Components of Diabetes Prevalence in Denmark

SDC, Knut Borch-Johnsen
April 2013
Version 2
http://BendixCarstensen.com/DMreg/Prevalence

Compiled Friday 5th April, 2013, 12:31
from: C:/Bendix/Steno/DM-register/NDR/projects/Prevalence/DMprev.tex

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

Contents

Chapter 1

Introduction

1.1 Problem

This note addresses the problem of partitioning the changes in prevalence of diabetes in the Danish population over the last 15 years to:

1. changes in mortality rates among diabetes patients
2. changes in incidence rates of diabetes in the population

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

1.2 Formalization

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

1.2.1 Conceptual

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

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

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

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

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

3. only incidence rates had remained stable, but mortality rates had developed as observed

The difference between observed prevalences and the predicted under scenario

1. is the combined effect of changes in the rates as seen since the starting point chosen.
2. is the effect of changing mortality rates alone, which is what is of interest.

For the sake of completeness we shall compute both.

Chapter 2

Analysis

2.1 Statistical framework

First we consider the setup as outlined in figure 2.1:

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

The aim is to provide a precise formula for the age-specific prevalences at calendar time t, $p(a,t)$, given that we know the age-specific prevalence at some reference point t_0, $p(a,t_0)$ (in this case 1995, 2000, or 2005), and the transition rates $\lambda(a,p)$, $\mu_{\mathrm{W}}(a,p)$ and $\mu_{\mathrm{DM}}(a,p)$.

We can derive analytical expressions for this, but the easiest approach is to acquire parametric expressions for the transition rates and then update the age-specific prevalences by applying the transition probability matrix to a $A \times 2$ matrix of number of persons in each of the states Well and DM.

For the given transition rates we can compute transition probabilities between states corresponding to a given (small) interval, δ, say, but first deriving the cumulative

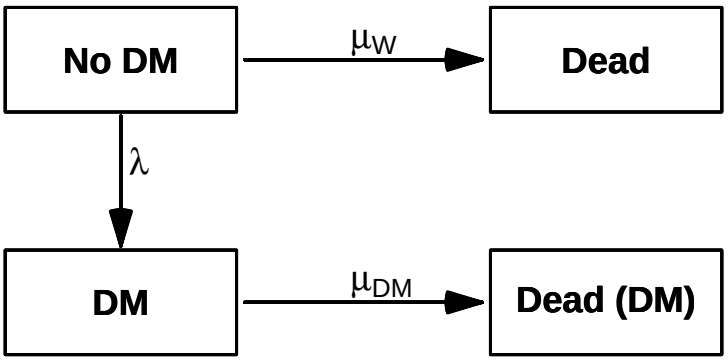

Figure 2.1: *The four states and transitions between them we consider*

intensities for intervals of this length

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

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

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

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

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

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

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

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

```
> library(Epi)
> for( yy in 2000+0:3 )
+ for( aa in 40+0:3 )
+ {
+ pdf( paste("./graph/DMpr-",yy,"-",aa,".pdf", sep="" ),
+      height=7, width=7 )
+ par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
+ Lexis.diagram( age=40+c(0,5), date=2000+c(0,5), int=1 )
+ w <- 0.6
+ d <- 0.3
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1,1+w,NA,2,2+w), col="forestgreen", lwd=9, lend="butt", ljoin="bevel" )
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1+w,1+w+d,NA,2+w,2+w+d), col="red", lwd=9, lend="butt", ljoin="bevel" )
+ lines( yy+c(1,1,NA,2,2),
+          aa-1+c(1+w+d,2,NA,2+w+d,3), col="black", lwd=9, lend="butt", ljoin="bevel" )
+ for( an in 1:17 )
+ arrows( yy+1.1, aa+0.6, yy+1.9, aa+1.4, lwd=3, angle=an )
+ dev.off()
+ }
```

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

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

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

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

$$q(a+\delta, p+\delta) = Q(a+\delta, p+\delta) \Big/ \sum_{\mathrm{W,DM}} Q(a+\delta, p+\delta)$$

so we update the prevalences at every step.

2.1.1 Births

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

Chapter 3

Practical implementation in R

3.1 Prerequisites

We are going to make models for the rates 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 labelled 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 timepoints `t.pt`

```
> int <- 0.1
> a.pt <- seq(0,100,int)[-1] - int/2
> t.pt <- seq(1995,2010,int)
> p.pt <- t.pt[-1] - int/2
```

We shall model all the rates by age-period-cohort models separately for men and women. As a sensitivity analysis we will also model the rates only by age-period models.

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 <- 10
> nk.p <- 5
> 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, 2 )
```

```
 [1] 0.05 0.15 0.25 0.35 0.45 0.55 0.65 0.75 0.85 0.95
```

```
> qn( 10, 5 )
```

```
 [1] 0.0200000 0.1266667 0.2333333 0.3400000 0.4466667 0.5533333 0.6600000
 [8] 0.7666667 0.8733333 0.9800000
```

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

Finally, to simplify the calling of the natural spline function we import a wrapper for the `ns` from the `splines` package:

```
> library( splines )
> source( "C:/stat/r/bxc/library.sources/useful/r/Ns.r" )
> Ns
```

```
function (x, df = NULL, knots = NULL, intercept = FALSE, Boundary.knots = NULL)
{
    if (is.null(Boundary.knots)) {
        if (!is.null(knots)) {
            knots <- sort(unique(knots))
            ok <- c(1, length(knots))
            Boundary.knots <- knots[ok]
            knots <- knots[-ok]
        }
    }
    ns(x, df = df, knots = knots, intercept = intercept, Boundary.knots = Boundary.knots)
}
```

3.2 Incidence rates

We model the incidence 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 (◸ and ◿):

```
> load( "C:/Bendix/Steno/DM-register/NDR/2009/data/inc.Rdata" )
> head( inc )
```

```
  sex         A        P        C D upper        Y      dm.Y   well.Y
1   K 0.6666667 1995.333 1994.667 2     1 17100.67 0.1300479 17100.54
2   K 0.3333333 1995.667 1995.333 0     0 17025.50 0.0000000 17025.50
3   K 0.6666667 1996.333 1995.667 4     1 17069.17 1.8617385 17067.30
4   K 0.3333333 1996.667 1996.333 4     0 16469.50 1.4455852 16468.05
5   K 0.6666667 1997.333 1996.667 2     1 16501.83 1.9835729 16499.85
6   K 0.3333333 1997.667 1997.333 0     0 16434.00 0.0000000 16434.00
```

```
> with( inc, addmargins( xtabs( D ~ floor(P) + sex ) ) )
```

```
        sex
floor(P)      K      M    Sum
    1995   7145   7748  14893
    1996   7392   8022  15414
    1997   7533   7927  15460
    1998   8044   8814  16858
    1999   8555   9307  17862
    2000   8883   9631  18514
    2001   9477  10211  19688
    2002  10912  11197  22109
    2003  11506  12477  23983
    2004  11615  12553  24168
    2005  10636  11663  22299
    2006  10930  12092  23022
    2007  11763  12703  24466
    2008  12641  13995  26636
    2009  12297  14242  26539
    Sum  149329 162582 311911
```

This dataset has the variable `D` for the number of DM-cases and the variable `well.Y` for the amount of follow-up among the non-diabetic part of the population, the latter is the correct denominator for analysis of the rates. The covariates of interest are of course `A`, `P` and `C` with values for age, period and cohort at the midpoints of the Lexis-triangles.

3.2.1 Models for incidence rates

Based on this we can now derive the location of the knots for this model:

```
> ( ki.a <- with( inc, quantile( rep(A,D), qn(nk.a) ) ) )
```

```
      5%      15%      25%      35%      45%      55%      65%      75%
31.66667 45.33333 52.33333 56.66667 60.66667 64.66667 68.66667 72.66667
     85%      95%
77.66667 84.66667
```

```
> ( ki.p <- with( inc, quantile( rep(P,D), qn(nk.p) ) ) )
```

```
     10%      30%      50%      70%      90%
1997.333 2000.667 2003.667 2006.333 2008.667
```

```
> ( ki.c <- with( inc, quantile( rep(C,D), qn(nk.c) ) ) )
```

```
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1918.667 1926.667 1933.333 1938.333 1943.667 1948.333 1955.667 1968.667
```

The model we set up is an model 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:

```
> m.inc.ap <- glm( D ~ Ns(A,knots=ki.a) + Ns(P,knots=ki.p),
+                  offset = log(well.Y), family=poisson,
+                  data = subset(inc,sex=="M") )
> f.inc.ap  <- update( m.inc.ap, data = subset(inc,sex=="K") )
> m.inc.apc <- update( m.inc.ap, . ~ . + detrend( Ns(P-A,kn=ki.c), P-A ) )
> f.inc.apc <- update( m.inc.apc, data = subset(inc,sex=="K") )
> m.inc.d   <- update( m.inc.ap, . ~ . -  Ns(P,knots=ki.p) + P )
> f.inc.d   <- update( f.inc.ap, . ~ . -  Ns(P,knots=ki.p) + P )
```

The avearge annual trends in incidence:

```
> inc.chg <- rbind( ci.exp(m.inc.d,subset="P")-1,
+                   ci.exp(f.inc.d,subset="P")-1 )*100
> rownames( inc.chg ) <- c("Incidence change:      Men",
+                          "                      Women")
> round( inc.chg, 1 )
```

```
                         exp(Est.) 2.5% 97.5%
Incidence change:      Men       3.8  3.7   3.9
                     Women       3.9  3.8   4.1
```

3.2.2 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)),
+               well.Y = int, dm.Y = int, Y = int )
> head( nd )
```

```
     A       P well.Y dm.Y   Y
1 0.05 1995.05    0.1  0.1 0.1
2 0.15 1995.05    0.1  0.1 0.1
3 0.25 1995.05    0.1  0.1 0.1
4 0.35 1995.05    0.1  0.1 0.1
5 0.45 1995.05    0.1  0.1 0.1
6 0.55 1995.05    0.1  0.1 0.1
```

Then we set up an array to hold the predicted incidence rates from the different models, separtely for the two sexes

```
> rnam <- list( a = a.pt,
+               p = p.pt,
+             sex = c("M","F"),
+              tp = c("apc","ap") )
> Lambda <- array( NA, dimnames=rnam, dim=sapply(rnam,length) )
> str( Lambda )
```

```
 logi [1:1000, 1:150, 1:2, 1:2] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ a  : chr [1:1000] "0.05" "0.15" "0.25" "0.35" ...
  ..$ p  : chr [1:150] "1995.05" "1995.15" "1995.25" "1995.35" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ tp : chr [1:2] "apc" "ap"
```

```
> 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 rates that we need for two sexes, and for two different modeling approaches.

3.3 Mortality rates

We have the mortality-rates in 1×1 age×period A-sets (□)

```
> load( "C:/Bendix/Steno/DM-register/NDR/2009/data/pop.mort.Rdata" )
> str( pop.mort )
```

```
'data.frame':        3000 obs. of  9 variables:
 $ sex  : Factor w/ 2 levels "K","M": 1 1 1 1 1 1 1 1 1 ...
 $ A    : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P    : num  1995 1996 1997 1998 1999 ...
 $ dm.D : num  0 0 0 0 0 0 0 0 0 0 ...
 $ dm.Y : num  0.13 3.307 1.984 0 0.877 ...
 $ pop.D: num  153 157 163 148 112 149 156 125 121 130 ...
 $ pop.Y: num  34126 33539 32936 32534 32303 ...
 $ D    : num  153 157 163 148 112 149 156 125 121 130 ...
 $ Y    : num  34126 33535 32934 32534 32302 ...
```

```
> head( pop.mort )
```

```
  sex A    P dm.D      dm.Y pop.D    pop.Y   D        Y
1   K 0 1995    0 0.1300479   153 34126.17 153 34126.04
2   K 0 1996    0 3.3073238   157 33538.67 157 33535.36
3   K 0 1997    0 1.9835729   163 32935.83 163 32933.85
4   K 0 1998    0 0.0000000   148 32533.67 148 32533.67
5   K 0 1999    0 0.8767967   112 32303.33 112 32302.46
6   K 0 2000    0 0.7214237   149 32575.17 149 32574.45
```

thus we have the number of deaths and person-time among diabetes patients in `dm.D` and `dm.Y` respectively, and the same quantities for the non-diabetic part of the population in `D` and `Y`.

3.3.1 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( pop.mort, quantile( rep(A  ,dm.D), qn(nk.a) ) ) )
```

```
 5% 15% 25% 35% 45% 55% 65% 75% 85% 95%
 54  64  69  73  77  79  82  85  88  92
```

```
> ( kmd.p <- with( pop.mort, quantile( rep(P  ,dm.D), qn(nk.p) ) ) )
```

```
 10%  30%  50%  70%  90%
1996 2000 2003 2006 2008
```

```
> ( kmd.c <- with( pop.mort, quantile( rep(P-A,dm.D), qn(nk.c) ) ) )
```

```
 6.25% 18.75% 31.25% 43.75% 56.25% 68.75% 81.25% 93.75%
  1910   1916   1920   1923   1927   1931   1937   1947
```

```
> m.md.ap <- glm( dm.D ~ Ns(A,knots=kmd.a) + Ns(P,knots=kmd.p),
+                         offset = log(dm.Y), family = poisson,
+                         data = subset( pop.mort, sex=="M" & dm.Y>0 ) )
> m.md.apc <- update( m.md.ap, . ~ . + detrend( Ns(P-A,kn=kmd.c), P-A ) )
> f.md.ap  <- update( m.md.ap , data = subset(pop.mort,sex=="K" & dm.Y>0 ) )
> f.md.apc <- update( m.md.apc, data = subset(pop.mort,sex=="K" & dm.Y>0 ) )
> m.md.d   <- update( m.md.ap, . ~ . -  Ns(P,knots=kmd.p) + P )
> f.md.d   <- update( f.md.ap, . ~ . -  Ns(P,knots=kmd.p) + P )
> Mu.DM <- array( NA, dimnames=rnam, dim=sapply(rnam,length) )
> 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 )
```

3.3.2 The non-diabetic population

The mortality in the non-diabetic population are modeled in exactly the same way

```
> ( kmw.a <- with( pop.mort, quantile( rep(A  ,D), qn(nk.a) ) ) )

 5% 15% 25% 35% 45% 55% 65% 75% 85% 95%
 45  60  67  72  77  80  83  86  89  94

> ( kmw.p <- with( pop.mort, quantile( rep(P  ,D), qn(nk.p) ) ) )

 10%  30%  50%  70%  90%
1996 1999 2002 2005 2008

> ( kmw.c <- with( pop.mort, quantile( rep(P-A,D), qn(nk.c) ) ) )

 6.25% 18.75% 31.25% 43.75% 56.25% 68.75% 81.25% 93.75%
  1908   1913   1918   1922   1926   1931   1940   1954

> m.mw.ap <- glm( D ~ Ns(A,knots=kmw.a) + Ns(P,knots=kmw.p),
+                      offset = log(Y), family = poisson,
+                      data = subset(pop.mort,sex=="M" & Y>0 ) )
> m.mw.apc <- update( m.mw.ap, . ~ . + detrend( Ns(P-A,kn=kmw.c), P-A ) )
> f.mw.ap  <- update( m.mw.ap , data = subset(pop.mort,sex=="K" & Y>0 ) )
> f.mw.apc <- update( m.mw.apc, data = subset(pop.mort,sex=="K" & Y>0 ) )
> m.mw.d   <- update( m.mw.ap, . ~ . -  Ns(P,knots=kmw.p) + P )
> f.mw.d   <- update( f.mw.ap, . ~ . -  Ns(P,knots=kmw.p) + P )
> Mu.W <- array( NA, dimnames=rnam, dim=sapply(rnam,length) )
> 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 )
```

The average annual trends in mortalities:

```
> mort.chg <- rbind( ci.exp(m.md.d,subset="P")-1,
+                    ci.exp(f.md.d,subset="P")-1,
+                    ci.exp(m.mw.d,subset="P")-1,
+                    ci.exp(f.mw.d,subset="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%
Incidence change:      Men        3.8  3.7   3.9
                     Women        3.9  3.8   4.1
Mortality change,   DM: Men      -3.8 -3.9  -3.6
                       Women     -3.4 -3.6  -3.2
Mortality change, Well: Men      -2.5 -2.5  -2.4
                       Women     -1.8 -1.8  -1.7
```

For convenience of calculations 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="estimates.etc.Rdata" )
```

Chapter 4

Prevalence predictions

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("estimates.etc.Rdata")
```

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 prevalences at 1.1.1995–1.1.2010, smoothed by natural splines.

`TR` — array of transition probabilities between states Well and DM. Transition probabilities are computed under the 4 different scenarios combining mortality and incidence rates either as they actually developed 1995–2009 or assuming they were constant at the 1995 level. These refer to intervals of length `int` and are therefore labelled 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 1.1.1995 and the transition probabilities as put in `TR`. The scenario dimension refers to the 4 scenarios: "obs", "m-fix", "i-fix" and "all-f", but this dimension in the array is expanded by 3 extra levels "mort", "inc" and "const" that are to be filled with the part of the prevalences that are attributable to decrease in mortality, increase in incidence and the disequilibrium between rates and prevalence in 1995. Likewise the period dimension is expanded by one relative to that in `TR`, since this refer to points in time and not time intervals.

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

4.1 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:

```
> 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 [1:1000, 1:150, 1:2, 1:2, 1:2, 1:2, 1:4] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 7
  ..$ a    : chr [1:1000] "0.05" "0.15" "0.25" "0.35" ...
  ..$ p    : chr [1:150] "1995.05" "1995.15" "1995.25" "1995.35" ...
  ..$ sex  : chr [1:2] "M" "F"
  ..$ tp   : chr [1:2] "apc" "ap"
  ..$ 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 1995 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 0.1 year):

```
> TR[,,,,"Well","Well","obs"] <- exp(-Lambda-Mu.W)
> TR[,,,,"Well","DM"  ,"obs"] <- exp(-Lambda-Mu.W)*Lambda
> TR[,,,,"DM"  ,"Well","obs"] <- 0
> TR[,,,,"DM"  ,"DM"  ,"obs"] <- exp(-Mu.DM)
```

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

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

```
> np <- dim(Lambda)[2]
> TR[,,,,"Well","Well","m-fix"] <- exp(-Lambda-Mu.W[,rep(1,np),,])
> TR[,,,,"Well","DM"  ,"m-fix"] <- exp(-Lambda-Mu.W[,rep(1,np),,])*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"] <- exp(-Lambda[,rep(1,np),,]-Mu.W)*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"] <- exp(-Lambda[,rep(1,np),,]-Mu.W[,rep(1,np),,])*Lambda[,rep(1,np),,
> TR[,,,,"DM"  ,"Well","all-f"] <- 0
> TR[,,,,"DM"  ,"DM"  ,"all-f"] <- exp(-Mu.DM[,rep(1,np),,])
```

qWe 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`:

```
> str( TR )
```

```
 num [1:1000, 1:150, 1:2, 1:2, 1:2, 1:2, 1:4] 1 1 1 1 1 ...
 - attr(*, "dimnames")=List of 7
  ..$ a    : chr [1:1000] "0.05" "0.15" "0.25" "0.35" ...
  ..$ p    : chr [1:150] "1995.05" "1995.15" "1995.25" "1995.35" ...
  ..$ sex  : chr [1:2] "M" "F"
  ..$ tp   : chr [1:2] "apc" "ap"
  ..$ 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] 9600000
```

4.2 Population prevalences

Finally, in order to get the machinery working, we need the observed prevalences and population size at the starting point, that is at 1.1.1995. These are available from the same tabulation of the diabetes register as before:

```
> load( "C:/Bendix/Steno/DM-register/NDR/2009/data/pr.Rdata" )
> levels( pr$sex ) <- c("F","M")
> summary( pr )
```

```
 sex            A               P               X                N
 F:1600   Min.   : 0.50   Min.   :1995   Min.   :   0.0   Min.   :   77
 M:1600   1st Qu.:25.25   1st Qu.:1999   1st Qu.: 134.8   1st Qu.:18275
          Median :50.00   Median :2002   Median : 571.5   Median :31829
          Mean   :50.00   Mean   :2002   Mean   : 882.4   Mean   :26871
          3rd Qu.:74.75   3rd Qu.:2006   3rd Qu.:1380.0   3rd Qu.:36800
          Max.   :99.50   Max.   :2010   Max.   :5127.0   Max.   :46208
```

```
> head( pr )
```

```
   sex   A    P  X     N
1    F 0.5 1995  0 34094
17   F 1.5 1995  4 32967
33   F 2.5 1995  6 33198
49   F 3.5 1995  5 31738
65   F 4.5 1995 13 31361
81   F 5.5 1995 13 30440
```

These are emprical prevalences (`X`—no. of cases of DM, `N`—population size) for each of the 16 dates 1.1.1995 – 1.1.2010 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 1995—2010, as a smooth function of age, and use the predicted prevalences to produce the prevalence of diabete 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 20 knots:

```
> ( kp.a <- c( 10, with( pr, quantile( rep(A,X), qn(15) ) ) ) )
```

```
          3.333333%       10% 16.66667% 23.33333%       30% 36.66667% 43.33333%
     10.0      28.5      40.5      47.5      52.5      55.5      58.5      61.5
      50% 56.66667% 63.33333%       70% 76.66667% 83.33333%       90% 96.66667%
     64.5      67.5      69.5      72.5      75.5      78.5      82.5      87.5
```

```
> pr.fit <- NArray( c( dimnames(Lambda)[c(1,3)],
+                      list( t = sort(unique(pr$P)) ) ) )
> str( pr.fit )
```

```
 logi [1:1000, 1:2, 1:16] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ a  : chr [1:1000] "0.05" "0.15" "0.25" "0.35" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ t  : chr [1:16] "1995" "1996" "1997" "1998" ...
```

```
> prod( dim(pr.fit) )
```

```
[1] 32000
```

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)[["t"]] )
+ pr.fit[,sx,dt] <- predict( glm( cbind(X,N-X) ~ Ns( A, kn=kp.a ),
+                                 family = binomial(link="log"),
+                                   data = subset( pr, sex==sx & P==as.numeric(dt) ) ),
+                            newdata = data.frame( A=as.numeric(dimnames(pr.fit)[["a"]]) ),
+                               type = "response" )
> round( ftable( pr.fit[c(1:2,NA,700+1:5),,], row.vars=2:1 )*100, 1 )
```

```
          t 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010
sex a
M   0.05     0.0  0.0  0.0  0.0  0.0  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
    0.15     0.0  0.0  0.0  0.0  0.0  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1  0.1
    NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
    70.05    6.0  6.5  7.1  7.6  8.2  8.9  9.6 10.4 11.3 12.3 13.3 14.0 14.5 14.9 15.7 16.4
    70.15    6.1  6.5  7.2  7.6  8.2  8.9  9.7 10.4 11.3 12.3 13.3 14.1 14.5 15.0 15.7 16.5
    70.25    6.1  6.6  7.2  7.6  8.2  8.9  9.7 10.4 11.3 12.3 13.4 14.1 14.6 15.1 15.8 16.5
    70.35    6.1  6.6  7.2  7.6  8.2  8.9  9.7 10.4 11.3 12.3 13.4 14.1 14.6 15.1 15.8 16.5
    70.45    6.1  6.6  7.2  7.7  8.3  8.9  9.7 10.4 11.3 12.3 13.4 14.2 14.7 15.2 15.9 16.6
F   0.05     0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.1  0.1  0.1  0.1  0.1  0.1  0.1
    0.15     0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.1  0.1  0.1  0.1  0.1  0.1  0.1
    NA        NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
    70.05    5.0  5.3  5.7  6.2  6.7  7.1  7.5  8.0  8.9  9.8 10.5 10.9 11.2 11.6 12.1 12.6
    70.15    5.0  5.3  5.7  6.2  6.7  7.2  7.6  8.1  8.9  9.8 10.6 11.0 11.3 11.7 12.1 12.7
    70.25    5.0  5.3  5.8  6.2  6.7  7.2  7.6  8.1  8.9  9.8 10.6 11.0 11.4 11.8 12.2 12.7
    70.35    5.0  5.4  5.8  6.2  6.7  7.2  7.6  8.1  9.0  9.8 10.6 11.1 11.4 11.8 12.2 12.8
    70.45    5.1  5.4  5.8  6.2  6.7  7.2  7.7  8.2  9.0  9.9 10.7 11.1 11.5 11.9 12.3 12.8
```

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

```
> par( mfrow=c(1,2), mar=c(1,0,1,0), mgp=c(3,1,0)/1.6, las=1,
+      oma=c(2,3,0,1), bty="n" )
> matplot( a.pt, pr.fit[,"M",]*100,
+          ylim=c(0,20), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", lty=1, col="blue", lwd=c(1,2) )
> text( 35, 19, "Men", adj=0, col="blue", cex=1.2 )
> text( 89, pr.fit["89.05","M","1995"]* 99, "1995", col="blue", adj=c(1,1) )
> text( 80, pr.fit["79.95","M","2010"]*101, "2010", col="blue", adj=c(0,0) )
> axis( side=1 )
> axis( side=2 )
> 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) )
> text( 35, 19, "Women", adj=0, col="red", cex=1.2 )
> text( 89, pr.fit["89.05","F","1995"]* 99, "1995", col="red", adj=c(1,1) )
> text( 80, pr.fit["79.95","F","2010"]*101, "2010", 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 )
```

For the calculations we shall only use the estimated prevalences as of 1.1.1995 as starting point for the simulation.

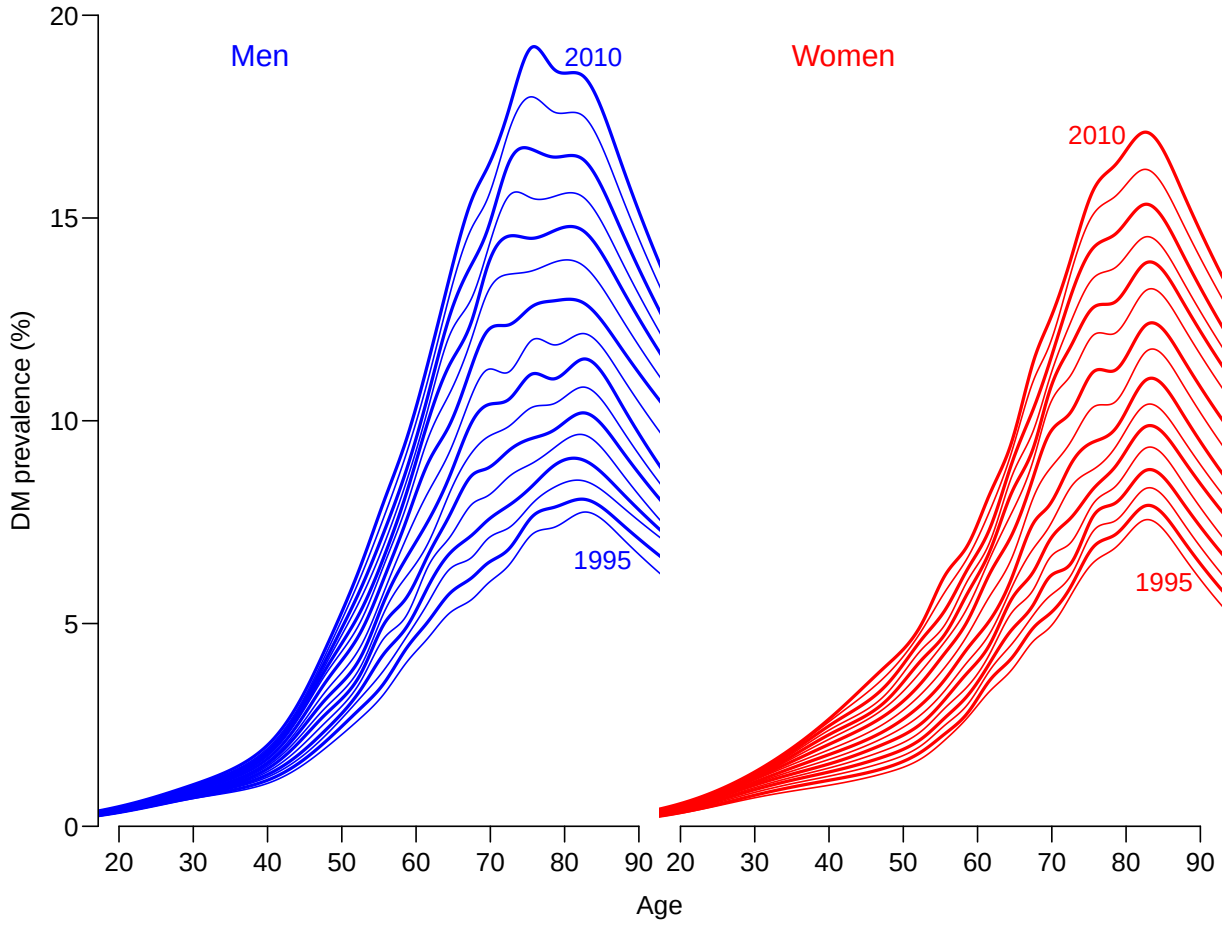

Figure 4.1: *Smoothed age-specific prevalences for the years 1995–2010. Blue is men, red is women.*

4.3 Prediction of the observed prevalences

Note that we do not need to predict the population size; we only predict the prevalences as fractions. When we multiply the fraction of persons in states (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 dead ones), 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 perido too (or, to put it differently, we need an extra first level to hold the starting prevalences as of 1.1.1995).

```
> dpr <- c( dimnames(Lambda)[1:4],
+           list( c(dimnames(TR)[["scene"]],"mort","inc","const") ) )
> names( dpr )[c(2,5)] <- c("t","what")
> dpr[["t"]] <- t.pt
> prv <- NArray(dpr)
> str( prv )
```

```
 logi [1:1000, 1:151, 1:2, 1:2, 1:7] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1000] "0.05" "0.15" "0.25" "0.35" ...
  ..$ t   : chr [1:151] "1995" "1995.1" "1995.2" "1995.3" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ tp  : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

```
> prod( dim(prv) )
```

```
[1] 4228000
```

Thus we must make a loop that updates the prevalences at 1.1.1995 to those at subsequent times, but first we must initialize the prevalences as modeled on 1.1.1995, as well as the prevalences at ages 0 (the new-borns) that we set to 0:

```
> # Smoothed prevalences at 1.1.1995 - the starting values
> prv[,1,,,] <- pr.fit[,,1]
> # Prevalences at age 0
> prv[1, ,"M",,] <- 0
> prv[1, ,"F",,] <- 0
> round( ftable( prv[c(1:2,700+1:3),1:2,,,], col.vars=4:5 )*100, 1 )
```

```
                tp   apc                                    ap
                what obs m-fix i-fix all-f mort inc const obs m-fix i-fix all-f mort inc const
a     t      sex
0.05  1995   M        0.0   0.0   0.0   0.0  0.0 0.0   0.0 0.0   0.0   0.0   0.0   0.0 0.0   0.0
             F        0.0   0.0   0.0   0.0  0.0 0.0   0.0 0.0   0.0   0.0   0.0   0.0 0.0   0.0
      1995.1 M        0.0   0.0   0.0   0.0  0.0 0.0   0.0 0.0   0.0   0.0   0.0   0.0 0.0   0.0
             F        0.0   0.0   0.0   0.0  0.0 0.0   0.0 0.0   0.0   0.0   0.0   0.0 0.0   0.0
0.15  1995   M        0.0   0.0   0.0   0.0  0.0 0.0   0.0 0.0   0.0   0.0   0.0   0.0 0.0   0.0
             F        0.0   0.0   0.0   0.0  0.0 0.0   0.0 0.0   0.0   0.0   0.0   0.0 0.0   0.0
      1995.1 M         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
70.05 1995   M        6.0   6.0   6.0   6.0  6.0 6.0   6.0 6.0   6.0   6.0   6.0   6.0 6.0   6.0
```

```
             F        5.0   5.0   5.0   5.0  5.0 5.0   5.0 5.0   5.0   5.0   5.0  5.0 5.0   5.0
      1995.1 M         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
70.15 1995   M        6.1   6.1   6.1   6.1  6.1 6.1   6.1 6.1   6.1   6.1   6.1  6.1 6.1   6.1
             F        5.0   5.0   5.0   5.0  5.0 5.0   5.0 5.0   5.0   5.0   5.0  5.0 5.0   5.0
      1995.1 M         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
70.25 1995   M        6.1   6.1   6.1   6.1  6.1 6.1   6.1 6.1   6.1   6.1   6.1  6.1 6.1   6.1
             F        5.0   5.0   5.0   5.0  5.0 5.0   5.0 5.0   5.0   5.0   5.0  5.0 5.0   5.0
      1995.1 M         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
```

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)["t"]-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
  10.94    0.00   10.95


> # Predicted prevalences (%):
> round( prv[1:10,1:5,1,1,1]*100, 3 )


      t
a       1995 1995.1 1995.2 1995.3 1995.4
  0.05 0.000  0.000  0.000  0.000  0.000
  0.15 0.039  0.001  0.001  0.001  0.001
  0.25 0.040  0.040  0.002  0.002  0.002
  0.35 0.040  0.041  0.041  0.002  0.002
  0.45 0.041  0.041  0.042  0.042  0.003
  0.55 0.041  0.042  0.042  0.042  0.043
  0.65 0.042  0.042  0.042  0.043  0.043
  0.75 0.042  0.043  0.043  0.043  0.044
  0.85 0.043  0.043  0.043  0.044  0.044
  0.95 0.043  0.043  0.044  0.044  0.045


> save( a.pt, prv, file="prv.Rdata" )
> load(            file="prv.Rdata" )
```

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

4.3.1 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 1.1.2010.

In the array `prv` are all the prevalences as predicted from the prevalence in 1995 using the estimated incidences and mortalities; predicted at intervals of `inc` whereas we have the smoothed empirical prevalences in 1995, 2000, 2005 and 2010 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:

```
> ( wh <- paste(1995+1:3*5) )
```

```
[1] "2000" "2005" "2010"
```

```
> 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"],
+                      pr.fit[,"M",wh])*100,
+          xlim=c(10,95), ylim=c(0,20), yaxs="i",
+          xlab="Age", ylab="Prevalence (%)",
+          type="l", col="blue", lty=rep(c(0,1),c(3,4)), lwd=2 )
> matlines( a.pt, prv[,wh,"M","apc","obs"]*100,
+          type="l", col="blue", lty="42", lwd=3 )
> matplot( a.pt, cbind(prv[,wh,"F","apc","obs"],
+                      pr.fit[,"F",wh])*100,
+          xlim=c(10,95), ylim=c(0,20), yaxs="i",
+          xlab="Age", ylab="", yaxt="n",
+          type="l", col="red", lty=rep(c(0,1),c(3,4)), lwd=2 )
> matlines( a.pt, prv[,wh,"F","apc","obs"]*100,
+           type="l", col="red", lty="42", lwd=3 )
> mtext( "Prevalence of DM (%)", side=2, line=2, las=0, outer=TRUE )
> mtext( "Age", side=1, line=2, las=0, outer=TRUE )
```

```
> ( wh <- paste(1995+1:3*5) )
```

```
[1] "2000" "2005" "2010"
```

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1 )
> matplot( a.pt, cbind(prv[,wh,"M","ap" ,"obs"],
+                      prv[,wh,"M","apc","obs"],
+                      pr.fit[,"M",wh])*100,
+          xlim=c(10,90), ylim=c(0,20), xlab="Age", ylab="Prevalence (%)",
+          type="l", col="blue", lty=rep(c(0,1),c(6,4)), 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(3,3)), lwd=3 )
> matplot( a.pt, cbind(prv[,wh,"F","ap" ,"obs"],
+                      prv[,wh,"F","apc","obs"],
+                      pr.fit[,"F",wh])*100,
+          xlim=c(10,90), ylim=c(0,20), xlab="Age", ylab="Prevalence (%)",
+          type="l", col="red", lty=rep(c(0,1),c(6,4)), 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(3,3)), lwd=3 )
```

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

We now compare the predicted prevalences under the four scenarios at 1.1.2010:

```
> str( prv )
```

```
 num [1:1000, 1:151, 1:2, 1:2, 1:7] 0 0.000395 0.000399 0.000404 0.000408 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1000] "0.05" "0.15" "0.25" "0.35" ...
  ..$ t   : chr [1:151] "1995" "1995.1" "1995.2" "1995.3" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ tp  : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

```
> prv[900+1:10,np,"M","apc",]*100
```

```
       what
a            obs    m-fix    i-fix    all-f mort inc const
  90.05 15.54911 12.75953 12.83256 10.33783   NA  NA    NA
  90.15 15.49596 12.71754 12.78888 10.30426   NA  NA    NA
```

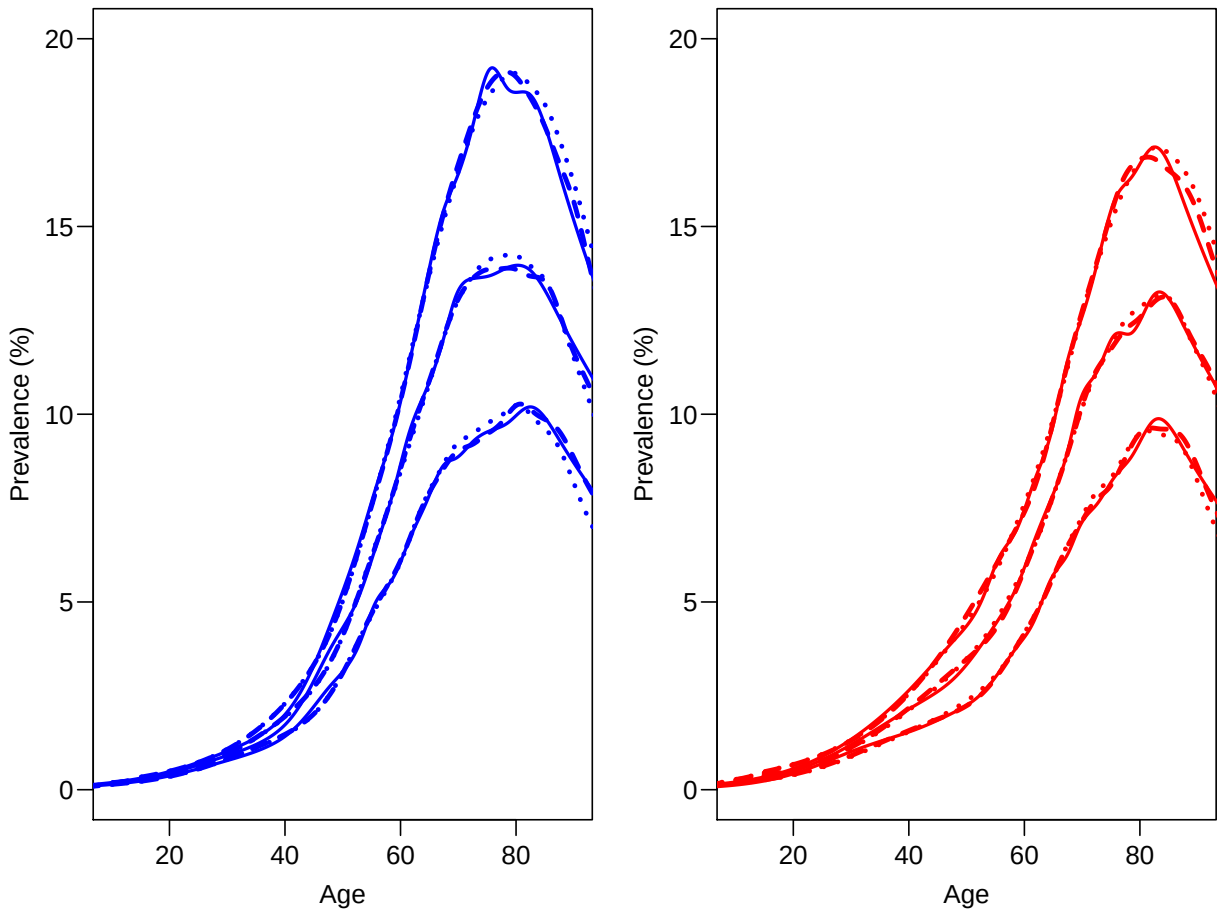

Figure 4.2: *Plot of observed (full lines) and predicted prevalences in 2000, 2005 and 2010 using simple age-period-models (dotted lines) or age-period-cohort models (broken lines). Clearly the broken lines gives a better approximation to the smoothed empirical rates.*

```
  90.25 15.44149 12.67524 12.74387 10.27029   NA  NA    NA
  90.35 15.38571 12.63260 12.69755 10.23592   NA  NA    NA
  90.45 15.32864 12.58960 12.64993 10.20113   NA  NA    NA
  90.55 15.27033 12.54626 12.60106 10.16593   NA  NA    NA
  90.65 15.21086 12.50260 12.55103 10.13035   NA  NA    NA
  90.75 15.15031 12.45864 12.49991 10.09442   NA  NA    NA
  90.85 15.08876 12.41442 12.44779 10.05817   NA  NA    NA
  90.95 15.02629 12.36995 12.39475 10.02162   NA  NA    NA
```

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1 )
> matplot( a.pt, cbind(prv[,np,"M","apc",],prv[,1,"M","apc",1])*100,
+          xlim=c(20,90), ylim=c(0,20), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
> matlines( a.pt, prv[,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" )
> matplot( a.pt, cbind(prv[,np,"F","apc",],prv[,1,"F","apc",1])*100,
+          xlim=c(20,90), ylim=c(0,20), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
> matlines( a.pt, prv[,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" )
```

```
> scen <- c("Mort obs, Inc obs","Mort 1995, Inc obs","Mort obs, Inc 1995","Mort 1995, Inc 1995")
> hts <- prv["74.95",np,"M","apc",1:4]*100
> cau.exp <-
+ function( wh=1:4, fill=FALSE )
+ {
+ pdf( paste( "./graph/DMpr-", paste(wh,collapse=""), if( fill ) "F",
+             ".pdf", sep="" ), height=8, width=11 )
+ par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1 )
+ matplot( a.pt, cbind(prv[,np,"M","apc",],prv[,1,"M","apc",1])*100,
+          xlim=c(20,90), ylim=c(0,20), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
+ matlines( a.pt, prv[,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(74.95,4)[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(0,20), xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
+ matlines( a.pt, prv[,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" )
+ 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 4.3 shows the predicted prevalences under 4 different scenarios compared to the observed prevalences as of 1.1.1995.

4.4 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.

Similary 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]
```

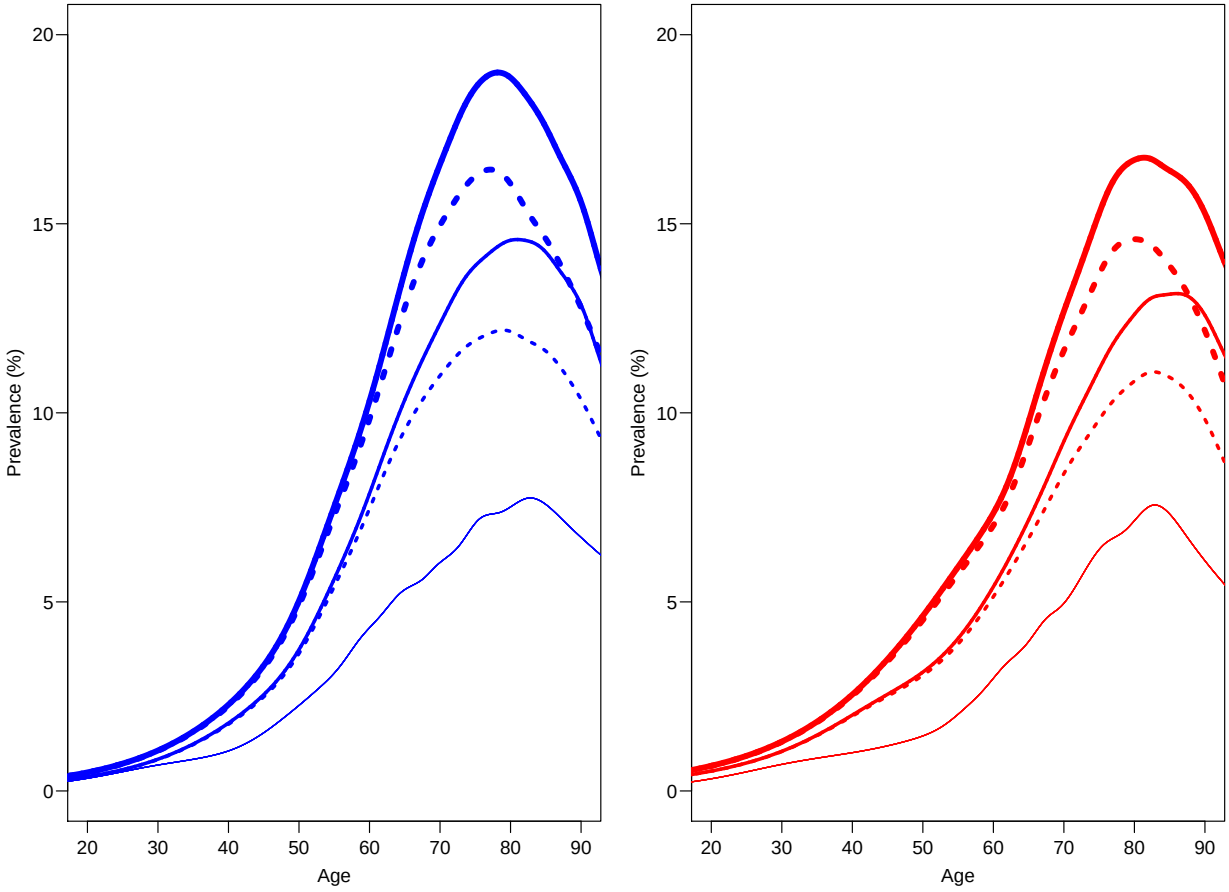

Figure 4.3: *The predicted prevalences under different scenarios:*
Full lines: Mortality rates evolve as observed, Broken lines: Mortality rates remain as 1995. Thick lines: Incidence rates evolve as observed, Thin lines: Incidence rates remain as in 1995.
The very thin lines lowest in the two displays are the observed prevalences in 1995.

```
$what
[1] "obs"    "m-fix" "i-fix" "all-f" "mort"   "inc"    "const"
```

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1 )
> 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" )
> 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", ylab="Prevalence (%)",
+          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" )
```

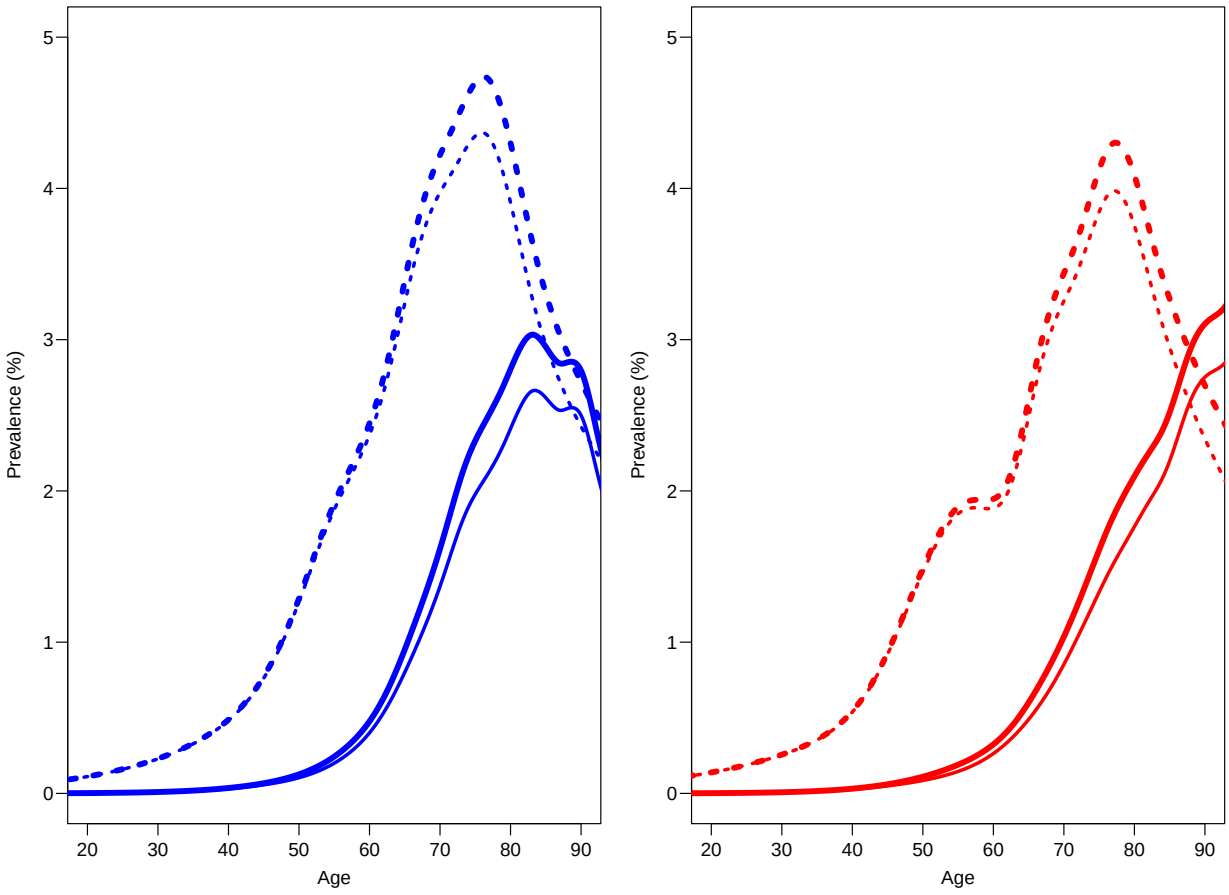

Figure 4.4: *Suggested contributions to age-specific prevalences from increasing mortalities over the last 15 years; the thick lines are obtained by subtracting the prediction based on fixing one rate from the one using the observed rates; thin lines based on subtracting the prediction based on fixing both rates from that where one is fixed. Full lines are for differences attributabel to changes in mortality rates, broken lines are for changes attributabel to cahnges in incidence rates.*

From figure 4.4 we see that the two possible ways of computing the contribution give pretty much the same results — the differences never exceed some 0.3%. Therefore, if we want to attribute fractions of the prevalence in 2010 to decreasing mortality and increasing incidence, we would want two measures that had a sum equal the the difference between the scenario with observed mortality and incidence rates ("`obs`"), and the scenario with rates fixed to those from 1995 ("`all-f`").

The thin lines at the bottom of figure 4.3 represent sthe prevalence at 1.1.1995, so it is pretty clear that the incidence an mortality rates as observed by 1995 did not provide for at steady state.

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

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

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

```
> prv[,,,,"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(3,5), mar=c(0,0,0,0), oma=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
+ for(j in 1:15)
+ {
+ poly.parts( a.pt, cbind(prv[,1   ,sex,"apc","obs"],
+                         prv[,j*10,sex,"apc",c("const","inc","mort")])*100,
+             col=clr, xlim=c(20,90), ylim=c(0,20), txt=(1996:2010)[j] )
+ abline(h=0)
+ if( j==1 ) text( rep(25,3), c(13,15,17),
+                  c("Original","Incidence","Mortality"),
+                  col=clr[2:4], font=2, adj=0, cex=1.2 )
```

```
+ if( j %in% c(1,6,11) ) axis( side=2 )
+ if( j %in% 11:15 ) 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 )
+ }
+ }


> one.comp( "M", rgb(c(3,2,1.5,0)/3,c(3,2,1.5,0)/3,1) )


> one.comp( "F", rgb(1,c(3,2,1.5,0)/3,c(3,2,1.5,0)/3) )


> 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) )
> 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(0,20) )
> axis( side=1 )
> axis( side=2 )
> abline(h=0:20,v=2:9*10,col=gray(0.9))
> text( rep(25,3), 17:19+0.5,
+       c("Original","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=0.8 )
> 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(0,20) )
> axis( side=1 )
> # axis( side=2 )
> abline(h=0:20,v=2:9*10,col=gray(0.9))
> text( rep(25,3), 17:19+0.5,
+       c("Original","Incidence","Mortality"),
+       col=clr[2:4], font=2, adj=0, cex=0.8 )
> mtext( "Age", side=1, outer=TRUE, line=0.8, font=1, las=0 )
> mtext( "Prevalence of DM (%)", side=2, outer=TRUE, line=2, font=1, las=0 )
```

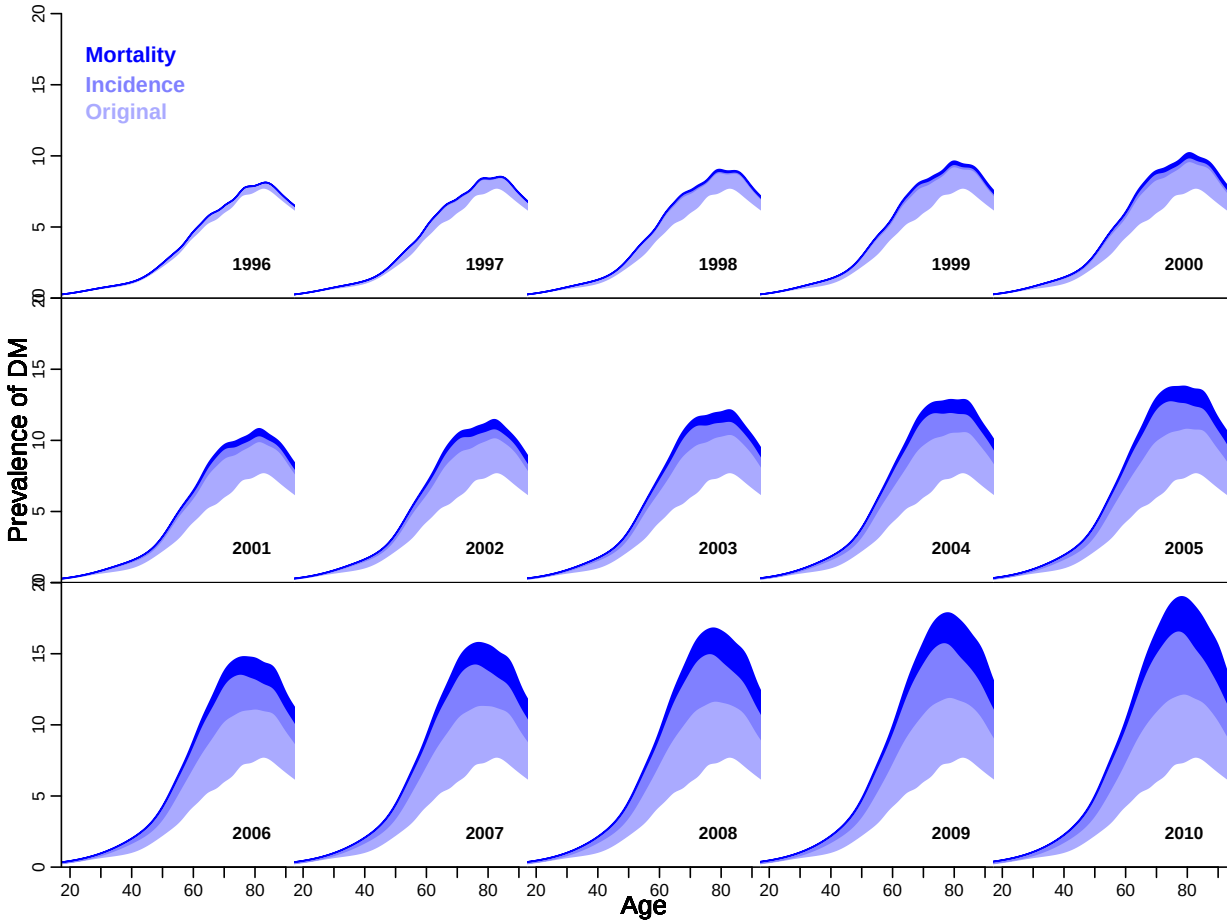

Figure 4.5: *Predicted increases age-specific prevalences of DM since 1995 in Denmark among men, partitioned by the contribution from rates as they were in 1995 ("Original"), increases in incidence and decrease in mortality, respectively.*

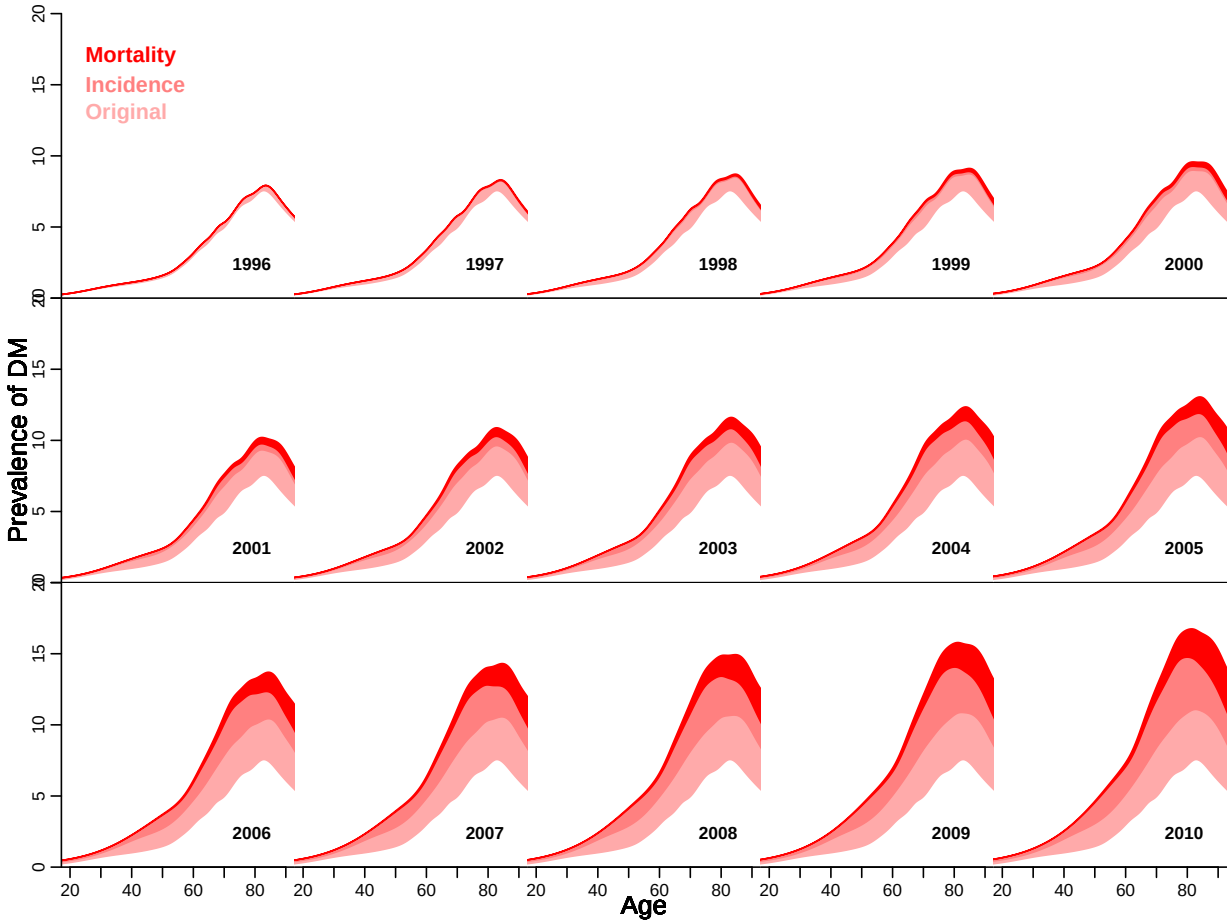

Figure 4.6: *Predicted increases age-specific prevalences of DM since 1995 in Denmark among women, partitioned by the contribution from rates as they were in 1995 ("Original"), increases in incidence and decrease in mortality, respectively.*

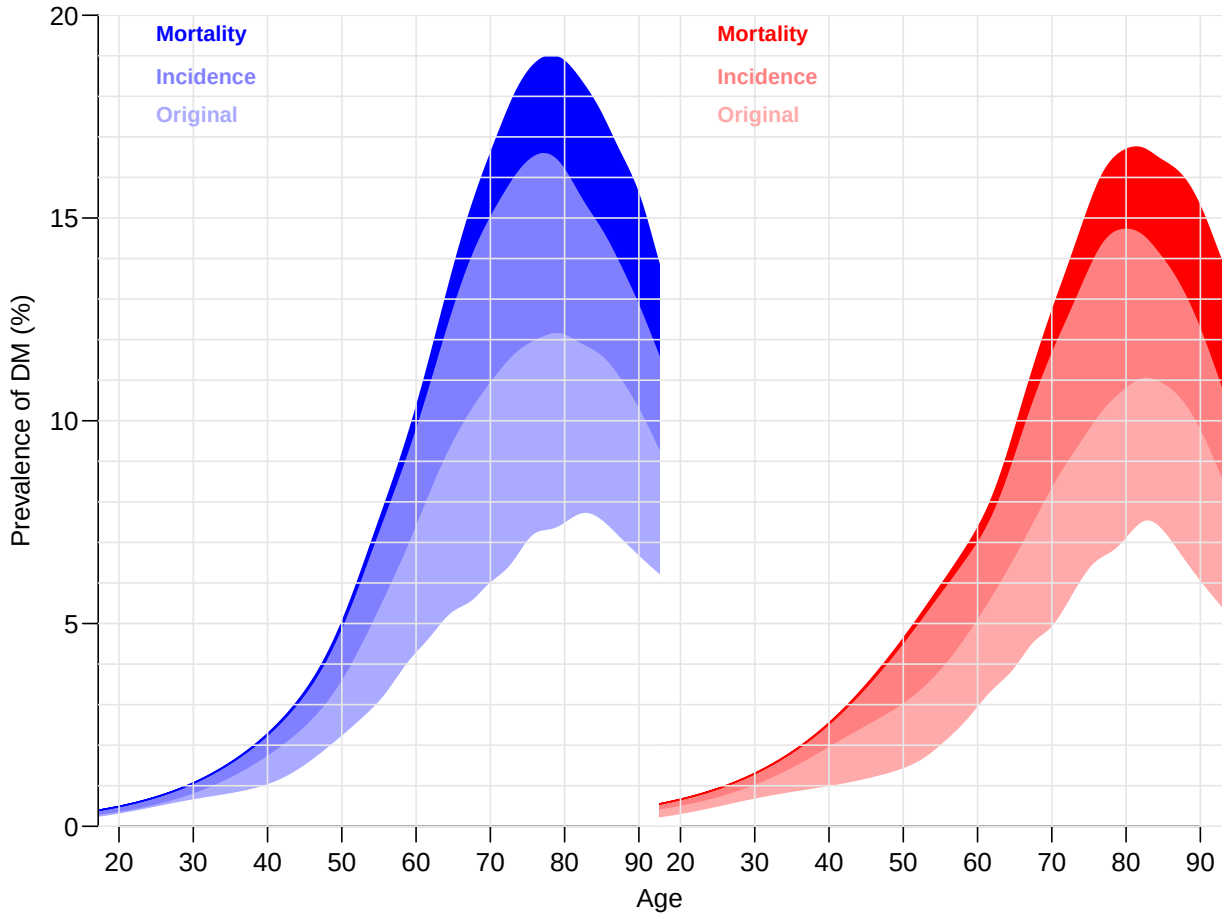

Figure 4.7: *Predicted age-specific prevalences of DM in Denmark 2010 among men (blue) and women (red), partitioned by the contribution from rates as they were in 1995 ("Original"), increases in incidence and decrease in mortality, respectively.*

4.5 The actual numbers of diabetes patients in Denmark

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

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

To show these effects we set up an array `prn` with structure like `prv` 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 100, and only 16 dates rater than 151 as `prv`.

```
> dn <- dimnames(prv)
> dn[[1]] <- 0:99
> dn[[2]] <- 1995:2010
> dn[[5]] <- dn[[5]][c(5:7,1)]
> prn <- NArray( dn )
> str( prv )
```

```
 num [1:1000, 1:151, 1:2, 1:2, 1:7] 0 0.000395 0.000399 0.000404 0.000408 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1000] "0.05" "0.15" "0.25" "0.35" ...
  ..$ t   : chr [1:151] "1995" "1995.1" "1995.2" "1995.3" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ tp  : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

```
> str( prn )
```

```
 logi [1:100, 1:16, 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 [1:16] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ tp  : chr [1:2] "apc" "ap"
  ..$ what: chr [1:4] "mort" "inc" "const" "obs"
```

In order to fill in the numbers we must have the age-specific prevalences at 1st January each year, that is at the dates 1.1.1996,...,1.1.2010 in the entries along the `t`-dimension of `prv`. Moreover we want the prevalences for a 1 year age class rather than an ageclass of lenght `int`. So we take the average prevalences from `prv` over each one-year interval.

```
> prv.a <- as.numeric( dimnames(prv)[["a"]] )
> wh.a <- c( diff( (prv.a-floor(prv.a)) > 0.5 ) == 1, FALSE )
> dimnames(prv)[["a"]][wh.a]
```

```
  [1] "0.45"  "1.45"  "2.45"  "3.45"  "4.45"  "5.45"  "6.45"  "7.45"  "8.45"
 [10] "9.45"  "10.45" "11.45" "12.45" "13.45" "14.45" "15.45" "16.45" "17.45"
 [19] "18.45" "19.45" "20.45" "21.45" "22.45" "23.45" "24.45" "25.45" "26.45"
 [28] "27.45" "28.45" "29.45" "30.45" "31.45" "32.45" "33.45" "34.45" "35.45"
 [37] "36.45" "37.45" "38.45" "39.45" "40.45" "41.45" "42.45" "43.45" "44.45"
 [46] "45.45" "46.45" "47.45" "48.45" "49.45" "50.45" "51.45" "52.45" "53.45"
 [55] "54.45" "55.45" "56.45" "57.45" "58.45" "59.45" "60.45" "61.45" "62.45"
 [64] "63.45" "64.45" "65.45" "66.45" "67.45" "68.45" "69.45" "70.45" "71.45"
 [73] "72.45" "73.45" "74.45" "75.45" "76.45" "77.45" "78.45" "79.45" "80.45"
 [82] "81.45" "82.45" "83.45" "84.45" "85.45" "86.45" "87.45" "88.45" "89.45"
 [91] "90.45" "91.45" "92.45" "93.45" "94.45" "95.45" "96.45" "97.45" "98.45"
[100] "99.45"
```

```
> ( wh.p <- dimnames(prn)[["t"]] )
```

```
 [1] "1995" "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004"
[11] "2005" "2006" "2007" "2008" "2009" "2010"
```

```
> comp <- c("mort","inc","const")
> str( prn[,,,,comp ] )
```

```
 logi [1:100, 1:16, 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 [1:16] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ tp  : chr [1:2] "apc" "ap"
  ..$ what: chr [1:3] "mort" "inc" "const"
```

```
> str( prv[wh.a,wh.p,,,comp] )
```

```
 num [1:100, 1:16, 1:2, 1:2, 1:3] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:100] "0.45" "1.45" "2.45" "3.45" ...
  ..$ t   : chr [1:16] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ tp  : chr [1:2] "apc" "ap"
  ..$ what: chr [1:3] "mort" "inc" "const"
```

```
> prn[,,,,comp ] <- prv[wh.a,wh.p,,,comp]
> prn[,,,,"obs"] <- prv[wh.a,rep(1,dim(prn)[2]),,,"obs"]
```

Now `prn` contians the prevalences (as fractions) for 100 age classes and 16 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
1    F 0.5 1995  0 34094
17   F 1.5 1995  4 32967
33   F 2.5 1995  6 33198
49   F 3.5 1995  5 31738
65   F 4.5 1995 13 31361
81   F 5.5 1995 13 30440
```

```
> subset(pr,A<1 & P<1997)
```

```
     sex   A    P X     N
1      F 0.5 1995 0 34094
2      F 0.5 1996 0 34051
1601   M 0.5 1995 3 35612
1602   M 0.5 1996 1 36055
```

```
> pop <- xtabs( N ~ A + P + sex, data=pr )[,,2:1]
> dmp <- xtabs( X ~ A + P + sex, data=pr )[,,2:1]
> str( pop )
```

```
 num [1:100, 1:16, 1:2] 35612 34747 35082 33330 32974 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0.5" "1.5" "2.5" "3.5" ...
  ..$ P  : chr [1:16] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
```

```
> str( prn )
```

```
 num [1:100, 1:16, 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 [1:16] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ tp  : chr [1:2] "apc" "ap"
  ..$ 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 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 <- "2010"
> oo <- c("mort","inc","const","obs")
> lim <- 5
> clr <- c("red","blue")
> draw.dmp <-
+ function(pp)
+ {
+ par( mar=c(3,3,3,0), mgp=c(3,1,0)/1.6, las=1 )
+ barplot( height=t( cbind( -dmp[,pp,"M"],
+                           dmp[,pp,"M"],
+                           dmp[,pp,"F"] ) )/ 1000,
+          horiz=TRUE, col=clr,
+          border=NA,space=0,axes=FALSE,names.arg=rep("",dim(prn)[1]),
+          xlim=c(-1,1)*lim*1.05,xlab="Persons in 1 year class (1000s)",ylab="Age")
+ abline(h=seq(0,100,5),
+        v=seq(-lim,lim,0.5),
+        col="white")
+ axis( side=1, at=seq(-lim,lim,1), labels=abs(seq(-lim,lim,1)) )
+ axis( side=2, at=seq(0,100,20) )
```

```
+ mtext( pp, at=-lim, adj=1.4, cex=1.3, font=2 )
+ mtext( formatC(sum(dmp[,pp,"M"]),0,format="f",big.mark=","), at=-1, col="blue", line=0, cex=0.99 )
+ mtext( formatC(sum(dmp[,pp,"F"]),0,format="f",big.mark=","), at= 1, col="red" , line=0, cex=0.99 )
+ mtext( "N", at=0, line=0, cex=0.99 )
+ }
> pdf( "./graph/DMpr-obs-film.pdf", width=8, height=6 )
> for( pp in paste(1996:2010) ) draw.dmp(pp)
> dev.off()


null device
          1


> for( pp in paste(1995:2010) )
+ {
+ pdf( paste("./graph/DMpr-obs-", pp, ".pdf", sep=""), width=8, height=6 )
+ draw.dmp(pp)
+ dev.off()
+ }
```

Now we can also draw a population pyramid using colors that range from very light to full:

```
> shd <- c(0.0, 1.5, 2.0, 2.8) / 3
> een <- c(1,1,1,1)
> clr <- rgb( c(een,rev(shd)),
+             c(shd,rev(shd)),
+             c(shd,    een ) )
> clr


[1] "#FF0000" "#FF8080" "#FFAAAA" "#FFEEEE" "#EEEEFF" "#AAAAFF" "#8080FF"
[8] "#0000FF"


> # Note: 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 <- 5
> 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=NA,space=0,axes=FALSE,names.arg=rep("",dim(prn)[1]),
+          xlim=c(-1,1)*lim*1.05,xlab="Persons in 1 year class (1000s)",ylab="Age")
+ abline(h=seq(0,100,5),
+        v=seq(-lim,lim,0.5),
+        col="white")
+ axis( side=1, at=seq(-lim,lim,1), labels=abs(seq(-lim,lim,1)) )
+ axis( side=2, at=seq(0,100,20) )
+ tt <- addmargins( apply( prn[,pp,,"apc",],2:3, sum ), 2 )
+ nn <- tt / tt[,5] * 100
+ ppos <- 1:5-0.1
+ npos <- -rev(ppos)
+ mtext( pp, at=-lim, adj=1.8, line=2, cex=1.2, font=2 )
+ mtext( c(lg<- c("Mort","Inc","Const","Org","All"),rev(lg)),
+        at=c(npos,ppos), col="black", cex=0.99, line=2 )
+ mtext( formatC(tt["M",1:5],0,,"f",,,","), at=npos, col="blue", line=1, cex=0.99 )
```

```
+ mtext( formatC(tt["F",5:1],0,,"f",,,","), at=ppos, col="red" , line=1, cex=0.99 )
+ mtext( formatC(nn["M",1:4],1,4,"f"), at=npos[1:4], col="blue", line=0, cex=0.99 )
+ mtext( formatC(nn["F",4:1],1,4,"f"), at=ppos[2:5], col="red" , line=0, cex=0.99 )
+ mtext( "N", at=0, line=1, cex=0.99 )
+ mtext( "%", at=0, line=0, cex=0.99 )
+ }
> pdf( "./graph/DMpr-film.pdf", width=9, height=6 )
> for( pp in paste(1996:2010) ) draw.pyr(pp)
> dev.off()


null device
          1


> for( pp in paste(1996:2010) )
+ {
+ pdf( paste("./graph/DMpr-", pp, ".pdf", sep=""), width=8, height=6 )
+ draw.pyr(pp)
+ dev.off()
+ }
```

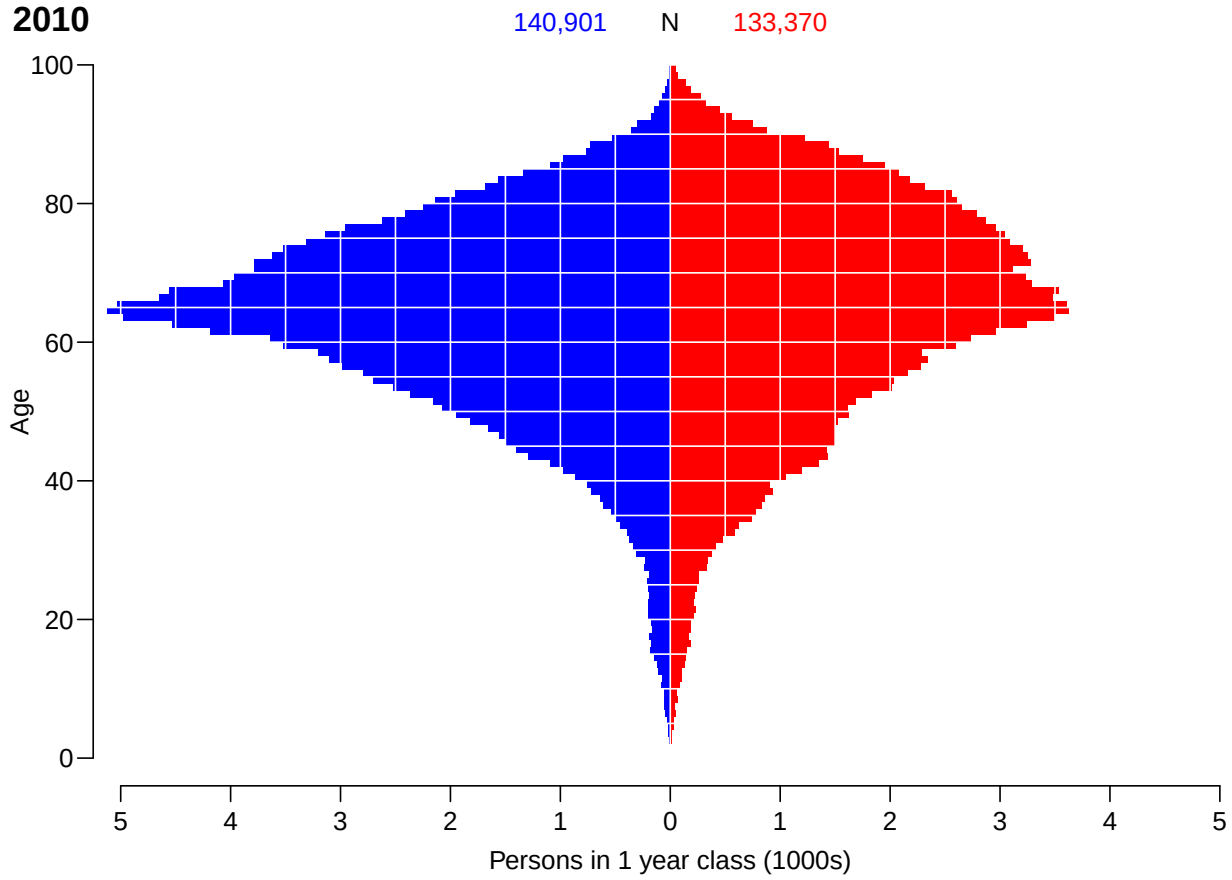

Figure 4.8: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2010.*

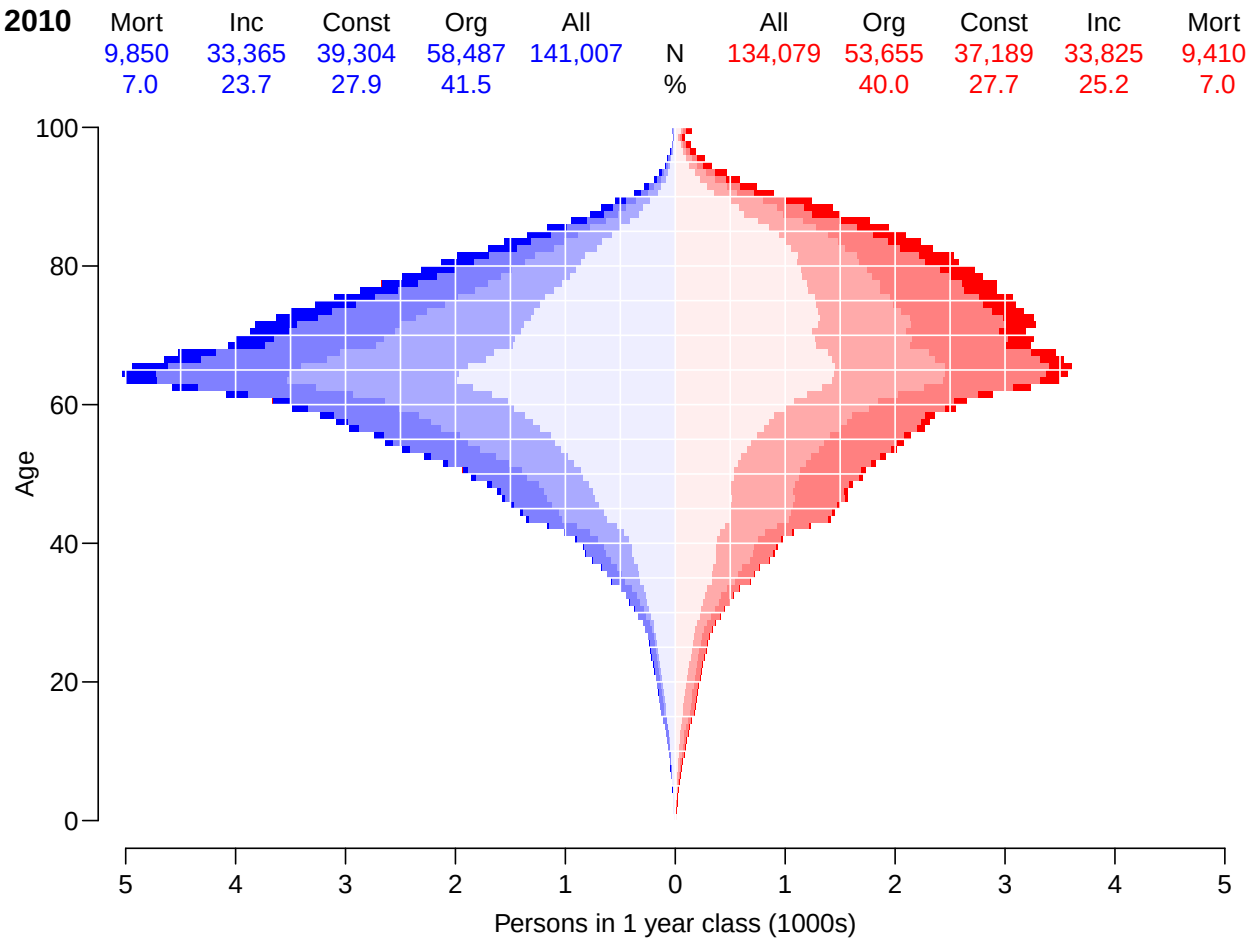

Figure 4.9: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2010, subdivided by the contribution from various causes: Mort: decreas in mortality, Inc: increase in incidence, Const: constant rates from 1995, Org: age-specific prevalence in 1995.*

Chapter 5

Article sections

5.1 Background

The motivation for this work is the research letter in the Lancet by Støvring *et al.* [1].

The core of the message is that there is no apparent change in diabetes incidence but an increase in prevalnece and a decrease in mortality. By the classical relationship between prevalence, incidence and duration (=survival with the disease):

$$\text{prevalence} = \text{incidence} \times \text{duration}$$

they suggest that the rise in prevalence is attributable to a decrease in mortality.

5.2 Relationships between prevalence, incidence and mortality

If we set aside the issues of age and sex in the population and just consider a homogeneous population we are dealing with the following situation: Suppose that mortality and incidence rates are constnt over time, and let p_0 be the prevalence of DM at time 0. Thus the prevalence of "Well" is $1 - p_0$. As a fraction of the entire population at time 0, the influx to the "DM" state over a timespan of length 1 is computed as the prevalence of "Well" times the probability of acquiring DM:

$$(1 - p_0) \times (1 - \exp(-\lambda \times 1))$$

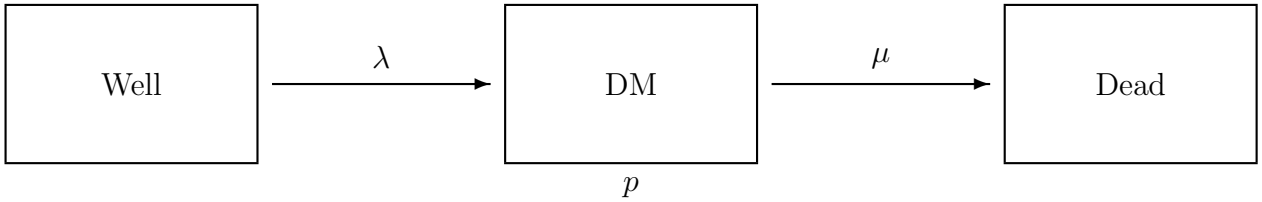

Figure 5.1: *States and transition rates used: λ: Incidence rate, p: prevalence, μ: mortality rate for diabetic patients.*

Similarly, as a fraction of the entire poulation, that remains in the "DM" state is computes as the prevalence of "DM" times the probaility of remaing in "DM" (i.e. the survival):

$$p_0 \times \exp(-\mu \times 1)$$

so the prevalence after a period of length 1 is:

$$p_1 = (1-p_0)\times(1-\exp(-\lambda\times 1))+p_0\times\exp(-\mu\times 1) \approx (1-p_0)\lambda+p_0(1-\mu) = \lambda-p_0(1-\mu-\lambda)$$

Since the mortality is larger than the incidence (at least in the older age-classes, $\mu \approx 20 \times \lambda$), the approximate formula for the prevalence shows that the prevalence increases by increasing incidence and by decreasing mortality.

This is under the assumption that the rates do not change. If the population is in a steady state, then $p_1 = p_0$, which is equivalent to:

$$\begin{aligned}
p &= \frac{1-\mathrm{e}^{-\lambda}}{2-\mathrm{e}^{-\lambda}-\mathrm{e}^{-\mu}} &&\approx \frac{\lambda}{\lambda+\mu} \\
\lambda &= -\ln\left(1-\frac{p}{1-p}(1-\mathrm{e}^{-\mu})\right) &&\approx \frac{p}{1-p}\mu \\
\mu &= -\ln\left(1-\frac{1-p}{p}(1-\mathrm{e}^{-\lambda})\right) &&\approx \frac{1-p}{p}\lambda
\end{aligned}$$

5.3 Comments on the Lancet paper

5.3.1 Prevalence and mortality

As an example, if we take the roughly stable incidence rates for males aged 65+, which is about 7%, and compute the steady-state prevalence for mortality rates in the range seem over the years 1993–99, we see expected prevalences rougly in the range 50–70 per 1000, somewhat above what is seen in the material.

```
> pim <-
+   function( prev, inc, mort )
+   {
+     # Computes the missing quantity among prevalence, incidence and
+     # mortality under a steady state assumption.
+     # BxC, 09/2003
+     if( missing(prev) + missing(inc) + missing(mort) != 1 )
+       stop( "You must provide excatly two of prev, inc and mort" )
+     if( missing( prev ) ) prev <- (1-exp(-inc))/(2-exp(-mort)-exp(-inc))
+     if( missing( mort ) ) mort <- -log( 1-((1-prev)/prev)*(1-exp(-inc)) )
+     if( missing( inc  ) ) inc  <- -log( 1-(prev/(1-prev))*(1-exp(-mort)) )
+     res <- list( prev=prev, inc=inc, mort=mort )
+     if( max( sapply( res, length ) ) == 1 ) res <- unlist( res )
+     res
+   }
> p.ch <-
+   function( prev, inc, mort )
+   {
+     # Computes the annual change in prevalence for given values of
+     # prevalence, incidence and mortality.
+     # BxC, 09/2003
+     exp( - mort ) + ((1-prev)/prev) * ( 1 - exp( -inc ) )
```

```
+   }
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="l" )
> inc <- 7/1000
> mort <- 140:100/1000
> plot( mort*1000, pim( inc=inc, mort=mort )$prev*1000,
+       xlab="Mortality per 1000 PY", ylab="Prevalence per 1000",
+       type="l", lwd=3, ylim=c(0,100) )
> abline( v=seq(100,140,5), h=0:10*10, col=gray(0.8) )
> box()
> for( i in 4:10/1000 )
+   lines( mort*1000, pim( inc=i, mort=mort )$prev*1000, lwd=3 )
> text( rep(141.5,7), pim( inc=4:10/1000, mort=140/1000)$prev*1000,
+       paste( 4:10 ), adj=1 )
> text( 141.5, 80, "Incidence rate\nper 1000 PY", adj=c(1,0.5) )
```

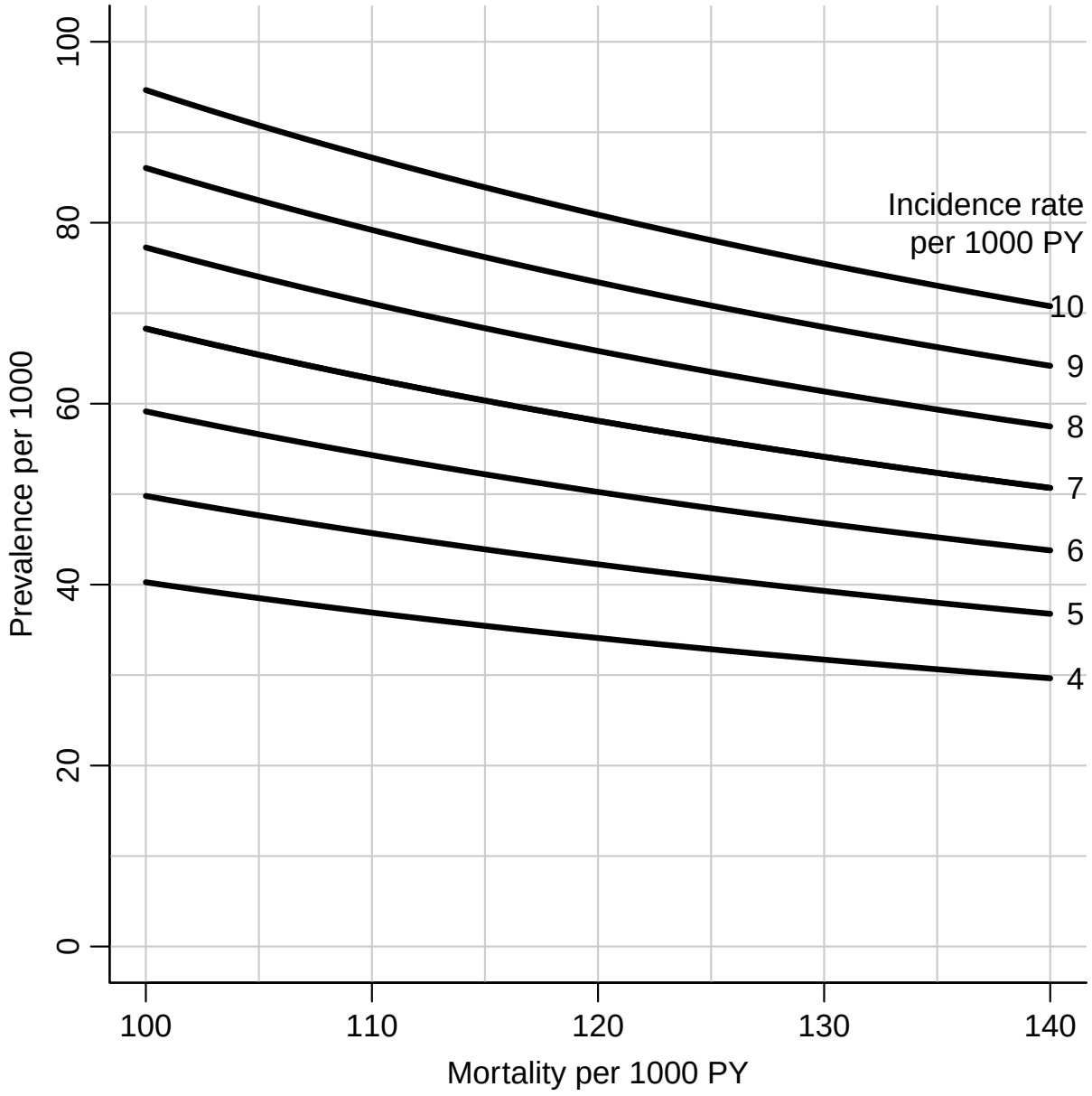

Figure 5.2: *Relationship of prevalence and mortality for different levels of incidence.*

5.4 Constant incidence rates?

The incidence rates for all antidiabetics (i.e. users of any antidiabetic medicine) are graphically reported to be in the vicinity of 6.5% for females and 7% for males.

The annual changes are estimated for females $1.0\%(-0.7; 2.7)\%$ and for males $1.0\%(-0.0; 3.1)\%$

5.5 Suggestions for article sections

The following are suggestions for *parts of* sections in a scientific paper reporting the results.

5.5.1 Material and methods

5.5.1.1 Data

5.5.1.2 Statistical methods

5.5.2 Results

5.5.3 Discussion

Bibliography

[1] H Støvring, M Andersen, H Beck-Nielsen, A Green, and W Vach. Rising prevalence of diabetes: evidence from a danish pharmacoepidemiological database. *Lancet*, 362:537–38, August 2003.