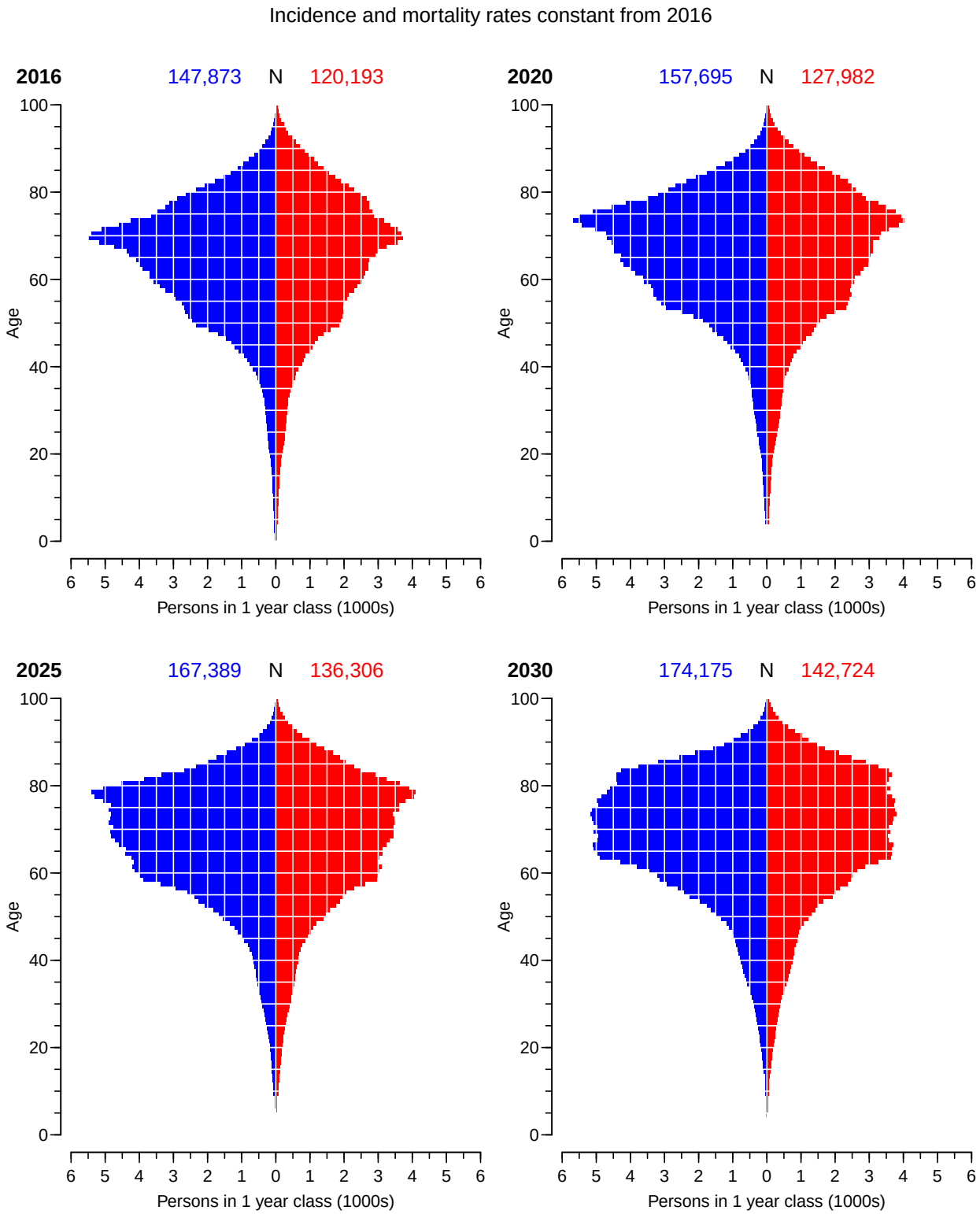

Figure 6.4: *Empirical age-distribution of the diabetes cases in Denmark based on a unchanged mortality and incidence rates from 2016.*

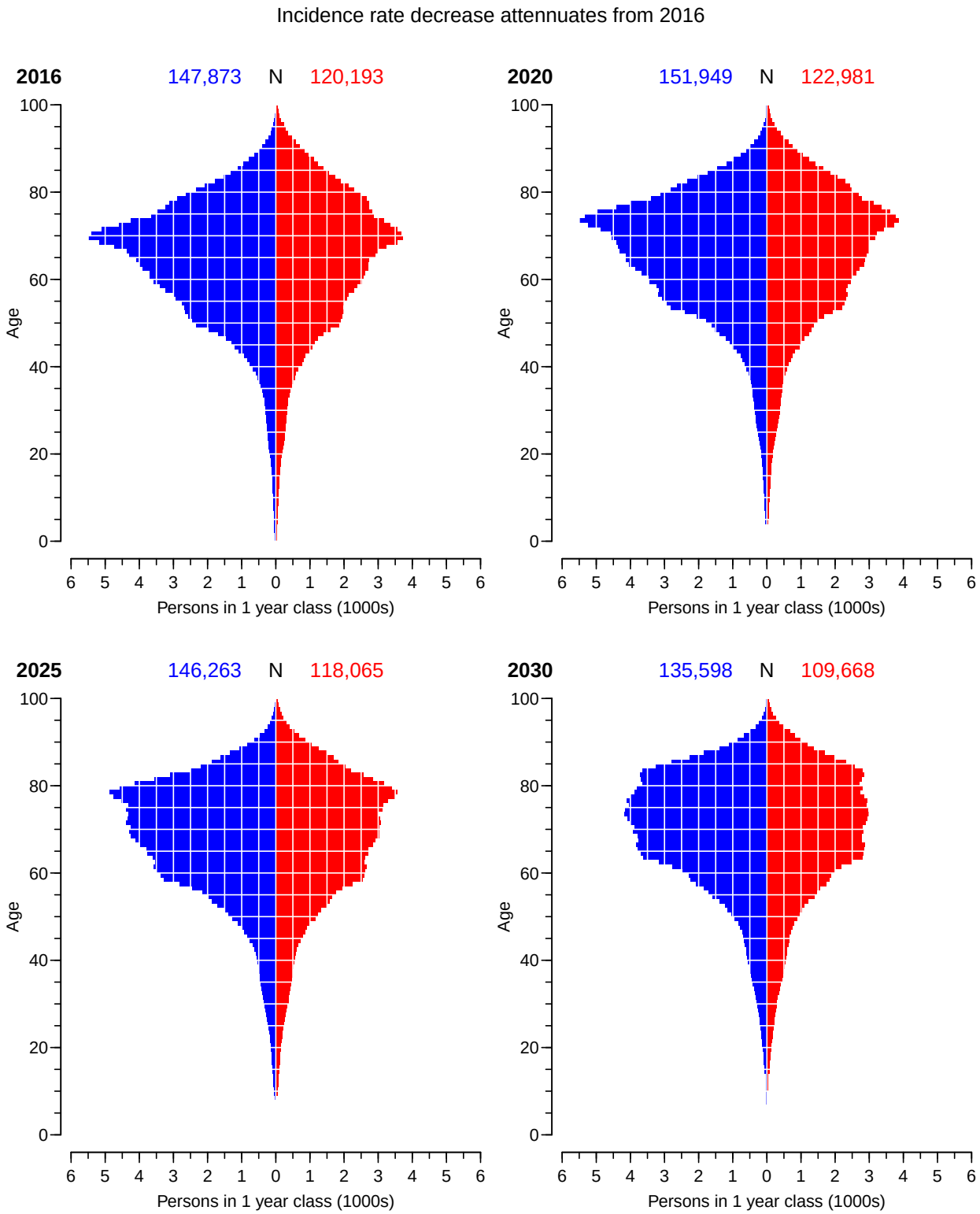

Figure 6.5: *Empirical age-distribution of the diabetes cases in Denmark based on an* attenuated *linear prediction of mortality and incidence rates.*

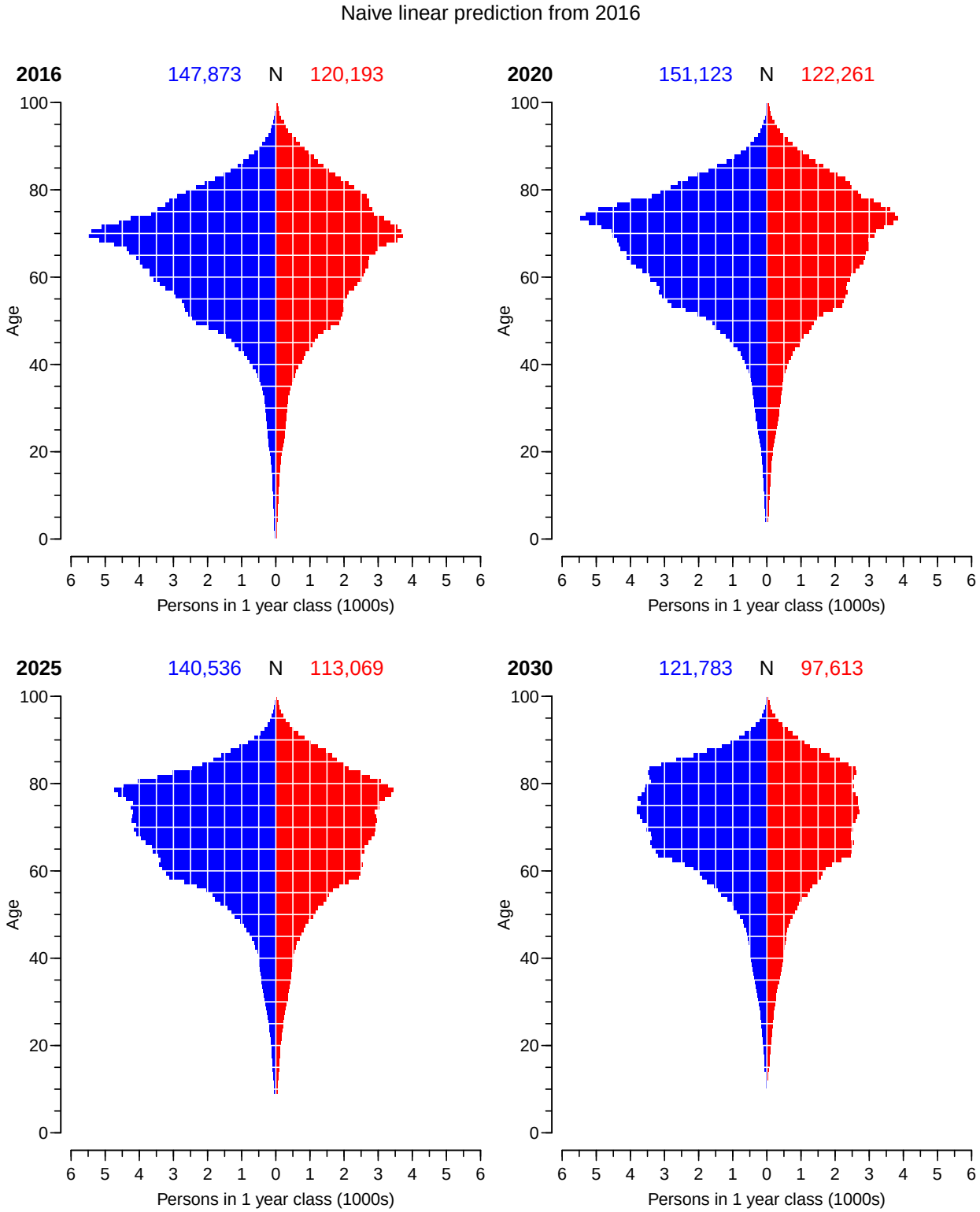

Figure 6.6: *Empirical age-distribution of the diabetes cases in Denmark based on a linear prediction of mortality and incidence rates — mainly decreasing.*

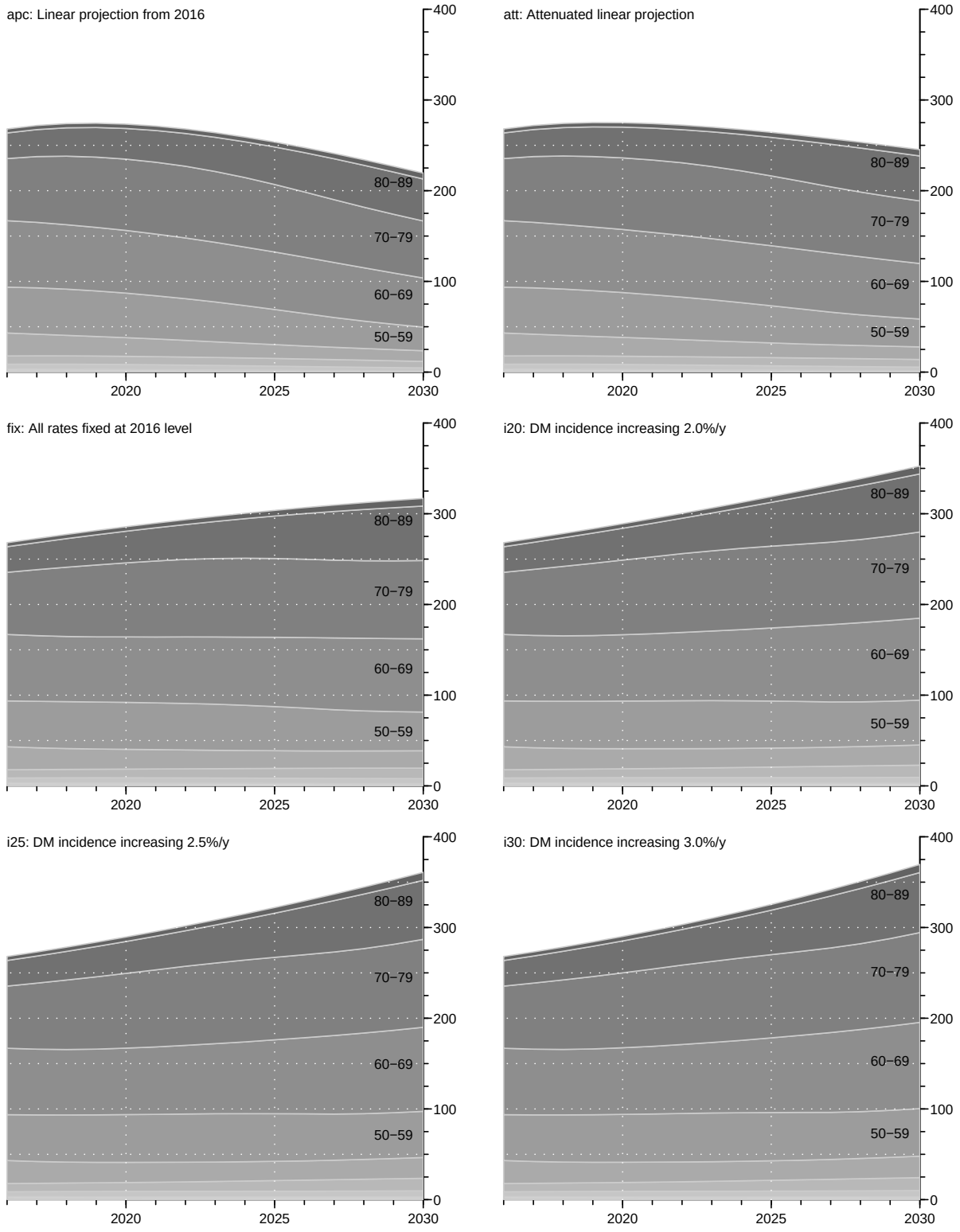

Figure 6.7: *Predicted number of diabetes patients in Denmark under different scenarios.*

References

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

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

[3] Anders Green, Camilla Sorts, Peter Bjdstrup Jensen, and Martha Emneus. Validation of the Danish National Diabetes Register. *Clinical Epidemiology*, 7:5–15, 2015.

Chapter 7

SAS programs

7.1 Rationale and overview

The following documented programs sequentially construct datasets with dates of dignosis of DM according to different criteria, then merge these to pick the earliest.

All created data will be in the data folder as SAS-datasets, and

We have the following programs:

`00-base` Collects the base information on all persons in Denmark, that is sex, date of birth and date of death, and stores it in the dataset `bef`.

`01-npr` Uses the national patient register (NPR) to generate three datasets, all with `pnr` as key:

- a dataset `npr` with the earliest date of DM diagnosis in the NPR, `doNPR`, as well as a variable `nprtyp` with values T1 (ICD10: E10) or T2 (ICD10: E11) or NA (anything else), based on whether E10 or E11 or neither is recorded on more than half of the person's NPR entries. Thus formally some of the follow-up will be based on type-information from future recording.
- a dataset `pcos` with the earliest date of registered PCOS, `doPCOS`
- a dataset `gdm` with recorded dates of GDM that are at least 200 days apart, `doGDM1`, `doGDM2`, ...

`02-dvdd` Uses the DADD to identify persons from outpatient clinics (and in due course from GPs) and to seek out persons deemed to be T1D patients. It creates a dataset, `dvdd` with key `pnr` and a variable for type of diabetes `dvdtyp`, based on whether T1 resp. T2 is recorded on more than half of the available clinical records. Thus formally some of the follow-up will be based on type-information from future recording.

`03-ndr` Uses the NDR to define data of diagnosis ignoring the originally implemented blood-glucose criteria and foot-therapy, and creates a datset `ndr` with a revised version of the "old" NDR.

`04-rmps` Generates a datset with `pnr` as key with one record per person, with dates of first dispenstion of each of a number of drugs, `donlMet`, `danyMet` etc. as well as `doOAD` and `doIns`, which are the two criteria dates that are carried on to the register. Note that this is at variance with the original NDR that used the *second* dispensation date.

`06-define` Collects data from the previously created datasets and defines date of diagnosis and type of diabetes (T1/T2), and thus generates a DM-register with sex, date of birth, date of death, date of inclusion, the latter being the smaller of date of OAD, date of insulin, date of DADD recording and date of NPR recording.

For persons with a record from DADD with type of diabetes defined, this is used. For other persons with a record from NPR the type is taken as that derived from the NPR if T1 or T2. For persons with no classification from DADD or from NPR, the person is classified as T1 if any type of medication has been taken out before age 15 or if insulin has been taken out before age 30, otherwise as T2.

`06a-comp` Compares the prevalent cases as of 1996, 2004 and 2012 from the olde revised NDR and from the reconstructed DMreg.

7.2 Program execution

All data analyses are run on the servers at Statistics Denmark. In order to have a thorough documentation of the data processing all SAS-programs have been run in sequence as batch jobs from the command prompt (`cmd`), where the progran in the file `xxx.sas`, say, is run and produces the files `xxx.log` and `xxx.lst`. Since the code from `xxx.sas` is contained in `xxx.log`, it suffices to show the files `xxx.log` and `xxx.lst` to provide full documenation of the data acquistion process.

The practical execution of the SAS-programs is done using the `cmd`-script `sj.bat` which reads:

```
start "sas job" /min sjx %~n1
```

The running of the program `xxx.sas` is started by issuing "`sj xxx`" at the command prompt.

The script `sj.bat` just starts a new process which in turn runss the script `sjx.bat`, which reads:

```
"C:\Program Files\SASHome\SASFoundation\9.4\sas.exe"^
 -CONFIG "C:\Program Files\SASHome\SASFoundation\9.4\nls\en\sasv9.cfg"^
 -$lognote1 "Program %~n1.sas" -autoexec optslibs.sas -sysin %~n1.sas
copy %~n1.log + ctrll + %~n1.lst %~n1.yt
exit
```

The second last line in the script simply copies the two result-files from SAS into one for convenence of inspection. It is the two result files that are transferred from DST to a local computer for inclusion in a documentation report.

This way there is a reasonable documentation that the results are actually produced by the listed code (in the `.log` file). Hopefully the program code is written in a reasonably human-readable way.

7.3 Program documentation

The following is a listing of the SAS-programs and -results (that is the `.log` and `.lst` files) used to generate the base datasets. Each is preceded by a very brief description; main technical points are included as comments in the program code, found in the `.log` files.

Note that according to rules of DST, all table entries of 5 or less in `.lst` file are masked as a "`*`".

7.3.1 optslibs.sas

This is common set of declarative commands that defines a couple of options, the location of the raw and the derived datasets and some global macro variables. It is included as autoexec file in all runs, with `options nonotes` for brevity of output:

```
* options used thoughout ;
options nocenter nonotes nomprint nosource2
        ps = 10000 /* 105 */
        ls = 90    /* 160 */
       obs = max
  formchar = '           '
/* format libraries we use */
 fmtsearch = ( dsfmt.times_personstatistik
               dsfmt.brancher
               dsfmt.uddannelser
               dsfmt.geokoder
              daffmt.drugfmt
               ttfmt.ttformat00
               ttfmt.ttformat10
               ttfmt.ttformat20 ) ;

* data libraries ;
libname DELPOP15 'E:\rawdata\705093\Delpopulationer\' ;
libname EKST15   'E:\rawdata\705093\Eksterne data\' ;
libname EKST16   'E:\rawdata\705093\Eksterne data\Diabasen 28-11-2016' ;
libname GRUND15  'E:\rawdata\705093\Grunddata\' ;
libname POPUL15  'E:\rawdata\705093\Population\' ;
libname TTDATA  'E:\workdata\705093\BxC\demoDM\DATA\' ;

* format libraries ;
libname  TTfmt 'E:\workdata\705093\QSN_MLiJ\' ;
libname daffmt 'E:\workdata\705093\BxC\daffodil\data' ;
libname  DSfmt 'E:\Formater\SAS formater i Danmarks Statistik\FORMATKATALOG';

* useful constants ;
%let primo   = '01JAN1995'd ;
%let cutdate = '01JAN2016'd ; * we have data including 2015 ;
%let gdmint  = 200 ; * minimal distance between GDM dates to constitute 2 GDM events ;

* macro to exclude observations with dates in GDM grace period ;
%macro xgdm( xdate, gdmwin = 365 ) ;
%do n = 1 %to 12 ;
    if doGDM&n. < &xdate. < ( doGDM&n. + &gdmwin. ) then delete ;
%end ;
%mend ;

* page ;
options notes ;
```

7.3.2 xgdm.sas

Note that the `optslibs.sas` also contains the definition of the `xgdm` macro: For each of the criteria it is necessary to exclude dates of meeting the criterion which fall within a grace period after a diagnosis of GDM. This is what the macro `xgdm` is for; it relies on the structure of the GDM dataset constructed in the `01-npr` program, which has the GDM dates in the wide form for person with at least one date of GDM. It iterates up to 12 in order to produce a note from the SAS system, that documents that only 11 GDM dataes are needed.

7.4 00-base

Reads the files with all person ids (pnr), for each calendar year of data, and forms a total roster of all pnr with demographic information (sex, date of birth, date of death).

7.4.1 00-base.lst

7.5 01-npr

Processes the records from the NPR, and produces records with GDM diagnoses and PCOS diagnoses.

Persons cannot enter on any criterion in a 365 days grace period after each GDM diagnosis. GDM diagnoses occurring within 200 days of another one is not counted, though. Thus all GDM diagnoses in the same person are at least 200 days apart.

Outputs the earliest NPR diagnosis clear of GDM and PCOS, and derives a tentative T1/T2 classification in the variable nprtyp.

```
1                                                          "Program 01-npr.sas"

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_SRV12  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.06 seconds
      cpu time            0.07 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          * read the NPR datasets in the two different formats and combine them ;
2
3          %macro mold ;
4          data all_npr1977_93 ;
5            set %do i = 1977 %to 1993 ;
6                GRUND15.LPRHEL&i. (keep = pnr recnum c_adiag d_inddto )
7                %end ; ;
8            * the ICD-8 codes incl. GDM / PCOS ;
9            if c_adiag in('24900','24901','24902','24903','24904',
10                         '24905','24906','24907','24908','24909',
11                         '25000','25001','25002','25003','25004',
12                         '25005','25006','25007','25008','25009',
13                         '63474','Y6449','61520','61521') ;
14           if substr(c_adiag,1,3) eq '249' then nprtyp = 'T1' ;
15           if substr(c_adiag,1,3) eq '250' then nprtyp = 'T2' ;
16         run ;
17         %mend ;
18         %mold ;
```

```
NOTE: There were 536220 observations read from the data set GRUND15.LPRHEL1977.
NOTE: There were 576400 observations read from the data set GRUND15.LPRHEL1978.
NOTE: There were 593234 observations read from the data set GRUND15.LPRHEL1979.
NOTE: There were 603085 observations read from the data set GRUND15.LPRHEL1980.
NOTE: There were 602068 observations read from the data set GRUND15.LPRHEL1981.
NOTE: There were 625449 observations read from the data set GRUND15.LPRHEL1982.
NOTE: There were 651393 observations read from the data set GRUND15.LPRHEL1983.
NOTE: There were 669830 observations read from the data set GRUND15.LPRHEL1984.
NOTE: There were 694317 observations read from the data set GRUND15.LPRHEL1985.
NOTE: There were 727270 observations read from the data set GRUND15.LPRHEL1986.
NOTE: There were 754514 observations read from the data set GRUND15.LPRHEL1987.
NOTE: There were 792682 observations read from the data set GRUND15.LPRHEL1988.
NOTE: There were 821261 observations read from the data set GRUND15.LPRHEL1989.
NOTE: There were 852349 observations read from the data set GRUND15.LPRHEL1990.
NOTE: There were 874069 observations read from the data set GRUND15.LPRHEL1991.
NOTE: There were 922350 observations read from the data set GRUND15.LPRHEL1992.
NOTE: There were 989322 observations read from the data set GRUND15.LPRHEL1993.
NOTE: The data set WORK.ALL_NPR1977_93 has 239128 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           4.79 seconds
      cpu time            1.21 seconds


19
20         %macro mnew ;
21         data all_npr1994_15 ;
22           set %do i = 1994 %to 2015 ;
23               GRUND15.LPRPOP&i. (keep = pnr recnum c_adiag d_inddto )
24               %end ;
25               %do i = 14 %to 15 ;
26               GRUND15.LPRPOP_UAFAMB&i. (keep = pnr recnum c_adiag d_inddto )
27               %end ; ;
28           * the ICD-10 codes incl GDM / PCOS ;
29           if substr(c_adiag,2,3) in ('E10','E11','E12','E13','E14','O24') or
30              substr(c_adiag,2,4) in ('H360','E748','E891','R730','R819','E282' ) ;
31           if substr(c_adiag,2,3) eq 'E10' then nprtyp = 'T1' ;
32           if substr(c_adiag,2,3) eq 'E11' then nprtyp = 'T2' ;
33         run ;
34         %mend ;
35         %mnew ;

WARNING: Multiple lengths were specified for the variable C_ADIAG by input data set(s).
         This can cause truncation of data.
NOTE: There were 2257703 observations read from the data set GRUND15.LPRPOP1994.
NOTE: There were 3099164 observations read from the data set GRUND15.LPRPOP1995.
NOTE: There were 3254161 observations read from the data set GRUND15.LPRPOP1996.
NOTE: There were 3348359 observations read from the data set GRUND15.LPRPOP1997.
NOTE: There were 3438023 observations read from the data set GRUND15.LPRPOP1998.
NOTE: There were 3544774 observations read from the data set GRUND15.LPRPOP1999.
NOTE: There were 3554711 observations read from the data set GRUND15.LPRPOP2000.
NOTE: There were 3676892 observations read from the data set GRUND15.LPRPOP2001.
NOTE: There were 3734587 observations read from the data set GRUND15.LPRPOP2002.
NOTE: There were 3714168 observations read from the data set GRUND15.LPRPOP2003.
NOTE: There were 4771057 observations read from the data set GRUND15.LPRPOP2004.
NOTE: There were 4970581 observations read from the data set GRUND15.LPRPOP2005.
NOTE: There were 5146887 observations read from the data set GRUND15.LPRPOP2006.
NOTE: There were 5169427 observations read from the data set GRUND15.LPRPOP2007.
NOTE: There were 5544338 observations read from the data set GRUND15.LPRPOP2008.
NOTE: There were 5835150 observations read from the data set GRUND15.LPRPOP2009.
NOTE: There were 7287169 observations read from the data set GRUND15.LPRPOP2010.
NOTE: There were 7664856 observations read from the data set GRUND15.LPRPOP2011.
NOTE: There were 6069159 observations read from the data set GRUND15.LPRPOP2012.
NOTE: There were 6227306 observations read from the data set GRUND15.LPRPOP2013.
NOTE: There were 6386162 observations read from the data set GRUND15.LPRPOP2014.
NOTE: There were 6798419 observations read from the data set GRUND15.LPRPOP2015.
NOTE: There were 317697 observations read from the data set GRUND15.LPRPOP_UAFAMB14.
NOTE: There were 1860412 observations read from the data set GRUND15.LPRPOP_UAFAMB15.
NOTE: The data set WORK.ALL_NPR1994_15 has 1333078 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           38.53 seconds
```

```
      cpu time            12.46 seconds


36
37         data all_npr ;
38           set all_npr1977_93
39               all_npr1994_15 ;
40         run ;

NOTE: There were 239128 observations read from the data set WORK.ALL_NPR1977_93.
NOTE: There were 1333078 observations read from the data set WORK.ALL_NPR1994_15.
NOTE: The data set WORK.ALL_NPR has 1572206 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           0.17 seconds
      cpu time            0.17 seconds


41
42         proc sort data = all_npr ; by pnr d_inddto ; run ;

NOTE: There were 1572206 observations read from the data set WORK.ALL_NPR.
NOTE: The data set WORK.ALL_NPR has 1572206 observations and 5 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.29 seconds
      cpu time            0.67 seconds


43
44         * only records from persons in the base population -
45           GDM & PCOS diagnoses are put in separate file ;
46         data DM
47              gdm
48              pcos ;
49            merge all_npr    ( in = npr )
50                  TTDATA.pop ( in = pop ) ;
51            by pnr ;
52            length diaggr $ 4 ;
53            if npr and pop ;
54            * GDM / PCOS (excluding men) or 'real' DM diagnoses ;
55            if substr(c_adiag,2,3) in('O24'          ) or
56                     c_adiag      in('63474','Y6449')
57               then do ;
58               if sex eq 2 then diaggr = 'GDM' ; else delete ;
59               end ;
60            else
61            if substr(c_adiag,2,4) in('E282'         ) or
62                     c_adiag      in('61520','61521')
63               then do ;
64               if sex eq 2 then diaggr = 'PCOS' ; else delete ;
65               end ;
66            else
67               diaggr = 'DM' ;
68            if diaggr eq  'DM'  then output  DM  ;
69            if diaggr eq 'GDM'  then output gdm  ;
70            if diaggr eq 'PCOS' then output pcos ;
71         run ;

NOTE: There were 1572206 observations read from the data set WORK.ALL_NPR.
NOTE: There were 7361669 observations read from the data set TTDATA.POP.
NOTE: The data set WORK.DM has 1371306 observations and 9 variables.
NOTE: The data set WORK.GDM has 53671 observations and 9 variables.
NOTE: The data set WORK.PCOS has 32167 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           1.78 seconds
      cpu time            1.56 seconds


72
73         title1 'PCOS records - id and first date of PCOS' ;
74         proc sort data = pcos ;  by pnr d_inddto ; run ;
```

```
NOTE: There were 32167 observations read from the data set WORK.PCOS.
NOTE: The data set WORK.PCOS has 32167 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


75         data TTdata.pcos ( keep = pnr doPCOS ) ;
76           set pcos ;
77           by pnr d_inddto ;
78           if first.pnr ;
79           doPCOS = d_inddto ;
80         run ;

NOTE: There were 32167 observations read from the data set WORK.PCOS.
NOTE: The data set TTDATA.PCOS has 20773 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.03 seconds
      cpu time            0.00 seconds


81
82         proc contents  data = TTdata.pcos ; run ;

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

NOTE: The PROCEDURE CONTENTS printed page 1.

83         proc tabulate  data = TTdata.pcos  missing  noseps ;
84           class doPCOS ;
85           table doPCOS, n * f=comma10.
86               / rts = 8 ;
87           format doPCOS year4. ;
88           run ;

NOTE: There were 20773 observations read from the data set TTDATA.PCOS.
NOTE: The PROCEDURE TABULATE printed page 2.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


89         proc tabulate  data = TTdata.pcos  missing  noseps ;
90           where doPCOS ge '01JAN2013'd ;
91           class doPCOS ;
92           table doPCOS, n * f=comma10.
93               / rts = 10 ;
94           format doPCOS yymms8. ;
95           run ;

NOTE: There were 2354 observations read from the data set TTDATA.PCOS.
      WHERE doPCOS>='01JAN2013'D;
NOTE: The PROCEDURE TABULATE printed page 3.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


96
97         title1 'GDM records - id and any date of GDM except if too close' ;
98         proc sort data = gdm ;  by pnr d_inddto ; run ;

NOTE: There were 53671 observations read from the data set WORK.GDM.
NOTE: The data set WORK.GDM has 53671 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
```

```
99
100        data gdm ( keep = pnr doGDM dno ) ;
101          set gdm ( rename = ( d_inddto = doGDM ) ) ;
102          by pnr doGDM ;
103          retain prevGDM ;
104          if  first.pnr then do ;
105              dno = 1 ;
106              prevGDM = doGDM ;
107              output ;
108              end ;
109          if ^first.pnr and ( doGDM - prevGDM ) gt &gdmint. then do ;
110              dno + 1 ;
111              output ;
112              prevGDM = doGDM ;
113              end ;
114        run ;

NOTE: There were 53671 observations read from the data set WORK.GDM.
NOTE: The data set WORK.GDM has 27882 observations and 3 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


115
116        proc transpose  data = gdm
117                         out = TTdata.gdm ( drop = _NAME_ _LABEL_ )
118                      prefix = doGDM ;
119          by pnr ;
120          var doGDM ;
121          id dno ;
122        run ;

NOTE: There were 27882 observations read from the data set WORK.GDM.
NOTE: The data set TTDATA.GDM has 21146 observations and 12 variables.
NOTE: PROCEDURE TRANSPOSE used (Total process time):
      real time           0.06 seconds
      cpu time            0.06 seconds


123
124        %let doGDMn = doGDM2 doGDM3 doGDM4 doGDM5 doGDM6 doGDM7 doGDM8 doGDM9 doGDM10
124      ! doGDM11 ;
125        title 'The recorded dates of Gestational diabetes' ;
126        proc contents  data = TTdata.gdm  ; run ;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

NOTE: The PROCEDURE CONTENTS printed page 4.

127        proc tabulate  data = TTdata.gdm  missing  noseps ;
128          class doGDM1 ;
129          var &doGDMn. ;
130          table doGDM1 all &doGDMn.,
131                n * f=comma10. / rts=9 ;
132          format doGDM1 year4. ;
133        run ;

NOTE: There were 21146 observations read from the data set TTDATA.GDM.
NOTE: The PROCEDURE TABULATE printed page 5.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


134        proc tabulate  data = TTdata.gdm  missing  noseps ;
```

```
135          where doGDM1 ge '01JAN2013'd ;
136          class doGDM1 ;
137          var &doGDMn. ;
138          table doGDM1 all &doGDMn.,
139                n * f=comma10. / rts=9 ;
140          format doGDM1 yymms7. ;
141        run ;

NOTE: There were 4312 observations read from the data set TTDATA.GDM.
      WHERE doGDM1>='01JAN2013'D;
NOTE: The PROCEDURE TABULATE printed page 6.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds


142        title1 ;
143
144        options mprint ;
145        data npr ( keep = pnr sex d_inddto nprtyp c_adiag ) ;
146          merge DM ( in = DM )
147                TTdata.gdm ;
148          by pnr ;
149          if DM ;
150          %xgdm( d_inddto ) ;
MPRINT(XGDM):   if ( doGDM1 - 30 ) < d_inddto < ( doGDM1 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM2 - 30 ) < d_inddto < ( doGDM2 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM3 - 30 ) < d_inddto < ( doGDM3 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM4 - 30 ) < d_inddto < ( doGDM4 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM5 - 30 ) < d_inddto < ( doGDM5 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM6 - 30 ) < d_inddto < ( doGDM6 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM7 - 30 ) < d_inddto < ( doGDM7 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM8 - 30 ) < d_inddto < ( doGDM8 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM9 - 30 ) < d_inddto < ( doGDM9 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM10 - 30 ) < d_inddto < ( doGDM10 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM11 - 30 ) < d_inddto < ( doGDM11 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM12 - 30 ) < d_inddto < ( doGDM12 + 365 ) then delete ;
151        run ;

NOTE: Variable doGDM12 is uninitialized.
NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      1300548 at 150:18   1300548 at 150:54   1332935 at 150:20   1332935 at 150:56
      1352536 at 150:20   1352536 at 150:56   1359343 at 150:20   1359343 at 150:56
      1360953 at 150:20   1360953 at 150:56   1361516 at 150:20   1361516 at 150:56
      1361597 at 150:20   1361597 at 150:56   1361626 at 150:20   1361626 at 150:56
      1361626 at 150:20   1361626 at 150:56   1361626 at 150:20   1361626 at 150:56
      1361626 at 150:20   1361626 at 150:56   1361627 at 150:20   1361627 at 150:56
NOTE: There were 1371306 observations read from the data set WORK.DM.
NOTE: There were 21146 observations read from the data set TTDATA.GDM.
NOTE: The data set WORK.NPR has 1361627 observations and 5 variables.
NOTE: DATA statement used (Total process time):
      real time           3.62 seconds
      cpu time            3.62 seconds


152        options nomprint ;
153
154        title1 'NPR diagnoses used - records' ;
155        proc tabulate  data = npr  missing  noseps ;
156          class c_adiag nprtyp d_inddto ;
157          table all c_adiag d_inddto,
158                nprtyp * f=comma10.
159                / rts = 10 ;
160          format d_inddto year4. ;
161        title1 ;
162

NOTE: There were 1361627 observations read from the data set WORK.NPR.
```

```
NOTE: The PROCEDURE TABULATE printed page 7.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.29 seconds
      cpu time            0.71 seconds


163        data npr1 ;
164          set npr ( keep = pnr d_inddto
165                  rename = ( d_inddto = doNPR ) ) ;
166          by pnr ;
167          if first.pnr ;
168        run ;

NOTE: There were 1361627 observations read from the data set WORK.NPR.
NOTE: The data set WORK.NPR1 has 223402 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.12 seconds
      cpu time            0.12 seconds


169
170        * Classifiy persons according to the most Frequently occurring type ;
171        data TTdata.npr ;
172          merge npr npr1 ;
173          by pnr ;
174          retain nT1 nT2 ;
175          if first.pnr then do ;
176             nT1 = 0 ;
177             nT2 = 0 ;
178             nRc = 0 ;
179             end ;
180          nT1 + ( nprtyp eq 'T1' ) ;
181          nT2 + ( nprtyp eq 'T2' ) ;
182          nRc + 1 ;
183          * If more than half of records agree on one type ;
184          if last.pnr then do ;
185             nprtyp = 'NA' ;
186             if nT1 > nRc/2 then nprtyp = 'T1' ;
187             if nT2 > nRc/2 then nprtyp = 'T2' ;
188             output ;
189             end ;
190          label doNPR = ' ' ;
191        run ;

NOTE: There were 1361627 observations read from the data set WORK.NPR.
NOTE: There were 223402 observations read from the data set WORK.NPR1.
NOTE: The data set TTDATA.NPR has 223402 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.26 seconds
      cpu time            0.23 seconds


192
193        title1 'Diagnoses of DM accepted from NPR' ;
194        proc contents  data = TTdata.npr ; run ;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

NOTE: The PROCEDURE CONTENTS printed page 8.

195
196        proc tabulate  data = TTdata.npr  missing  noseps ;
197          class doNPR nprtyp sex ;
198          table all doNPR,
199              ( all sex nprtyp ) * f=comma10.
200              / rts = 7 ;
201          format doNPR year4.
202                  sex koen_t. ;
```

```
203         run ;

NOTE: There were 223402 observations read from the data set TTDATA.NPR.
NOTE: The PROCEDURE TABULATE printed page 9.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.12 seconds
      cpu time            0.12 seconds


204
205        title2 '- only from 1 January 2013 - checking seasonality' ;
206        proc tabulate  data = TTdata.npr  missing  noseps ;
207          where doNPR ge '01JAN2013'd ;
208          class doNPR nprtyp sex ;
209          table all doNPR,
210              ( all sex nprtyp ) * f=comma10.
211              / rts = 10 ;
212          format doNPR yymms8.
213                  sex koen_t. ;
214          run ;

NOTE: There were 24630 observations read from the data set TTDATA.NPR.
      WHERE doNPR>='01JAN2013'D;
NOTE: The PROCEDURE TABULATE printed page 10.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds


215        title1 ;

NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           50.45 seconds
      cpu time            21.26 seconds
```

7.5.1 01-npr.lst

```
PCOS records - id and first date of PCOS              17:15 Thursday, January 12, 2017   1

The CONTENTS Procedure

Data Set Name        TTDATA.PCOS                   Observations          20773
Member Type          DATA                          Variables             *
Engine               V9                            Indexes               0
Created              12/01/2017 17:15:49           Observation Length    24
Last Modified        12/01/2017 17:15:49           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    8
First Data Page             *
Max Obs per Page            2715
Obs in First Data Page      2664
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\pcos.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12
```

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
2	doPCOS	Num	8			
1	pnr	Char	12	$12.	$10.	Personnummer

doPCOS	N
1976	*
1977	577
1978	607
1979	590
1980	648
1981	741
1982	682
1983	757
1984	717
1985	757
1986	742
1987	610
1988	637
1989	630
1990	589
1991	574
1992	535
1993	459
1994	109
1995	108
1996	153
1997	153
1998	188
1999	229
2000	193
2001	212
2002	272
2003	378
2004	479
2005	532
2006	538
2007	631
2008	566
2009	695
2010	732
2011	705
2012	691
2013	789
2014	832
2015	733

doPCOS	N
2013/01	73
2013/02	54
2013/03	58
2013/04	66
2013/05	68
2013/06	65
2013/07	45
2013/08	67
2013/09	74
2013/10	85
2013/11	67

```
 2013/12          67
 2014/01          82
 2014/02          68
 2014/03          60
 2014/04          61
 2014/05          71
 2014/06          69
 2014/07          49
 2014/08          68
 2014/09          82
 2014/10          78
 2014/11          74
 2014/12          70
 2015/01          71
 2015/02          49
 2015/03          79
 2015/04          64
 2015/05          72
 2015/06          79
 2015/07          38
 2015/08          58
 2015/09          67
 2015/10          61
 2015/11          54
 2015/12          41


The CONTENTS Procedure

Data Set Name        TTDATA.GDM                    Observations          21146
Member Type          DATA                          Variables             12
Engine               V9                            Indexes               0
Created              12/01/2017 17:15:49           Observation Length    104
Last Modified        12/01/2017 17:15:49           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    34
First Data Page             *
Max Obs per Page            629
Obs in First Data Page      607
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\gdm.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12


              Alphabetic List of Variables and Attributes

 #    Variable    Type    Len    Format       Informat    Label

 *    doGDM1      Num       8    DDMMYY10.
 *    doGDM2      Num       8    DDMMYY10.
 *    doGDM3      Num       8    DDMMYY10.
 *    doGDM4      Num       8    DDMMYY10.
 6    doGDM5      Num       8    DDMMYY10.
 7    doGDM6      Num       8    DDMMYY10.
 8    doGDM7      Num       8    DDMMYY10.
 9    doGDM8      Num       8    DDMMYY10.
10    doGDM9      Num       8    DDMMYY10.
11    doGDM10     Num       8    DDMMYY10.
```

```
12    doGDM11     Num       8    DDMMYY10.
 *    pnr         Char     12    $12.         $10.        Personnummer
```

	N
doGDM1	
1980	*
1987	51
1988	82
1989	108
1990	96
1991	84
1992	89
1993	160
1994	474
1995	435
1996	604
1997	591
1998	681
1999	546
2000	583
2001	585
2002	590
2003	774
2004	949
2005	1,020
2006	1,041
2007	1,112
2008	1,250
2009	1,354
2010	1,231
2011	1,245
2012	1,098
2013	1,347
2014	1,478
2015	1,487
All	21,146
doGDM2	5,108
doGDM3	1,218
doGDM4	292
doGDM5	82
doGDM6	23
doGDM7	8
doGDM8	*
doGDM9	*
doGDM10	*
doGDM11	*

	N
doGDM1	
2013/01	110
2013/02	108
2013/03	100
2013/04	128
2013/05	104
2013/06	126
2013/07	101
2013/08	123
2013/09	107
2013/10	104
2013/11	113
2013/12	123
2014/01	143
2014/02	110

2014/03	133
2014/04	133
2014/05	150
2014/06	128
2014/07	123
2014/08	121
2014/09	116
2014/10	101
2014/11	109
2014/12	111
2015/01	164
2015/02	131
2015/03	139
2015/04	103
2015/05	122
2015/06	148
2015/07	136
2015/08	140
2015/09	109
2015/10	107
2015/11	118
2015/12	70
All	4,312
doGDM2	240
doGDM3	12
doGDM4	0
doGDM5	0
doGDM6	0
doGDM7	0
doGDM8	0
doGDM9	0
doGDM10	0
doGDM11	0

	nprtyp		
		T1	T2
	N	N	N
All	217,741	482,966	660,920
Aktions-diagnose			
24900	.	5,914	.
24901	.	2,260	.
24902	.	366	.
24903	.	224	.
24904	.	96	.
24905	.	715	.
24906	.	70	.
24907	.	1,325	.
24908	.	2,512	.
24909	.	21,160	.
25000	.	.	19,663
25001	.	.	4,017
25002	.	.	320
25003	.	.	642
25004	.	.	86
25005	.	.	919
25006	.	.	113
25007	.	.	1,879
25008	.	.	4,768
25009	.	.	46,640
DE10	.	1,314	.
DE100	.	3,842	.
DE100A	.	25	.
DE100B	.	402	.

DE100C	.	34	.
DE100D	.	87	.
DE100E	.	504	.
DE100F	.	236	.
DE101	.	20,757	.
DE102	.	16,513	.
DE103	.	29,194	.
DE104	.	9,241	.
DE105	.	14,990	.
DE105A	.	258	.
DE105B	.	3,631	.
DE105C	.	1,186	.
DE105D	.	27	.
DE106	.	3,622	.
DE107	.	40,553	.
DE108	.	42,135	.
DE109	.	252,402	.
DE109A	.	7,371	.
DE11	.	.	836
DE110	.	.	3,014
DE110A	.	.	21
DE110B	.	.	167
DE110C	.	.	54
DE110D	.	.	46
DE110E	.	.	45
DE111	.	.	1,668
DE112	.	.	34,802
DE113	.	.	15,172
DE114	.	.	20,538
DE115	.	.	21,169
DE115A	.	.	446
DE115B	.	.	10,122
DE115C	.	.	1,387
DE115D	.	.	61
DE116	.	.	5,600
DE117	.	.	44,029
DE118	.	.	68,759
DE119	.	.	341,690
DE119A	.	.	12,247
DE12	8	.	.
DE120	225	.	.
DE120A	10	.	.
DE120B	10	.	.
DE120C	*	.	.
DE121	156	.	.
DE122	111	.	.
DE123	87	.	.
DE124	80	.	.
DE125	404	.	.
DE125A	*	.	.
DE125B	40	.	.
DE125C	42	.	.
DE126	39	.	.
DE127	91	.	.
DE128	134	.	.
DE129	298	.	.
DE13	45	.	.
DE130	77	.	.
DE131	510	.	.
DE132	328	.	.
DE133	1,667	.	.
DE134	309	.	.
DE135	213	.	.
DE135A	*	.	.
DE135B	141	.	.
DE135C	35	.	.
DE135D	*	.	.
DE136	191	.	.
DE137	521	.	.
DE138	869	.	.

DE139	5,543	.	.
DE14	378	.	.
DE140	726	.	.
DE140A	53	.	.
DE140B	12	.	.
DE140C	20	.	.
DE140D	20	.	.
DE141	2,262	.	.
DE142	898	.	.
DE143	3,458	.	.
DE144	3,909	.	.
DE145	5,941	.	.
DE145A	30	.	.
DE145B	2,437	.	.
DE145C	449	.	.
DE145D	17	.	.
DE146	476	.	.
DE147	1,375	.	.
DE148	5,738	.	.
DE149	39,751	.	.
DE748	213	.	.
DE748A	6	.	.
DE748B	41	.	.
DE748C	9	.	.
DE748E	9	.	.
DE748G	*	.	.
DE891	310	.	.
DE891A	80	.	.
DE891B	*	.	.
DH360	113,008	.	.
DH360A	1,506	.	.
DH360B	2,877	.	.
DH360C	1,724	.	.
DH360D	1,471	.	.
DH360E	1,271	.	.
DH360F	2,543	.	.
DH360H	3,230	.	.
DH360J	3,150	.	.
DH360K	4,501	.	.
DR730	1,048	.	.
DR730A	342	.	.
DR819	253	.	.
Indl-elsesda-to			
1941	.	*	.
1968	.	*	.
1970	.	9	.
1971	.	9	.
1972	.	14	*
1973	.	46	*
1974	.	37	*
1975	.	28	*
1976	.	28	48
1977	.	10	3,228
1978	.	63	4,094
1979	.	83	4,398
1980	*	104	4,700
1981	.	118	5,039
1982	.	90	5,598
1983	*	147	5,566
1984	*	107	6,381
1985	*	145	7,411
1986	6	281	7,777
1987	*	4,347	3,592
1988	57	5,194	3,509
1989	69	5,717	3,498
1990	59	5,774	3,459
1991	63	6,570	3,941
1992	117	6,952	4,265

1993	651	13,835	7,484
1994	2,475	13,872	9,790
1995	3,624	14,623	11,248
1996	4,006	14,301	13,100
1997	5,041	15,576	14,039
1998	7,681	15,301	16,039
1999	7,572	17,561	18,330
2000	6,537	15,504	18,510
2001	7,167	16,213	19,139
2002	7,495	16,556	19,620
2003	9,580	19,142	23,752
2004	10,608	19,287	22,988
2005	11,653	26,729	28,494
2006	11,939	21,130	30,991
2007	12,545	17,671	26,508
2008	19,695	43,773	59,929
2009	18,985	34,241	53,635
2010	13,674	19,625	33,534
2011	13,707	23,212	33,566
2012	11,736	15,259	24,028
2013	12,108	23,470	42,929
2014	10,824	13,700	27,644
2015	8,052	16,507	29,110

The CONTENTS Procedure

Data Set Name	TTDATA.NPR	Observations	223402
Member Type	DATA	Variables	9
Engine	V9	Indexes	0
Created	12/01/2017 17:15:53	Observation Length	72
Last Modified	12/01/2017 17:15:53	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	65536
Number of Data Set Pages	247
First Data Page	*
Max Obs per Page	908
Obs in First Data Page	881
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	E:\workdata\705093\BxC\demoDM\DATA\npr.sas7bdat
Release Created	9.0401M3
Host Created	X64_SRV12

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
2	C_ADIAG	Char	6	$20.	$20.	Aktionsdiagnose
3	D_INDDTO	Num	8	DDMMYY10.	DATE9.	Indlelsesdato
6	doNPR	Num	8	DDMMYY10.	DATE9.	
9	nRc	Num	8			
7	nT1	Num	8			
8	nT2	Num	8			
4	nprtyp	Char	*			
1	pnr	Char	12	$12.	$10.	Personnummer
5	sex	Num	8			

		sex			nprtyp	
	All	Mand	Kvinde	NA	T1	T2
	N	N	N	N	N	N
All	223,402	126,834	96,568	37,839	44,535	141,028
doNPR						
1941	*	.	*	.	*	.
1968	*	*	.	.	*	.
1970	*	*	*	.	*	.
1971	*	.	*	*	*	.
1972	9	*	7	*	7	*
1973	19	12	7	*	15	*
1974	18	9	9	*	12	*
1975	16	10	6	8	7	*
1976	69	33	36	14	41	14
1977	2,414	1,170	1,244	465	1,241	708
1978	2,289	1,145	1,144	448	1,148	693
1979	2,066	1,045	1,021	362	1,017	687
1980	1,987	1,041	946	381	896	710
1981	1,835	972	863	320	805	710
1982	2,014	1,071	943	311	839	864
1983	1,881	950	931	321	745	815
1984	2,084	992	1,092	327	847	910
1985	2,166	1,145	1,021	342	827	997
1986	2,265	1,218	1,047	336	943	986
1987	2,314	1,206	1,108	312	975	1,027
1988	2,430	1,340	1,090	319	1,040	1,071
1989	2,605	1,362	1,243	352	1,048	1,205
1990	2,760	1,508	1,252	351	1,049	1,360
1991	3,062	1,644	1,418	393	1,139	1,530
1992	3,240	1,723	1,517	423	1,113	1,704
1993	5,127	2,927	2,200	774	1,706	2,647
1994	5,996	3,288	2,708	1,027	1,520	3,449
1995	6,457	3,627	2,830	1,120	1,466	3,871
1996	6,791	3,832	2,959	1,137	1,403	4,251
1997	7,038	3,925	3,113	1,119	1,403	4,516
1998	7,899	4,406	3,493	1,316	1,466	5,117
1999	8,021	4,439	3,582	1,398	1,340	5,283
2000	8,109	4,525	3,584	1,278	1,313	5,518
2001	8,094	4,580	3,514	1,440	1,290	5,364
2002	8,307	4,706	3,601	1,607	1,271	5,429
2003	9,356	5,378	3,978	1,732	1,245	6,379
2004	9,276	5,333	3,943	1,723	1,296	6,257
2005	9,041	5,108	3,933	1,604	1,236	6,201
2006	9,341	5,413	3,928	1,797	1,259	6,285
2007	9,210	5,317	3,893	1,667	1,210	6,333
2008	8,742	5,167	3,575	1,486	1,196	6,060
2009	8,624	5,158	3,466	1,469	1,178	5,977
2010	8,635	5,202	3,433	1,314	1,181	6,140
2011	8,860	5,286	3,574	1,399	1,129	6,332
2012	8,297	4,871	3,426	1,462	916	5,919
2013	8,314	4,958	3,356	1,429	935	5,950
2014	8,114	4,873	3,241	1,396	902	5,816
2015	8,202	4,914	3,288	1,352	913	5,937

\- only from * January 2013 - checking seasonality

		sex			nprtyp	
	All	Mand	Kvinde	NA	T1	T2
	N	N	N	N	N	N
All	24,630	14,745	9,885	4,177	2,750	17,703
doNPR						
2013/01	823	501	322	139	107	577

2013/02	678	409	269	116	89	473
2013/03	730	452	278	129	91	510
2013/04	774	478	296	145	87	542
2013/05	775	451	324	137	74	564
2013/06	658	395	263	110	74	474
2013/07	535	296	239	72	57	406
2013/08	641	372	269	107	73	461
2013/09	671	389	282	111	58	502
2013/10	730	430	300	130	71	529
2013/11	712	416	296	124	85	503
2013/12	587	369	218	109	69	409
2014/01	825	507	318	155	91	579
2014/02	629	403	226	117	79	433
2014/03	790	470	320	148	89	553
2014/04	622	390	232	105	74	443
2014/05	671	375	296	112	92	467
2014/06	618	348	270	108	50	460
2014/07	532	326	206	101	58	373
2014/08	567	356	211	84	72	411
2014/09	734	427	307	137	79	518
2014/10	760	471	289	136	70	554
2014/11	757	437	320	101	87	569
2014/12	609	363	246	92	61	456
2015/01	743	481	262	93	84	566
2015/02	699	438	261	110	82	507
2015/03	774	463	311	133	77	564
2015/04	628	388	240	106	65	457
2015/05	708	412	296	121	81	506
2015/06	729	424	305	116	74	539
2015/07	525	303	222	93	63	369
2015/08	610	339	271	108	61	441
2015/09	792	483	309	131	82	579
2015/10	719	434	285	136	87	496
2015/11	727	424	303	132	75	520
2015/12	548	325	223	73	82	393

7.6 02-dvdd

The DVDD contains annual records for diabetes patients, mostly from out-patient clinics, but (eventually) also from GPs. There records contain type and date of diagnosis. The program chooses the earliest reported date of diagnosis and the type of diabetes reported more than half of the times (`dvdtyp`).

Uses the GDM dates to exclude possible inclusion dates in GDM grace periods.

```
1                                                    "Program 02-dvdd.sas"

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_SRV12  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.06 seconds
      cpu time            0.09 seconds
```

```
NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          proc sort data = EKST15.DVDD  out = dvdd ;
2            by pnr status_dato diag_dato diag_type ;
3          run ;

NOTE: There were 349662 observations read from the data set EKST15.DVDD.
NOTE: The data set WORK.DVDD has 349662 observations and 53 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           10.89 seconds
      cpu time            0.70 seconds


4
5          * check number of *persons* in the data set ;
6          proc sort data = dvdd  out = pers  nodupkey ;
7            by pnr ;
8          run ;

NOTE: There were 349662 observations read from the data set WORK.DVDD.
NOTE: 260882 observations with duplicate key values were deleted.
NOTE: The data set WORK.PERS has 88780 observations and 53 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.32 seconds
      cpu time            0.40 seconds


9
10         options mprint ;
11         * only persons in base and included before 1.1.2015 ;
12         data dvdd     ;
13           merge dvdd       ( in = dvdd )
14                 TTDATA.pop ( in = pop )
15                 TTdata.gdm ;
16           by pnr ;
17           if pop and dvdd ;
18           * remove status records after the cut date ;
19           if status_dato > &cutdate. then delete ;
20           * Do not count diagnosis or status dates in GDM grace period ;
21           %xgdm(   diag_dato ) ;
MPRINT(XGDM):   if ( doGDM1 - 30 ) < diag_dato < ( doGDM1 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM2 - 30 ) < diag_dato < ( doGDM2 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM3 - 30 ) < diag_dato < ( doGDM3 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM4 - 30 ) < diag_dato < ( doGDM4 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM5 - 30 ) < diag_dato < ( doGDM5 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM6 - 30 ) < diag_dato < ( doGDM6 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM7 - 30 ) < diag_dato < ( doGDM7 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM8 - 30 ) < diag_dato < ( doGDM8 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM9 - 30 ) < diag_dato < ( doGDM9 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM10 - 30 ) < diag_dato < ( doGDM10 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM11 - 30 ) < diag_dato < ( doGDM11 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM12 - 30 ) < diag_dato < ( doGDM12 + 365 ) then delete ;
22           %xgdm( status_dato ) ;
MPRINT(XGDM):   if ( doGDM1 - 30 ) < status_dato < ( doGDM1 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM2 - 30 ) < status_dato < ( doGDM2 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM3 - 30 ) < status_dato < ( doGDM3 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM4 - 30 ) < status_dato < ( doGDM4 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM5 - 30 ) < status_dato < ( doGDM5 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM6 - 30 ) < status_dato < ( doGDM6 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM7 - 30 ) < status_dato < ( doGDM7 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM8 - 30 ) < status_dato < ( doGDM8 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM9 - 30 ) < status_dato < ( doGDM9 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM10 - 30 ) < status_dato < ( doGDM10 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM11 - 30 ) < status_dato < ( doGDM11 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM12 - 30 ) < status_dato < ( doGDM12 + 365 ) then delete ;
```

```
23         run ;

NOTE: Variable doGDM12 is uninitialized.
NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      328133 at 21:18   328133 at 21:54   338200 at 21:20   338200 at 21:56
      344872 at 21:20   344872 at 21:56   347392 at 21:20   347392 at 21:56
      348024 at 21:20   348024 at 21:56   348206 at 21:20   348206 at 21:56
      348239 at 21:20   348239 at 21:56   348244 at 21:20   348244 at 21:56
      348244 at 21:20   348244 at 21:56   348244 at 21:20   348244 at 21:56
      348244 at 21:20   348244 at 21:56   348244 at 21:20   348244 at 21:56
      328133 at 22:18   328133 at 22:54   337789 at 22:20   337789 at 22:56
      343805 at 22:20   343805 at 22:56   345987 at 22:20   345987 at 22:56
      346491 at 22:20   346491 at 22:56   346636 at 22:20   346636 at 22:56
      346662 at 22:20   346662 at 22:56   346666 at 22:20   346666 at 22:56
      346666 at 22:20   346666 at 22:56   346666 at 22:20   346666 at 22:56
      346666 at 22:20   346666 at 22:56   346666 at 22:20   346666 at 22:56
NOTE: There were 349662 observations read from the data set WORK.DVDD.
NOTE: There were 7361669 observations read from the data set TTDATA.POP.
NOTE: There were 21146 observations read from the data set TTDATA.GDM.
NOTE: The data set WORK.DVDD has 346666 observations and 68 variables.
NOTE: DATA statement used (Total process time):
      real time           6.98 seconds
      cpu time            4.34 seconds


24         options nomprint ;
25
26         * clean out multiple status dates and return a date of diagnosis ;
27         data dvdd                                /* All records */
28              dvdd_fix ( keep = pnr doDVDD ) ; /* one per pnr with revised date of DM
28       ! diagnosis */
29           set dvdd    ( keep = pnr status_dato diag_dato diag_type doBth doDth ) ;
30           by pnr status_dato diag_dato diag_type ;
31           retain doDVDD ;
32           * use only the first among identical status dates within each person ;
33           if first.status_dato ;
34           * set the revised DM date to the earlier of diag_dato and status dates ;
35           if first.pnr then doDVDD = min(         diag_dato, status_dato ) ;
36                        else doDVDD = min( doDVDD, diag_dato, status_dato ) ;
37                            output dvdd ;
38           if last.pnr then output dvdd_fix ;
39         run ;

NOTE: There were 346666 observations read from the data set WORK.DVDD.
NOTE: The data set WORK.DVDD has 344543 observations and 7 variables.
NOTE: The data set WORK.DVDD_FIX has 87975 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.40 seconds
      cpu time            0.23 seconds


40
41         * add the computed doDVDD date to the status records ;
42         data dvdd     ;
43           merge dvdd
44                 dvdd_fix ;
45           by pnr ;
46         run ;

NOTE: There were 344543 observations read from the data set WORK.DVDD.
NOTE: There were 87975 observations read from the data set WORK.DVDD_FIX.
NOTE: The data set WORK.DVDD has 344543 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.12 seconds
      cpu time            0.10 seconds


47
```

```
48         * DVDD will provide classification of follow-up as T1 / *not* T1 (=T2) ;
49         * tabulation of the sequences of type classifications occurring ;
50         proc sort  data = dvdd  out = dvdd_type ;
51           by pnr status_dato ;
52         run ;

NOTE: There were 344543 observations read from the data set WORK.DVDD.
NOTE: The data set WORK.DVDD_TYPE has 344543 observations and 7 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.07 seconds
      cpu time            0.07 seconds


53
54         data dvdd_type ( keep = pnr doDVDD status_dato typ )
55              dvdd_hist ( keep = pnr hist ) ;
56           set dvdd_type ;
57           by pnr ;
58           length typ $ 4  hist $ 80 ;
59           retain hist ;
60           typ = substr( diag_type, 1, 2 ) ;
61           if typ eq "Ty" then typ = "T" || substr( diag_type, 6, 1 ) ;
62           if  first.pnr then hist = typ ;
63           if ^first.pnr and ( diag_type ne lag(diag_type) )
64                         then hist = trim(hist) || " " || typ ;
65           output dvdd_type ;
66           if last.pnr then output dvdd_hist ;
67         run ;

NOTE: There were 344543 observations read from the data set WORK.DVDD_TYPE.
NOTE: The data set WORK.DVDD_TYPE has 344543 observations and 4 variables.
NOTE: The data set WORK.DVDD_HIST has 88398 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           0.14 seconds
      cpu time            0.11 seconds


68
69         * classification rule: if more than half of registrations T1 then T1 ;
70         *                       if more than half of registrations T2 then T2 ;
71         data dvdd ( keep = pnr doDVDD dvdtyp nT1 nT2 nRc ) ;
72           set dvdd_type ;
73           by pnr status_dato ;
74           retain nT1 nT2 ;
75           if first.pnr then do ;
76              nT1 = 0 ;
77              nT2 = 0 ;
78              nRc = 0 ;
79              end ;
80           nT1 + ( typ eq "T1" ) ;
81           nT2 + ( typ eq "T2" ) ;
82           nRc + 1 ;
83           * If most and more than half of records agree on one type ;
84           if last.pnr then do ;
85              if nT1 > nRc/2 then dvdtyp = 'T1' ; else
86              if nT2 > nRc/2 then dvdtyp = 'T2' ; else dvdtyp = 'NA' ;
87              output ;
88              end ;
89         run ;

NOTE: There were 344543 observations read from the data set WORK.DVDD_TYPE.
NOTE: The data set WORK.DVDD has 88398 observations and 6 variables.
NOTE: DATA statement used (Total process time):
      real time           0.07 seconds
      cpu time            0.07 seconds


90
91         data TTdata.dvdd ;
92           merge dvdd
```

```
93                dvdd_hist ( keep = pnr hist ) ;
94            by pnr ;
95          run ;

NOTE: There were 88398 observations read from the data set WORK.DVDD.
NOTE: There were 88398 observations read from the data set WORK.DVDD_HIST.
NOTE: The data set TTDATA.DVDD has 88398 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           0.20 seconds
      cpu time            0.03 seconds


96
97          title1 'Dates and types from DVDD' ;
98          proc contents  data = TTdata.dvdd ; run ;

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

NOTE: The PROCEDURE CONTENTS printed page 1.

99          title2 'Classification based on most frequent type recorded in DVDD - persons'
99        ! ;
100         proc tabulate  data = TTdata.dvdd   missing  noseps ;
101           class dvdtyp doDVDD nT1 nT2 NRc ;
102           table all doDVDD,
103               ( all dvdtyp ) * f=comma9.
104               / rts = 8 ;
105           table nRc * nT1,
106                 nT2 * f=5.
107                 / rts = 5  indent = 1 ;
108           format doDVDD year4. ;
109         run ;

NOTE: There were 88398 observations read from the data set TTDATA.DVDD.
NOTE: The PROCEDURE TABULATE printed pages 2-3.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.07 seconds
      cpu time            0.03 seconds


110
111         proc tabulate  data = TTdata.dvdd   missing  noseps  order = freq ;
112           class dvdtyp hist ;
113           table all hist="sequence of different types",
114               ( all dvdtyp ) * f=comma7.
115                 / rts = 30 ;
116         run ;

NOTE: There were 88398 observations read from the data set TTDATA.DVDD.
NOTE: The PROCEDURE TABULATE printed page 4.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.01 seconds
      cpu time            0.04 seconds


117         title1 ;

NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           19.64 seconds
      cpu time            6.39 seconds
```

7.6.1 02-dvdd.lst

```
Dates and types from DVDD                          14:17 Tuesday, January 24, 2017   1
```

```
The CONTENTS Procedure

Data Set Name        TTDATA.DVDD                   Observations          88398
Member Type          DATA                          Variables             7
Engine               V9                            Indexes               0
Created              24/01/2017 14:18:15           Observation Length    128
Last Modified        24/01/2017 14:18:15           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    174
First Data Page             *
Max Obs per Page            511
Obs in First Data Page      498
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\dvdd.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12


            Alphabetic List of Variables and Attributes

#    Variable    Type    Len    Format    Informat    Label

2    doDVDD      Num       8
6    dvdtyp      Char      *
7    hist        Char     80
5    nRc         Num       8
3    nT1         Num       8
4    nT2         Num       8
1    pnr         Char     12    $12.      $10.        Personnummer

Dates and types from DVDD                              14:17 Tuesday, January 24, 2017   2
Classification based on most frequent type recorded in DVDD - persons


                              dvdtyp

           All        NA         T1         T2

            N          N          N          N

 All       88,398     3,555     23,947     60,896
 doDVDD
 1890           *         .          .          *
 1899           *         .          .          *
 1900          57         *         11         45
 1901           *         .          .          *
 1909           *         *          .          .
 1919           *         .          .          *
 1920          13         .          *          9
 1923           *         *          .          .
 1927           *         .          *          .
 1931           *         .          .          *
 1933           *         .          *          .
 1934           *         .          *          .
 1936           *         .          *          .
 1937           *         .          *          .
 1938           *         .          *          .
 1939           *         .          *          *
 1940           6         .          6          .
 1941           *         .          *          .
```

1942	8	.	7	*
1943	*	.	*	.
1944	9	.	9	.
1945	16	.	15	*
1946	13	.	12	*
1947	21	.	21	.
1948	28	*	26	*
1949	26	.	26	.
1950	37	.	34	*
1951	34	.	34	.
1952	36	.	31	*
1953	40	*	38	*
1954	51	*	47	*
1955	80	*	74	*
1956	117	.	114	*
1957	86	.	84	*
1958	109	.	104	*
1959	101	*	94	*
1960	176	*	163	12
1961	173	*	162	6
1962	182	*	169	11
1963	191	*	179	10
1964	182	*	165	15
1965	185	*	170	13
1966	200	*	179	17
1967	211	*	188	22
1968	213	*	191	21
1969	237	*	207	28
1970	350	*	283	65
1971	301	*	261	36
1972	386	*	309	74
1973	352	6	291	55
1974	390	6	313	71
1975	430	7	316	107
1976	434	*	343	86
1977	468	7	351	110
1978	567	6	414	147
1979	545	9	392	144
1980	884	19	460	405
1981	590	7	411	172
1982	714	18	410	286
1983	711	8	417	286
1984	755	13	417	325
1985	1,058	19	420	619
1986	970	12	467	491
1987	1,055	22	476	557
1988	1,144	24	470	650
1989	1,170	30	499	641
1990	2,015	41	590	1,384
1991	1,407	36	543	828
1992	1,860	31	535	1,294
1993	1,708	30	501	1,177
1994	1,967	45	599	1,323
1995	2,664	56	594	2,014
1996	2,381	55	588	1,738
1997	2,499	65	634	1,800
1998	3,079	53	635	2,391
1999	2,825	62	556	2,207
2000	3,847	86	627	3,134
2001	3,143	79	668	2,396
2002	3,267	85	588	2,594
2003	3,317	96	544	2,677
2004	3,646	114	584	2,948
2005	3,889	104	566	3,219
2006	4,061	181	613	3,267
2007	3,989	177	576	3,236
2008	4,037	218	614	3,205
2009	3,649	290	533	2,826
2010	3,503	231	527	2,745
2011	2,999	231	432	2,336

```
2012       2,533       286       401     1,846
2013       2,180       324       341     1,515
2014       1,639       282       237     1,120
2015         154        38        18        98
```

Dates and types from DVDD 14:17 Tuesday, January 24, 2017 3
Classification based on most frequent type recorded in DVDD - persons

	nT2													
	0	*	*	*	*	*	6	7	8	9	10	11	12	13
	N	N	N	N	N	N	N	N	N	N	N	N	N	N
*														
0	1288	20100	.	.	.	.	.	.	.	.	.	.	.	.
*	2259	.	.	.	.	.	.	.	.	.	.	.	.	.
*														
0	445	220	10634	.	.	.	.	.	.	.	.	.	.	.
*	58	173	.	.	.	.	.	.	.	.	.	.	.	.
*	1934	.	.	.	.	.	.	.	.	.	.	.	.	.
*														
0	251	79	160	7239	.	.	.	.	.	.	.	.	.	.
*	24	7	133	.	.	.	.	.	.	.	.	.	.	.
*	29	107	.	.	.	.	.	.	.	.	.	.	.	.
*	1992	.	.	.	.	.	.	.	.	.	.	.	.	.
*														
0	109	32	53	104	5229	.	.	.	.	.	.	.	.	.
*	8	*	10	146	.	.	.	.	.	.	.	.	.	.
*	9	8	79	.	.	.	.	.	.	.	.	.	.	.
*	37	116	.	.	.	.	.	.	.	.	.	.	.	.
*	1819	.	.	.	.	.	.	.	.	.	.	.	.	.
*														
0	68	14	28	22	72	4614	.	.	.	.	.	.	.	.
*	*	*	*	*	100	.	.	.	.	.	.	.	.	.
*	8	*	*	85	.	.	.	.	.	.	.	.	.	.
*	17	*	82	.	.	.	.	.	.	.	.	.	.	.
*	30	119	.	.	.	.	.	.	.	.	.	.	.	.
*	2194	.	.	.	.	.	.	.	.	.	.	.	.	.
6														
0	51	14	*	18	26	66	3534	.	.	.	.	.	.	.
*	*	*	*	*	*	122	.	.	.	.	.	.	.	.
*	10	*	.	*	96	.	.	.	.	.	.	.	.	.
*	7	.	*	80	.	.	.	.	.	.	.	.	.	.
*	9	*	59	.	.	.	.	.	.	.	.	.	.	.
*	35	84	.	.	.	.	.	.	.	.	.	.	.	.
6	2472	.	.	.	.	.	.	.	.	.	.	.	.	.
7														
0	60	10	9	7	6	16	55	2822	.	.	.	.	.	.
*	6	.	*	*	*	*	101	.	.	.	.	.	.	.
*	7	.	*	*	*	104	.	.	.	.	.	.	.	.
*	6	*	*	*	65	.	.	.	.	.	.	.	.	.
*	*	*	*	66	.	.	.	.	.	.	.	.	.	.
*	9	*	83	.	.	.	.	.	.	.	.	.	.	.
6	27	101	.	.	.	.	.	.	.	.	.	.	.	.
7	3143	.	.	.	.	.	.	.	.	.	.	.	.	.
8														
0	40	9	*	*	6	16	9	25	2382	.	.	.	.	.
*	*	.	.	*	.	.	*	53	.	.	.	.	.	.
*	.	*	.	*	.	*	50	.	.	.	.	.	.	.
*	*	*	*	*	*	57	.	.	.	.	.	.	.	.
*	*	.	.	*	64	.	.	.	.	.	.	.	.	.
*	6	*	*	79	.	.	.	.	.	.	.	.	.	.
6	12	*	60	.	.	.	.	.	.	.	.	.	.	.
7	24	67	.	.	.	.	.	.	.	.	.	.	.	.
8	2963	.	.	.	.	.	.	.	.	.	.	.	.	.
9														
0	46	*	*	*	*	*	*	7	17	2103	.	.	.	.

```
 *      *      .      .      *      *      .      .      *     59      .      .      .      .      .
 *      *      .      *      .      .      *      *     34      .      .      .      .      .      .
 *      *      .      .      .      .      *     49      .      .      .      .      .      .      .
 *      *      .      .      *      *     41      .      .      .      .      .      .      .      .
 *      *      *      *      *     55      .      .      .      .      .      .      .      .      .
 6      6      *      .     39      .      .      .      .      .      .      .      .      .      .
 7      7      *     33      .      .      .      .      .      .      .      .      .      .      .
 8     24     70      .      .      .      .      .      .      .      .      .      .      .      .
 9   3255      .      .      .      .      .      .      .      .      .      .      .      .      .
10
 0      *      .      .      .      .      .      *      *      .      *    257      .      .      .
 *      .      .      .      .      .      .      .      *      .      7      .      .      .      .
 *      .      *      .      .      .      .      .      .      *      .      .      .      .      .
 *      .      .      .      .      .      .      .      *      .      .      .      .      .      .
 *      .      .      .      .      .      .      *      .      .      .      .      .      .      .
 *      .      .      .      .      .      *      .      .      .      .      .      .      .      .
 6      .      .      .      *      *      .      .      .      .      .      .      .      .      .
 7      .      .      *      *      .      .      .      .      .      .      .      .      .      .
 8      *      *      6      .      .      .      .      .      .      .      .      .      .      .
 9     11      8      .      .      .      .      .      .      .      .      .      .      .      .
 10   319      .      .      .      .      .      .      .      .      .      .      .      .      .
11
 0      .      .      .      .      .      *      .      .      .      .      *     20      .      .
 *      .      .      .      .      .      .      .      .      .      .      *      .      .      .
 *      .      .      .      .      .      .      .      .      .      *      .      .      .      .
 9      .      .      *      .      .      .      .      .      .      .      .      .      .      .
 10     *      *      .      .      .      .      .      .      .      .      .      .      .      .
 11    30      .      .      .      .      .      .      .      .      .      .      .      .      .
12
 0      .      .      .      .      .      .      .      .      .      .      .      .      *      .
 10     .      .      *      .      .      .      .      .      .      .      .      .      .      .
 12     *      .      .      .      .      .      .      .      .      .      .      .      .      .
13
 0      .      .      .      .      .      .      .      .      .      .      .      .      .      *
 13     *      .      .      .      .      .      .      .      .      .      .      .      .      .
```

```
Dates and types from DVDD                                14:17 Tuesday, January 24, 2017   4
Classification based on most frequent type recorded in DVDD - persons
```

	dvdtyp			
	All	T2	T1	NA
	N	N	N	N
All	88,398	60,896	23,947	3,555
sequence of different types				
T2	58,940	58,940	.	.
T1	22,384	.	22,384	.
An	2,308	.	.	2,308
T2 T1	1,287	404	693	190
T1 T2	933	502	262	169
An T2	585	328	.	257
T2 T1 T2	369	342	15	12
T2 An	344	133	.	211
T1 T2 T1	274	18	248	8
An T1	192	.	129	63
T1 An	159	.	79	80
T2 An T2	111	107	.	*
T1 An T1	55	.	48	7
T2 T1 T2 T1	44	20	13	11
An T2 An	35	*	.	32
-I T2	35	30	.	*
T1 T2 T1 T2	28	13	6	9
-I	27	.	.	27
-I An	25	.	.	25
-I T1	21	.	19	*
An T2 T1	17	*	*	13

```
T1 T2 An                    15      *      *      11
An T1 T2                    15      *      *       9
T2 An T1                    14      *      *       9
An T1 An                    14      .      *      13
T2 T1 An                    14      *      .      13
T2 T1 T2 T1 T2              13     12      .       *
T1 An T2                    11      *      .       7
T2 -I T2                    10     10      .       .
T1 -I T1                     9      .      9       .
T2 T1 An T1                  7      .      *       *
T2 T1 T2 An                  6      *      *       *
An T2 An T2                  6      *      .       *
-I An T2                     6      .      .       6
T1 T2 An T1                  *      .      *       *
T1 T2 T1 T2 T1               *      .      *       *
T2 T1 T2 T1 T2 T1            *      *      *       .
An T1 An T1                  *      .      *       *
T2 -I                        *      *      .       *
An T2 T1 T2                  *      *      .       *
T1 An T2 T1                  *      .      *       *
An -I An                     *      .      .       *
-I T2 T1                     *      .      *       *
T2 An T2 An                  *      .      .       *
T1 An T1 An                  *      .      *       *
An T2 An T1                  *      .      .       *
An T1 T2 T1                  *      .      *       *
-I T1 T2 T1                  *      .      *       .
T1 T2 T1 An                  *      .      *       *
-I An T1                     *      .      *       *
T1 An T1 An T1               *      .      *       *
T2 T1 An T2                  *      *      .       *
An T2 T1 An                  *      .      .       *
-I T1 T2                     *      *      *       .
T2 An T1 T2                  *      *      .       .
An T1 An T2                  *      .      .       *
T1 An T2 An                  *      .      .       *
-I T2 T1 T2                  *      .      .       *
-I An -I An T2               *      .      .       *
T1 -I An                     *      .      .       *
T1 T2 An T2                  *      .      .       *
An T2 An T2 An               *      .      .       *
An T1 T2 An                  *      .      .       *
T1 -I An T1 T2               *      .      .       *
-I An T2 An                  *      .      .       *
T1 T2 T1 T2 T1 T2 T1 T2      *      *      .       .
T1 An T1 T2 An               *      .      .       *
T2 An T1 T2 T1 An T1         *      .      *       .
T2 -I T1 T2 T1               *      .      .       *
An -I An T2 An               *      .      .       *
T2 -I An T2 An T2 T1         *      .      .       *
T1 -I An T2                  *      .      .       *
An T2 T1 T2 T1               *      .      *       .
T1 T2 An T2 An T1            *      .      .       *
T1 T2 T1 T2 T1 T2 T1         *      .      *       .
-I T2 T1 T2 T1               *      *      .       .
T2 An T1 T2 T1               *      *      .       .
-I An T1 An                  *      .      .       *
T2 An T1 An                  *      .      .       *
```

7.7 03-ndr

Constucts a version of the old NDR without using the blood-glucose criteria and the foot-therapy criterion.

```
1                                                  "Program 03-ndr.sas"
```

```
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_SRV12  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.07 seconds
      cpu time            0.06 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          ************************************************************************
2          * NDR ;
3          proc sort  data = EKST15.ndr  out = ndr ;
4             by pnr ;
5          run ;

NOTE: Input data set is already sorted; it has been copied to the output data set.
NOTE: There were 523770 observations read from the data set EKST15.NDR.
NOTE: The data set WORK.NDR has 523770 observations and 9 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.54 seconds
      cpu time            0.07 seconds


6
7          * Only persons present in the base (TTDATA.bef_grund) ;
8          data TTDATA.NDR ( keep = pnr sex doNDR datNDR doBth doDth inklaars ) ;
9             merge ndr ( in = ndr
10                       rename = ( d_inkldto = DoNDR ) )
11                  TTDATA.pop ( in = pop ) ;
12            by pnr ;
13            if ndr and pop ;
14            * revised date of inclusion (ignoring blood glucose criteria) ;
15            doNDR = min( d_lpr, /*d_fodt,*/ d_ins, d_oad ) ;
16            if doNDR eq d_lpr  then inklaars = 'lpr' ;
17         *  if doNDR eq d_fodt then inklaars = 'fod' ;
18            if doNDR eq d_oad  then inklaars = 'oad' ;
19            if doNDR eq d_ins  then inklaars = 'ins' ;
20            * To avoid small cell entries in overview table ;
21            datNDR = max( doNDR, '01JAN1990'd ) ;
22            label datNDR="doDM" ;
23            if doNDR gt .z ;
24         run ;

NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      106885 at 15:12
NOTE: There were 523770 observations read from the data set WORK.NDR.
NOTE: There were 7361669 observations read from the data set TTDATA.POP.
NOTE: The data set TTDATA.NDR has 389304 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           2.67 seconds
      cpu time            1.28 seconds
```

```
25
26         title1 'Persons from NDR with valid corrected DM dates' ;
27         title2 'datNDR is the doNDR left censored at 1.1.1990' ;
28         proc tabulate  data = TTdata.NDR  missing  noseps  formchar='           ' ;
29           class datNDR inklaars ;
30           table all datNDR,
31               ( all inklaars ) * f=comma7.
32               ( all * f= 8.1
33                     inklaars * f=4.1 ) * pctn< all inklaars >
34                 / rts = 8 ;
35           format datNDR year4. ;
36           keylabel n = " " ;
37         run ;

NOTE: There were 389304 observations read from the data set TTDATA.NDR.
NOTE: At least one W.D format was too small for the number to be printed. The decimal may
      be shifted by the "BEST" format.
NOTE: The PROCEDURE TABULATE printed page 1.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.15 seconds
      cpu time            0.15 seconds


38         title2 ;
39
40         proc contents  data = TTdata.ndr ;
41         run ;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

NOTE: The PROCEDURE CONTENTS printed page 2.

NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           3.53 seconds
      cpu time            1.59 seconds
```

7.7.1 03-ndr.lst

```
Persons from NDR with valid corrected DM dates         10:42 Tuesday, January 24, 2017   1
datNDR is the doNDR left censored at 1.1.1990
```

	inklaars				All	inklaars		
	All	ins	lpr	oad	PctN	ins PctN	lpr PctN	oad PctN
All	389,304	10,163	181,868	197,273	100.0	100	100	100
doDM								
1990	12,842	.	12,842	.	100.0	.	100	.
1991	8,854	.	8,854	.	100.0	.	100	.
1992	8,236	.	8,236	.	100.0	.	100	.
1993	12,362	.	12,362	.	100.0	.	100	.
1994	37,349	6,094	7,037	24,218	100.0	16.3	18.8	64.8
1995	12,458	598	5,877	5,983	100.0	4.8	47.2	48.0
1996	12,476	317	6,013	6,146	100.0	2.5	48.2	49.3
1997	12,027	187	5,964	5,876	100.0	1.6	49.6	48.9
1998	13,175	177	6,468	6,530	100.0	1.3	49.1	49.6
1999	13,636	191	6,723	6,722	100.0	1.4	49.3	49.3
2000	14,123	148	7,338	6,637	100.0	1.0	52.0	47.0
2001	14,996	139	7,646	7,211	100.0	0.9	51.0	48.1
2002	15,386	163	8,042	7,181	100.0	1.1	52.3	46.7
2003	17,208	148	8,659	8,401	100.0	0.9	50.3	48.8

```
 2004    17,749      156   8,440   9,153    100.0  0.9 47.6 51.6
 2005    17,439      163   8,236   9,040    100.0  0.9 47.2 51.8
 2006    18,062      177   8,313   9,572    100.0  1.0 46.0 53.0
 2007    18,966      241   8,153  10,572    100.0  1.3 43.0 55.7
 2008    20,088      264   7,738  12,086    100.0  1.3 38.5 60.2
 2009    20,951      262   7,652  13,037    100.0  1.3 36.5 62.2
 2010    22,396      236   7,485  14,675    100.0  1.1 33.4 65.5
 2011    24,997      265   7,192  17,540    100.0  1.1 28.8 70.2
 2012    23,528      237   6,598  16,693    100.0  1.0 28.0 70.9


Persons from NDR with valid corrected DM dates         10:42 Tuesday, January 24, 2017   2

The CONTENTS Procedure

Data Set Name        TTDATA.NDR                    Observations          389304
Member Type          DATA                          Variables             7
Engine               V9                            Indexes               0
Created              24/01/2017 10:42:45           Observation Length    56
Last Modified        24/01/2017 10:42:45           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    334
First Data Page             *
Max Obs per Page            1167
Obs in First Data Page      1136
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\ndr.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12


          Alphabetic List of Variables and Attributes

#    Variable    Type    Len    Format    Informat    Label

2    DoNDR       Num       8    DATE9.    DATE9.      D_INKLDTO
7    datNDR      Num       8                          doDM
4    doBth       Num       8
5    doDth       Num       8
6    inklaars    Char      *
1    pnr         Char     12    $12.      $10.        V_CPR
3    sex         Num       8
```

7.8 04-rmps

Processes the records from the RMPS with other target medications and creates a file (pRMPS) with one record per person with at least one prescription of either OAD or insulin.

Uses the GDM dates to exclude possible inclusion dates in GDM grace periods.

```
1                                                      "Program 04-rmps.sas"

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_SRV12  platform.
```

```
NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.07 seconds
      cpu time            0.09 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          %macro med ;
2          data rmps ;
3            length druggr $ 7 ;
4            set %do i = 1995 %to 2015 ;
5                GRUND15.lmdb&i. ( keep = pnr ATC eksd apk volume packsize doso
6                                  where = ( substr(atc,1,3) eq 'A10' ) )
7                %end ; ;
8            * Grouping of drugs as a character ;
9                                           druggr = put(        atc     , $atc5grp. ) ;
10           if ( druggr eq "Other" ) then druggr = put( substr(atc,1,5), $atc4grp. ) ;
11         run ;
12         %mend ;
13         %med ;

NOTE: There were 561599 observations read from the data set GRUND15.LMDB1995.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 626172 observations read from the data set GRUND15.LMDB1996.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 670759 observations read from the data set GRUND15.LMDB1997.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 732363 observations read from the data set GRUND15.LMDB1998.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 795049 observations read from the data set GRUND15.LMDB1999.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 826889 observations read from the data set GRUND15.LMDB2000.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 906757 observations read from the data set GRUND15.LMDB2001.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 985120 observations read from the data set GRUND15.LMDB2002.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1066934 observations read from the data set GRUND15.LMDB2003.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1179915 observations read from the data set GRUND15.LMDB2004.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1290865 observations read from the data set GRUND15.LMDB2005.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1407844 observations read from the data set GRUND15.LMDB2006.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1519880 observations read from the data set GRUND15.LMDB2007.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1659373 observations read from the data set GRUND15.LMDB2008.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1751002 observations read from the data set GRUND15.LMDB2009.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 1875270 observations read from the data set GRUND15.LMDB2010.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 2003071 observations read from the data set GRUND15.LMDB2011.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 2101642 observations read from the data set GRUND15.LMDB2012.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 2130915 observations read from the data set GRUND15.LMDB2013.
```

```
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 2152130 observations read from the data set GRUND15.LMDB2014.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: There were 2164414 observations read from the data set GRUND15.LMDB2015.
      WHERE SUBSTR(atc, 1, 3)='A10';
NOTE: The data set WORK.RMPS has 28407963 observations and 8 variables.
NOTE: DATA statement used (Total process time):
      real time           4:43.26
      cpu time            43.21 seconds


14
15         proc sort  data = rmps ;  by pnr druggr eksd ;  run ;

NOTE: There were 28407963 observations read from the data set WORK.RMPS.
NOTE: The data set WORK.RMPS has 28407963 observations and 8 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           6.21 seconds
      cpu time            12.67 seconds


16
17         * Generate one record per person with a drug dispensation
18           but excluding dispensations in the time after GDM and PCOS ;
19         options mprint ;
20         data rmps
21              pospcos ( keep = pnr pospcos ) ;
22           merge rmps        ( in = r )
23                 TTdata.GDM
24                 TTdata.PCOS ( in = p )
25                 TTdata.pop  ( in = b  keep = pnr doBth ) ;
26           by pnr ;
27           if r and b ;
28           pcos = p ;
29           retain pospcos ;
30           * exclude dispensations in GDM-window ;
31           %xgdm( eksd ) ;
MPRINT(XGDM):   if ( doGDM1 - 30 ) < eksd < ( doGDM1 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM2 - 30 ) < eksd < ( doGDM2 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM3 - 30 ) < eksd < ( doGDM3 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM4 - 30 ) < eksd < ( doGDM4 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM5 - 30 ) < eksd < ( doGDM5 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM6 - 30 ) < eksd < ( doGDM6 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM7 - 30 ) < eksd < ( doGDM7 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM8 - 30 ) < eksd < ( doGDM8 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM9 - 30 ) < eksd < ( doGDM9 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM10 - 30 ) < eksd < ( doGDM10 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM11 - 30 ) < eksd < ( doGDM11 + 365 ) then delete ;
MPRINT(XGDM):   if ( doGDM12 - 30 ) < eksd < ( doGDM12 + 365 ) then delete ;
32           * exclude metformin dispenstions after PCOS if before age 40 ;
33           if pcos and
34              druggr eq '11' and
35              ( doPCOS - 30 ) < eksd < ( doBth  + 365.25*40 ) then delete ;
36           output rmps ;
37           * if metformin (druggr=11) is the only dispensation between age 20 and 40,
38             it is possibly PCOS so the person is flagged by pospcos ;
39           if first.pnr then pospcos =
40              ( doBth + 365.25*20 ) < eksd < ( doBth + 365.25*40 ) and druggr eq '11' ;
41           if ( doBth + 365.25*20 ) < eksd < ( doBth + 365.25*40 ) and druggr ne '11'
42                       then pospcos = 0 ;
43           if last.pnr and pospcos then output pospcos ;
44         run ;

NOTE: Variable doGDM12 is uninitialized.
NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      27625444 at 31:18   27625444 at 31:54   28022030 at 31:20   28022030 at 31:56
      28236064 at 31:20   28236064 at 31:56   28309326 at 31:20   28309326 at 31:56
      28328127 at 31:20   28328127 at 31:56   28333628 at 31:20   28333628 at 31:56
```

```
      28334561 at 31:20   28334561 at 31:56   28334796 at 31:20   28334796 at 31:56
      28334831 at 31:20   28334831 at 31:56   28334831 at 31:20   28334831 at 31:56
      28334831 at 31:20   28334831 at 31:56   28334831 at 31:20   28334831 at 31:56
NOTE: There were 28407963 observations read from the data set WORK.RMPS.
NOTE: There were 21146 observations read from the data set TTDATA.GDM.
NOTE: There were 20773 observations read from the data set TTDATA.PCOS.
NOTE: There were 7361669 observations read from the data set TTDATA.POP.
NOTE: The data set WORK.RMPS has 28295912 observations and 24 variables.
NOTE: The data set WORK.POSPCOS has 19885 observations and 2 variables.
NOTE: DATA statement used (Total process time):
      real time           1:28.60
      cpu time            1:22.62


45         options nomprint ;
46
47         * Exclude metformin dispensation between 20 and 40 for those who start
47       ! metformin
48           after 20 oand only fill metformin before age 40 ;
49         data rmps ;
50           merge rmps pospcos ;
51           by pnr ;
52         drop pospcos ;
53           if pospcos and
54              ( doBth + 365.25*20 ) < eksd < ( doBth + 365.25*40 )
55              then delete ;
56         run ;

NOTE: There were 28295912 observations read from the data set WORK.RMPS.
NOTE: There were 19885 observations read from the data set WORK.POSPCOS.
NOTE: The data set WORK.RMPS has 27943194 observations and 23 variables.
NOTE: DATA statement used (Total process time):
      real time           16.07 seconds
      cpu time            7.70 seconds


57
58         * Naming convention for date variables :
59           "donlXX" is first date of single drug medication with XX
60           "danyXX" is first date of *any* medication with drug XX
61           Thus "donl" is for selection on mono-users, otherwise "dany" is used.
62           We put the names of these data variables into a macro variable and
63           later in an array --- note that insulins are the last 5 entries in the array
63       ! ;
64         %let drdates = donlMet   danyMet
65                        donlSU    danySU
66                        donlTZD   danyTZD
67                        donlDPP4  danyDPP4
68                                  danyGLP1
69                        donlSGLT2 danySGLT2
70                        donlAca   danyAca
71                                  danyMeg
72                        dofIns doiIns domIns dolIns doIns ;
73
74         data TTdata.pRMPS ( keep = pnr &drdates. doOAD doRMPS ) ;
75           set rmps ;
76           by pnr druggr eksd ;
77           retain          &drdates. ;
78           array drd [19] &drdates. ;
79           * inintiate all drugdates to missing ;
80           if first.pnr then do i = 1 to dim(drd) ; drd(i) = . ; end ;
81           * dates of first exposure ;
82           if first.druggr then do ;
83              if              druggr eq  '11'   then donlMet  =     eksd ;
84              if index(substr(druggr,2,2),'1')  then danyMet  = min(eksd,danyMet) ;
85              if              druggr eq  '12'   then donlSU   =     eksd ;
86              if index(substr(druggr,2,2),'2')  then danySU   = min(eksd,danySU ) ;
87              if              druggr eq  '13'   then donlTZD  =     eksd ;
88              if index(substr(druggr,2,2),'3')  then danyTZD  = min(eksd,danyTZD) ;
89              if              druggr eq  '14'   then donlDPP4 =     eksd ;
```

```
90              if index(substr(druggr,2,2),'4')  then danyDPP4 = min(eksd,danyDPP4) ;
91              if              druggr eq  '15'   then danyGLP1 =     eksd ;
92              if              druggr eq  '16'   then donlSGLT2=     eksd ;
93              if index(substr(druggr,2,2),'6')  then danySGLT2= min(eksd,danySGLT2) ;
94              if              druggr eq  '18'   then donlAca  =     eksd ;
95              if index(substr(druggr,2,2),'8')  then danyAca  = min(eksd,danyAca) ;
96              if              druggr eq  '19'   then danyMeg  =     eksd ;
97              if              druggr eq 'fastIns' then dofIns =     eksd ;
98              if              druggr eq  'intIns' then doiIns =     eksd ;
99              if              druggr eq  'mixIns' then domIns =     eksd ;
100             if              druggr eq 'longIns' then dolIns =     eksd ;
101             end ;
102           * Date of any drugexposure / resp. any OAD (i.e. non-ins) exposure ;
103           if last.pnr then do ;       doRMPS = min(     of drd[*] ) ;
104             do i = 1 to dim(drd)-5 ; doOAD  = min( doOAD, drd[i] ) ; end ;
105             doIns = min( dofIns, doiIns, domIns, dolIns ) ;
106             output ;
107             end ;
108         run ;

NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      881239 at 104:40   260604 at 105:14
NOTE: There were 27943194 observations read from the data set WORK.RMPS.
NOTE: The data set TTDATA.PRMPS has 409844 observations and 22 variables.
NOTE: DATA statement used (Total process time):
      real time           10.76 seconds
      cpu time            6.65 seconds


109
110         title1 "Persons from the RPMS" ;
111         proc contents  data = TTdata.pRMPS ; run ;

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

NOTE: The PROCEDURE CONTENTS printed page 1.

112         proc tabulate  data = TTdata.pRMPS  missing  noseps ;
113           class &drdates. doOAD doRMPS ;
114           table all &drdates. doOAD doRMPS,
115                 n * f=comma10.
116                 / rts = 15 ;
117           table all doOAD,
118               ( all doINS ) * f=comma7.
119                 / rts = 7 ;
120           format &drdates. doOAD doRMPS year4. ;
121           keylabel n = ' ' ;
122         run ;

NOTE: There were 409844 observations read from the data set TTDATA.PRMPS.
NOTE: The PROCEDURE TABULATE printed pages 2-5.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.42 seconds
      cpu time            1.11 seconds


123         title1 ;
124
125

NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           6:45.54
      cpu time            2:34.10
```

7.8.1 04-rmps.lst

```
Persons from the RPMS                                14:04 Tuesday, January 24, 2017   1

The CONTENTS Procedure

Data Set Name        TTDATA.PRMPS                  Observations          409844
Member Type          DATA                          Variables             22
Engine               V9                            Indexes               0
Created              24/01/2017 14:10:53           Observation Length    184
Last Modified        24/01/2017 14:10:53           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    1155
First Data Page             *
Max Obs per Page            355
Obs in First Data Page      338
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\prmps.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12


             Alphabetic List of Variables and Attributes

 #    Variable     Type    Len    Format    Informat    Label

14    danyAca      Num       8
 9    danyDPP4     Num       8
10    danyGLP1     Num       8
15    danyMeg      Num       8
 *    danyMet      Num       8
12    danySGLT2    Num       8
 *    danySU       Num       8
 7    danyTZD      Num       8
20    doIns        Num       8
22    doOAD        Num       8
21    doRMPS       Num       8
16    dofIns       Num       8
17    doiIns       Num       8
19    dolIns       Num       8
18    domIns       Num       8
13    donlAca      Num       8
 8    donlDPP4     Num       8
 *    donlMet      Num       8
11    donlSGLT2    Num       8
 *    donlSU       Num       8
 6    donlTZD      Num       8
 *    pnr          Char     12    $12.      $10.        Personnummer

Persons from the RPMS                                14:04 Tuesday, January 24, 2017   2


 All              409,844
 donlMet
 .                117,064
 1995              13,332
 1996               4,358
 1997               4,065
 1998               4,948
 1999               6,813
 2000               8,906
```

```
2001              9,388
2002              9,974
2003             11,755
2004             13,278
2005             13,319
2006             14,407
2007             16,728
2008             19,263
2009             21,005
2010             22,782
2011             27,064
2012             23,187
2013             16,641
2014             15,096
2015             16,471
danyMet
.               115,985
1995             13,332
1996              4,358
1997              4,065
1998              4,948
1999              6,813
2000              8,906
2001              9,388
2002              9,974
2003             11,755
2004             13,339
2005             13,382
2006             14,554
2007             16,856
2008             19,372
2009             21,139
2010             22,897
2011             27,106
2012             23,267
2013             16,709
2014             15,149
2015             16,550
donlSU
.               214,656
1995             45,851
1996             10,198
1997              9,615
1998             10,402
1999             10,093
2000              9,110
2001              9,389
2002              8,772
2003             10,016
2004              9,589
2005              9,099
2006              8,298
2007              8,183
2008              7,272
2009              6,004
2010              5,907
2011              4,633
2012              3,637
2013              3,255
2014              2,911
2015              2,954
danySU
.               214,655
1995             45,851
1996             10,198
1997              9,615
1998             10,402
1999             10,093
2000              9,110
2001              9,389
```

```
2002              8,772
2003             10,016
2004              9,589
2005              9,099
2006              8,298
2007              8,184
2008              7,273
2009              6,005
2010              5,906
2011              4,632
2012              3,637
2013              3,255
2014              2,911
2015              2,954
donlTZD
.               405,171
2000                100
2001                652
2002                424
2003                333
2004                426
2005                468
2006                584
2007                573
2008                390
2009                303
2010                178
2011                 90
2012                 52
2013                 39
2014                 36
2015                 25
danyTZD
.               399,747
2000                100
2001                652
2002                424
2003                334
2004              1,114
2005              1,422
2006              2,126
2007              1,847
2008                952
2009                617
2010                312
2011                 56
2012                 43
2013                 36
2014                 37
2015                 25
donlDPP4
.               369,776
2007              2,290
2008              4,452
2009              3,239
2010              4,425
2011              4,669
2012              4,471
2013              5,055
2014              5,418
2015              6,049
danyDPP4
.               349,498
2007              2,290
2008              5,153
2009              5,464
2010              7,541
2011              7,764
2012              7,453
2013              7,830
```

```
2014                8,125
2015                8,726
danyGLP1
.                 377,256
2007                  176
2008                1,030
2009                2,217
2010                7,256
2011                6,471
2012                5,235
2013                3,566
2014                2,969
2015                3,668
donlSGLT2
.                 400,423
2012                   20
2013                1,948
2014                2,958
2015                4,495
danySGLT2
.                 400,183
2012                   20
2013                1,948
2014                3,041
2015                4,652
donlAca
.                 402,362
1995                  685
1996                1,232
1997                1,174
1998                1,172
1999                  712
2000                  453
2001                  427
2002                  246
2003                  221
2004                  193
2005                  173
2006                  134
2007                  115
2008                  111
2009                   72
2010                   95
2011                   57
2012                   75
2013                   55
2014                   43
2015                   37
danyAca
.                 402,362
1995                  685
1996                1,232
1997                1,174
1998                1,172
1999                  712
2000                  453
2001                  427
2002                  246
2003                  221
2004                  193
2005                  173
2006                  134
2007                  115
2008                  111
2009                   72
2010                   95
2011                   57
2012                   75
2013                   55
2014                   43
```

```
2015                37
danyMeg
.                  409,844
dofIns
.                  333,788
1995                17,490
1996                 3,203
1997                 2,427
1998                 2,234
1999                 2,276
2000                 2,240
2001                 2,206
2002                 2,311
2003                 2,325
2004                 2,536
2005                 2,610
2006                 2,697
2007                 2,832
2008                 2,983
2009                 3,189
2010                 3,308
2011                 3,407
2012                 3,531
2013                 3,856
2014                 4,065
2015                 4,330
doiIns
.                  308,325
1995                29,190
1996                 4,509
1997                 3,616
1998                 3,742
1999                 4,027
2000                 4,188
2001                 4,102
2002                 4,421
2003                 4,844
2004                 4,684
2005                 4,400
2006                 4,206
2007                 4,143
2008                 3,011
2009                 2,590
2010                 2,568
2011                 2,579
2012                 2,535
2013                 3,075
2014                 2,740
2015                 2,349
domIns
.                  348,584
1995                 7,828
1996                 2,766
1997                 2,215
1998                 2,610
1999                 2,753
2000                 2,688
2001                 2,746
2002                 2,705
2003                 3,533
2004                 4,116
2005                 4,025
2006                 4,321
2007                 4,448
2008                 3,152
2009                 2,582
2010                 1,934
2011                 1,761
2012                 1,539
2013                 1,406
```

```
2014                1,191
2015                  941
dolIns
.                 349,929
1996                   10
1997                    6
1998                    *
1999                    *
2000                    *
2004                2,590
2005                4,688
2006                3,149
2007                2,943
2008                6,758
2009                6,778
2010                5,630
2011                5,239
2012                5,210
2013                4,747
2014                5,486
2015                6,671
doIns
.                 260,604
1995                33,463
1996                 4,865
1997                 4,009
1998                 4,347
1999                 4,804
2000                 4,977
2001                 4,895
2002                 5,204
2003                 5,947
2004                 6,395
2005                 6,231
2006                 6,305
2007                 6,552
2008                 6,194
2009                 6,257
2010                 5,957
2011                 6,282
2012                 6,303
2013                 6,616
2014                 6,641
2015                 6,996
doOAD
.                   52,998
1995                48,996
1996                10,882
1997                10,141
1998                11,113
1999                11,421
2000                11,548
2001                12,223
2002                12,126
2003                14,339
2004                14,722
2005                14,646
2006                15,150
2007                16,588
2008                18,298
2009                19,160
2010                21,329
2011                25,032
2012                21,853
2013                15,984
2014                14,747
2015                16,548
doRMPS
1995                79,747
1996                12,916
```

1997	11,321
1998	12,245
1999	12,435
2000	12,605
2001	13,250
2002	13,132
2003	15,361
2004	15,909
2005	15,757
2006	16,377
2007	17,715
2008	19,427
2009	20,194
2010	22,314
2011	25,900
2012	22,673
2013	17,089
2014	15,865
2015	17,612

Persons from the RPMS 14:04 Tuesday, January 24, 2017 3

		doIns							
	All	.	1995	1996	1997	1998	1999	2000	2001	2002
All	409,844	260,604	33,463	4,865	4,009	4,347	4,804	4,977	4,895	5,204
doOAD										
.	52,998	.	27,298	2,044	1,214	1,151	1,137	1,172	1,147	1,169
1995	48,996	26,300	2,712	2,170	1,970	1,993	2,069	1,855	1,415	1,416
1996	10,882	5,917	375	286	259	277	329	365	329	338
1997	10,141	5,587	273	48	279	252	282	274	337	307
1998	11,113	6,070	278	26	64	325	298	299	336	333
1999	11,421	6,354	293	33	30	75	381	284	289	327
2000	11,548	6,609	275	29	24	41	67	407	292	276
2001	12,223	7,163	241	17	22	26	27	77	460	300
2002	12,126	7,449	226	24	19	23	36	40	63	470
2003	14,339	9,283	179	19	20	20	32	26	35	61
2004	14,722	10,191	130	17	7	16	21	23	29	32
2005	14,646	10,402	122	9	9	14	18	19	29	31
2006	15,150	11,280	101	15	12	15	16	13	18	23
2007	16,588	12,824	127	17	15	24	17	14	17	21
2008	18,298	14,886	110	15	9	11	11	23	12	15
2009	19,160	16,022	135	17	9	17	11	18	10	18
2010	21,329	18,556	120	10	8	12	16	11	19	18
2011	25,032	22,582	127	14	9	16	7	17	19	17
2012	21,853	19,831	127	12	6	16	8	17	12	*
2013	15,984	14,429	91	8	8	14	8	9	9	11
2014	14,747	13,451	63	20	9	*	*	*	14	7
2015	16,548	15,418	60	15	7	*	9	9	*	9

(Continued)

Persons from the RPMS 14:04 Tuesday, January 24, 2017 4

	doIns									
	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
All	5,947	6,395	6,231	6,305	6,552	6,194	6,257	5,957	6,282	6,303
doOAD										
.	1,147	1,227	1,210	1,287	1,287	1,316	1,306	1,291	1,239	1,188
1995	1,403	1,201	931	772	629	438	364	311	290	239
1996	357	363	282	277	243	176	151	121	118	91
1997	377	351	318	263	233	187	139	141	139	114
1998	425	407	360	314	319	240	199	179	168	146

1999	407	419	351	338	341	279	235	188	184	173
2000	385	381	381	350	362	280	271	212	217	198
2001	377	394	402	365	414	328	300	259	243	242
2002	322	382	362	352	341	306	322	251	261	226
2003	480	413	408	424	447	368	375	308	332	292
2004	63	559	383	348	346	388	353	295	343	300
2005	22	64	605	370	371	323	336	310	325	354
2006	26	31	60	575	355	333	373	311	336	334
2007	19	34	41	61	617	338	340	288	347	328
2008	29	31	21	32	73	689	317	293	284	342
2009	19	28	23	31	36	64	712	312	303	312
2010	21	26	24	40	47	37	75	709	321	277
2011	23	23	25	27	25	27	31	73	723	320
2012	15	20	18	28	17	23	31	59	54	727
2013	12	13	10	20	11	18	14	15	22	61
2014	11	10	7	19	19	21	*	20	16	16
2015	7	18	9	12	19	15	8	11	17	23

(Continued)

Persons from the RPMS 14:04 Tuesday, January 24, 2017 5

	doIns		
	2013	2014	2015
All	6,616	6,641	6,996
doOAD			
.	1,380	1,380	1,408
1995	200	175	143
1996	83	76	69
1997	80	82	78
1998	144	92	91
1999	164	143	133
2000	160	168	163
2001	229	176	161
2002	229	223	199
2003	281	276	260
2004	320	278	280
2005	317	284	312
2006	312	307	304
2007	358	381	360
2008	359	361	375
2009	346	344	373
2010	283	310	389
2011	288	313	326
2012	299	262	266
2013	705	235	261
2014	57	709	259
2015	22	66	786

7.9 05-diab

Uses eye-screening dates from the national eye-screening database to supplement persons with diabetes and update dates of diabetes diagosis. Uses the clinically recorded type of diabetes to define the status of the patients.

Uses the dataset with GDM dates to exclude inclusion dates in GDM grace periods.

```
1                                                    "Program 05-diab.sas"

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_SRV12  platform.


NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.07 seconds
      cpu time            0.09 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          title1 'Original Diabase' ;
2          proc contents  data = ekst15.diabasen ; run ;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.03 seconds
      cpu time            0.01 seconds

NOTE: The PROCEDURE CONTENTS printed page 1.

3          title1 ;
4
5          proc sort  data = ekst15.diabasen
6                     out = diab ;
7            by pnr Report_EyeScreeningDate ;
8          run ;

NOTE: There were 98929 observations read from the data set EKST15.DIABASEN.
NOTE: The data set WORK.DIAB has 98929 observations and 13 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.15 seconds
      cpu time            0.10 seconds


9
10         options mprint ;
11         data diab ;
12           merge diab ( in = d )
13                 TTdata.GDM ;
14           by pnr ;
15           * exclude visits in GDM grace period ;
16           %xgdm( Report_EyeScreeningDate ) ;
MPRINT(XGDM):   if ( doGDM1 - 30 ) < Report_EyeScreeningDate < ( doGDM1 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM2 - 30 ) < Report_EyeScreeningDate < ( doGDM2 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM3 - 30 ) < Report_EyeScreeningDate < ( doGDM3 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM4 - 30 ) < Report_EyeScreeningDate < ( doGDM4 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM5 - 30 ) < Report_EyeScreeningDate < ( doGDM5 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM6 - 30 ) < Report_EyeScreeningDate < ( doGDM6 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM7 - 30 ) < Report_EyeScreeningDate < ( doGDM7 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM8 - 30 ) < Report_EyeScreeningDate < ( doGDM8 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM9 - 30 ) < Report_EyeScreeningDate < ( doGDM9 + 365 ) then
```

```
delete ;
MPRINT(XGDM):   if ( doGDM10 - 30 ) < Report_EyeScreeningDate < ( doGDM10 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM11 - 30 ) < Report_EyeScreeningDate < ( doGDM11 + 365 ) then
delete ;
MPRINT(XGDM):   if ( doGDM12 - 30 ) < Report_EyeScreeningDate < ( doGDM12 + 365 ) then
delete ;
17         run ;

NOTE: Variable doGDM12 is uninitialized.
NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      89780 at 16:18    89780 at 16:54    107772 at 16:20   107772 at 16:56
      112844 at 16:20   112844 at 16:56   114221 at 16:20   114221 at 16:56
      114499 at 16:20   114499 at 16:56   114577 at 16:20   114577 at 16:56
      114616 at 16:20   114616 at 16:56   114625 at 16:20   114625 at 16:56
      114626 at 16:20   114626 at 16:56   114626 at 16:20   114626 at 16:56
      114626 at 16:20   114626 at 16:56   114627 at 16:20   114627 at 16:56
NOTE: There were 98929 observations read from the data set WORK.DIAB.
NOTE: There were 21146 observations read from the data set TTDATA.GDM.
NOTE: The data set WORK.DIAB has 114627 observations and 25 variables.
NOTE: DATA statement used (Total process time):
      real time           0.35 seconds
      cpu time            0.34 seconds


18
19         data TTdata.diab  ( keep = pnr doDIAB )
20                    diab ;
21           set diab ;
22           by pnr Report_EyeScreeningDate ;
23           doDiab = Report_EyeScreeningDate ;
24           visit = "later" ;
25           if first.pnr then do ;
26              output TTdata.diab ;
27              visit = "first" ;
28              end ;
29           output diab ;
30         run ;

NOTE: There were 114627 observations read from the data set WORK.DIAB.
NOTE: The data set TTDATA.DIAB has 57676 observations and 2 variables.
NOTE: The data set WORK.DIAB has 114627 observations and 27 variables.
NOTE: DATA statement used (Total process time):
      real time           0.09 seconds
      cpu time            0.06 seconds


31
32         title1 'Distribution of first and later visits' ;
33         proc tabulate  data = diab  missing  noseps ;
34           class doDIAB visit ;
35           table doDIAB,
36                 visit * f=comma9.
37                 / rts = 20 ;
38           format doDIAB  yymms7. ;
39         run ;

NOTE: There were 114627 observations read from the data set WORK.DIAB.
NOTE: The PROCEDURE TABULATE printed page 2.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.04 seconds
      cpu time            0.07 seconds


40         title1 ;
41
42         title1 'First visits from Diabasen' ;
43         proc contents  data = TTdata.diab ; run ;
```

```
NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

NOTE: The PROCEDURE CONTENTS printed page 3.

NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           0.82 seconds
      cpu time            0.73 seconds
```

7.9.1 05-diab.lst

```
Original Diabase                                    17:36 Thursday, January 12, 2017   1

The CONTENTS Procedure

Data Set Name        EKST15.DIABASEN               Observations          98929
Member Type          DATA                          Variables             13
Engine               V9                            Indexes               0
Created              27/11/2015 14:29:35           Observation Length    336
Last Modified        27/11/2015 14:29:35           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                YES
Label                Diabasen
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          32768
Number of Data Set Pages    1020
First Data Page             *
Max Obs per Page            97
Obs in First Data Page      88
Number of Data Set Repairs  0
Filename                    E:\rawdata\705093\Eksterne data\diabasen.sas7bdat
Release Created             9.0401M3
Host Created                X64_ES08R2


                     Alphabetic List of Variables and Attributes

 #  Variable                 Type  Len  Format       Informat     Label

13  BlurredMediaLeftEyeCode  Char   25               $25.         Slrede medier venstre je
12  BlurredMediaRightEyeCode Char   25               $25.         Slrede medier hjre je
 *  EyeScreeningIndication   Char   10               $10.         enscreeningsindikation
    Code
10  MaculopatiStatus         Char   53               $53.         Maculopati-status
    LeftEyeCode                                                  venstre je
11  MaculopatiStatus         Char   53               $53.         Maculopati-status hjre je
    RightEyeCode
 *  PriorEyeSurgery          Char   23               $23.         Tidligere jenkirurgi
    LeftEyeCode                                                  venstre je
 *  PriorEyeSurgery          Char   23               $23.         Tidligere jenkirurgi
    RightEyeCode                                                 hjre je
 *  Report_EyeScreeningDate  Num     8  IS8601DA10.  IS8601DA10.  enscreeningsdato
 8  RetinaStatusLeftEyeCode  Char   23               $23.         Retina status venstre je
 9  RetinaStatusRightEyeCode Char   23               $23.         Retina status hjre je
 6  VisualAcuityLeftEyeCode  Char   28               $28.         Synsstyrke venstre je
 7  VisualAcuityRightEyeCode Char   28               $28.         Synsstyrke hjre je
 *  pnr                      Char   12  $12.         $10.         Personnummer


 Sort Information
```

```
Sortedby       pnr
Validated      YES
Character Set  ANSI
```

Distribution of first and later visits 17:36 Thursday, January 12, 2017 2

	visit	
	first	later
	N	N
doDiab		
.	18,197	.
2009/01	385	.
2009/02	409	.
2009/03	389	*
2009/04	354	*
2009/05	282	7
2009/06	337	19
2009/07	149	15
2009/08	328	47
2009/09	359	64
2009/10	357	53
2009/11	354	51
2009/12	302	54
2010/01	625	121
2010/02	431	159
2010/03	543	287
2010/04	412	262
2010/05	444	255
2010/06	695	342
2010/07	604	219
2010/08	980	434
2010/09	1,246	498
2010/10	1,304	488
2010/11	1,269	538
2010/12	779	449
2011/01	1,251	645
2011/02	1,247	601
2011/03	1,475	798
2011/04	628	548
2011/05	993	748
2011/06	1,062	726
2011/07	454	369
2011/08	724	936
2011/09	877	1,076
2011/10	794	1,058
2011/11	885	1,264
2011/12	554	961
2012/01	752	1,293
2012/02	560	1,001
2012/03	766	1,307
2012/04	619	1,002
2012/05	624	1,113
2012/06	632	1,223
2012/07	295	602
2012/08	439	1,003
2012/09	399	1,149
2012/10	397	1,219
2012/11	357	1,157
2012/12	177	693
2013/01	507	1,252
2013/02	325	966
2013/03	332	929
2013/04	416	1,192
2013/05	331	1,157
2013/06	348	1,099

```
 2013/07                  100       412
 2013/08                  293       910
 2013/09                  250     1,108
 2013/10                  267     1,209
 2013/11                  238     1,252
 2013/12                  178       778
 2014/01                  234     1,390
 2014/02                  393     1,327
 2014/03                  548     1,501
 2014/04                  486     1,285
 2014/05                  425     1,324
 2014/06                  409     1,303
 2014/07                  268       589
 2014/08                  359     1,128
 2014/09                  509     1,418
 2014/10                  414     1,451
 2014/11                  469     1,360
 2014/12                  384     1,259
 2015/01                  362     1,359
 2015/02                  336     1,134

First visits from Diabasen                          17:36 Thursday, January 12, 2017   3

The CONTENTS Procedure

Data Set Name        TTDATA.DIAB                   Observations          57676
Member Type          DATA                          Variables             *
Engine               V9                            Indexes               0
Created              12/01/2017 17:36:50           Observation Length    24
Last Modified        12/01/2017 17:36:50           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    22
First Data Page             *
Max Obs per Page            2715
Obs in First Data Page      2664
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\diab.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12


            Alphabetic List of Variables and Attributes

#    Variable    Type    Len    Format    Informat    Label

2    doDiab      Num       8
1    pnr         Char     12    $12.      $10.        Personnummer
```

7.10 06-define

Collects records from the processed registers and defines a diabetes register and the relevant dates in it.

The inclusion date will be the smaller of the earliest dates from NPR, RMPS, DADD and DIAB, and the inclusion criterion will be the the one that triggered the inclusion.

Diabetes type is derived as described above.

Also derives a diabetes register exclusively based on drug information only.

```
1                                                          "Program 06-define.sas"

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_SRV12  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.06 seconds
      cpu time            0.07 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          /* get the TTdata.gdm, and exclude all registration in the year after the gdm
1        ! */
2          /* not here but in all other programs generating dates of inclusion */
3
4          data DMreg ;
5            merge TTdata.npr   ( in = npr   keep = pnr doNPR  nprtyp )
6                  TTdata.DVDD  ( in = dvdd  keep = pnr doDVDD dvdtyp )
7                  TTdata.pRMPS ( in = rmps  keep = pnr doRMPS doOAD doIns )
8                  TTdata.DIAB  ( in = diab  keep = pnr doDIAB )
9                  TTdata.pop   ( in = pop ) ;
10           by pnr ;
11           format doBth doDM doDth doNPR doDVDD doOAD doIns doDIAB  ddmmyy10. ;
12           if pop and ( npr or dvdd or rmps or diab ) ;
13           * Date of diagnosis - GDM and PCOS - taken care of in RMPS / DVDD ;
14           doDM = min( doNPR, doDVDD, doOAD, doIns, doDIAB ) ;
15           if doDM > doBth ;
16           if doDM eq doNPR  then inCr = "NPR"  ;
17           if doDM eq doDVDD then inCr = "DVDD" ;
18           if doDM eq doOAD  then inCr = "OAD"  ;
19           if doDM eq doIns  then inCr = "Ins"  ;
20           if doDM eq doDIAB then inCr = "DIAB" ;
21           * Type of diabetes is T1 if known from DVDD or NPR or
22             if any dispensation before age 15 or insulin before 30 ;
23           if ( nprtyp eq 'T1' and dvdtyp ne 'T2' )  or
24                                  dvdtyp eq 'T1'    or
25              .z < (doOAD - doBth) < ( 365.25 * 15 ) or
26              .z < (doIns - doBth) < ( 365.25 * 30 ) then DMtp = 'T1' ;
27           else DMtp = 'T2' ;
28           * never override DVDD verdict of T2 ;
29           if dvdtyp eq 'T2' then DMtp = 'T2' ;
30           * for simplification of tables ;
31           doDMx = max( doDM, '31DEC1994'd ) ;
32         run ;

NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      14385 at 14:10    36528 at 25:18    274757 at 26:18
NOTE: There were 223402 observations read from the data set TTDATA.NPR.
NOTE: There were 88398 observations read from the data set TTDATA.DVDD.
NOTE: There were 409844 observations read from the data set TTDATA.PRMPS.
NOTE: There were 57676 observations read from the data set TTDATA.DIAB.
```

```
NOTE: There were 7361669 observations read from the data set TTDATA.POP.
NOTE: The data set WORK.DMREG has 428251 observations and 16 variables.
NOTE: DATA statement used (Total process time):
      real time           3.93 seconds
      cpu time            2.57 seconds


33
34         title1 'Overview of typing of DM' ;
35         proc tabulate  data = DMreg  missing  noseps ;
36           class dvdtyp nprtyp DMtp inCr doDMx ;
37           table ( all   dmTP ) *
38                 ( all nprtyp ),
39                 ( all dvdtyp ) * f=comma7.
40                 / rts = 15 ;
41           keylabel n = ' ' ;
42           format doDMx year4. ;
43           run ;

NOTE: There were 428251 observations read from the data set WORK.DMREG.
NOTE: The PROCEDURE TABULATE printed page 1.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.12 seconds
      cpu time            0.17 seconds


44
45         title1 'The reconstructed diabetes register' ;
46         data TTdata.DMreg ;
47           set DMreg ;
48           keep pnr sex DMtp inCr /* inklaars */
49                doBth doDM doDth doNPR doDVDD doOAD doIns doDIAB ;
50           label DMtp = 'Type of DM'
51                 inCr = 'Inclusion criterion' ;
52         run ;

NOTE: There were 428251 observations read from the data set WORK.DMREG.
NOTE: The data set TTDATA.DMREG has 428251 observations and 12 variables.
NOTE: DATA statement used (Total process time):
      real time           0.20 seconds
      cpu time            0.09 seconds


53         proc contents  data = TTdata.DMreg ; run ;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.00 seconds
      cpu time            0.00 seconds

NOTE: The PROCEDURE CONTENTS printed page 2.

54
55         * add temporary variables for the tabulation ;
56         data a ;
57           set TTdata.DMreg ;
58           a1 = floor( ( doDM - doBth ) /  365.25    ) ;
59           a5 = floor( ( doDM - doBth ) / (365.25*5) ) * 5 ;
60           doDM = max( doDM, '31DEC1994'd ) ;
61         run ;

NOTE: There were 428251 observations read from the data set TTDATA.DMREG.
NOTE: The data set WORK.A has 428251 observations and 14 variables.
NOTE: DATA statement used (Total process time):
      real time           0.25 seconds
      cpu time            0.07 seconds


62
63         proc tabulate  data = a   missing  noseps ;
64           class sex doDM DMtp a1 a5 inCr ; * inklaars ;
```

```
65          table all doDM,
66              ( all DMtp ) * f = comma8.
67                 all * pctn< DMtp all > * f = 5.1
68                DMtp * pctn< DMtp all > * f = 4.1
69              / rts = 9  condense ;
70          table all doDM,
71              ( all inCr ) * f = comma7.
72                 all * pctn< inCr all > * f = 5.1
73                inCr * pctn< inCr all > * f = 4.1
74              / rts = 9  condense ;
75          table all a1,
76                sex * (
77                all * f = comma7.
78                 all * pctn< inCr all > * f = 5.1
79                inCr * pctn< inCr all > * f = 4.1 )
80              / rts = 9  condense ;
81          table sex, all a1 = "DM age",
82              ( all inCr ) * f = comma7.
83                 all * pctn< inCr all > * f = 5.1
84                inCr * pctn< inCr all > * f = 4.1
85              / rts = 9  condense ;
86          table all DMtp,
87                all a5,
88                 all * f = comma7.
89                doDM * f = comma6.
90              / rts = 6  condense ;
91          format doDM year4.
92                   sex koen_t. ;
93        run ;

NOTE: There were 428251 observations read from the data set WORK.A.
NOTE: At least one W.D format was too small for the number to be printed. The decimal may
      be shifted by the "BEST" format.
NOTE: The PROCEDURE TABULATE printed pages 3-7.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.15 seconds
      cpu time            0.31 seconds


94
95        title2 'Diagnoses 2012 ff. only' ;
96        proc tabulate  data = TTdata.DMreg  missing  noseps ;
97          where  doDM > '31DEC2011'd ;
98          class sex doDM inCr DMtp ;
99          table all doDM,
100             ( all DMtp ) * f = comma8.
101                all * pctn< DMtp all > * f = 5.1
102               DMtp * pctn< DMtp all > * f = 4.1
103             / rts = 9  condense ;
104         table all doDM,
105             ( all inCr ) * f = comma7.
106                all * pctn< inCr all > * f = 5.1
107               inCr * pctn< inCr all > * f = 4.1
108             / rts = 9  condense ;
109         format doDM yymms7.
110                  sex koen_t. ;
111       run ;

NOTE: There were 73169 observations read from the data set TTDATA.DMREG.
      WHERE doDM>'31DEC2011'D;
NOTE: The PROCEDURE TABULATE printed pages 8-9.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.04 seconds
      cpu time            0.10 seconds


112       title1 ;
113
114       title1 'The reconstructed diabetes *drug* register' ;
115       data TTdata.DMdreg ;
```

```
116          merge TTdata.pRMPS ( in = rmps  keep = pnr doOAD doIns )
117                TTdata.pop   ( in = pop ) ;
118          by pnr ;
119          keep pnr sex DMtp inCr
120                 doBth doDM doDth doOAD doIns ;
121          format doBth doDM doDth doOAD doIns  ddmmyy10. ;
122          if pop and rmps ;
123          * Date of diagnosis - GDM and PCOS - taken care of in RMPS / DVDD ;
124          doDM = min( doOAD, doIns ) ;
125          if doDM > doBth ;
126          * Type of diabetes is T1 if known from DVDD or
127            if any dispensation before age 15 or insulin before 30 ;
128          if .z < doOAD - doBth < ( 365.25 * 15 ) or
129             .z < doIns - doBth < ( 365.25 * 30 ) then DMtp = 'T1' ;
130          if DMtp ne 'T1' then DMtp = 'T2' ;
131          if doDM eq doOAD  then inCr = "OAD " ;
132          if doDM eq doIns  then inCr = "Ins " ;
133        run ;

NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      52995 at 128:17    260453 at 129:17
NOTE: There were 409844 observations read from the data set TTDATA.PRMPS.
NOTE: There were 7361669 observations read from the data set TTDATA.POP.
NOTE: The data set TTDATA.DMDREG has 409841 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           1.45 seconds
      cpu time            1.31 seconds


134
135        proc contents  data = TTdata.DMdreg ; run ;

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

NOTE: The PROCEDURE CONTENTS printed page 10.

136        proc tabulate  data = TTdata.DMdreg  missing  noseps ;
137          class sex doDM DMtp inCr ;
138          table all doDM,
139              ( all DMtp inCr ) * f = comma7.
140                 all * pctn<  all > * f = 6.1
141              ( DMtp * pctn< DMtp >
142                inCr * pctn< inCr > ) * f = 4.1
143              / rts = 6 ;
144          keylabel n = ' ' ;
145          format doDM year4.
146                  sex koen_t. ;
147        run ;

NOTE: There were 409841 observations read from the data set TTDATA.DMDREG.
NOTE: The PROCEDURE TABULATE printed page 11.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.17 seconds
      cpu time            0.23 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           6.51 seconds
      cpu time            4.99 seconds
```

7.10.1 06-define.lst

```
Overview of typing of DM                          14:29 Tuesday, January 24, 2017   1
```

		dvdtyp				
		All	NA	T1	T2	
All	All	428,251	339,982	3,552	23,886	60,831
	nprtyp					
		204,986	204,207	66	25	688
	NA	37,821	25,998	1,174	2,687	7,962
	T1	44,479	21,086	702	19,833	2,858
	T2	140,965	88,691	1,610	1,341	49,323
DMtp	All	47,339	22,602	851	23,886	.
T1	nprtyp					
		665	638	*	25	.
	NA	3,434	638	109	2,687	.
	T1	41,621	21,086	702	19,833	.
	T2	1,619	240	38	1,341	.
T2	All	380,912	317,380	2,701	.	60,831
	nprtyp					
		204,321	203,569	64	.	688
	NA	34,387	25,360	1,065	.	7,962
	T1	2,858	.	.	.	2,858
	T2	139,346	88,451	1,572	.	49,323

The reconstructed diabetes register 14:29 Tuesday, January 24, 2017 2

The CONTENTS Procedure

Data Set Name	TTDATA.DMREG	Observations	428251
Member Type	DATA	Variables	12
Engine	V9	Indexes	0
Created	24/01/2017 14:29:14	Observation Length	96
Last Modified	24/01/2017 14:29:14	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	65536
Number of Data Set Pages	629
First Data Page	*
Max Obs per Page	681
Obs in First Data Page	658
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	E:\workdata\705093\BxC\demoDM\DATA\dmreg.sas7bdat
Release Created	9.0401M3
Host Created	X64_SRV12

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
12	DMtp	Char	*			Type of DM
8	doBth	Num	8	DDMMYY10.		
10	doDM	Num	8	DDMMYY10.		
*	doDVDD	Num	8	DDMMYY10.		
6	doDiab	Num	8	DDMMYY10.		
9	doDth	Num	8	DDMMYY10.		
*	doIns	Num	8	DDMMYY10.		
*	doNPR	Num	8	DDMMYY10.	DATE9.	
*	doOAD	Num	8	DDMMYY10.		
11	inCr	Char	*			Inclusion criterion

```
*    pnr          Char     12    $12.          $10.        Personnummer
7    sex          Num       8
```

The reconstructed diabetes register 14:29 Tuesday, January 24, 2017 3

	Type of DM			Type of DM		
	All	T1	T2	All	T1	T2
	N	N	N	PctN	PctN	PctN
All	428,251	47,339	380,912	100.0	11.1	88.9
doDM						
1994	56,578	21,640	34,938	100.0	38.2	61.8
1995	38,055	3,151	34,904	100.0	8.3	91.7
1996	12,117	1,323	10,794	100.0	10.9	89.1
1997	11,589	1,306	10,283	100.0	11.3	88.7
1998	12,883	1,335	11,548	100.0	10.4	89.6
1999	12,788	1,183	11,605	100.0	9.3	90.7
2000	13,784	1,243	12,541	100.0	9.0	91.0
2001	13,685	1,260	12,425	100.0	9.2	90.8
2002	13,882	1,181	12,701	100.0	8.5	91.5
2003	15,817	1,129	14,688	100.0	7.1	92.9
2004	16,286	1,154	15,132	100.0	7.1	92.9
2005	16,124	1,094	15,030	100.0	6.8	93.2
2006	16,719	1,170	15,549	100.0	7.0	93.0
2007	17,861	1,126	16,735	100.0	6.3	93.7
2008	19,553	1,167	18,386	100.0	6.0	94.0
2009	20,034	1,097	18,937	100.0	5.5	94.5
2010	21,853	1,083	20,770	100.0	5.0	95.0
2011	25,474	1,015	24,459	100.0	4.0	96.0
2012	22,346	936	21,410	100.0	4.2	95.8
2013	17,049	935	16,114	100.0	5.5	94.5
2014	16,012	927	15,085	100.0	5.8	94.2
2015	17,762	884	16,878	100.0	5.0	95.0

The reconstructed diabetes register 14:29 Tuesday, January 24, 2017 4

	Inclusion criterion						Inclusion criterion					
	All	DIA	DVD	Ins	NPR	OAD	All	DIA	DVD	Ins	NPR	OAD
	N	N	N	N	N	N	PctN	PctN	PctN	PctN	PctN	PctN
All	428,251	509	69,363	12,080	84,063	262,236	100.0	0.1	16.2	2.8	19.6	61.2
doDM												
1994	56,578	.	23,636	.	32,942	.	100.0	.	41.8	.	58.2	.
1995	38,055	.	2,358	3,493	2,561	29,643	100.0	.	6.2	9.2	6.7	77.9
1996	12,117	.	1,995	478	2,389	7,255	100.0	.	16.5	3.9	19.7	59.9
1997	11,589	.	2,102	364	2,373	6,750	100.0	.	18.1	3.1	20.5	58.2
1998	12,883	.	2,535	387	2,643	7,318	100.0	.	19.7	3.0	20.5	56.8
1999	12,788	.	2,265	394	2,638	7,491	100.0	.	17.7	3.1	20.6	58.6
2000	13,784	.	3,086	396	2,714	7,588	100.0	.	22.4	2.9	19.7	55.0
2001	13,685	.	2,541	358	2,667	8,119	100.0	.	18.6	2.6	19.5	59.3
2002	13,882	.	2,573	345	2,851	8,113	100.0	.	18.5	2.5	20.5	58.4
2003	15,817	.	2,587	361	3,116	9,753	100.0	.	16.4	2.3	19.7	61.7
2004	16,286	.	2,671	384	2,826	10,405	100.0	.	16.4	2.4	17.4	63.9
2005	16,124	.	2,856	389	2,572	10,307	100.0	.	17.7	2.4	16.0	63.9
2006	16,719	.	2,858	397	2,523	10,941	100.0	.	17.1	2.4	15.1	65.4
2007	17,861	.	2,724	425	2,544	12,168	100.0	.	15.3	2.4	14.2	68.1
2008	19,553	.	2,623	477	2,366	14,087	100.0	.	13.4	2.4	12.1	72.0
2009	20,034	6	2,306	473	2,246	15,003	100.0	0.0	11.5	2.4	11.2	74.9
2010	21,853	48	2,265	431	2,005	17,104	100.0	0.2	10.4	2.0	9.2	78.3
2011	25,474	266	1,936	419	1,924	20,929	100.0	1.0	7.6	1.6	7.6	82.2
2012	22,346	80	1,522	440	1,791	18,513	100.0	0.4	6.8	2.0	8.0	82.8
2013	17,049	37	1,167	544	1,809	13,492	100.0	0.2	6.8	3.2	10.6	79.1

2014	16,012	63	733	539	2,074	12,603	100.0	0.4	4.6	3.4	13.0	78.7
2015	17,762	9	24	586	2,489	14,654	100.0	0.1	0.1	3.3	14.0	82.5

The reconstructed diabetes register 14:29 Tuesday, January 24, 2017 5

	sex												
	Mand							Kvinde					
	Inclusion criterion							Inclusion criterion					
	All	All	DIA	DVD	Ins	NPR	OAD	All	All	DIA	DVD	Ins	NPR	OAD
	N	PctN	PctN	PctN	PctN	PctN	PctN	N	PctN	PctN	PctN	PctN	PctN	PctN
All	236,401	100.0	0.1	17.3	2.5	19.5	60.5	191,850	100.0	0.1	14.8	3.2	19.7	62.1
a1														
0	124	100.0	.	51.6	4.0	42.7	1.6	126	100.0	.	60.3	2.4	34.1	3.2
*	280	100.0	.	50.7	1.1	47.1	1.1	235	100.0	.	47.2	2.6	48.9	1.3
*	298	100.0	.	49.7	2.0	47.3	1.0	226	100.0	.	53.5	0.4	45.1	0.9
*	303	100.0	.	51.5	0.7	46.9	1.0	291	100.0	.	50.5	1.4	47.4	0.7
*	319	100.0	.	52.0	1.6	45.8	0.6	297	100.0	.	50.8	0.7	47.8	0.7
*	334	100.0	.	55.4	0.9	42.8	0.9	368	100.0	.	56.0	0.5	43.5	.
6	387	100.0	.	54.5	0.8	44.2	0.5	367	100.0	.	61.0	0.8	37.9	0.3
7	366	100.0	.	54.4	0.8	44.8	.	374	100.0	.	60.2	0.3	39.3	0.3
8	384	100.0	.	53.4	0.5	45.6	0.5	435	100.0	.	62.8	0.5	35.9	0.9
9	479	100.0	.	58.7	0.6	40.3	0.4	507	100.0	.	60.7	0.6	38.3	0.4
10	450	100.0	.	60.0	0.9	39.1	.	526	100.0	.	62.5	0.8	35.6	1.1
11	563	100.0	.	65.2	1.1	33.2	0.5	585	100.0	.	59.8	1.4	37.4	1.4
12	649	100.0	.	60.4	0.6	37.8	1.2	559	100.0	.	62.8	0.7	35.1	1.4
13	635	100.0	.	61.7	1.3	35.7	1.3	453	100.0	.	60.0	1.5	34.7	3.8
14	604	100.0	.	61.8	0.7	35.6	2.0	409	100.0	.	62.3	1.0	31.5	5.1
15	519	100.0	.	64.0	0.8	33.9	1.3	437	100.0	.	54.9	2.7	27.2	15.1
16	475	100.0	.	64.8	2.1	30.7	2.3	438	100.0	.	56.4	1.8	20.8	21.0
17	483	100.0	.	72.9	1.9	23.0	2.3	522	100.0	.	47.7	1.5	21.6	29.1
18	488	100.0	.	63.9	1.8	30.3	3.9	555	100.0	.	41.6	1.6	17.7	39.1
19	518	100.0	.	66.8	2.3	26.3	4.6	598	100.0	.	37.3	1.5	18.6	42.6
20	449	100.0	.	62.4	3.3	32.1	2.2	436	100.0	.	55.3	4.8	28.2	11.7
21	509	100.0	.	67.0	3.9	25.9	3.1	472	100.0	0.4	49.6	3.4	31.6	15.0
22	575	100.0	.	63.8	3.1	29.6	3.5	522	100.0	0.2	48.5	4.0	30.7	16.7
23	542	100.0	.	64.2	2.2	29.0	4.6	567	100.0	.	51.3	4.1	30.3	14.3
24	633	100.0	.	64.3	2.8	29.2	3.6	607	100.0	.	46.5	4.6	29.2	19.8
25	594	100.0	.	58.4	3.9	32.7	5.1	644	100.0	.	43.5	4.2	30.6	21.7
26	722	100.0	.	60.2	3.7	29.4	6.6	703	100.0	0.1	41.8	6.8	33.1	18.1
27	653	100.0	0.2	58.3	2.8	28.3	10.4	705	100.0	0.3	42.4	4.5	27.0	25.8
28	798	100.0	.	53.9	4.0	32.6	9.5	744	100.0	0.1	40.6	6.3	29.6	23.4
29	838	100.0	0.1	55.3	4.3	28.4	11.9	777	100.0	0.3	41.7	6.3	29.1	22.7
30	862	100.0	0.1	52.7	3.7	29.9	13.6	826	100.0	0.2	39.5	7.4	29.5	23.4
31	923	100.0	0.2	49.6	3.0	31.4	15.7	844	100.0	0.4	40.2	8.3	28.6	22.6
32	1,039	100.0	0.1	48.4	3.6	31.9	16.1	880	100.0	0.2	39.2	9.0	29.4	22.2
33	1,149	100.0	0.1	49.1	3.1	31.4	16.3	879	100.0	.	37.7	9.9	30.3	22.2
34	1,177	100.0	0.1	46.9	3.7	32.0	17.3	974	100.0	0.3	39.0	10.0	28.6	22.1
35	1,302	100.0	0.2	45.1	3.1	31.3	20.4	981	100.0	0.1	36.8	9.5	30.3	23.3
36	1,451	100.0	0.1	43.7	3.6	33.8	18.8	994	100.0	0.2	40.2	9.1	29.1	21.4
37	1,524	100.0	0.1	45.0	2.9	32.8	19.2	1,077	100.0	0.2	38.0	9.5	28.6	23.8
38	1,593	100.0	0.1	44.6	3.3	33.7	18.3	1,146	100.0	0.3	40.8	9.2	29.4	20.2
39	1,819	100.0	.	47.2	3.1	33.2	16.5	1,270	100.0	0.4	44.7	7.8	29.4	17.7
40	3,216	100.0	0.0	25.4	1.6	17.0	56.0	3,457	100.0	0.1	14.4	3.5	9.1	72.9
41	2,656	100.0	0.1	32.3	3.0	23.1	41.5	2,035	100.0	0.1	24.9	5.3	15.0	54.7
42	2,725	100.0	0.1	31.0	2.5	22.6	43.9	2,039	100.0	.	25.6	5.0	16.2	53.3
43	3,044	100.0	0.0	30.8	2.2	21.6	45.4	2,006	100.0	0.1	26.3	3.5	18.5	51.6
44	3,264	100.0	.	29.8	2.4	22.1	45.7	2,083	100.0	0.0	26.2	3.6	18.7	51.4
45	3,436	100.0	0.1	27.6	2.2	22.3	47.8	2,244	100.0	0.1	26.0	3.6	17.9	52.5
46	3,723	100.0	0.1	26.3	2.2	21.2	50.3	2,346	100.0	0.1	25.5	3.7	17.6	53.1
47	4,008	100.0	0.2	25.8	2.1	20.3	51.6	2,380	100.0	.	25.0	3.4	17.1	54.5
48	4,201	100.0	0.1	24.5	1.9	20.4	53.2	2,611	100.0	0.1	23.9	3.3	17.8	55.0
49	4,522	100.0	0.1	25.0	2.6	19.3	53.1	2,752	100.0	.	23.2	3.1	17.4	56.3
50	4,789	100.0	0.0	23.6	1.8	18.5	56.1	2,952	100.0	0.2	20.4	2.8	16.4	60.3

51	4,999	100.0	0.0	22.0	1.7	19.1	57.2	3,143	100.0	0.1	21.4	2.7	17.9	57.9
52	5,053	100.0	.	19.6	1.7	19.4	59.3	3,307	100.0	0.1	20.5	2.0	18.1	59.4
53	5,137	100.0	0.1	18.8	2.2	18.1	60.8	3,450	100.0	0.1	18.8	2.4	18.8	59.9
54	5,517	100.0	0.1	17.7	2.1	18.9	61.2	3,417	100.0	0.2	16.9	1.8	18.9	62.2
55	5,616	100.0	0.1	17.2	1.8	19.6	61.3	3,590	100.0	0.1	16.4	1.9	19.5	62.1
56	5,727	100.0	0.1	16.2	1.9	18.1	63.8	3,770	100.0	0.1	16.1	2.0	19.6	62.3
57	6,042	100.0	0.1	15.2	1.9	17.7	65.2	3,802	100.0	0.2	14.4	1.7	20.6	63.1
58	6,058	100.0	0.1	14.0	1.8	18.5	65.5	3,981	100.0	0.1	14.8	1.7	18.4	65.0
59	6,097	100.0	0.2	13.9	2.0	18.1	65.8	4,067	100.0	0.1	15.4	1.8	18.5	64.2
60	6,370	100.0	0.2	12.7	2.0	16.8	68.4	4,234	100.0	0.1	12.2	2.0	19.4	66.2
61	6,547	100.0	0.2	10.6	2.0	17.1	70.2	4,276	100.0	0.1	12.2	2.2	18.7	66.7
62	6,400	100.0	0.1	10.3	2.0	16.6	70.9	4,324	100.0	0.2	10.2	2.1	19.3	68.2
63	6,342	100.0	0.2	9.1	2.3	16.7	71.7	4,465	100.0	0.2	10.2	2.2	19.4	68.0
64	6,450	100.0	0.2	8.6	2.2	16.7	72.4	4,610	100.0	0.1	8.9	2.3	18.0	70.7
65	6,170	100.0	0.1	7.6	2.5	16.3	73.5	4,622	100.0	0.0	7.7	2.6	18.5	71.2
66	5,983	100.0	0.1	7.2	2.2	16.7	73.8	4,582	100.0	0.1	7.5	2.7	19.1	70.5
67	5,869	100.0	0.2	6.5	2.3	16.0	75.0	4,549	100.0	0.3	7.2	2.4	18.7	71.4
68	5,662	100.0	0.2	6.1	2.8	17.0	74.0	4,582	100.0	0.2	5.5	2.3	20.3	71.8
69	5,614	100.0	0.2	5.1	2.9	16.1	75.8	4,367	100.0	0.1	5.5	2.5	19.9	72.0
70	5,189	100.0	0.2	5.1	2.5	16.3	75.9	4,375	100.0	0.2	5.1	2.5	19.3	72.9
71	4,941	100.0	0.2	4.4	2.8	15.8	76.7	4,299	100.0	0.1	3.8	2.9	19.3	73.9
72	4,669	100.0	0.1	4.1	2.9	17.8	75.0	4,315	100.0	0.1	4.3	3.0	17.8	74.8
73	4,645	100.0	0.2	3.6	3.1	15.9	77.2	4,353	100.0	0.2	3.0	3.1	18.6	75.2
74	4,273	100.0	0.3	3.1	3.1	16.2	77.3	4,151	100.0	0.1	3.4	2.9	17.8	75.8
75	3,910	100.0	0.1	3.2	3.2	16.4	77.0	3,902	100.0	0.2	3.0	3.4	18.6	74.7
76	3,885	100.0	0.2	2.6	3.3	17.2	76.8	3,860	100.0	0.1	2.1	2.7	19.4	75.7
77	3,445	100.0	0.1	1.9	2.8	17.3	77.8	3,633	100.0	0.1	2.4	2.7	18.9	75.9
78	3,190	100.0	0.3	1.9	3.2	16.6	78.0	3,479	100.0	0.2	1.5	3.2	18.1	77.0
79	2,924	100.0	0.2	1.9	3.6	15.6	78.6	3,340	100.0	0.0	1.8	3.8	18.6	75.8
80	2,651	100.0	0.1	1.8	3.5	15.5	79.0	3,134	100.0	0.1	1.2	3.8	17.2	77.7
81	2,315	100.0	0.2	1.5	3.5	15.5	79.4	2,899	100.0	0.1	1.1	4.1	16.5	78.1
82	2,068	100.0	0.1	1.9	4.2	14.4	79.4	2,695	100.0	0.1	0.7	3.8	16.2	79.1
83	1,810	100.0	.	1.3	4.5	15.4	78.8	2,440	100.0	0.2	1.1	4.3	17.5	77.0
84	1,501	100.0	0.1	1.1	3.5	15.8	79.5	2,165	100.0	0.1	0.8	4.3	18.2	76.6
85	1,352	100.0	0.1	1.0	4.6	15.4	78.9	1,997	100.0	0.1	0.7	3.8	15.7	79.8
86	1,111	100.0	0.2	0.3	4.3	16.1	79.1	1,708	100.0	0.1	0.6	4.1	17.4	77.7
87	890	100.0	0.1	0.4	4.6	14.3	80.6	1,514	100.0	0.1	0.2	4.2	15.1	80.4
88	701	100.0	0.1	0.7	4.7	15.3	79.2	1,236	100.0	0.2	0.4	4.5	16.5	78.4
89	514	100.0	.	0.2	3.7	14.0	82.1	974	100.0	.	0.5	4.9	17.2	77.3
90	404	100.0	.	.	5.4	14.6	80.0	815	100.0	.	0.5	4.0	14.0	81.5
91	297	100.0	.	0.3	5.1	19.5	75.1	604	100.0	.	.	5.3	13.2	81.5
92	230	100.0	.	0.4	7.0	17.0	75.7	456	100.0	.	0.2	7.7	11.0	81.1
93	138	100.0	.	0.7	3.6	16.7	79.0	364	100.0	.	0.3	6.3	13.7	79.7
94	112	100.0	.	.	5.4	18.8	75.9	267	100.0	0.4	0.4	3.7	13.9	81.6
95	87	100.0	.	.	5.7	11.5	82.8	169	100.0	.	.	11.2	13.0	75.7
96	38	100.0	.	.	5.3	5.3	89.5	126	100.0	.	.	7.1	15.1	77.8
97	29	100.0	.	.	3.4	10.3	86.2	74	100.0	.	.	4.1	24.3	71.6
98	17	100.0	.	.	5.9	5.9	88.2	54	100.0	.	.	3.7	24.1	72.2
99	11	100.0	.	.	18.2	18.2	63.6	24	100.0	.	.	12.5	12.5	75.0
100	*	100.0	.	.	.	.	100	15	100.0	.	.	6.7	20.0	73.3
101	*	100.0	.	.	.	.	100	9	100.0	.	.	11.1	11.1	77.8
102	*	100.0	.	.	50.0	.	50.0	*	100.0	.	.	.	.	100
104	.	.	.	.	.	.	.	*	100.0	.	.	.	.	100

The reconstructed diabetes register 14:29 Tuesday, January 24, 2017 6

sex Mand

		Inclusion criterion						Inclusion criterion				
	All	DIA	DVD	Ins	NPR	OAD	All	DIA	DVD	Ins	NPR	OAD
	N	N	N	N	N	N	PctN	PctN	PctN	PctN	PctN	PctN
All	236,401	277	40,999	5,896	46,213	143,016	100.0	0.1	17.3	2.5	19.5	60.5
DM age												
0	124	.	64	*	53	*	100.0	.	51.6	4.0	42.7	1.6
*	280	.	142	*	132	*	100.0	.	50.7	1.1	47.1	1.1
*	298	.	148	6	141	*	100.0	.	49.7	2.0	47.3	1.0

*	303	.	156	*	142	*	100.0	.	51.5	0.7	46.9	1.0
*	319	.	166	*	146	*	100.0	.	52.0	1.6	45.8	0.6
*	334	.	185	*	143	*	100.0	.	55.4	0.9	42.8	0.9
6	387	.	211	*	171	*	100.0	.	54.5	0.8	44.2	0.5
7	366	.	199	*	164	.	100.0	.	54.4	0.8	44.8	.
8	384	.	205	*	175	*	100.0	.	53.4	0.5	45.6	0.5
9	479	.	281	*	193	*	100.0	.	58.7	0.6	40.3	0.4
10	450	.	270	*	176	.	100.0	.	60.0	0.9	39.1	.
11	563	.	367	6	187	*	100.0	.	65.2	1.1	33.2	0.5
12	649	.	392	*	245	8	100.0	.	60.4	0.6	37.8	1.2
13	635	.	392	8	227	8	100.0	.	61.7	1.3	35.7	1.3
14	604	.	373	*	215	12	100.0	.	61.8	0.7	35.6	2.0
15	519	.	332	*	176	7	100.0	.	64.0	0.8	33.9	1.3
16	475	.	308	10	146	11	100.0	.	64.8	2.1	30.7	2.3
17	483	.	352	9	111	11	100.0	.	72.9	1.9	23.0	2.3
18	488	.	312	9	148	19	100.0	.	63.9	1.8	30.3	3.9
19	518	.	346	12	136	24	100.0	.	66.8	2.3	26.3	4.6
20	449	.	280	15	144	10	100.0	.	62.4	3.3	32.1	2.2
21	509	.	341	20	132	16	100.0	.	67.0	3.9	25.9	3.1
22	575	.	367	18	170	20	100.0	.	63.8	3.1	29.6	3.5
23	542	.	348	12	157	25	100.0	.	64.2	2.2	29.0	4.6
24	633	.	407	18	185	23	100.0	.	64.3	2.8	29.2	3.6
25	594	.	347	23	194	30	100.0	.	58.4	3.9	32.7	5.1
26	722	.	435	27	212	48	100.0	.	60.2	3.7	29.4	6.6
27	653	*	381	18	185	68	100.0	0.2	58.3	2.8	28.3	10.4
28	798	.	430	32	260	76	100.0	.	53.9	4.0	32.6	9.5
29	838	*	463	36	238	100	100.0	0.1	55.3	4.3	28.4	11.9
30	862	*	454	32	258	117	100.0	0.1	52.7	3.7	29.9	13.6
31	923	*	458	28	290	145	100.0	0.2	49.6	3.0	31.4	15.7
32	1,039	*	503	37	331	167	100.0	0.1	48.4	3.6	31.9	16.1
33	1,149	*	564	36	361	187	100.0	0.1	49.1	3.1	31.4	16.3
34	1,177	*	552	43	377	204	100.0	0.1	46.9	3.7	32.0	17.3
35	1,302	*	587	40	408	265	100.0	0.2	45.1	3.1	31.3	20.4
36	1,451	*	634	52	490	273	100.0	0.1	43.7	3.6	33.8	18.8
37	1,524	*	686	44	500	292	100.0	0.1	45.0	2.9	32.8	19.2
38	1,593	*	711	53	537	291	100.0	0.1	44.6	3.3	33.7	18.3
39	1,819	.	859	56	603	301	100.0	.	47.2	3.1	33.2	16.5
40	3,216	*	818	50	547	1,800	100.0	0.0	25.4	1.6	17.0	56.0
41	2,656	*	857	81	613	1,102	100.0	0.1	32.3	3.0	23.1	41.5
42	2,725	*	845	67	616	1,195	100.0	0.1	31.0	2.5	22.6	43.9
43	3,044	*	937	68	657	1,381	100.0	0.0	30.8	2.2	21.6	45.4
44	3,264	.	972	79	721	1,492	100.0	.	29.8	2.4	22.1	45.7
45	3,436	*	950	74	765	1,643	100.0	0.1	27.6	2.2	22.3	47.8
46	3,723	*	978	82	789	1,871	100.0	0.1	26.3	2.2	21.2	50.3
47	4,008	7	1,036	83	815	2,067	100.0	0.2	25.8	2.1	20.3	51.6
48	4,201	*	1,030	80	855	2,233	100.0	0.1	24.5	1.9	20.4	53.2
49	4,522	*	1,129	118	871	2,401	100.0	0.1	25.0	2.6	19.3	53.1
50	4,789	*	1,129	85	887	2,687	100.0	0.0	23.6	1.8	18.5	56.1
51	4,999	*	1,100	83	957	2,857	100.0	0.0	22.0	1.7	19.1	57.2
52	5,053	.	989	86	980	2,998	100.0	.	19.6	1.7	19.4	59.3
53	5,137	*	968	115	928	3,123	100.0	0.1	18.8	2.2	18.1	60.8
54	5,517	*	975	118	1,045	3,374	100.0	0.1	17.7	2.1	18.9	61.2
55	5,616	7	966	99	1,101	3,443	100.0	0.1	17.2	1.8	19.6	61.3
56	5,727	6	927	107	1,036	3,651	100.0	0.1	16.2	1.9	18.1	63.8
57	6,042	6	917	114	1,067	3,938	100.0	0.1	15.2	1.9	17.7	65.2
58	6,058	8	849	111	1,121	3,969	100.0	0.1	14.0	1.8	18.5	65.5
59	6,097	10	847	120	1,106	4,014	100.0	0.2	13.9	2.0	18.1	65.8
60	6,370	11	806	125	1,073	4,355	100.0	0.2	12.7	2.0	16.8	68.4
61	6,547	10	692	132	1,117	4,596	100.0	0.2	10.6	2.0	17.1	70.2
62	6,400	7	662	130	1,062	4,539	100.0	0.1	10.3	2.0	16.6	70.9
63	6,342	11	578	146	1,059	4,548	100.0	0.2	9.1	2.3	16.7	71.7
64	6,450	10	553	143	1,074	4,670	100.0	0.2	8.6	2.2	16.7	72.4
65	6,170	9	467	153	1,004	4,537	100.0	0.1	7.6	2.5	16.3	73.5
66	5,983	7	428	133	1,002	4,413	100.0	0.1	7.2	2.2	16.7	73.8
67	5,869	11	380	137	938	4,403	100.0	0.2	6.5	2.3	16.0	75.0
68	5,662	10	344	156	962	4,190	100.0	0.2	6.1	2.8	17.0	74.0
69	5,614	11	286	160	902	4,255	100.0	0.2	5.1	2.9	16.1	75.8
70	5,189	8	264	132	848	3,937	100.0	0.2	5.1	2.5	16.3	75.9
71	4,941	9	219	140	783	3,790	100.0	0.2	4.4	2.8	15.8	76.7
72	4,669	*	193	136	832	3,504	100.0	0.1	4.1	2.9	17.8	75.0

73	4,645	8	169	144	740	3,584	100.0	0.2	3.6	3.1	15.9	77.2
74	4,273	13	133	131	691	3,305	100.0	0.3	3.1	3.1	16.2	77.3
75	3,910	*	125	125	643	3,012	100.0	0.1	3.2	3.2	16.4	77.0
76	3,885	6	100	128	667	2,984	100.0	0.2	2.6	3.3	17.2	76.8
77	3,445	*	64	98	597	2,681	100.0	0.1	1.9	2.8	17.3	77.8
78	3,190	8	62	103	530	2,487	100.0	0.3	1.9	3.2	16.6	78.0
79	2,924	6	57	106	457	2,298	100.0	0.2	1.9	3.6	15.6	78.6
80	2,651	*	48	94	412	2,094	100.0	0.1	1.8	3.5	15.5	79.0
81	2,315	*	34	80	359	1,838	100.0	0.2	1.5	3.5	15.5	79.4
82	2,068	*	40	86	298	1,641	100.0	0.1	1.9	4.2	14.4	79.4
83	1,810	.	23	82	279	1,426	100.0	.	1.3	4.5	15.4	78.8
84	1,501	*	16	52	237	1,194	100.0	0.1	1.1	3.5	15.8	79.5
85	1,352	*	13	62	208	1,067	100.0	0.1	1.0	4.6	15.4	78.9
86	1,111	*	*	48	179	879	100.0	0.2	0.3	4.3	16.1	79.1
87	890	*	*	41	127	717	100.0	0.1	0.4	4.6	14.3	80.6
88	701	*	*	33	107	555	100.0	0.1	0.7	4.7	15.3	79.2
89	514	.	*	19	72	422	100.0	.	0.2	3.7	14.0	82.1
90	404	.	.	22	59	323	100.0	.	.	5.4	14.6	80.0
91	297	.	*	15	58	223	100.0	.	0.3	5.1	19.5	75.1
92	230	.	*	16	39	174	100.0	.	0.4	7.0	17.0	75.7
93	138	.	*	*	23	109	100.0	.	0.7	3.6	16.7	79.0
94	112	.	.	6	21	85	100.0	.	.	5.4	18.8	75.9
95	87	.	.	*	10	72	100.0	.	.	5.7	11.5	82.8
96	38	.	.	*	*	34	100.0	.	.	5.3	5.3	89.5
97	29	.	.	*	*	25	100.0	.	.	3.4	10.3	86.2
98	17	.	.	*	*	15	100.0	.	.	5.9	5.9	88.2
99	11	.	.	*	*	7	100.0	.	.	18.2	18.2	63.6
100	*	.	.	.	.	*	100.0	.	.	.	.	100
101	*	.	.	.	.	*	100.0	.	.	.	.	100
102	*	.	.	*	.	*	100.0	.	.	50.0	.	50.0

sex Kvinde

		Inclusion criterion					Inclusion criterion					
	All	DIA	DVD	Ins	NPR	OAD	All	DIA	DVD	Ins	NPR	OAD
	N	N	N	N	N	N	PctN	PctN	PctN	PctN	PctN	PctN
All	191,850	232	28,364	6,184	37,850	119,220	100.0	0.1	14.8	3.2	19.7	62.1
DM age												
0	126	.	76	*	43	*	100.0	.	60.3	2.4	34.1	3.2
*	235	.	111	6	115	*	100.0	.	47.2	2.6	48.9	1.3
*	226	.	121	*	102	*	100.0	.	53.5	0.4	45.1	0.9
*	291	.	147	*	138	*	100.0	.	50.5	1.4	47.4	0.7
*	297	.	151	*	142	*	100.0	.	50.8	0.7	47.8	0.7
*	368	.	206	*	160	.	100.0	.	56.0	0.5	43.5	.
6	367	.	224	*	139	*	100.0	.	61.0	0.8	37.9	0.3
7	374	.	225	*	147	*	100.0	.	60.2	0.3	39.3	0.3
8	435	.	273	*	156	*	100.0	.	62.8	0.5	35.9	0.9
9	507	.	308	*	194	*	100.0	.	60.7	0.6	38.3	0.4
10	526	.	329	*	187	6	100.0	.	62.5	0.8	35.6	1.1
11	585	.	350	8	219	8	100.0	.	59.8	1.4	37.4	1.4
12	559	.	351	*	196	8	100.0	.	62.8	0.7	35.1	1.4
13	453	.	272	7	157	17	100.0	.	60.0	1.5	34.7	3.8
14	409	.	255	*	129	21	100.0	.	62.3	1.0	31.5	5.1
15	437	.	240	12	119	66	100.0	.	54.9	2.7	27.2	15.1
16	438	.	247	8	91	92	100.0	.	56.4	1.8	20.8	21.0
17	522	.	249	8	113	152	100.0	.	47.7	1.5	21.6	29.1
18	555	.	231	9	98	217	100.0	.	41.6	1.6	17.7	39.1
19	598	.	223	9	111	255	100.0	.	37.3	1.5	18.6	42.6
20	436	.	241	21	123	51	100.0	.	55.3	4.8	28.2	11.7
21	472	*	234	16	149	71	100.0	0.4	49.6	3.4	31.6	15.0
22	522	*	253	21	160	87	100.0	0.2	48.5	4.0	30.7	16.7
23	567	.	291	23	172	81	100.0	.	51.3	4.1	30.3	14.3
24	607	.	282	28	177	120	100.0	.	46.5	4.6	29.2	19.8

25	644	.	280	27	197	140	100.0	.	43.5	4.2	30.6	21.7
26	703	*	294	48	233	127	100.0	0.1	41.8	6.8	33.1	18.1
27	705	*	299	32	190	182	100.0	0.3	42.4	4.5	27.0	25.8
28	744	*	302	47	220	174	100.0	0.1	40.6	6.3	29.6	23.4
29	777	*	324	49	226	176	100.0	0.3	41.7	6.3	29.1	22.7
30	826	*	326	61	244	193	100.0	0.2	39.5	7.4	29.5	23.4
31	844	*	339	70	241	191	100.0	0.4	40.2	8.3	28.6	22.6
32	880	*	345	79	259	195	100.0	0.2	39.2	9.0	29.4	22.2
33	879	.	331	87	266	195	100.0	.	37.7	9.9	30.3	22.2
34	974	*	380	97	279	215	100.0	0.3	39.0	10.0	28.6	22.1
35	981	*	361	93	297	229	100.0	0.1	36.8	9.5	30.3	23.3
36	994	*	400	90	289	213	100.0	0.2	40.2	9.1	29.1	21.4
37	1,077	*	409	102	308	256	100.0	0.2	38.0	9.5	28.6	23.8
38	1,146	*	468	106	337	232	100.0	0.3	40.8	9.2	29.4	20.2
39	1,270	*	568	99	373	225	100.0	0.4	44.7	7.8	29.4	17.7
40	3,457	*	497	121	316	2,521	100.0	0.1	14.4	3.5	9.1	72.9
41	2,035	*	506	107	306	1,114	100.0	0.1	24.9	5.3	15.0	54.7
42	2,039	.	521	102	330	1,086	100.0	.	25.6	5.0	16.2	53.3
43	2,006	*	527	70	371	1,035	100.0	0.1	26.3	3.5	18.5	51.6
44	2,083	*	546	76	390	1,070	100.0	0.0	26.2	3.6	18.7	51.4
45	2,244	*	584	80	401	1,177	100.0	0.1	26.0	3.6	17.9	52.5
46	2,346	*	598	87	412	1,246	100.0	0.1	25.5	3.7	17.6	53.1
47	2,380	.	595	81	408	1,296	100.0	.	25.0	3.4	17.1	54.5
48	2,611	*	623	87	464	1,435	100.0	0.1	23.9	3.3	17.8	55.0
49	2,752	.	638	86	478	1,550	100.0	.	23.2	3.1	17.4	56.3
50	2,952	6	601	83	483	1,779	100.0	0.2	20.4	2.8	16.4	60.3
51	3,143	*	673	85	564	1,819	100.0	0.1	21.4	2.7	17.9	57.9
52	3,307	*	679	65	597	1,964	100.0	0.1	20.5	2.0	18.1	59.4
53	3,450	*	649	83	650	2,065	100.0	0.1	18.8	2.4	18.8	59.9
54	3,417	6	579	60	645	2,127	100.0	0.2	16.9	1.8	18.9	62.2
55	3,590	*	587	70	699	2,230	100.0	0.1	16.4	1.9	19.5	62.1
56	3,770	*	607	74	738	2,347	100.0	0.1	16.1	2.0	19.6	62.3
57	3,802	7	546	66	783	2,400	100.0	0.2	14.4	1.7	20.6	63.1
58	3,981	*	588	68	734	2,587	100.0	0.1	14.8	1.7	18.4	65.0
59	4,067	*	626	75	751	2,611	100.0	0.1	15.4	1.8	18.5	64.2
60	4,234	6	517	86	821	2,804	100.0	0.1	12.2	2.0	19.4	66.2
61	4,276	*	523	96	798	2,854	100.0	0.1	12.2	2.2	18.7	66.7
62	4,324	9	440	92	834	2,949	100.0	0.2	10.2	2.1	19.3	68.2
63	4,465	8	455	99	865	3,038	100.0	0.2	10.2	2.2	19.4	68.0
64	4,610	*	410	106	832	3,257	100.0	0.1	8.9	2.3	18.0	70.7
65	4,622	*	355	122	854	3,289	100.0	0.0	7.7	2.6	18.5	71.2
66	4,582	6	344	123	877	3,232	100.0	0.1	7.5	2.7	19.1	70.5
67	4,549	12	326	111	850	3,250	100.0	0.3	7.2	2.4	18.7	71.4
68	4,582	7	251	104	931	3,289	100.0	0.2	5.5	2.3	20.3	71.8
69	4,367	*	242	107	871	3,143	100.0	0.1	5.5	2.5	19.9	72.0
70	4,375	9	223	111	844	3,188	100.0	0.2	5.1	2.5	19.3	72.9
71	4,299	6	163	123	830	3,177	100.0	0.1	3.8	2.9	19.3	73.9
72	4,315	*	184	128	770	3,228	100.0	0.1	4.3	3.0	17.8	74.8
73	4,353	7	129	135	809	3,273	100.0	0.2	3.0	3.1	18.6	75.2
74	4,151	*	141	119	740	3,148	100.0	0.1	3.4	2.9	17.8	75.8
75	3,902	6	119	134	727	2,916	100.0	0.2	3.0	3.4	18.6	74.7
76	3,860	*	82	105	747	2,921	100.0	0.1	2.1	2.7	19.4	75.7
77	3,633	*	86	97	688	2,757	100.0	0.1	2.4	2.7	18.9	75.9
78	3,479	8	51	113	628	2,679	100.0	0.2	1.5	3.2	18.1	77.0
79	3,340	*	59	127	621	2,532	100.0	0.0	1.8	3.8	18.6	75.8
80	3,134	*	37	120	540	2,435	100.0	0.1	1.2	3.8	17.2	77.7
81	2,899	*	33	120	478	2,264	100.0	0.1	1.1	4.1	16.5	78.1
82	2,695	*	20	102	437	2,132	100.0	0.1	0.7	3.8	16.2	79.1
83	2,440	*	26	104	426	1,880	100.0	0.2	1.1	4.3	17.5	77.0
84	2,165	*	18	94	393	1,658	100.0	0.1	0.8	4.3	18.2	76.6
85	1,997	*	13	76	314	1,593	100.0	0.1	0.7	3.8	15.7	79.8
86	1,708	*	11	70	298	1,327	100.0	0.1	0.6	4.1	17.4	77.7
87	1,514	*	*	63	229	1,217	100.0	0.1	0.2	4.2	15.1	80.4
88	1,236	*	*	56	204	969	100.0	0.2	0.4	4.5	16.5	78.4
89	974	.	*	48	168	753	100.0	.	0.5	4.9	17.2	77.3
90	815	.	*	33	114	664	100.0	.	0.5	4.0	14.0	81.5
91	604	.	.	32	80	492	100.0	.	.	5.3	13.2	81.5
92	456	.	*	35	50	370	100.0	.	0.2	7.7	11.0	81.1
93	364	.	*	23	50	290	100.0	.	0.3	6.3	13.7	79.7
94	267	*	*	10	37	218	100.0	0.4	0.4	3.7	13.9	81.6

95	169	.	.	19	22	128	100.0	.	.	11.2	13.0	75.7
96	126	.	.	9	19	98	100.0	.	.	7.1	15.1	77.8
97	74	.	.	*	18	53	100.0	.	.	4.1	24.3	71.6
98	54	.	.	*	13	39	100.0	.	.	3.7	24.1	72.2
99	24	.	.	*	*	18	100.0	.	.	12.5	12.5	75.0
100	15	.	.	*	*	11	100.0	.	.	6.7	20.0	73.3
101	9	.	.	*	*	7	100.0	.	.	11.1	11.1	77.8
102	*	.	.	.	.	*	100.0	.	.	.	.	100
104	*	.	.	.	.	*	100.0	.	.	.	.	100

All

		doDM								
	All	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
	N	N	N	N	N	N	N	N	N	N	N
All	428,251	56,578	38,055	12,117	11,589	12,883	12,788	13,784	13,685	13,882	15,817
a5											
0	2,499	1,322	60	58	62	49	67	59	57	47	56
*	4,001	2,012	71	55	80	74	68	98	100	100	95
10	5,433	2,750	81	81	86	82	103	95	126	106	101
15	5,033	2,253	78	98	93	91	76	81	85	106	91
20	5,312	2,511	145	120	114	121	133	103	134	125	132
25	7,178	2,857	255	189	183	203	176	212	189	188	185
30	9,553	3,168	540	300	279	319	303	311	306	264	322
35	13,157	3,717	831	412	405	425	419	537	494	485	529
40	26,525	4,483	1,144	588	567	657	676	810	727	683	861
45	32,223	5,089	1,801	913	820	946	875	1,046	974	1,059	1,164
50	41,764	5,049	2,459	1,312	1,289	1,503	1,444	1,595	1,446	1,434	1,552
55	48,750	4,894	2,980	1,200	1,262	1,449	1,551	1,688	1,848	1,947	2,181
60	54,018	4,475	4,067	1,367	1,310	1,455	1,499	1,660	1,626	1,768	2,129
65	52,000	4,096	5,127	1,418	1,275	1,492	1,459	1,502	1,527	1,605	1,856
70	45,210	3,505	6,017	1,421	1,300	1,406	1,341	1,419	1,431	1,425	1,623
75	35,568	2,517	5,461	1,213	1,215	1,235	1,193	1,191	1,153	1,116	1,356
80	23,678	1,299	4,119	855	741	800	834	786	835	848	926
85	11,997	478	2,101	407	384	439	420	437	464	411	465
90	3,687	95	624	89	106	120	133	129	140	139	161
95	629	8	87	21	17	16	18	24	22	23	29
100	36	.	7	.	*	*	.	*	*	*	*

(Continued)

						doDM					
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
	N	N	N	N	N	N	N	N	N	N	N	N
All	16,286	16,124	16,719	17,861	19,553	20,034	21,853	25,474	22,346	17,049	16,012	17,762
a5												
0	42	58	54	49	64	64	49	62	49	64	50	57
*	105	117	81	92	91	110	112	117	88	116	115	104
10	139	123	163	143	159	146	179	143	156	150	154	167
15	91	137	143	137	193	167	187	189	212	175	183	167
20	95	113	140	135	149	164	159	141	152	132	136	158
25	223	246	225	212	242	222	209	207	207	168	193	187
30	295	304	320	351	327	341	296	304	256	221	218	208
35	484	512	462	446	449	411	421	399	339	350	322	308
40	985	977	1,134	1,252	1,269	1,315	1,479	1,528	1,438	1,288	1,329	1,335
45	1,152	1,137	1,202	1,283	1,477	1,612	1,662	1,900	1,767	1,448	1,391	1,505
50	1,585	1,493	1,683	1,738	1,841	1,998	2,176	2,454	2,201	1,711	1,749	2,052
55	2,298	2,223	2,034	2,215	2,326	2,390	2,628	3,020	2,607	2,048	1,837	2,124

60	2,210	2,287	2,549	2,704	3,124	3,077	3,424	3,785	3,172	2,141	1,993	2,196
65	1,884	1,913	2,015	2,181	2,563	2,672	3,046	4,014	3,385	2,406	2,163	2,401
70	1,727	1,626	1,675	1,902	2,011	2,145	2,340	2,919	2,529	1,786	1,672	1,990
75	1,358	1,252	1,299	1,377	1,553	1,489	1,680	2,128	1,876	1,346	1,186	1,374
80	962	944	910	937	997	1,000	1,067	1,296	1,107	864	739	812
85	451	480	431	504	511	528	541	623	588	456	438	440
90	169	154	166	162	177	154	153	212	184	155	118	147
95	29	28	32	39	29	26	43	31	31	21	26	29
100	*	.	*	*	*	*	*	*	*	*	.	*

Type of DM T1

	doDM									
	All	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
	N	N	N	N	N	N	N	N	N	N	N
All	47,339	21,640	3,151	1,323	1,306	1,335	1,183	1,243	1,260	1,181	1,129
a5											
0	2,338	1,231	54	54	58	47	63	58	52	42	52
*	3,878	1,945	71	53	79	73	66	97	97	100	92
10	5,215	2,626	80	79	81	79	100	91	123	103	100
15	3,603	1,949	65	86	77	76	58	65	67	71	63
20	3,519	1,865	114	82	81	85	83	64	93	86	75
25	3,677	1,803	164	103	114	114	83	101	100	87	72
30	3,061	1,565	125	89	86	104	95	83	87	81	64
35	2,805	1,317	119	82	88	75	61	102	86	78	89
40	2,724	1,202	108	90	77	100	86	79	78	70	65
45	2,526	1,103	173	88	76	102	59	68	67	63	56
50	2,461	975	199	99	94	89	87	90	79	68	58
55	2,359	920	256	73	80	69	78	82	69	82	71
60	2,244	840	289	86	70	75	58	57	66	45	75
65	2,132	819	371	75	61	66	67	60	74	60	41
70	1,919	690	403	71	77	71	58	46	49	50	55
75	1,484	495	286	61	55	52	41	48	39	41	53
80	895	216	200	32	33	39	25	24	19	35	27
85	367	63	63	18	15	15	9	23	13	12	16
90	111	14	10	*	*	*	6	*	*	*	*
95	20	*	*	.	*	.	.	.	*	*	.
100	*	.	.	.	.	.	.	.	.	.	.

(Continued)

	doDM											
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
	N	N	N	N	N	N	N	N	N	N	N	N
All	1,154	1,094	1,170	1,126	1,167	1,097	1,083	1,015	936	935	927	884
a5												
0	40	55	53	48	60	61	46	56	48	61	46	53
*	103	107	79	91	91	108	105	112	83	112	110	104
10	131	115	159	138	150	143	171	136	153	148	151	158
15	57	80	71	61	94	90	83	94	108	85	97	106
20	57	57	61	66	78	82	76	74	86	77	82	95
25	70	69	95	80	76	72	76	67	65	72	85	109
30	65	66	74	70	57	62	64	51	46	44	54	29
35	63	80	81	59	65	73	75	52	52	43	39	26
40	71	56	79	87	88	65	72	65	47	55	40	44
45	74	58	55	68	66	60	52	57	51	56	37	37
50	71	52	71	75	50	51	48	39	35	45	56	30

55	88	71	61	68	45	44	46	47	35	28	24	22
60	69	60	73	60	76	55	47	44	35	26	27	11
65	54	45	39	40	55	34	35	42	36	21	22	15
70	46	35	37	39	35	30	30	26	21	19	20	11
75	40	32	36	37	33	26	24	18	15	18	19	15
80	27	32	35	22	32	19	17	21	11	12	8	9
85	20	14	9	12	9	13	8	11	7	6	6	*
90	6	9	*	*	6	8	6	*	*	7	*	*
95	*	*	*	*	*	.	*	*	.	.	.	*
100	.	.	.	.	.	*	.	.	.	.	.	.

Type of DM T2

	doDM										
	All	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
	N	N	N	N	N	N	N	N	N	N	N
All	380,912	34,938	34,904	10,794	10,283	11,548	11,605	12,541	12,425	12,701	14,688
a5											
0	161	91	6	*	*	*	*	*	*	*	*
*	123	67	.	*	*	*	*	*	*	.	*
10	218	124	*	*	*	*	*	*	*	*	*
15	1,430	304	13	12	16	15	18	16	18	35	28
20	1,793	646	31	38	33	36	50	39	41	39	57
25	3,501	1,054	91	86	69	89	93	111	89	101	113
30	6,492	1,603	415	211	193	215	208	228	219	183	258
35	10,352	2,400	712	330	317	350	358	435	408	407	440
40	23,801	3,281	1,036	498	490	557	590	731	649	613	796
45	29,697	3,986	1,628	825	744	844	816	978	907	996	1,108
50	39,303	4,074	2,260	1,213	1,195	1,414	1,357	1,505	1,367	1,366	1,494
55	46,391	3,974	2,724	1,127	1,182	1,380	1,473	1,606	1,779	1,865	2,110
60	51,774	3,635	3,778	1,281	1,240	1,380	1,441	1,603	1,560	1,723	2,054
65	49,868	3,277	4,756	1,343	1,214	1,426	1,392	1,442	1,453	1,545	1,815
70	43,291	2,815	5,614	1,350	1,223	1,335	1,283	1,373	1,382	1,375	1,568
75	34,084	2,022	5,175	1,152	1,160	1,183	1,152	1,143	1,114	1,075	1,303
80	22,783	1,083	3,919	823	708	761	809	762	816	813	899
85	11,630	415	2,038	389	369	424	411	414	451	399	449
90	3,576	81	614	87	103	116	127	124	139	134	156
95	609	6	86	21	16	16	18	24	21	21	29
100	35	.	7	.	*	*	.	*	*	*	*

(Continued)

	doDM											
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
	N	N	N	N	N	N	N	N	N	N	N	N
All	15,132	15,030	15,549	16,735	18,386	18,937	20,770	24,459	21,410	16,114	15,085	16,878
a5												
0	*	*	*	*	*	*	*	6	*	*	*	*
*	*	10	*	*	.	*	7	*	*	*	*	.
10	8	8	*	*	9	*	8	7	*	*	*	9
15	34	57	72	76	99	77	104	95	104	90	86	61
20	38	56	79	69	71	82	83	67	66	55	54	63
25	153	177	130	132	166	150	133	140	142	96	108	78
30	230	238	246	281	270	279	232	253	210	177	164	179
35	421	432	381	387	384	338	346	347	287	307	283	282
40	914	921	1,055	1,165	1,181	1,250	1,407	1,463	1,391	1,233	1,289	1,291
45	1,078	1,079	1,147	1,215	1,411	1,552	1,610	1,843	1,716	1,392	1,354	1,468

50	1,514	1,441	1,612	1,663	1,791	1,947	2,128	2,415	2,166	1,666	1,693	2,022
55	2,210	2,152	1,973	2,147	2,281	2,346	2,582	2,973	2,572	2,020	1,813	2,102
60	2,141	2,227	2,476	2,644	3,048	3,022	3,377	3,741	3,137	2,115	1,966	2,185
65	1,830	1,868	1,976	2,141	2,508	2,638	3,011	3,972	3,349	2,385	2,141	2,386
70	1,681	1,591	1,638	1,863	1,976	2,115	2,310	2,893	2,508	1,767	1,652	1,979
75	1,318	1,220	1,263	1,340	1,520	1,463	1,656	2,110	1,861	1,328	1,167	1,359
80	935	912	875	915	965	981	1,050	1,275	1,096	852	731	803
85	431	466	422	492	502	515	533	612	581	450	432	435
90	163	145	165	160	171	146	147	210	182	148	114	144
95	27	27	31	36	28	26	41	30	31	21	26	27
100	*	.	*	*	*	*	*	*	*	*	.	*

		Type of DM			Type of DM	
	All	T1	T2	All	T1	T2
	N	N	N	PctN	PctN	PctN
All	73,169	3,682	69,487	100.0	5.0	95.0
doDM						
2012/01	2,955	298	2,657	100.0	10.1	89.9
2012/02	1,851	47	1,804	100.0	2.5	97.5
2012/03	2,211	57	2,154	100.0	2.6	97.4
2012/04	1,583	42	1,541	100.0	2.7	97.3
2012/05	1,950	66	1,884	100.0	3.4	96.6
2012/06	2,185	75	2,110	100.0	3.4	96.6
2012/07	1,271	51	1,220	100.0	4.0	96.0
2012/08	1,526	63	1,463	100.0	4.1	95.9
2012/09	1,695	63	1,632	100.0	3.7	96.3
2012/10	1,827	72	1,755	100.0	3.9	96.1
2012/11	1,813	61	1,752	100.0	3.4	96.6
2012/12	1,479	41	1,438	100.0	2.8	97.2
2013/01	2,435	250	2,185	100.0	10.3	89.7
2013/02	1,494	80	1,414	100.0	5.4	94.6
2013/03	1,468	72	1,396	100.0	4.9	95.1
2013/04	1,511	66	1,445	100.0	4.4	95.6
2013/05	1,363	52	1,311	100.0	3.8	96.2
2013/06	1,369	63	1,306	100.0	4.6	95.4
2013/07	1,029	41	988	100.0	4.0	96.0
2013/08	1,170	60	1,110	100.0	5.1	94.9
2013/09	1,342	51	1,291	100.0	3.8	96.2
2013/10	1,318	61	1,257	100.0	4.6	95.4
2013/11	1,320	73	1,247	100.0	5.5	94.5
2013/12	1,230	66	1,164	100.0	5.4	94.6
2014/01	1,840	174	1,666	100.0	9.5	90.5
2014/02	1,277	67	1,210	100.0	5.2	94.8
2014/03	1,475	75	1,400	100.0	5.1	94.9
2014/04	1,162	62	1,100	100.0	5.3	94.7
2014/05	1,304	80	1,224	100.0	6.1	93.9
2014/06	1,306	47	1,259	100.0	3.6	96.4
2014/07	1,049	62	987	100.0	5.9	94.1
2014/08	1,079	71	1,008	100.0	6.6	93.4
2014/09	1,414	76	1,338	100.0	5.4	94.6
2014/10	1,410	61	1,349	100.0	4.3	95.7
2014/11	1,342	81	1,261	100.0	6.0	94.0
2014/12	1,354	71	1,283	100.0	5.2	94.8
2015/01	1,421	87	1,334	100.0	6.1	93.9
2015/02	1,423	80	1,343	100.0	5.6	94.4
2015/03	1,670	82	1,588	100.0	4.9	95.1
2015/04	1,432	70	1,362	100.0	4.9	95.1
2015/05	1,387	71	1,316	100.0	5.1	94.9
2015/06	1,629	69	1,560	100.0	4.2	95.8
2015/07	1,168	65	1,103	100.0	5.6	94.4
2015/08	1,310	66	1,244	100.0	5.0	95.0

2015/09	1,601	83	1,518	100.0	5.2	94.8
2015/10	1,589	82	1,507	100.0	5.2	94.8
2015/11	1,515	65	1,450	100.0	4.3	95.7
2015/12	1,617	64	1,553	100.0	4.0	96.0

The reconstructed diabetes register 14:29 Tuesday, January 24, 2017 9
Diagnoses 2012 ff. only

			Inclusion criterion						Inclusion criterion			
	All	DIA	DVD	Ins	NPR	OAD	All	DIA	DVD	Ins	NPR	OAD
	N	N	N	N	N	N	PctN	PctN	PctN	PctN	PctN	PctN
All	73,169	189	3,446	2,109	8,163	59,262	100.0	0.3	4.7	2.9	11.2	81.0
doDM												
2012/01	2,955	18	965	34	180	1,758	100.0	0.6	32.7	1.2	6.1	59.5
2012/02	1,851	10	46	40	134	1,621	100.0	0.5	2.5	2.2	7.2	87.6
2012/03	2,211	8	59	35	155	1,954	100.0	0.4	2.7	1.6	7.0	88.4
2012/04	1,583	6	39	36	125	1,377	100.0	0.4	2.5	2.3	7.9	87.0
2012/05	1,950	*	58	30	172	1,685	100.0	0.3	3.0	1.5	8.8	86.4
2012/06	2,185	9	135	44	159	1,838	100.0	0.4	6.2	2.0	7.3	84.1
2012/07	1,271	*	26	31	114	1,096	100.0	0.3	2.0	2.4	9.0	86.2
2012/08	1,526	*	47	35	169	1,271	100.0	0.3	3.1	2.3	11.1	83.3
2012/09	1,695	6	34	41	140	1,474	100.0	0.4	2.0	2.4	8.3	87.0
2012/10	1,827	*	38	39	184	1,564	100.0	0.1	2.1	2.1	10.1	85.6
2012/11	1,813	7	34	40	159	1,573	100.0	0.4	1.9	2.2	8.8	86.8
2012/12	1,479	*	41	35	100	1,302	100.0	0.1	2.8	2.4	6.8	88.0
2013/01	2,435	*	776	35	168	1,452	100.0	0.2	31.9	1.4	6.9	59.6
2013/02	1,494	*	40	41	166	1,244	100.0	0.2	2.7	2.7	11.1	83.3
2013/03	1,468	*	40	44	166	1,214	100.0	0.3	2.7	3.0	11.3	82.7
2013/04	1,511	*	32	33	170	1,272	100.0	0.3	2.1	2.2	11.3	84.2
2013/05	1,363	*	29	46	165	1,122	100.0	0.1	2.1	3.4	12.1	82.3
2013/06	1,369	*	70	49	136	1,111	100.0	0.2	5.1	3.6	9.9	81.2
2013/07	1,029	.	20	64	117	828	100.0	.	1.9	6.2	11.4	80.5
2013/08	1,170	*	36	52	128	952	100.0	0.2	3.1	4.4	10.9	81.4
2013/09	1,342	*	25	35	141	1,138	100.0	0.2	1.9	2.6	10.5	84.8
2013/10	1,318	*	35	58	161	1,059	100.0	0.4	2.7	4.4	12.2	80.3
2013/11	1,320	*	36	39	152	1,088	100.0	0.4	2.7	3.0	11.5	82.4
2013/12	1,230	*	28	48	139	1,012	100.0	0.2	2.3	3.9	11.3	82.3
2014/01	1,840	*	411	55	182	1,187	100.0	0.3	22.3	3.0	9.9	64.5
2014/02	1,277	*	38	47	139	1,049	100.0	0.3	3.0	3.7	10.9	82.1
2014/03	1,475	6	34	34	185	1,216	100.0	0.4	2.3	2.3	12.5	82.4
2014/04	1,162	*	28	39	157	934	100.0	0.3	2.4	3.4	13.5	80.4
2014/05	1,304	*	33	41	161	1,064	100.0	0.4	2.5	3.1	12.3	81.6
2014/06	1,306	8	32	31	146	1,089	100.0	0.6	2.5	2.4	11.2	83.4
2014/07	1,049	*	29	36	145	837	100.0	0.2	2.8	3.4	13.8	79.8
2014/08	1,079	*	31	39	151	854	100.0	0.4	2.9	3.6	14.0	79.1
2014/09	1,414	7	25	45	198	1,139	100.0	0.5	1.8	3.2	14.0	80.6
2014/10	1,410	10	34	59	209	1,098	100.0	0.7	2.4	4.2	14.8	77.9
2014/11	1,342	6	27	49	218	1,042	100.0	0.4	2.0	3.7	16.2	77.6
2014/12	1,354	*	11	64	183	1,094	100.0	0.1	0.8	4.7	13.5	80.8
2015/01	1,421	*	21	49	204	1,145	100.0	0.1	1.5	3.4	14.4	80.6
2015/02	1,423	7	*	52	206	1,155	100.0	0.5	0.2	3.7	14.5	81.2
2015/03	1,670	.	.	55	236	1,379	100.0	.	.	3.3	14.1	82.6
2015/04	1,432	.	.	44	208	1,180	100.0	.	.	3.1	14.5	82.4
2015/05	1,387	.	.	50	207	1,130	100.0	.	.	3.6	14.9	81.5
2015/06	1,629	.	.	54	218	1,357	100.0	.	.	3.3	13.4	83.3
2015/07	1,168	.	.	41	168	959	100.0	.	.	3.5	14.4	82.1
2015/08	1,310	.	.	44	202	1,064	100.0	.	.	3.4	15.4	81.2
2015/09	1,601	.	.	58	219	1,324	100.0	.	.	3.6	13.7	82.7
2015/10	1,589	.	.	40	231	1,318	100.0	.	.	2.5	14.5	82.9
2015/11	1,515	.	.	49	205	1,261	100.0	.	.	3.2	13.5	83.2
2015/12	1,617	.	.	50	185	1,382	100.0	.	.	3.1	11.4	85.5

The CONTENTS Procedure

Data Set Name	TTDATA.DMDREG	Observations	409841
Member Type	DATA	Variables	9
Engine	V9	Indexes	0
Created	24/01/2017 14:29:15	Observation Length	72
Last Modified	24/01/2017 14:29:15	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	WINDOWS_64		
Encoding	wlatin1 Western (Windows)		

Engine/Host Dependent Information

Data Set Page Size	65536
Number of Data Set Pages	452
First Data Page	*
Max Obs per Page	908
Obs in First Data Page	881
Number of Data Set Repairs	0
ExtendObsCounter	YES
Filename	E:\workdata\705093\BxC\demoDM\DATA\dmdreg.sas7bdat
Release Created	9.0401M3
Host Created	X64_SRV12

Alphabetic List of Variables and Attributes

#	Variable	Type	Len	Format	Informat	Label
8	DMtp	Char	*			
5	doBth	Num	8	DDMMYY10.		
7	doDM	Num	8	DDMMYY10.		
6	doDth	Num	8	DDMMYY10.		
2	doIns	Num	8	DDMMYY10.		
3	doOAD	Num	8	DDMMYY10.		
9	inCr	Char	*			
1	pnr	Char	12	$12.	$10.	Personnummer
4	sex	Num	8			

The reconstructed diabetes *drug* register 14:29 Tuesday, January 24, 2017 11

		DMtp		inCr		All	DMtp		inCr	
	All	T1	T2	Ins	OAD	PctN	PctN	PctN	PctN	PctN
All	409,841	16,029	393,812	66,056	343,785	100.0	3.9	96.1	16.1	83.9
doDM										
1995	79,746	4,629	75,117	31,393	48,353	100.0	5.8	94.2	39.4	60.6
1996	12,915	1,268	11,647	2,519	10,396	100.0	9.8	90.2	19.5	80.5
1997	11,321	530	10,791	1,597	9,724	100.0	4.7	95.3	14.1	85.9
1998	12,245	475	11,770	1,605	10,640	100.0	3.9	96.1	13.1	86.9
1999	12,434	498	11,936	1,587	10,847	100.0	4.0	96.0	12.8	87.2
2000	12,605	472	12,133	1,660	10,945	100.0	3.7	96.3	13.2	86.8
2001	13,250	505	12,745	1,610	11,640	100.0	3.8	96.2	12.2	87.8
2002	13,132	507	12,625	1,601	11,531	100.0	3.9	96.1	12.2	87.8
2003	15,361	480	14,881	1,567	13,794	100.0	3.1	96.9	10.2	89.8
2004	15,909	506	15,403	1,714	14,195	100.0	3.2	96.8	10.8	89.2
2005	15,757	497	15,260	1,685	14,072	100.0	3.2	96.8	10.7	89.3
2006	16,377	527	15,850	1,759	14,618	100.0	3.2	96.8	10.7	89.3
2007	17,715	508	17,207	1,782	15,933	100.0	2.9	97.1	10.1	89.9
2008	19,427	550	18,877	1,793	17,634	100.0	2.8	97.2	9.2	90.8
2009	20,194	583	19,611	1,777	18,417	100.0	2.9	97.1	8.8	91.2
2010	22,314	583	21,731	1,792	20,522	100.0	2.6	97.4	8.0	92.0
2011	25,900	561	25,339	1,657	24,243	100.0	2.2	97.8	6.4	93.6

```
 2012  22,673     580  22,093   1,620  21,053  100.0  2.6 97.4  7.1 92.9
 2013  17,089     574  16,515   1,779  15,310  100.0  3.4 96.6 10.4 89.6
 2014  15,865     585  15,280   1,781  14,084  100.0  3.7 96.3 11.2 88.8
 2015  17,612     611  17,001   1,778  15,834  100.0  3.5 96.5 10.1 89.9
```

7.11 06a-comp

Compares the prevalence of DM as derived from the revised (shrunk) version of the NDR and the newly constructed DMreg, at 1 Jan 1996, 2004 and 2012.

```
1                                                          "Program 06a-comp.sas"

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_SRV12  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.07 seconds
      cpu time            0.09 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          data prev ;
2            merge TTdata.ndr    ( in = nn )
3                  TTdata.dmreg  ( in = dd ) ;
4            by pnr ;
5            ndr = nn ;
6            dmreg = dd ;
7            mdif = (doNDR - doDM )/30 ;
8            do pdat = 1996, 2004, 2012 ;
9               a5 = floor( (mdy(1,1,pdat)-doBth )/(365.25*5) ) * 5 ;
10              if ( ndr   and doNDR < mdy(1,1,pdat) < min(doDth,mdy(1,2,pdat)) ) or
11                 ( dmreg and doDM  < mdy(1,1,pdat) < min(doDth,mdy(1,2,pdat)) ) then
11       ! output ;
12           end ;
13         run ;

NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      95279 at 7:17
NOTE: There were 389304 observations read from the data set TTDATA.NDR.
NOTE: There were 428251 observations read from the data set TTDATA.DMREG.
NOTE: The data set WORK.PREV has 499300 observations and 20 variables.
NOTE: DATA statement used (Total process time):
      real time           1.03 seconds
      cpu time            0.64 seconds


14
15         proc tabulate  data = prev  noseps  missing ;
```

```
16          class a5 ndr dmreg pdat sex inklaars inCr ;
17          table all a5,
18                pdat * ( all * f=comma9.
19                         ndr * dmreg * pctn<ndr*dmreg> * f=5.1 )
20                / rts = 5 ;
21          table pdat * ( all a5 ),
22              ( all * f=comma7.
23                ndr * dmreg * inCr="DMreg" * pctn<ndr*dmreg*inCr> * f=5.1 )
24                / rts = 11  indent = 1 ;
25          table pdat * ( all sex ) *
26              ( all inklaars="NDR" ),
27              ( all inCr ) * f=comma7.
28                / rts = 19 ;
29          keylabel n = ' ' ;
30        run ;

NOTE: There were 499300 observations read from the data set WORK.PREV.
NOTE: The PROCEDURE TABULATE printed pages 1-3.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.09 seconds
      cpu time            0.26 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           1.34 seconds
      cpu time            0.99 seconds
```

7.11.1 06a-comp.lst

```
The SAS System                                        14:32 Tuesday, January 24, 2017   1
```

	pdat											
	1996				2004				2012			
		ndr				ndr				ndr		
	0	*			0	*			0	*		
	dmreg	dmreg			dmreg	dmreg			dmreg	dmreg		
	*	0	*		*	0	*		*	0	*	
	All	PctN	PctN	PctN	All	PctN	PctN	PctN	All	PctN	PctN	PctN
All	95,310	2.0	3.3	94.7	154,703	2.6	3.8	93.7	249,287	3.7	4.7	91.6
a5												
-10	*	.	100.0	.	.	.	.	.	.	.	.	.
0	102	9.8	10.8	79.4	121	4.1	7.4	88.4	122	13.9	5.7	80.3
*	368	11.4	2.4	86.1	498	6.4	2.6	91.0	553	3.6	4.2	92.2
10	571	8.4	0.5	91.1	915	4.7	2.5	92.8	1,129	3.9	2.1	94.0
15	1,153	5.6	1.1	93.2	1,207	6.4	2.6	91.1	1,951	6.1	3.4	90.5
20	1,658	3.6	1.7	94.7	1,515	6.4	2.2	91.4	2,288	11.1	4.7	84.2
25	2,356	3.8	2.9	93.3	2,527	5.9	3.6	90.5	2,642	13.9	7.1	79.0
30	3,187	3.7	2.1	94.2	3,721	6.0	6.0	87.9	4,002	12.6	9.1	78.2
35	4,100	4.7	2.2	93.1	5,828	4.7	5.5	89.8	6,280	7.8	8.8	83.4
40	5,113	3.3	2.2	94.5	7,776	4.9	3.7	91.4	10,526	6.1	6.1	87.8
45	6,908	2.3	2.3	95.4	10,400	3.7	2.7	93.6	16,189	5.2	4.0	90.8
50	8,551	1.6	1.5	96.9	13,254	2.7	2.6	94.8	20,269	4.2	3.1	92.6
55	8,599	1.2	1.3	97.5	19,605	1.9	2.0	96.1	26,043	3.6	3.0	93.5
60	9,492	1.1	1.9	97.1	19,253	1.7	2.4	95.8	32,840	2.7	3.1	94.2
65	10,369	1.1	2.8	96.1	18,258	1.5	2.9	95.6	39,434	2.4	3.4	94.1
70	10,800	1.3	3.6	95.1	16,585	1.8	3.8	94.4	30,531	2.6	4.0	93.4
75	9,716	1.1	5.1	93.8	14,492	1.7	5.0	93.3	24,249	2.4	5.4	92.1
80	7,129	1.6	6.5	91.9	10,724	1.9	6.6	91.5	16,804	2.9	7.0	90.2
85	3,823	2.7	9.1	88.2	5,598	2.3	8.4	89.3	9,239	3.2	10.2	86.6

90	1,139	4.3	10.4	85.3	2,030	2.8	10.7	86.5	3,459	3.2	13.1	83.7
95	159	8.2	11.3	80.5	363	3.3	11.0	85.7	673	4.9	15.9	79.2
100	15	13.3	6.7	80.0	31	3.2	12.9	83.9	60	3.3	18.3	78.3
105	*	100.0	.	.	*	50.0	.	50.0	*	.	25.0	75.0

		ndr										
		0					*					
		dmreg					dmreg					
		*					0			*		
		DMreg					DMreg			DMreg		
		DIA	DVD	Ins	NPR	OAD	DIA	DVD	Ins	NPR	OAD	
	All	PctN	PctN	PctN	PctN	PctN	PctN	PctN	PctN	PctN	PctN	
1996												
All	95,310	.	0.0	0.2	1.5	0.4	3.3	.	27.4	3.4	32.9	31.1
-10	*	.	.	.	.	.	100.0	.	.	.	.	.
0	102	.	.	.	9.8	.	10.8	.	51.0	.	28.4	.
*	368	.	0.5	.	10.9	.	2.4	.	70.4	.	15.5	0.3
10	571	.	0.7	.	7.7	.	0.5	.	76.0	0.2	14.9	.
15	1,153	.	0.3	0.3	4.8	0.3	1.1	.	72.3	0.9	19.9	0.1
20	1,658	.	0.1	0.8	2.6	0.1	1.7	.	71.5	1.3	21.4	0.5
25	2,356	.	0.1	0.5	2.8	0.4	2.9	.	65.7	2.9	23.0	1.6
30	3,187	.	0.2	0.4	2.9	0.3	2.1	.	61.2	6.0	23.8	3.2
35	4,100	.	0.1	0.5	3.6	0.5	2.2	.	54.3	9.0	25.8	4.0
40	5,113	.	0.0	0.4	2.7	0.2	2.2	.	51.6	7.7	28.0	7.1
45	6,908	.	0.0	0.3	1.7	0.3	2.3	.	50.9	4.4	27.9	12.2
50	8,551	.	0.0	0.2	1.1	0.2	1.5	.	44.2	2.3	30.4	19.9
55	8,599	.	0.0	0.1	0.7	0.4	1.3	.	35.5	2.1	33.7	26.2
60	9,492	.	.	0.1	0.7	0.3	1.9	.	25.4	2.1	35.4	34.2
65	10,369	.	.	0.1	0.6	0.4	2.8	.	13.8	2.6	37.0	42.6
70	10,800	.	.	0.1	0.7	0.5	3.6	.	5.7	3.0	38.6	47.8
75	9,716	.	.	0.1	0.7	0.4	5.1	.	1.2	3.0	38.0	51.5
80	7,129	.	.	0.1	1.1	0.4	6.5	.	0.3	3.6	35.7	52.4
85	3,823	.	.	0.0	2.1	0.6	9.1	.	.	2.9	34.8	50.5
90	1,139	.	.	0.1	4.0	0.3	10.4	.	.	3.0	32.7	49.6
95	159	.	.	.	6.3	1.9	11.3	.	.	1.9	30.2	48.4
100	15	.	.	.	13.3	.	6.7	.	.	13.3	13.3	53.3
105	*	.	.	.	100.0	.	.	.	.	.	.	.
2004												
All	154,703	.	0.0	0.5	0.8	1.3	3.8	0.0	29.6	2.3	20.8	40.9
0	121	.	.	.	2.5	1.7	7.4	.	.	4.1	83.5	0.8
*	498	.	0.4	1.6	3.0	1.4	2.6	.	29.5	1.8	59.6	.
10	915	.	0.3	0.8	3.3	0.3	2.5	.	73.1	0.7	18.9	0.1
15	1,207	.	0.4	0.7	4.1	1.2	2.6	.	72.3	1.1	16.1	1.6
20	1,515	.	0.4	0.9	3.0	2.1	2.2	.	67.7	2.1	19.3	2.2
25	2,527	.	0.2	0.9	2.9	1.9	3.6	.	63.9	2.3	20.6	3.8
30	3,721	.	0.3	1.6	1.9	2.2	6.0	.	56.6	3.3	20.3	7.8
35	5,828	.	0.1	1.0	1.6	2.0	5.5	.	52.2	3.5	23.0	11.1
40	7,776	.	0.1	1.1	1.8	1.9	3.7	.	46.1	4.8	21.6	18.9
45	10,400	.	0.1	0.6	1.5	1.6	2.7	.	43.7	5.0	21.4	23.5
50	13,254	.	0.0	0.6	1.1	1.0	2.6	.	40.9	3.3	19.6	30.9
55	19,605	.	0.0	0.4	0.5	0.9	2.0	0.0	36.6	2.0	20.0	37.4
60	19,253	.	0.0	0.3	0.4	0.9	2.4	.	30.9	1.4	20.3	43.4
65	18,258	.	0.0	0.2	0.2	1.0	2.9	0.0	25.2	1.5	20.4	48.4
70	16,585	.	0.0	0.3	0.3	1.3	3.8	0.0	17.5	1.5	20.7	54.7
75	14,492	.	.	0.2	0.3	1.3	5.0	.	10.6	2.2	21.1	59.4
80	10,724	.	.	0.3	0.3	1.3	6.6	0.0	5.0	1.7	21.1	63.6
85	5,598	.	.	0.3	0.4	1.5	8.4	.	1.3	2.1	21.0	64.9
90	2,030	.	.	0.3	0.6	1.9	10.7	.	0.5	2.1	19.2	64.6
95	363	.	.	0.3	1.4	1.7	11.0	.	.	0.8	17.1	67.8

100	31	.	.	.	3.2	.	12.9	.	.	6.5	12.9	64.5
105	*	.	.	.	.	50.0	.	.	.	.	.	50.0
2012												
All	249,287	0.1	0.1	0.4	0.6	2.5	4.7	0.1	24.5	1.5	13.7	51.8
0	122	.	.	1.6	7.4	4.9	5.7	.	.	4.1	76.2	.
*	553	.	.	0.5	1.4	1.6	4.2	.	.	0.7	91.1	0.4
10	1,129	.	.	0.8	1.8	1.3	2.1	.	1.1	1.4	89.9	1.6
15	1,951	.	0.7	0.6	1.4	3.4	3.4	.	47.6	1.1	36.2	5.6
20	2,288	.	0.5	0.8	2.1	7.7	4.7	0.0	58.0	1.3	14.6	10.4
25	2,642	.	0.6	0.9	2.5	10.0	7.1	0.1	54.2	2.0	15.0	7.7
30	4,002	0.0	0.5	1.1	2.1	8.9	9.1	0.1	47.3	2.2	17.6	11.1
35	6,280	0.0	0.5	1.4	1.6	4.3	8.8	0.1	45.8	2.0	18.0	17.5
40	10,526	0.1	0.2	0.8	1.0	3.9	6.1	0.0	36.6	1.8	15.6	33.7
45	16,189	0.0	0.2	0.6	0.9	3.5	4.0	0.0	32.4	2.2	14.2	42.0
50	20,269	0.0	0.2	0.5	0.7	2.7	3.1	0.0	29.6	2.4	13.3	47.2
55	26,043	0.0	0.1	0.5	0.7	2.3	3.0	0.1	27.0	2.4	12.7	51.3
60	32,840	0.1	0.1	0.3	0.4	1.8	3.1	0.1	25.5	1.5	12.4	54.7
65	39,434	0.1	0.1	0.3	0.3	1.7	3.4	0.1	22.7	1.1	12.2	58.1
70	30,531	0.1	0.1	0.2	0.3	2.0	4.0	0.1	19.6	0.9	11.9	60.9
75	24,249	0.1	0.1	0.2	0.2	1.9	5.4	0.1	17.1	0.9	12.2	61.8
80	16,804	0.1	0.0	0.2	0.2	2.4	7.0	0.0	12.5	1.2	12.6	63.8
85	9,239	0.1	.	0.3	0.3	2.5	10.2	0.1	8.5	1.5	12.4	64.2
90	3,459	.	.	0.2	0.2	2.7	13.1	.	4.9	1.6	12.9	64.3
95	673	.	.	.	0.7	4.2	15.9	.	1.9	2.2	12.0	63.0
100	60	.	.	.	3.3	.	18.3	.	.	3.3	11.7	63.3
105	*	.	.	.	.	.	25.0	.	.	.	25.0	50.0

			Inclusion criterion						
			All		DIA	DVD	Ins	NPR	OAD
pdat	All	All	95,310	3,116	.	26,109	3,388	32,728	29,969
1996		NDR							
			1,930	.	.	30	160	1,397	343
		ins	6,757	268	.	2,301	2,172	1,970	46
		lpr	53,238	2,611	.	16,966	818	25,177	7,666
		oad	33,385	237	.	6,812	238	4,184	21,914
	sex	All	49,701	1,275	.	14,714	1,341	17,189	15,182
	*	NDR							
			710	.	.	18	74	459	159
		ins	3,406	84	.	1,446	727	1,124	25
		lpr	28,427	1,094	.	9,412	421	13,534	3,966
		oad	17,158	97	.	3,838	119	2,072	11,032
	*	All	45,609	1,841	.	11,395	2,047	15,539	14,787
		NDR							
			1,220	.	.	12	86	938	184
		ins	3,351	184	.	855	1,445	846	21
		lpr	24,811	1,517	.	7,554	397	11,643	3,700
		oad	16,227	140	.	2,974	119	2,112	10,882
2004	All	All	154,703	5,835	*	45,915	4,328	33,318	65,303
		NDR							
			3,947	.	.	74	712	1,203	1,958
		ins	6,487	258	.	2,517	2,387	1,196	129
		lpr	78,037	5,439	*	28,964	1,065	28,704	13,861
		oad	66,232	138	.	14,360	164	2,215	49,355
	sex	All	83,439	2,340	*	26,477	1,790	18,350	34,480
	*	NDR							
			1,542	.	.	48	336	352	806
		ins	3,279	77	.	1,564	850	725	63
		lpr	42,821	2,196	*	16,538	531	16,159	7,395
		oad	35,797	67	.	8,327	73	1,114	26,216
	*	All	71,264	3,495	*	19,438	2,538	14,968	30,823
		NDR							
			2,405	.	.	26	376	851	1,152
		ins	3,208	181	.	953	1,537	471	66
		lpr	35,216	3,243	*	12,426	534	12,545	6,466

```
             oad    30,435      71       .   6,033      91   1,101  23,139
2012  All    All   249,287  11,609     321  61,435   4,846  35,535 135,541
             NDR
                     9,246       .     150     324   1,016   1,405   6,351
             ins     6,470     259       .   2,579   2,566     737     329
             lpr   103,985  10,925     117  39,835   1,099  32,206  19,803
             oad   129,586     425      54  18,697     165   1,187 109,058
       sex   All   136,043   4,685     180  35,929   2,094  19,447  73,708
       *     NDR
                     3,565       .      87     162     462     430   2,424
             ins     3,269      76       .   1,579     975     458     181
             lpr    56,796   4,333      67  23,163     584  17,946  10,703
             oad    72,413     276      26  11,025      73     613  60,400
       *     All   113,244   6,924     141  25,506   2,752  16,088  61,833
             NDR
                     5,681       .      63     162     554     975   3,927
             ins     3,201     183       .   1,000   1,591     279     148
             lpr    47,189   6,592      50  16,672     515  14,260   9,100
             oad    57,173     149      28   7,672      92     574  48,658
```

7.12 Tabulation of follow-up and prevalence

For analysis of incidence rates of diabetes and mortality rates for persons with and without diabetes we tabulate the events and person-years for the entire Danish population by sex, state (T1, T2 and no DM), age and date of follow-up and date of birth, the latter three in 1-year intervals. Since we have individual records for the entire population (with sex, date of birth, migration and death) we could in principle choose smaller intervals of say half a year of 3 months.

7.13 `08-mkFU`

Defines follow-up intervals for all persons in the population, taking emigration and immigration records into account. Tese intervals are subdivided in pre- and post-DM diagnosis, the latter marked as T1 or T2. This follow-up is then split by age and subsequently by date using the `Lexis` macro.

Subsequently the number of deaths and DM (T1 or T2) events are tabulated by sex, age and date of follow-up and date of birth too.

Finally all is merged to a dataset, `FUtab`, classified by sex, age and calendar time and upper/lower Lexis triangle (correponding to the two possible periods of date of birth for a given combination of period of age and date of follow-up) [1].

```
1                                                      "Program 08-mkFU.sas"

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_SRV12  platform.


NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server
```

```
NOTE: SAS initialization used:
      real time           0.06 seconds
      cpu time            0.10 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.


NOTE: AUTOEXEC processing completed.

1          * Begin and end of study period, respectively ;
2          %let bos = '01jan1995'd ;
3          %let eos = '31dec2015'd ;
4
5          proc sort  data = TTdata.pop    out = pop   ; by pnr      ; run ;

NOTE: There were 7361669 observations read from the data set TTDATA.POP.
NOTE: The data set WORK.POP has 7361669 observations and 4 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           2.96 seconds
      cpu time            2.23 seconds


6          proc sort  data = TTdata.xdk    out = xdk   ; by pnr doEm ; run ;

NOTE: There were 1825733 observations read from the data set TTDATA.XDK.
NOTE: The data set WORK.XDK has 1825733 observations and 3 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.51 seconds
      cpu time            0.59 seconds


7          proc sort  data = TTdata.DMreg  out = dmreg ; by pnr      ; run ;

NOTE: There were 428251 observations read from the data set TTDATA.DMREG.
NOTE: The data set WORK.DMREG has 428251 observations and 12 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           0.64 seconds
      cpu time            0.25 seconds


8
9          data FU ;
10           merge pop ( in = p )
11                 xdk ( in = x ) ;
12           by pnr ;
13           format doBth doDth doEm doIm entry exit ddmmyy10. ;
14           if ^p  then delete ;
15           retain entry exit ;
16           if first.pnr then do ;
17              entry = max( &bos., doBth ) ;
18               exit = min( &eos., doDth ) ;
19              end ;
20           * if no migration records just output ;
21           if ^x then output ;
22           else do ;
23             if doEm > .z then do ;
24                 exit = min( exit, doEm ) ;
25                 output ;
26                 firstout = 1 ;
27                 end ;
28             if doIm > .z then do ;
29                entry = max( entry, doIm ) ;
30                 exit = min( &eos., doDth ) ;
31                firstout = 0 ;
32                end ;
33             if last.pnr and ^firstout then output ;
34           end ;
35         run ;
```

```
NOTE: There were 7361669 observations read from the data set WORK.POP.
NOTE: There were 1825733 observations read from the data set WORK.XDK.
NOTE: The data set WORK.FU has 7825474 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           2.46 seconds
      cpu time            2.10 seconds


36
37         data TTdata.FU ;
38           merge FU DMreg ( in = dm
39                            keep = pnr doDM DMtp ) ;
40           by pnr ;
41           state = "Well" ;
42           if ^dm then output ;
43           if  dm then do ;
44               if exit < doDM         then output ;
45               if        doDM < entry then do ;
46                  state = DMtp ;
47                  output ;
48                  end ;
49               if entry < doDM < exit then do ;
50                  oldex = exit ;
51                   exit = doDM ;
52                  output ;
53                  entry = doDM ;
54                   exit = oldex ;
55                  state = DMtp ;
56                  output ;
57                  end ;
58           end ;
59         run ;

NOTE: There were 7825474 observations read from the data set WORK.FU.
NOTE: There were 428251 observations read from the data set WORK.DMREG.
NOTE: The data set TTDATA.FU has 8191903 observations and 13 variables.
NOTE: DATA statement used (Total process time):
      real time           5.86 seconds
      cpu time            2.09 seconds


60
61         data FU ;
62           set TTdata.FU ;
63           fail = ( doDth eq exit ) ;
64         run ;

NOTE: There were 8191903 observations read from the data set TTDATA.FU.
NOTE: The data set WORK.FU has 8191903 observations and 14 variables.
NOTE: DATA statement used (Total process time):
      real time           5.67 seconds
      cpu time            1.28 seconds


65
66         options nosource2 ;
67         %inc '../../../sas/lexis.sas' ;
224
225        * tabulation interval for risk time and events ;
226        %let int = 1 ;
227
228        %lexis( data = FU,
229                 out = FUa,
230              breaks = 0 to 100 by &int.,
231              origin = doBth,
232               scale = 365.25,
233                left = A ) ;

NOTE: There were 8191903 observations read from the data set WORK.FU.
```

```
NOTE: The data set WORK.DISCRD has 91 observations and 17 variables.
NOTE: The data set WORK.FUA has 121849450 observations and 17 variables.
NOTE: DATA statement used (Total process time):
      real time           55.10 seconds
      cpu time            39.75 seconds


234
235         %lexis( data = FUa,
236                  out = FUap,
237               breaks = 190 to 220 by &int.,
238               origin = '01jan1800'd,
239                scale = 365.25,
240                 left = P ) ;

NOTE: There were 121849450 observations read from the data set WORK.FUA.
NOTE: The data set WORK.DISCRD has 0 observations and 18 variables.
NOTE: The data set WORK.FUAP has 235985275 observations and 18 variables.
NOTE: DATA statement used (Total process time):
      real time           3:27.73
      cpu time            2:47.21


241
242         data FU ;
243           set FUap ;
244           * set the counters of death and DM incidence ;
245            D = ( doDth eq exit ) ;
246           T1 = 0 ;
247           T2 = 0 ;
248           if ( doDM  eq exit ) then do ;
249              if ( DMtp eq 'T1' ) then T1 = 1 ;
250              if ( DMtp eq 'T2' ) then T2 = 1 ;
251              end ;
252           C = floor( (doBth-'01jan1800'd)/(365.25*&int.) ) * &int. + 1800 ;
253           P = P + 1800 ;
254           Y = risk/1000 ;
255           label A = "Age at FU"
256                 P = "Date of FU" ;
257         run ;

NOTE: There were 235985275 observations read from the data set WORK.FUAP.
NOTE: The data set WORK.FU has 235985275 observations and 23 variables.
NOTE: DATA statement used (Total process time):
      real time           1:45.43
      cpu time            41.15 seconds


258
259         * The datset FU now contains the follow-up classified by sex and state ;
260         * The deaths are those where doDth equals exit ;
261         * The dm diagnoses are those where doDM eq exit ;
262
263         proc summary  data = FU  nway ;
264           class sex state A P C ;
265           var  D T1 T2 Y ;
266           output out = FUtab ( drop = _FREQ_ _TYPE_ )
267                  sum = ;
268         run ;

NOTE: There were 235985275 observations read from the data set WORK.FU.
NOTE: The data set WORK.FUTAB has 25723 observations and 9 variables.
NOTE: PROCEDURE SUMMARY used (Total process time):
      real time           1:11.65
      cpu time            2:37.65


269
270         data TTdata.FUtab ;
271           set FUtab ;
```

```
272          U = P - A - C ;
273        run ;

NOTE: There were 25723 observations read from the data set WORK.FUTAB.
NOTE: The data set TTDATA.FUTAB has 25723 observations and 10 variables.
NOTE: DATA statement used (Total process time):
      real time           0.07 seconds
      cpu time            0.00 seconds


274
275        title1 'Tabulation of the population follow-up in DK' ;
276        proc contents  data = TTdata.FUtab ; run ;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.03 seconds
      cpu time            0.01 seconds

NOTE: The PROCEDURE CONTENTS printed page 1.

277        proc tabulate  data = TTdata.FUtab  missing  noseps ;
278          class sex state A P U ;
279          var D T1 T2 Y ;
280          table ( all sex U P ),
281                ( all state ) *
282                ( D * f=comma9.
283                 T1 * f=comma6.
284                 T2 * f=comma7.
285                  Y * f=comma10.2 )
286                / rts = 16  condense ;
287          keylabel sum = ' ' ;
288          format sex koen_t. ;
289        run ;

NOTE: There were 25723 observations read from the data set TTDATA.FUTAB.
NOTE: The PROCEDURE TABULATE printed page 2.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.09 seconds
      cpu time            0.01 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           7:38.45
      cpu time            6:54.51
```

7.13.1 08-mkFU.lst

```
Tabulation of the population follow-up in DK          14:39 Tuesday, January 24, 2017   1

The CONTENTS Procedure

Data Set Name        TTDATA.FUTAB                  Observations          25723
Member Type          DATA                          Variables             10
Engine               V9                            Indexes               0
Created              24/01/2017 14:47:26           Observation Length    80
Last Modified        24/01/2017 14:47:26           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    32
```

```
First Data Page             1
Max Obs per Page            817
Obs in First Data Page      793
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\futab.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12


Alphabetic List of Variables and Attributes

 #    Variable    Type    Len    Label

 *    A           Num       8    Age at FU
 *    C           Num       8
 6    D           Num       8
 *    P           Num       8    Date of FU
 7    T1          Num       8
 8    T2          Num       8
10    U           Num       8
 9    Y           Num       8
 *    sex         Num       8
 *    state       Char      4
```

Tabulation of the population follow-up in DK 14:39 Tuesday, January 24, 2017 2

	state							
	All				T1			
	D	T1	T2	Y	D	T1	T2	Y
All	1,166,535	24,842	342,446	114,424.04	15,668	0	0	568.89
sex								
Mand	575,622	14,247	188,969	56,682.35	9,061	0	0	321.98
Kvinde	590,913	10,595	153,477	57,741.69	6,607	0	0	246.91
U								
0	587,188	14,971	174,206	58,540.14	7,936	0	0	292.31
*	579,347	9,871	168,240	55,883.90	7,732	0	0	276.58
Date of FU								
1994	215	38	461	25.02	*	0	0	0.10
1995	62,935	3,064	34,162	5,234.42	779	0	0	22.94
1996	59,924	1,336	10,679	5,263.38	697	0	0	23.72
1997	59,276	1,295	10,510	5,284.40	749	0	0	24.32
1998	57,903	1,236	11,256	5,302.56	757	0	0	24.84
1999	58,766	1,193	12,086	5,318.76	782	0	0	25.23
2000	57,100	1,259	12,017	5,336.44	813	0	0	25.66
2001	57,737	1,165	12,338	5,356.07	782	0	0	26.07
2002	58,000	1,123	12,607	5,374.31	828	0	0	26.43
2003	57,138	1,108	14,616	5,388.37	851	0	0	26.65
2004	55,029	1,107	15,011	5,402.09	778	0	0	26.95
2005	54,329	1,133	14,927	5,417.62	810	0	0	27.21
2006	54,842	1,107	15,397	5,436.67	805	0	0	27.55
2007	55,133	1,085	16,694	5,461.82	784	0	0	27.83
2008	53,699	1,095	18,014	5,496.39	754	0	0	28.21
2009	54,190	1,053	18,803	5,530.71	787	0	0	28.52
2010	53,523	1,057	20,496	5,560.95	777	0	0	28.82
2011	51,904	977	24,150	5,590.64	712	0	0	29.09
2012	51,420	866	21,083	5,618.59	631	0	0	29.36
2013	51,546	859	15,717	5,648.79	632	0	0	29.61
2014	50,551	812	14,703	5,687.60	585	0	0	29.88
2015	51,375	874	16,719	5,688.46	572	0	0	29.90

(Continued)

	state							
	T2				Well			
	D	T1	T2	Y	D	T1	T2	Y
All	141,075	0	0	2,870.52	1,009,792	24,842	342,446	110,984.63
sex								
Mand	77,164	0	0	1,555.30	489,397	14,247	188,969	54,805.06
Kvinde	63,911	0	0	1,315.22	520,395	10,595	153,477	56,179.56
U								
0	71,554	0	0	1,485.06	507,698	14,971	174,206	56,762.77
*	69,521	0	0	1,385.46	502,094	9,871	168,240	54,221.86
Date of FU								
1994	8	0	0	0.17	204	38	461	24.75
1995	4,607	0	0	58.19	57,549	3,064	34,162	5,153.29
1996	4,922	0	0	67.10	54,305	1,336	10,679	5,172.57
1997	5,048	0	0	72.42	53,479	1,295	10,510	5,187.65
1998	5,132	0	0	78.32	52,014	1,236	11,256	5,199.40
1999	5,506	0	0	84.47	52,478	1,193	12,086	5,209.06
2000	5,564	0	0	91.06	50,723	1,259	12,017	5,219.72
2001	5,767	0	0	97.64	51,188	1,165	12,338	5,232.36
2002	6,029	0	0	104.23	51,143	1,123	12,607	5,243.64
2003	6,274	0	0	111.44	50,013	1,108	14,616	5,250.28
2004	6,217	0	0	120.02	48,034	1,107	15,011	5,255.12
2005	6,460	0	0	128.44	47,059	1,133	14,927	5,261.97
2006	6,743	0	0	137.01	47,294	1,107	15,397	5,272.11
2007	7,024	0	0	145.92	47,325	1,085	16,694	5,288.07
2008	7,010	0	0	156.43	45,935	1,095	18,014	5,311.75
2009	7,686	0	0	167.24	45,717	1,053	18,803	5,334.95
2010	7,895	0	0	179.10	44,851	1,057	20,496	5,353.02
2011	8,018	0	0	193.43	43,174	977	24,150	5,368.11
2012	8,347	0	0	208.02	42,442	866	21,083	5,381.21
2013	8,639	0	0	217.68	42,275	859	15,717	5,401.50
2014	8,948	0	0	223.92	41,018	812	14,703	5,433.80
2015	9,231	0	0	228.27	41,572	874	16,719	5,430.29

7.14 09-mkPr

Uses the follow-up data to count the persons that are alive at the 1 January 1995,...,2016, classified by sex, state (T1/T2/noDM) and age in 1-year classes. The resulting dataset, `prv` is classified by sex, state, age and date has only the extra variable `n` with the counts

```
1                                                    "Program 09-mkPr.sas"

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_SRV12  platform.



NOTE: Updated analytical products:

      SAS/STAT 14.1

NOTE: Additional host information:

 X64_SRV12 WIN 6.2.9200  Server

NOTE: SAS initialization used:
      real time           0.09 seconds
      cpu time            0.10 seconds


NOTE: AUTOEXEC processing beginning; file is E:\workdata\705093\BXC\demoDM\sas\optslibs.sas.
```

```
NOTE: AUTOEXEC processing completed.

1          data prv  ( keep = state sex age pdat ) ;
2            set TTdata.FU ;
3            * To make sure we get the last minute exiters in too ;
4            if exit ge '31DEC2015'd then exit = '02JAN2016'd ;
5            do pdat = 1995 to 2016 ;
6               if entry <= mdy(1,1,pdat) < exit then do ;
7                  age = floor( (mdy(1,1,pdat)-doBth)/365.25 ) ;
8                  output ;
9                  end ;
10              end ;
11         run ;

NOTE: Missing values were generated as a result of performing an operation on missing
      values.
      Each place is given by: (Number of times) at (Line):(Column).
      2002 at 7:15   2002 at 7:36   2002 at 7:43
NOTE: There were 8191903 observations read from the data set TTDATA.FU.
NOTE: The data set WORK.PRV has 120001859 observations and 4 variables.
NOTE: DATA statement used (Total process time):
      real time           28.21 seconds
      cpu time            21.85 seconds


12
13         proc summary  data = prv  nway ;
14           class pdat state sex age ;
15           output  out = TTdata.prv ( keep = pdat state sex age _freq_
16                                     rename = ( _freq_ = n ) ) ;
17         run ;

NOTE: There were 120001859 observations read from the data set WORK.PRV.
NOTE: The data set TTDATA.PRV has 13545 observations and 5 variables.
NOTE: PROCEDURE SUMMARY used (Total process time):
      real time           17.36 seconds
      cpu time            33.96 seconds


18
19         proc contents  data = TTdata.prv ; run ;

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

NOTE: The PROCEDURE CONTENTS printed page 1.

20
21         proc tabulate  data = TTdata.prv ( rename = ( n = x ) )  missing  noseps ;
22           class pdat state ;
23           var x ;
24           table pdat,
25               ( x = ' ' * sum = 'no. persons alive'   * f=comma9.
26                 x = ' ' * n   = 'no. records in data' * f=comma6. ) * ( all state )
27                 / rts = 6 ;
28         run ;

NOTE: There were 13545 observations read from the data set TTDATA.PRV.
NOTE: The PROCEDURE TABULATE printed page 2.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.03 seconds
      cpu time            0.03 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           45.82 seconds
```

```
      cpu time            56.04 seconds
```

7.14.1 09-mkPr.lst

```
The SAS System                                        14:51 Tuesday, January 24, 2017   1

The CONTENTS Procedure

Data Set Name        TTDATA.PRV                    Observations          13545
Member Type          DATA                          Variables             5
Engine               V9                            Indexes               0
Created              24/01/2017 14:52:19           Observation Length    40
Last Modified        24/01/2017 14:52:19           Deleted Observations  0
Protection                                         Compressed            NO
Data Set Type                                      Sorted                NO
Label
Data Representation  WINDOWS_64
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          65536
Number of Data Set Pages    9
First Data Page             1
Max Obs per Page            1632
Obs in First Data Page      1595
Number of Data Set Repairs  0
ExtendObsCounter            YES
Filename                    E:\workdata\705093\BxC\demoDM\DATA\prv.sas7bdat
Release Created             9.0401M3
Host Created                X64_SRV12


Alphabetic List of Variables and Attributes

#    Variable    Type    Len

4    age         Num       8
5    n           Num       8
1    pdat        Num       8
3    sex         Num       8
2    state       Char      4

The SAS System                                        14:51 Tuesday, January 24, 2017   2


                no. persons alive                no. records in data

                           state                              state

         All       T1        T2       Well     All     T1     T2    Well

 pdat
 1995 5,222,317    21,263    34,435 5,166,619    605    192    198    215
 1996 5,253,550    23,600    64,444 5,165,506    612    194    203    215
 1997 5,276,124    24,251    70,243 5,181,630    609    193    200    216
 1998 5,294,854    24,792    75,672 5,194,390    611    194    199    218
 1999 5,312,778    25,254    81,831 5,205,693    613    194    201    218
 2000 5,328,484    25,668    88,379 5,214,437    608    194    199    215
 2001 5,348,316    26,095    94,799 5,227,422    608    195    199    214
 2002 5,367,470    26,459   101,325 5,239,686    612    196    200    216
 2003 5,383,257    26,726   107,841 5,248,690    612    196    200    216
 2004 5,396,881    26,958   116,142 5,253,781    618    200    202    216
 2005 5,411,220    27,261   124,841 5,259,118    619    199    202    218
 2006 5,428,440    27,564   133,148 5,267,728    615    197    201    217
 2007 5,449,469    27,865   141,703 5,279,901    614    198    201    215
 2008 5,480,236    28,175   151,253 5,300,808    614    198    200    216
```

2009	5,518,953	28,533	162,218	5,328,202	613	197	199	217
2010	5,548,845	28,816	173,245	5,346,784	614	198	200	216
2011	5,579,763	29,079	185,682	5,365,002	617	197	202	218
2012	5,608,449	29,340	201,528	5,377,581	623	199	204	220
2013	5,637,470	29,561	214,139	5,393,770	624	197	205	222
2014	5,670,648	29,784	221,087	5,419,777	625	197	204	224
2015	5,713,553	30,005	226,703	5,456,845	628	197	205	226
2016	5,768,780	30,278	234,098	5,504,404	631	198	206	227