Components of Diabetes Prevalence in Denmark

SDC
October 2015
http://BendixCarstensen.com/DMreg/Prevalence
Version 5.1

Compiled Tuesday 13th October, 2015, 08:36
from: `/home/bendix/sdc/DMreg/NDR-demo/prev/Xprev.tex`

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

Contents

Article

Authors:
Bendix Carstensen & **Marit Eika Jørgensen**
Steno Diabetes Center, Clinical Epidemiology

This chapter is the text for the article intended for publication ; the remaining chapters contains the background material for the article, including all R-code generating all analyses and graphs.

Research in Context

Evidence before this study

We searched Google scholar for articles with terms “diabetes”, “prevalence”, “incidence” and “mortality”. Qualitatively it appears well known that mortality influences prevalence negatively and incidence positively and may papers have pointed this out; but only one study has made an attempt to quantify the effects [1], however not subdivided by age, and found that the major driver of changes in prevalence was the existing imbalance between incidence and mortality rates. A precise quantification of the effect of actually observed (changes in) diabetes incidence and mortality on prevalence is lacking.

Added value of this study

This study is the first to precisely quantify the relative contributions of changes in mortality and incidence as well as that of epidemiological imbalance to the changes in prevalence of diabetes, both in terms of prevalence and absolute number of diabetes patients in Denmark.

Moreover it describes a precise formal framework for this type of calculations that is readily applicable in other settings where incidence and mortality rates of diabetes are available.

Implications of all the available evidence

The major driver of diabetes prevalence in Denmark is the increasing incidence rates and the epidemiological imbalance already present in 1995 each contributing about a quarter of all prevalent cases in 2012; the decreasing mortality among diabetes patients only accounts for about 7.5% of the present number of diabetes patients.

Abstract

The prevalence of diabetes has been increasing dramatically the last decades. It is also known that incidence rates of diabetes has been increasing and that mortality — both among persons with and without diabetes — has been decreasing. But it is not known to which extent these two components have influenced the prevalence of diabetes.

This paper defines the relative contribution of incidence and mortality to changes in prevalence in proper demographic terms, and uses Danish register data for the period 1995–2011 to address the question statistically precise in a Danish context, using age- and period-specific prevalence, incidence and mortality rates, based on data from the Danish National Diabetes Register.

The average annual increase in diabetes incidence was some 4%, the average decrease in mortality among non-diabetic persons about 2.5% whereas the average annual decrease in mortality among diabetes patients was 4%. The fraction of prevalent cases of DM at the end of 2011 attributable to declining mortality was 7.5%, the fraction attributable to increasing incidence rates was 30% and the fraction attributable to the non-steady state imbalance between incidence and mortality rates in 1995 was 25%.

In conclusion, the major driver of the increasing prevalence of DM in Denmark over the period 1995–2011 is the increasing incidence rates, whereas the declining mortality rates play a smaller albeit not insignificant role.

Keywords: Demography, Diabetes, Incidence, Mortality, Prevalence, Trends.

Introduction

Diabetes (DM) is among the leading causes of death in Europe with diabetic macro- and microvascular complications resulting in increased disability and enormous health care costs [2]. As a consequence, systematic and timely diabetes surveillance is essential for planning and implementing health policy. Many countries have faced a rapid increase in diabetes prevalence [3]. Data from the US Center for Disease Control show a near quadrupling of diagnosed diabetes from 5.5 million persons in 1980 to 21.1 million in 2010 [4]. The most recent Scottish Diabetes Survey (2013) estimates that 5.1% of the population had a diagnosis of diabetes in Scotland at the start of 2013, compared to 4.1% in 2007 [5]. The large increase in prevalence is presumably caused by an increase in incidence due to underlying risk factors, primarily obesity and aging of the population. However, a significant decline in mortality has been reported, both in type 1 [6, 7, 8] and type 2 diabetes [9, 10, 11]. Also in the non-diabetic background populations there has been a decrease in mortality, but not to the same extent as among patients with diabetes. Based on regional registers, the notion of a rise in diabetes incidence in Denmark has been challenged in a paper by Støvring *et al.*. [12]. The core message by these authors is that there is no apparent change in diabetes incidence but an increase in prevalence and a decrease in mortality. By the classical relationship between prevalence, incidence and duration (=survival with diabetes):

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

the authors suggest that the rise in prevalence is largely attributable to a decrease in mortality, that is increasing duration. The formula is however only a rough approximation to reality as it concerns a steady-state situation and does not take age into account. As such, it should be more regarded as a *qualitative* statement about the relation ship between the quantities.

A number of papers have pointed this qualitative relationship out too, for example [13, 14], The only study that has attempted a quantification is Evans *et al.* [1], however not quantifying effects by age, and using a very crude age adjustment with age-classes 20 years wide.

The motivation for the present work is 1) a wish to place the discussion of contributions from changing incidence and mortality rates on prevalence of diabetes at a firm theoretical foundation, 2) to quantify the relative sizes of these contributions to prevalence and thereby 3) to show how to use this machinery in a practical context.

Specifically, we use the age-specific prevalences of diabetes as of 1 January 1995 as starting point and apply age- and period-specific incidence and mortality rates to predict the prevalences in the period 1 January 1995 through 1 January 2012. By using either the observed rates or rates assumed constant after 1995 we obtain predicted prevalences in 2012 under different scenarios. Differences between scenarios are then defined as the fraction of the prevalences attributable to different factors.

Methods

Prevalence, incidence and mortality

The following is to some extent a repetition of standard theory from demography / probability theory, and the extension to several age-classes and time-varying incidence and mortality rates is if not straight-forward, then well-known from any curriculum in demography and probability theory.

Diabetes incidence and mortality in the population is fully described by a 3-state model, with three transition rates (Figure 1). If each of these rates is assumed to depend on sex, and continuously on age, calendar time and date of birth, it is possible to use the continuous age-distribution of prevalent diabetes patients at the start of the observation period (1 January 1995) in conjunction with the incidence and mortality rates for the period to predict the prevalence at the end of the period, 1 January 2012.

This is simply done by using a sex-, age- and period-specific transitions between the three states “No DM”, “DM” and “Dead” (Figure 1). In each step, the population at a given time in a given (say 1-year) age-class with and without diabetes is updated for one year, so that we know how many at that time are in the three states the year after.

Specifically, we considered transitions over a small interval of length ℓ and using the notation $\mathrm{P}_{\text{No DM},\text{DM}}(\ell)$ for $\mathrm{P}\left\{\text{DM at } (a+\ell, p+\ell) \mid \text{No DM at } (a,p)\right\}$, we used the

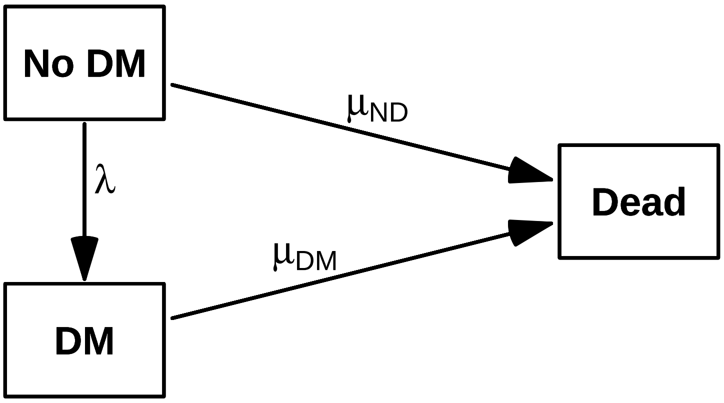

Figure 1: *States and transition rates used:* λ*: Incidence rate,* μ_{ND}*: mortality rate in persons without diabetes,* μ_{DM}*: mortality rate in persons with diabetes. Prevalence of diabetes is the fraction in state DM relative to all in states No DM and DM.*
Each rate is modeled separately for men and women, using an age-period-cohort model with continuous effects. The prevalence of diabetes is the fraction of persons in "DM" relative to all persons in "DM" and "No DM".

following transition probabilities:

$$\begin{aligned}
\mathrm{P}_{\text{No DM,No DM}}(\ell) &= \exp\bigl(-(\lambda+\mu_{\text{ND}})\ell\bigr)\\
\mathrm{P}_{\text{DM,DM}}(\ell) &= \exp\bigl(-\mu_{\text{DM}}\ell\bigr)\\
\mathrm{P}_{\text{No DM,DM}}(\ell) &= \frac{\lambda}{\lambda+\mu_{\text{ND}}}\Bigl(1-\exp\bigl(-(\lambda+\mu_{\text{ND}})\ell\bigr)\Bigr) \approx \lambda\ell\\
\mathrm{P}_{\text{No DM,Dead}}(\ell) &= \frac{\mu_{\text{ND}}}{\lambda+\mu_{\text{ND}}}\Bigl(1-\exp\bigl(-(\lambda+\mu_{\text{ND}})\ell\bigr)\Bigr) \approx \mu_{\text{ND}}\ell\\
\mathrm{P}_{\text{DM,Dead}}(\ell) &= 1-\exp\bigl(-\mu_{\text{DM}}\ell\bigr) \approx \mu_{\text{DM}}\ell
\end{aligned}$$

The reason to choose ℓ small is that the above approximations should be valid, notably that the probability of two transitions No DM→DM→Dead occurring in one interval should be negligible. This will render the updating machinery reasonably accurate to predict the observed prevalences at the end of the study period.

Data

We modeled incidence and mortality rates using data from the Danish National Diabetes Register (NDR) [15, 16], the Human Mortality Database [17] and the data-bank of Statistics Denmark to provide

- the number of events of diabetes,
- number of deaths among persons with and without diabetes and
- follow-up time (risk time, person-years)

classified by sex, age, date of follow-up and date of birth (Lexis triangles). The midpoint of the Lexis triangles were coded for each entry for continuous modeling of effects [18].

We also used the NDR to enumerate the number of prevalent cases of diabetes in 1-year age-classes at each of the dates 1 January 1995 through 1 January 2012, and from Statistics Denmark we obtained population size at each of these dates, all subdivided by sex and 1-year age-class.

All analyses were done seperately for men and women.

Statistical methods

We fitted separate log-link binomial models for men and women to the no. of prevalent cases at each of the dates 1 January 1995 through 1 January 2012, using natural splines (restricted cubic splines) to describe the age-dependence. These models thus provided estimates of diabetes prevalence as a continuous function of age.

We fitted age-period-cohort models [18] for diabetes incidence rates and mortality among persons with and without diabetes, separately for each sex. Effects of age, date of follow-up (period) and date of birth (cohort) were modelled continuously by natural splines (restricted cubic splines), and the models thus provide predicted incidence and mortality rates as continuous functions of age and date of follow-up.

Since we only use the age-period-cohort (APC) models for prediction of rates, the usual identification problem of the parametrization of effects in APC models is not relevant here. We estimated the average time trend from the APC models using the observed number of events as weights as described in Carstensen [18].

Demographic methods

We used the models fitted to predict the incidence and mortality rates at the midpoint of all months from 1 January 1995 through 1 January 2012 at the start of each of the 1200 age-classes, *i.e.* we used $\ell = 1$ month (formally 365.25/12 days). For updating the prevalence in age class $(a, a + \ell)$ at time p to the prevalence in age-class $(a + \ell, a + 2\ell)$ at time $p + \ell$, we used rates predicted at age $a + \ell$ at time $p + \frac{\ell}{2}$. As a check on the appropriateness of the calculations, the predicted prevalences at the end of the study period is compared with the actual observed.

The same exercise was then repeated in scenarios where we fixed the (age-specific) incidence and/or mortality rates to be as in 1995. The difference between predicted prevalences under these scenarios and the actually observed will then represent the contributions to the prevalence in 2012 from increasing incidence and decreasing mortality respectively.

Specifically, the contribution from changing incidence rates were computed in two different ways:

1. Difference between results with 1995-fixed resp. observed incidence rates using the observed mortality rates over the period.

2. Difference between results with 1995-fixed resp. observed incidence rates using the mortality rates fixed to the 1995 level.

— and vice versa for the contribution from the changing mortality rates.

Table 1: *Events and person-years (in 1000s) in the Danish population in the 17 year study period 1995–2011.*

		No diabetes			Diabetes	
		DM diag	Deaths	P-years	Deaths	P-years
Men	1995-2000	41,775	129,184	12,740.5	19,403	301.0
	2001-2006	67,423	138,354	15,425.5	28,592	542.0
	2007-2011	85,117	124,886	15,565.0	35,261	801.3
	Total	194,315	392,424	43,731.1	83,256	1,644.3
Women	1995-2000	38,617	133,223	13,068.4	18,002	294.4
	2001-2006	62,505	147,536	15,793.7	25,970	522.3
	2007-2011	75,792	132,416	15,900.7	31,161	758.9
	Total	176,914	413,175	44,762.9	75,133	1,575.6
M+W	Total	371,229	805,599	88,494.0	158,389	3,219.9

The contributions from changing incidence resp. mortality were taken as the average of the two approaches for each.

Finally, we took the difference between the observed prevalences in 1995 and those predicted for 2012 by fixing *both* incidence and mortality rates to the 1995 level throughout, as the component of prevalence attributable to the demographic imbalance in 1995. This is thus the change in prevalence occurring from the fact that incidence and mortality rates in 1995 were *not* in a steady-state equilibrium with equal number of incident cases of DM and deaths among DM patients.

A full account of all calculations is available as:
`http://bendixcarstensen.com/DMreg/Prevalence/Xprev.pdf`

Results

In the study period there were a bit more than 90 mil. person-years and over 370,000 new cases of diabetes and some 960,000 deaths, of which almost 160,000 among diabetes patients (Tables 1 , **??**). Even though the last of the periods is only 5 years long, there are more diagnoses and deaths in this, partly due to the increasing diabetes incidence, but primarily due to the aging population in general.

Trends in incidence and mortality rates

From the fitted age-period-cohort model we extracted the average annual trend in rates; as seen from table 2 there was a clear increase in incidence of diabetes of 4% per year, while mortality was *de*creasing; 2.5% per year for persons without diabetes, but 4% per year for persons with diabetes.

Table 2: *Average annual change (%) in incidence and mortality rates in Denmark over the period 1995–2011.*

		% change (95%c.i.)
No diabetes:		
DM incidence	Men	3.9 (3.8; 4.0)
	Women	4.0 (3.9; 4.1)
Mortality	Men	−2.7 (−2.8;−2.6)
	Women	−2.3 (−2.4;−2.2)
Diabetes:		
Mortality	Men	−4.0 (−4.2;−3.9)
	Women	−3.8 (−4.0;−3.7)

Components of prevalence

The prevalences at 1 January 2012, as predicted from the age-specific prevalences in 1995 and the fitted incidence and mortality rates from the age-period-cohort models for the rates, showed a very good agreement with the observed prevalences in 2012, as is seen from figure 2. Thus the prediction method is sufficiently accurate to yield credible results for the scenarios considered.

The components of the prevalences as derived from the models are shown in figure 3, where it is seen that the fraction of the diabetes prevalence attributable to decreasing mortality is quite substantial in older ages. However it is equally clear that the dominant components in the changing diabetes prevalence are the increasing incidence and the fact that the prevailing incidence and mortality rates in 1995 were not in equilibrium with the prevalences.

We multiplied the components of prevalence as derived with the number of persons in the Danish population in different ages to produce the *number* of diabetes patients in the population attributable to each of the causes. This is shown in figure 4.

It is seen that the mortality decrease has a comparatively small impact on the number of cases, because the effect is largest in the older ages where the number of prevalent diabetes cases is limited anyway.

The fraction of diabetes cases attributable to declining mortality is about 7.5%, whereas the fraction attributable to increasing incidence of diabetes is almost 30%, and 25% attributable to the imbalance between incidence, mortality and prevalence already present in 1995. The number of cases present in the Danish population using the age-specific prevalences from 1 January 1995 with the age-distribution as of 1 January 2012 represent the remaining 37%. These figures are only slightly different between men and women, as seen in figure 4.

The development of the components as a fraction of all prevalent diabetes cases in different ages is shown in figure 5, and not surprisingly, the mortality decrease has the largest impact on the prevalences in older ages. We found that the fraction attributable to mortality decline was less than 5% in age 60, a bit larger in age 70 and almost 15% at age 80 for men, and slightly less in women (Figure 5).

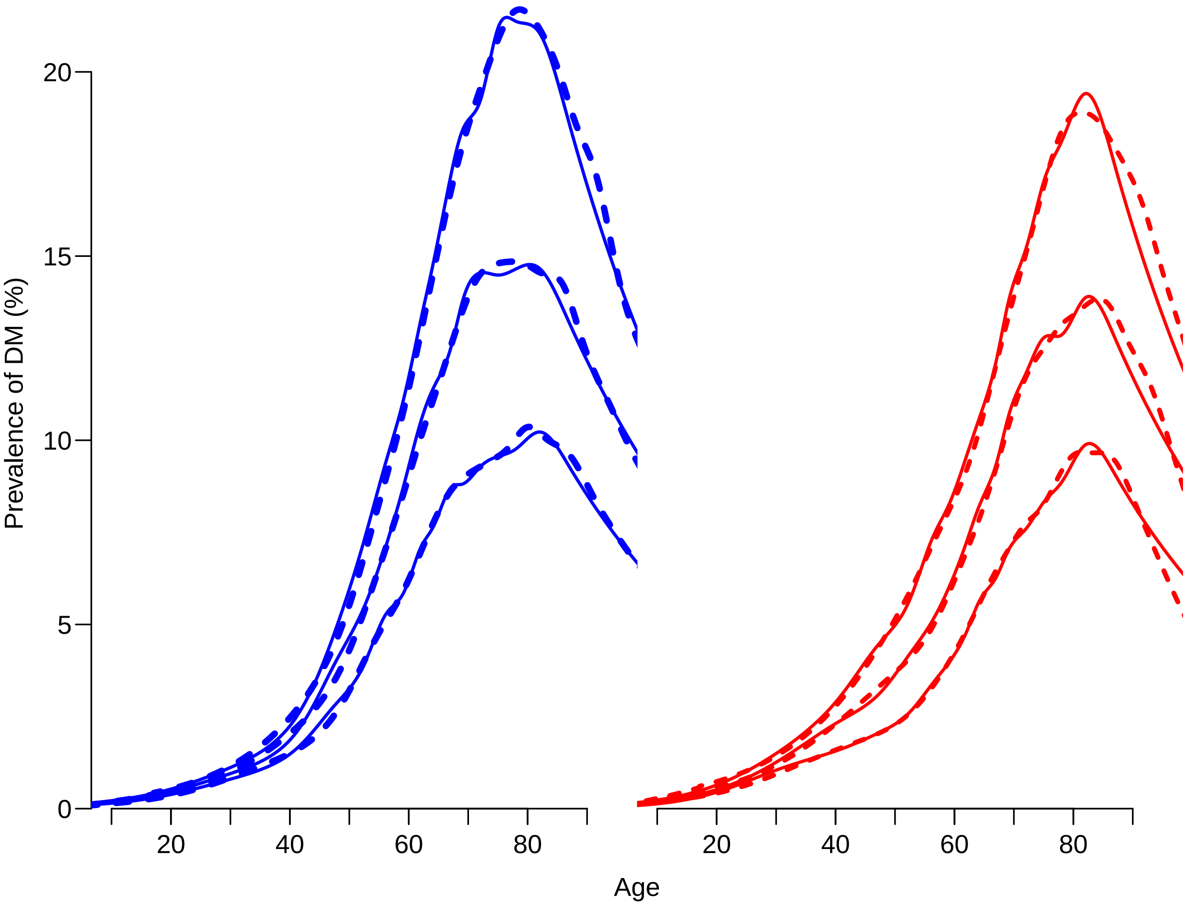

Figure 2: *Observed (full lines) and predicted (broken lines) prevalence of DM in Denmark (from low to high) 2000, 2006 and 2012. The observed prevalences are smoothed using natural splines. The predicted prevalences are based on the prevalences as of 1995 and estimated rates from age-period-cohort models for the incidence and mortality rates for the transitions in figure 1. Men in blue, women in red.*

Discussion

We have shown that over the last decades in Denmark, the decline in mortality has had some impact on the increasing prevalence of diabetes, whereas the major drivers of the prevalence increase have been the increase in diabetes incidence as well as the imbalance between incidence and mortality already present in 1995.

The absolute number of cases attributable to the different components is of course heavily dependent on the particular age-distribution in the Danish population. The patterns of birth has been very similar throughout Europe in the 1940es and 50es, with a baby-boom in the late 1940es, thus rendering a large part of today's population in the age-range where diabetes incidence is increasing. But even so these absolute numbers are less generalizable than the age-specific components of prevalence.

The finding of a decline in diabetes-related mortality is encouraging, although the resulting increase in diabetes prevalence obviously challenges the health care system. A larger number of older people will survive with diabetes complications with increased costs of diabetes treatment, as well as costs related to screening for and treatment of

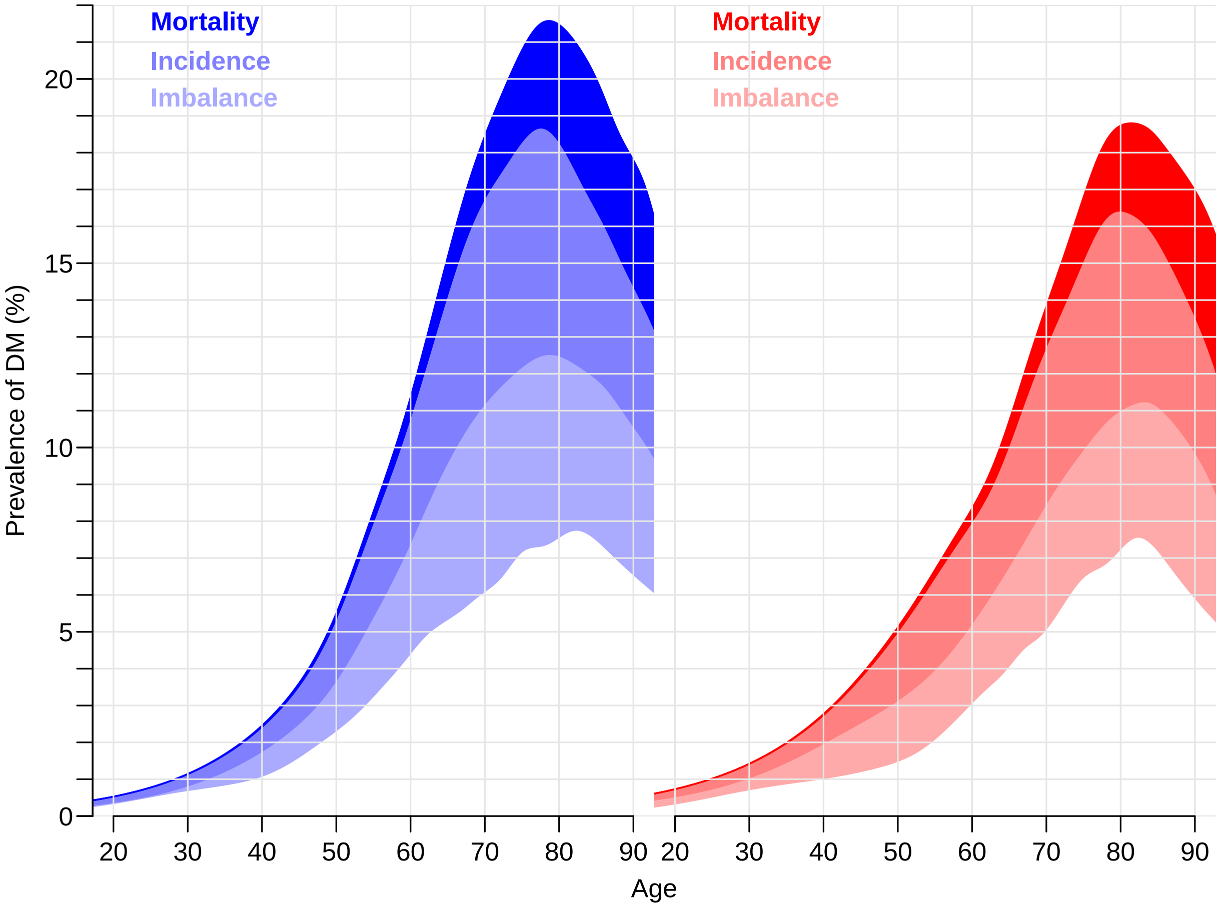

Figure 3: *Components of the changes in diabetes prevalence 1995–2012, based on prevalence in 1995 and models for incidence and mortality in the period. Men in blue, women in red.*

complications. Contrary, the observed increase in diabetes incidence calls for intensified preventive strategies in non-diabetic individuals. Thus, the increasing diabetes prevalence has different public health consequences according to the contributing prevalence components, a finding that underscores the value of such a detailed examination.

Comparison with other studies

Few studies have addressed the relative contribution of mortality vs incidence to diabetes prevalence. A recent study from Israel observed a deceleration in the upward trend in diabetes prevalence despite declining mortality [19].

Støvring *et al.* [12] merely analyzed relative annual changes in incidence, prevalence and mortality, and no formal quantification of the relative impact of mortality and incidence changes were made, so it is not possible to make a precise comparison. But the authors concluded that “Although our data do not allow a firm conclusion as to why prevalence is rising, we believe that the decrease in mortality should be taken into account. Otherwise, incorrect conclusions could be drawn about the relation between the western lifestyle and the rising number of diabetics.” This is indeed confirmed by this study as we estimate that as much as 7.5% of the current diabetes cases can be ascribed to the last 17 years’ decreasing mortality, less for ages under 70, somewhat more for older ages.

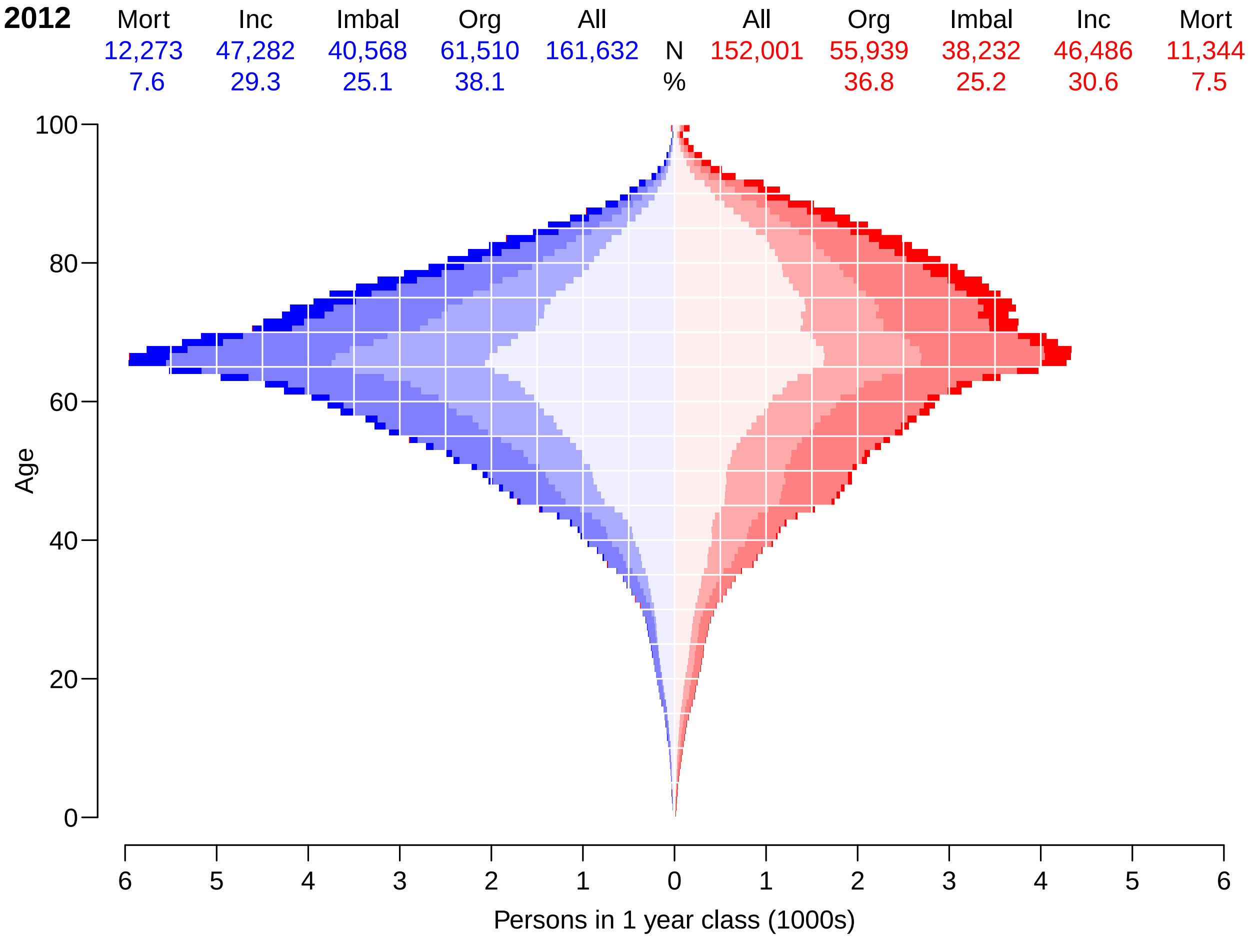

Figure 4: *Age-distribution of persons with diabetes according to components of the changes in diabetes prevalence 1995–2012. Men in blue, women in red.*

Evans *et al.* [1] used Tayside (Scotland) data to attempt a quantification of the relative contributions of incidence and mortality. They showed that 60% of the increase in diabetes prevalence over the period 1993–2004 was attributable to the initial imbalance between incidence and mortality; 25% to the increasing incidence and only 11% to decreasing mortality, which in very broad terms can be said to be qualitatively similar to our results. However the paper use very broad age-classes (divided only at 45, 65 and 85) leaving ample space for residual confounding by changing age-structure of the population of diabetes patients, so understandably the paper presents no age-specific effects, and thus their conclusions for the overall prevalence may be strongly dependent on the particular age-distribution of the Scottish population and changes in it. Moreover it is not clear from the paper to what extent the authors took the mortality change in the non-diabetic population (as those available for new diagnoses) under different scenarios into account. Finally the comparison made was only with the scenario with incidence and mortality fixed at the initial level and not with the actually observed rates, so the results are not entirely comparable with ours although the populations in Scotland and Denmark in many respects are quite similar.

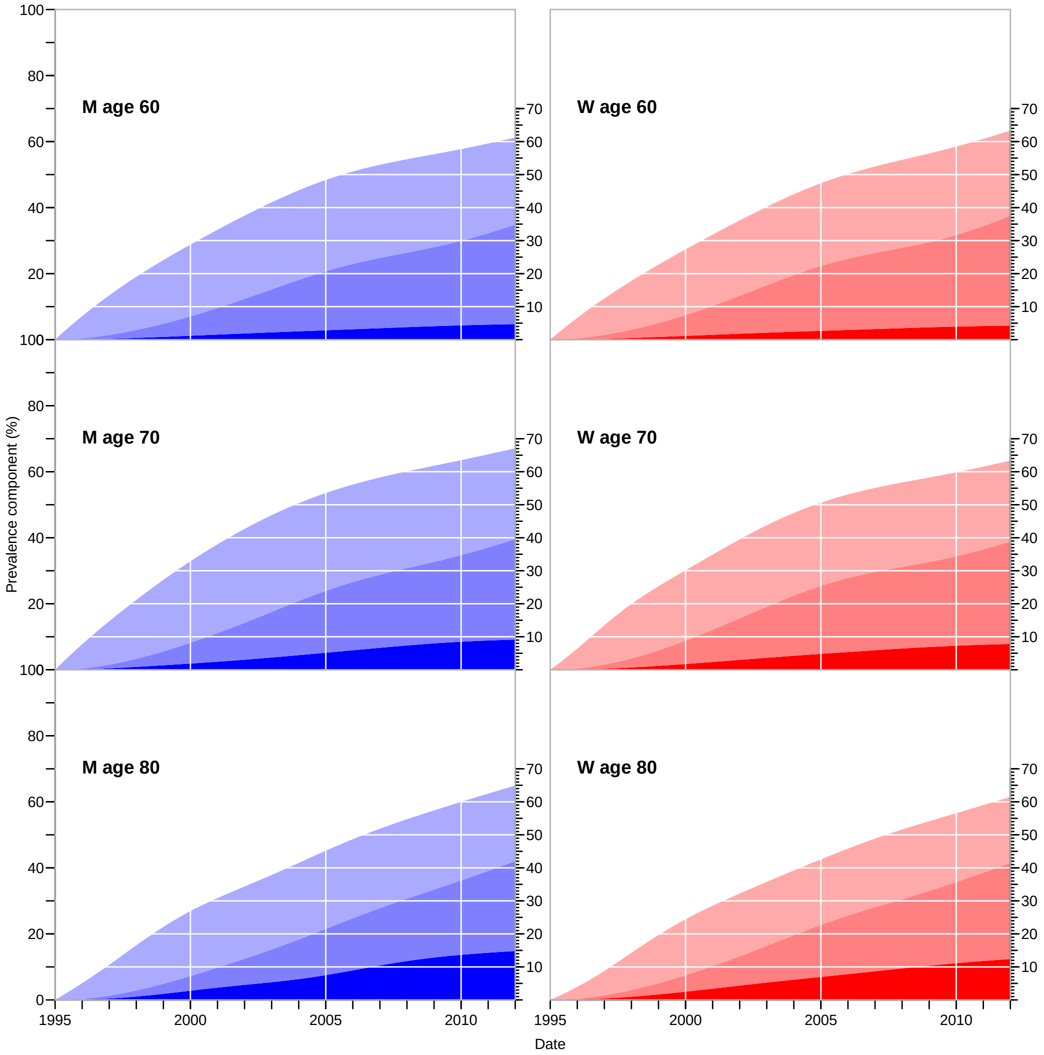

Figure 5: *Fraction of the prevalent cases at different times attributable to a) declining mortality (bottom, full color), b) increasing incidence (middle, pale color) and c) prevalence/mortality imbalance 1995 (top, weak color). The white area above the curves corresponds to the fraction of the cases that would have been alive if incidence and mortality rates had remained as in 1995. Men in blue, women in red.*

Strengths and limitations

We developed a model for partitioning prevalence changes in four contributions, which was based on application of well-known demographic concepts and classical epidemiological modeling of occurrence rates. While this machinery in principle is straight-forward to use in any context, it does rely on the availability of detailed register data of diabetes incidence and mortality.

Moreover, it is important that the predictive capability of the models be checked against the observed development in diabetes prevalence. This was done using the observed prevalences in 1995 and estimated incidence and mortality rates to project prevalences through the period, and showed excellent agreement between observed and predicted prevalences throughout. Therefore we conclude that the modeling of rates have been sufficiently accurate to give credible results.

The register-based approach in our study has clear limitations. Firstly, the study cannot distinguish between type 1 and type 2 diabetes. Secondly, it is not possible to determine whether the observed increase in diabetes incidence reflects a true change in incidence or whether it is caused by intensified diagnostic activity, resulting in more low-risk people with diabetes being included in the national diabetes register. Finally, a recent study has suggested that some misclassification exists with occurrence of false-positive cases in the register [20], primarily due to the bloodglucose criteria used (see [16], 3rd & 4th criterion, p. 2188). Conversely, there may be people diagnosed with diabetes who are not detected in the register, but the extent of this problem is unknown. All three mentioned limitations may give rise to imprecision of incidence and mortality estimates, but it is not possible to disentangle how these potential biases may influence the direction of the findings in the current study.

It is however possible to repeat all analyses using a definition of inclusion in the NDR without the blood-glucose criteria. The number of diabetes patients at 1 January 2012 will then shrink from 313,633 to 248,393, but more interestingly, the annual increase in incidence will be 3.6% (and not 4%) and the annual decrease in mortality for DM patients will be 3.7% (and not 4%), so lesser changes in both.

Despite this, the fraction of cases attributable to declining mortality will change from 7.6%(M) and 7.5%(W) to 7.9%(M) and 8.4%(W), and the fractions attributable to incidence changes from 29.3%(M) and 30.6%(W) to 24.8%(M) and 22.9%(W). Thus it seems that a diabetes definition with smaller sensitivity and larger specificity of diabetes diagnosis will attribute a larger fraction of prevalence change to changes in mortality and less to the changes in diabetes incidence. It is likely that the modified definition of DM include persons with more severe disease, and thus that it is for the more severe cases of DM that changes in mortality have had the largest impact on prevalence. This is however not possible to analyze in further detail because more detailed information on severity of DM is not available for all patients.

A full account of both sets of calculations with further details is available as: `http://bendixcarstensen.com/DMreg/Prevalence/Xprev.pdf`

Conclusion

This demographic study of diabetes prevalence in Denmark shows that the increasing prevalence of type 2 diabetes is substantially influenced both by the decline in mortality affecting primarily the oldest part of the population, as well as the increase in incidence with effects across the age-spectrum. The major drivers of the prevalence increase have been the increase in diabetes incidence as well as the imbalance between incidence and mortality already present in 1995.

The prevalence components have differential consequences for the health care system, and deserve a more detailed examination, in particular when prediction of future prevalences is undertaken.

Key messages

- The major driver of changes in prevalence of diabetes in Denmark is the increase in incidence rates of diabetes.

- The decrease in mortality rates over the period 1995–2011 contributes less than 10% of the prevalent cases at 1 January 2012, however still represents a success in treatment of diabetes.
- A formal way of partitioning contributions to prevalence change based on knowledge of changes in incidence and mortality rates is presented.

Acknowledgments

This work was initiated by a bright idea by the former head of the Epidemiology group at Steno Diabetes Center, Knut Borch-Johnsen who literally commissioned a proper analysis of the relative effect of incidence and mortality changes on the trends in prevalence. We are grateful for the visionary prompting of our demographic fidgeting of concepts.

Disclosures

Both authors are employees of Steno Diabetes Center, a clinic and research institution owned by NovoNordisk; both authors own shares in NovoNordisk. MEJ is PI on trial sponsored by AstraZeneca.

References

[1] JMM Evans, KN Barnett, SA Ogston, and AD Morris. Increasing prevalence of type 2 diabetes in a scottish population: effect of increasing incidence or decreasing mortality? *Diabetologia*, 50(4):729–732, 2007.

[2] *IDF diabetes atlas. 6th edn.* IDF, Brussels, `www.idf.org/diabetesatlas`, 2013.

[3] T. Tamayo, J. Rosenbauer, S. H. Wild, A. M. Spijkerman, C. Baan, N. G. Forouhi, C. Herder, and W. Rathmann. Diabetes in Europe: an update. *Diabetes Res. Clin. Pract.*, 103(2):206–217, Feb 2014.

[4] L. S. Geiss, J. Wang, Y. J. Cheng, T. J. Thompson, L. Barker, Y. Li, A. L. Albright, and E. W. Gregg. Prevalence and incidence trends for diagnosed diabetes among adults aged 20 to 79 years, United States, 1980-2012. *JAMA*, 312(12):1218–1226, Sep 2014.

[5] `http://www.diabetesinscotland.org.uk/Publications/SDS2013.pdf`. Technical report.

[6] M. E. Jørgensen, T. P. Almdal, and B. Carstensen. Time trends in mortality rates in type 1 diabetes from 2002 to 2011. *Diabetologia*, 56(11):2401–2404, Nov 2013.

[7] M. Lind, A. M. Svensson, M. Kosiborod, S. Gudbjornsdottir, A. Pivodic, H. Wedel, S. Dahlqvist, M. Clements, and A. Rosengren. Glycemic control and excess mortality in type 1 diabetes. *N. Engl. J. Med.*, 371(21):1972–1982, Nov 2014.

[8] S. J. Livingstone, D. Levin, H. C. Looker, R. S. Lindsay, S. H. Wild, N. Joss, G. Leese, P. Leslie, R. J. McCrimmon, W. Metcalfe, J. A. McKnight, A. D. Morris, D. W. Pearson, J. R. Petrie, S. Philip, N. A. Sattar, J. P. Traynor, and H. M. Colhoun. Estimated life expectancy in a Scottish cohort with type 1 diabetes, 2008-2010. *JAMA*, 313(1):37–44, Jan 2015.

[9] M. Lind, L. A. Garcia-Rodriguez, G. L. Booth, L. Cea-Soriano, B. R. Shah, G. Ekeroth, and L. L. Lipscombe. Mortality trends in patients with and without diabetes in Ontario, Canada and the UK from 1996 to 2009: a population-based study. *Diabetologia*, 56(12):2601–2608, Dec 2013.

[10] K. Færch, B. Carstensen, T. P. Almdal, and M. E. Jørgensen. Improved survival among patients with complicated type 2 diabetes in Denmark: A prospective study (2002-2010). *J. Clin. Endocrinol. Metab.*, page jc20133210, Jan 2014.

[11] J. J. Walker, S. J. Livingstone, H. M. Colhoun, R. S. Lindsay, J. A. McKnight, A. D. Morris, J. R. Petrie, S. Philip, N. Sattar, and S. H. Wild. Effect of socioeconomic status on mortality among people with type 2 diabetes: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetes Care*, 34(5):1127–1132, May 2011.

[12] 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.

[13] A Green, H Støvring, M Andersen, and H Beck-Nielsen. The epidemic of type 2 diabetes is a statistical artefact. *Diabetologia*, 48(8):1456–1458, Aug 2005.

[14] S Colagiuri, K Borch-Johnsen, C Glümer, and D Vistisen. There really is an epidemic of type 2 diabetes. *Diabetologia*, 2005.

[15] 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.

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

[17] Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). available at `www.mortality.org` or `www.humanmortality.de` (data downloaded 28 november 2014). Technical report.

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

[19] T. Karpati, C. J. Cohen-Stavi, M. Leibowitz, M. Hoshen, B. S. Feldman, and R. D. Balicer. Towards a subsiding diabetes epidemic: trends from a large population-based study in Israel. *Popul Health Metr*, 12(1):32, 2014.

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

Chapter 1

Introduction and framework

1.1 Problem

This report addresses the problem of partitioning the changes in prevalence of diabetes in the Danish population over the period 1.1.1995–31.12.2011 in components that are attributable to:

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

This measure will naturally be sex- and age-specific, and hence independent of the demographic changes in the population. But of course the age- and sex-specific components of prevalence can be multiplied by the corresponding population size to yield *numbers* of persons attributable to different contributors.

1.2 Formalization

First we formalize the problem conceptually, then statistically, and finally we 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 prevalences under different scenarios

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.

1.3 Statistical framework

First we consider the setup as outlined in figure 1.1:

```
> library( splines )
> 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
> bbb <-
+ 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[N][D]),
+                               expression(mu[D][M]) ) )


> bbb$Boxes$col.border[1:2] <- "transparent"
> bbb$Boxes$col.bg[1:2] <- c("forestgreen","red")
> bbb$Boxes$col.txt[1:2] <- "white"
> bbb$Boxes$ht <- bbb$Boxes$ht * 1.5
> boxes( bbb )
```

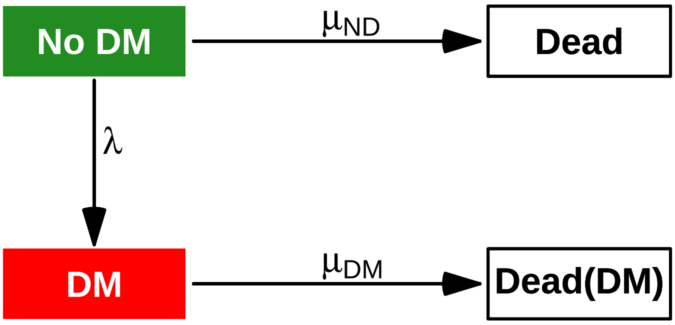

Figure 1.1: *The four states and transitions between them we consider*

```
> boxes( bbb )
> rect( 5, 5, 35, 95, lwd=3, col="transparent", border="blue" )
```

```
> bbb$Boxes[2,"col.bg"] <- grey(0.7)
> boxes(bbb)
> rect( 5, 5, 35, 95, lwd=3, col="transparent", border="blue" )
```

```
> tm <- matrix(NA,3,3)
> rownames(tm) <- colnames(tm) <- c("No DM","DM","Dead")
> tm[1,2] <- tm[1,3] <- tm[2,3] <- 1
> boxes.Lexis( tm, boxpos = list( x=c(20,20,80),
+                                 y=c(80,20,50) ),
+                   wmult=1.3, hmult=4,
+                   txt.arr = c( expression(lambda),
+                                expression(mu[N][D]),
+                                expression(mu[D][M]) ) )
```

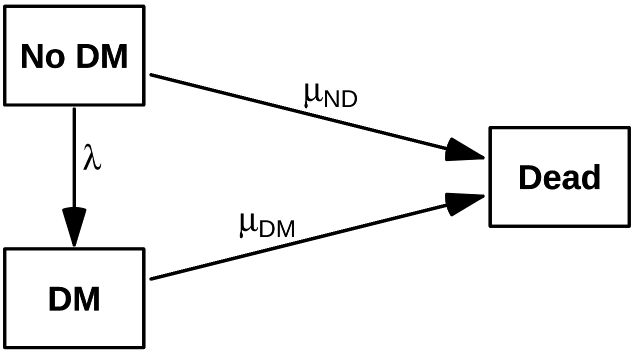

Figure 1.2: *The three states and transitions between them we consider*

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

We can derive analytic 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 an $A \times 2$ matrix of number of persons in each of the states “No DM” and “DM”.

For the given transition rates we can compute transition probabilities between states corresponding to a given (small) interval, of length ℓ, say. For convenience we first define the cumulative intensities for intervals of this length:

$$\Lambda(a,p) = \lambda(a,p) \times \ell, \qquad M_{\text{ND}}(a,p) = \mu_{\text{ND}}(a,p) \times \ell, \qquad M_{\text{DM}}(a,p) = \mu_{\text{DM}}(a,p) \times \ell$$

The probability of remaining in the “No DM” state for a (small) length of ℓ is — omitting the (a,p) dependence and assuming the the rates are constant over the small interval:

$$\text{P}\left\{\text{No DM at } (a+\ell, p+\ell) \mid \text{No DM at } (a,p)\right\} = \text{P}_{\text{No DM,No DM}}(\ell) = \exp\bigl(-(\lambda + \mu_{\text{ND}})\ell\bigr)$$

and the probabilities of going to “Dead” resp. “DM” are:

$$\begin{aligned}
\mathrm{P}_{\text{No DM,Dead}}(\ell) &= \frac{\mu_{\text{ND}}}{\lambda+\mu_{\text{ND}}}\Big(1-\exp\big(-(\lambda+\mu_{\text{ND}})\ell\big)\Big)\\
\mathrm{P}_{\text{No DM,DM}}(\ell) &= \frac{\lambda}{\lambda+\mu_{\text{ND}}}\Big(1-\exp\big(-(\lambda+\mu_{\text{ND}})\ell\big)\Big)
\end{aligned}$$

but since

$$1-\exp\big(-(\lambda+\mu_{\text{ND}})\ell\big) \approx (\lambda+\mu_{\text{ND}})\ell$$

we have the approximation for the transition probabilities:

$$\begin{aligned}
\mathrm{P}_{\text{No DM,Dead}}(\ell) &\approx \mu_{\text{ND}}\ell = M_{\text{ND}}\\
\mathrm{P}_{\text{DM,Dead}}(\ell) &\approx \mu_{\text{DM}}\ell = M_{\text{DM}}\\
\mathrm{P}_{\text{No DM,DM}}(\ell) &\approx \lambda\ell = \Lambda
\end{aligned}$$

So by this approximation we have the transition matrix $\mathbf{T}_{a,p}(\ell)$:

$$\mathbf{T}_{a,p}(\ell) = \begin{pmatrix} \mathrm{e}^{-\Lambda-M_{\text{ND}}} & \Lambda & M_{\text{ND}}\\ 0 & \mathrm{e}^{-M_{\text{DM}}} & M_{\text{DM}}\\ 0 & 0 & 1 \end{pmatrix}$$

So we see that the rates only enter via the cumulative rates over the intervals, so this is what we eventually must compute from the models.

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:

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

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 1.3:

```
> for( yy in 2000+0:3 )
+ for( aa in 40+0:3 )
+ {
+ pdf( paste("./graph/DMpr-",yy,"-",aa,"i.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),
+           aa-1+c(1,1+w), col="forestgreen", lwd=9, lend="butt", ljoin="bevel" )
+ lines( yy+c(1,1),
+           aa-1+c(1+w,2), col="red", lwd=9, lend="butt", ljoin="bevel" )
+ dev.off()
+ 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,2,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(NA,NA,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()
+ pdf( paste("./graph/DMpr-",yy,"-",aa,"-dot.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,2,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(NA,NA,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 )
+ points( yy+1.5, aa+1, pch=3, lwd=2, cex=2, col="blue")
+ dev.off()
+ }
```

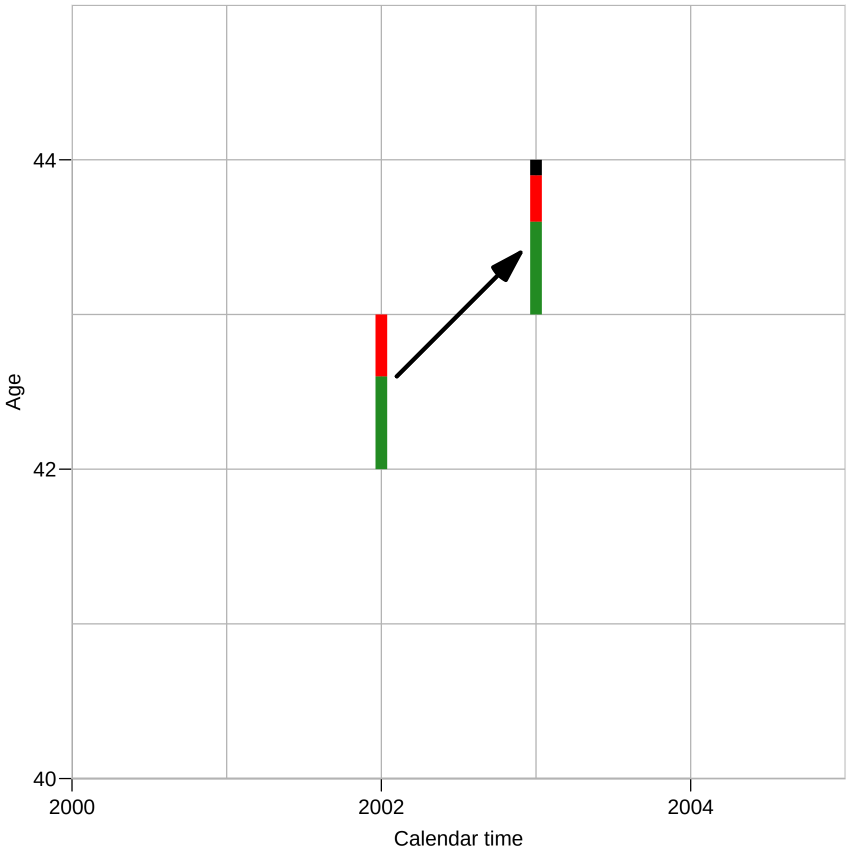

Figure 1.3: *Updating of* number *of persons in different states over a period of length $\ell = 1$ year.*

Now, if we instead have the *fraction* of living persons in states (No DM,DM) in the vector $q(a,p)$ (which is now just a 2-vector with sum 1) then the fractions in the *three* states after a period of ℓ is:

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

Here we have the fraction of the persons in age a at time p who at time $p+\ell$ (and hence in age $a+\ell$) who are in states (No DM,DM,Dead). But since we are only interested in the progression of prevalences, we instead use:

$$\begin{aligned} Q(a+\ell, p+\ell) &= q(a,p)\mathbf{T}_{a,p}(\ell)[\texttt{1 : 2, 1 : 2}] \\ q(a+\ell, p+\ell) &= Q(a+\ell, p+\ell) \Big/ \sum_{\text{W,DM}} Q(a+\ell, p+\ell) \end{aligned}$$

so we just update the prevalences at every step.

1.3.1 Births

Note that for every step in the updating we will lose estimates in the initial age-class; in order for the machinery to work we therefore need to feed in the number of births in each age-group with some assumption about the distribution between DM/non-DM. We will assume the distribution to be 0:1, that is we assume that all new-born are free of diabetes — presumably a fair approximation.

Chapter 2

Practical implementation in R

2.1 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 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`

```
> int <- 1/12
> a.pt <- seq(0,100,int)[-1] - int/2
> t.pt <- seq(1995,2012,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 0.7666667 0.8733333
[10] 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, so that the outer intervals are $1/(\texttt{nk} \times \texttt{bd}$.

2.2 Rates from registers

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( "/home/bendix/sdc/DMreg/NDR-demo/2012/data/FU-o.Rda" )
> str(TT)
'data.frame':        6732 obs. of  9 variables:
 $ sex : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
 $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P   : num  1995 1995 1996 1996 1997 ...
 $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
 $ Y.nD: num  17026 17101 16468 17067 16434 ...
 $ Y.DM: num  0 0.13 1.44 1.86 0 ...
 $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...
 $ D.nD: num  137 16 134 23 152 14 132 16 95 17 ...
 $ X   : int  0 2 4 4 0 2 0 0 0 1 ...
 - attr(*, "Variables")= chr [1:9, 1] "Sex" "1-year age class" "1-year period" "Indicator of upper L
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr  "sex" "A" "P" "U" ...
  .. ..$ : chr "Data frame using the original definintion of DM from NDR"
```

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

     Data frame using the original definintion of DM from NDR
sex  "Sex"
A    "1-year age class"
P    "1-year period"
U    "Indicator of upper Lexis triangle"
Y.nD "P-Y among non-diabetics"
Y.DM "P-Y among diabetes patients"
D.DM "Deaths among non-diabetics"
D.nD "Deaths among diabetes patients"
X    "Diabetes diagnoses among non-diabetics"
```

```
> head( TT )

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

```
> summary( TT )

 sex            A            P              U            Y.nD               Y.DM
 M:3366   Min.   : 0   Min.   :1995   Min.   :0.0   Min.   :   46.38   Min.   :   0.0
 F:3366   1st Qu.:24   1st Qu.:1999   1st Qu.:0.0   1st Qu.: 8375.00   1st Qu.:  73.0
          Median :49   Median :2003   Median :0.5   Median :15921.06   Median : 307.6
          Mean   :49   Mean   :2003   Mean   :0.5   Mean   :13145.27   Mean   : 478.3
          3rd Qu.:74   3rd Qu.:2007   3rd Qu.:1.0   3rd Qu.:17938.03   3rd Qu.: 743.1
          Max.   :98   Max.   :2011   Max.   :1.0   Max.   :22937.66   Max.   :3004.6
      D.DM             D.nD             X
 Min.   :  0.00   Min.   : -1.0   Min.   :  0.00
 1st Qu.:  0.00   1st Qu.:  8.0   1st Qu.:  8.00
 Median :  6.00   Median : 55.0   Median : 34.00
 Mean   : 23.53   Mean   :119.7   Mean   : 55.14
 3rd Qu.: 41.00   3rd Qu.:200.0   3rd Qu.: 92.00
 Max.   :137.00   Max.   :576.0   Max.   :352.00
```

```
> TT[TT$D.nD<0,]

    sex  A    P U     Y.nD     Y.DM D.DM D.nD X
393   F 19 2004 0 14080.07 67.76044    1   -1 7
```

When fitting models we need the mean age and period, and we need to remove the negative number of non-DM deaths. We see from the tabulation that we truly have data in Lexis-triangles:

```
> TT <- transform( TT, A = A + (1+U)/3,
+                      P = P + (2-U)/3,
+                   D.nD = pmax(D.nD,0) )
> with( subset( TT, A<5 & P<1999 ),
+       print( table( Age=round(A,2),
+                     Per=round(P,2) ),
+              zero.print="." ) )
      Per
Age    1995.33 1995.67 1996.33 1996.67 1997.33 1997.67 1998.33 1998.67
  0.33       .       2       .       2       .       2       .       2
  0.67       2       .       2       .       2       .       2       .
  1.33       .       2       .       2       .       2       .       2
  1.67       2       .       2       .       2       .       2       .
  2.33       .       2       .       2       .       2       .       2
  2.67       2       .       2       .       2       .       2       .
  3.33       .       2       .       2       .       2       .       2
  3.67       2       .       2       .       2       .       2       .
  4.33       .       2       .       2       .       2       .       2
  4.67       2       .       2       .       2       .       2       .
```

A brief overview of the number of events and PY:

```
> tt <- xtabs( cbind(X,D.nD,D.DM,Y.nD/1000,Y.DM/1000) ~ sex + floor(P),
+              data = TT )
> round( ftable( addmargins(tt,1:2), row.vars=1:2 ), 1 )
                     X     D.nD     D.DM       V4       V5
sex floor(P)
M   1995        7743.0  27434.0   3747.0   2531.0     51.7
    1996        8008.0  26616.0   3697.0   2542.8     55.7
    1997        7916.0  25633.0   3826.0   2550.3     59.9
    1998        8808.0  25022.0   3895.0   2556.1     64.4
    1999        9300.0  24479.0   4238.0   2560.4     69.3
    2000        9608.0  23913.0   4272.0   2564.6     74.5
    2001       10206.0  23915.0   4408.0   2569.0     80.1
    2002       11170.0  23458.0   4790.0   2572.0     86.2
    2003       12364.0  22460.0   4940.0   2573.0     93.1
    2004       12462.0  23055.0   5015.0   2573.0    100.5
    2005       11613.0  21553.0   5167.0   2573.9    107.5
    2006       12090.0  21803.0   5335.0   2577.0    114.1
    2007       12709.0  21415.0   5532.0   2583.6    120.9
    2008       13994.0  21135.0   5628.0   2593.3    128.8
    2009       14290.0  20653.0   6189.0   2600.5    136.9
    2010       14962.0  20425.0   6195.0   2604.3    145.4
    2011       17072.0  19455.0   6382.0   2606.3    155.1
    Sum       194315.0 392424.0  83256.0  43731.1   1644.3
F   1995        7131.0  27919.0   3630.0   2599.1     51.0
    1996        7377.0  26892.0   3450.0   2609.4     54.6
    1997        7522.0  26580.0   3540.0   2615.5     58.6
    1998        8034.0  25617.0   3562.0   2620.4     62.7
    1999        8553.0  26215.0   3820.0   2624.1     67.4
    2000        8872.0  25408.0   3970.0   2627.7     72.1
    2001        9469.0  25471.0   4113.0   2631.6     77.3
    2002       10778.0  25611.0   4282.0   2633.7     83.2
    2003       11348.0  23435.0   4487.0   2633.7     89.8
    2004       11465.0  24606.0   4406.0   2633.2     96.8
    2005       10573.0  23005.0   4712.0   2633.8    103.2
    2006       10911.0  22935.0   4892.0   2636.0    109.1
    2007       11768.0  23079.0   5012.0   2640.5    115.3
    2008       12654.0  22233.0   5019.0   2647.7    122.5
    2009       12349.0  22147.0   5266.0   2654.5    129.8
    2010       12808.0  21444.0   5572.0   2659.7    136.9
```

```
    2011       15302.0  20578.0   5400.0   2662.4    145.3
    Sum       176914.0 413175.0  75133.0  44762.9   1575.6
Sum 1995       14874.0  55353.0   7377.0   5130.1    102.7
    1996       15385.0  53508.0   7147.0   5152.1    110.3
    1997       15438.0  52213.0   7366.0   5165.8    118.6
    1998       16842.0  50639.0   7457.0   5176.4    127.1
    1999       17853.0  50694.0   8058.0   5184.5    136.7
    2000       18480.0  49321.0   8242.0   5192.3    146.6
    2001       19675.0  49386.0   8521.0   5200.6    157.4
    2002       21948.0  49069.0   9072.0   5205.7    169.4
    2003       23712.0  45895.0   9427.0   5206.8    182.9
    2004       23927.0  47661.0   9421.0   5206.2    197.3
    2005       22186.0  44558.0   9879.0   5207.7    210.7
    2006       23001.0  44738.0  10227.0   5213.0    223.1
    2007       24477.0  44494.0  10544.0   5224.1    236.2
    2008       26648.0  43368.0  10647.0   5241.0    251.4
    2009       26639.0  42800.0  11455.0   5255.0    266.8
    2010       27770.0  41869.0  11767.0   5263.9    282.4
    2011       32374.0  40033.0  11782.0   5268.7    300.4
    Sum       371229.0 805599.0 158389.0  88494.0   3219.9
```

```
> tt <- xtabs( cbind(X,D.nD,D.DM,Y.nD/1000,Y.DM/1000) ~ sex + gP,
+              data = transform( TT,
+                                gP = factor( (P>2000)+(P>2006),
+                                                 labels=c("1995-2000",
+                                                          "2001-2006",
+                                                          "2006-2011") ) ) )
> round( ftable( addmargins(tt,1:2), row.vars=1:2 ), 1 )
```

```
                      X     D.nD     D.DM       V4       V5
sex gP
M   1995-2000   41775.0 129184.0  19403.0  12740.5    301.0
    2001-2006   67423.0 138354.0  28592.0  15425.5    542.0
    2006-2011   85117.0 124886.0  35261.0  15565.0    801.3
    Sum        194315.0 392424.0  83256.0  43731.1   1644.3
F   1995-2000   38617.0 133223.0  18002.0  13068.4    294.4
    2001-2006   62505.0 147536.0  25970.0  15793.7    522.3
    2006-2011   75792.0 132416.0  31161.0  15900.7    758.9
    Sum        176914.0 413175.0  75133.0  44762.9   1575.6
Sum 1995-2000   80392.0 262407.0  37405.0  25809.0    595.4
    2001-2006  129928.0 285890.0  54562.0  31219.2   1064.3
    2006-2011  160909.0 257302.0  66422.0  31465.7   1560.3
    Sum        371229.0 805599.0 158389.0  88494.0   3219.9
```

Then 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("apc","ap") ) )
> str( Lambda )
```

```
 logi [1:1200, 1:204, 1:2, 1:2] 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 [1:204] "1995.04166666667" "1995.125" "1995.20833333333" "1995.29166666667" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:2] "apc" "ap"
```

```
> prod( dim(Lambda) )
```

```
[1] 979200
```

2.3 Incidence rates

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

```
> ( ki.a <- with( TT, quantile( rep(  A,X), qn(nk.a) ) ) )
      5%      15%      25%      35%      45%      55%      65%      75%      85%      95% 
31.66667 45.66667 52.33333 56.66667 60.66667 64.66667 68.33333 72.66667 77.66667 84.33333 
> ( ki.p <- with( TT, quantile( rep(P  ,X), qn(nk.p) ) ) )
     10%      30%      50%      70%      90% 
1997.333 2001.667 2004.667 2008.333 2010.667 
> ( ki.c <- with( TT, quantile( rep(P-A,X), qn(nk.c) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75% 
1919.667 1928.333 1934.667 1939.667 1944.667 1949.667 1956.667 1970.333 
```

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. However we want to extract the avearge trend from the APC-model, so we also fit the model with the parametrization that allows us to extract the trend by using the information (in this case `X`) as weights:

```
> m.inc.ap <- glm( X ~ Ns(A,knots=ki.a) + Ns(P,knots=ki.p),
+                      offset = log(Y.nD),
+                      family = poisson,
+                        data = subset(TT,sex=="M") )
> m.inc.aPC <- update( m.inc.ap, . ~ . - Ns(P  ,kn=ki.p) + I(P) +
+                               detrend( Ns(P  ,kn=ki.p), P  , X ) +
+                               detrend( Ns(P-A,kn=ki.c), P-A, X ) )
> m.inc.apc <- update( m.inc.ap, . ~ . + Ns(P-A,kn=ki.c) )
> f.inc.ap  <- update( m.inc.ap , data = subset(TT,sex=="F") )
> f.inc.apc <- update( m.inc.apc, data = subset(TT,sex=="F") )
> f.inc.aPC <- update( m.inc.aPC, data = subset(TT,sex=="F") )
```

The average annual trends in incidence:

```
> 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       3.9  3.8   4.0
                      Women       4.0  3.9   4.1
```

2.3.1 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.nD = int, Y.DM = int )
> head( nd )
```

```
           A        P       Y.nD       Y.DM
1 0.04166667 1995.042 0.08333333 0.08333333
2 0.12500000 1995.042 0.08333333 0.08333333
3 0.20833333 1995.042 0.08333333 0.08333333
4 0.29166667 1995.042 0.08333333 0.08333333
5 0.37500000 1995.042 0.08333333 0.08333333
6 0.45833333 1995.042 0.08333333 0.08333333
> dim( nd )
[1] 244800      4
```

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.

2.4 Mortality rates

We have the mortality-rates in `TT` too, by using `D.nD` and `Y.nD` for mortality in the population without diabetes and `D.DM` and `Y.DM` for the mortality in persons with diabetes.

2.4.1 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( TT, quantile( rep(A  ,D.DM), qn(nk.a) ) ) )
      5%      15%      25%      35%      45%      55%      65%      75%      85%      95%
55.33333 64.33333 69.66667 73.66667 77.33333 79.66667 82.66667 85.33333 88.33333 92.66667
> ( kmd.p <- with( TT, quantile( rep(P  ,D.DM), qn(nk.p) ) ) )
     10%      30%      50%      70%      90%
1997.333 2001.333 2004.667 2007.667 2010.667
> ( kmd.c <- with( TT, quantile( rep(P-A,D.DM), qn(nk.c) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1910.667 1916.667 1920.667 1924.333 1927.667 1932.333 1938.667 1948.333
> m.md.ap <- glm( D.DM ~ Ns(A,knots=kmd.a) + Ns(P,knots=kmd.p),
+                        offset = log(Y.DM),
+                        family = poisson,
+                          data = subset( TT, sex=="M" & Y.DM>0 ) )
> m.md.aPC <- update( m.md.ap, . ~ . - Ns(P  ,kn=kmd.p) + I(P) +
+                             detrend( Ns(P  ,kn=kmd.p), P  , D.DM ) +
+                             detrend( Ns(P-A,kn=kmd.c), P-A, D.DM ) )
> m.md.apc <- update( m.md.ap, . ~ . + Ns(P-A,kn=kmd.c) )
> f.md.ap  <- update( m.md.ap , data = subset( TT, sex=="F" & Y.DM>0 ) )
> f.md.apc <- update( m.md.apc, data = subset( TT, sex=="F" & Y.DM>0 ) )
> f.md.aPC <- update( m.md.aPC, data = subset( TT, sex=="F" & Y.DM>0 ) )
> 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 )
```

2.4.2 Persons without diabetes

The mortality in the population without diabetes is modeled in exactly the same way:

```
> ( kmw.a <- with( TT, quantile( rep(A  ,D.nD), qn(nk.a) ) ) )
      5%      15%      25%      35%      45%      55%      65%      75%      85%      95%
45.66667 60.33333 67.33333 72.66667 76.66667 80.33333 83.33333 86.33333 89.66667 93.66667
```

```
> ( kmw.p <- with( TT, quantile( rep(P  ,D.nD), qn(nk.p) ) ) )
     10%      30%      50%      70%      90%
1996.333 1999.667 2002.667 2006.333 2010.333
```

```
> ( kmw.c <- with( TT, quantile( rep(P-A,D.nD), qn(nk.c) ) ) )
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75%
1908.333 1914.333 1918.667 1922.333 1926.667 1932.333 1940.667 1955.333
```

```
> m.mw.ap <- glm( D.nD ~ Ns(A,knots=kmw.a) + Ns(P,knots=kmw.p),
+                   offset = log(Y.nD),
+                   family = poisson,
+                     data = subset( TT, sex=="M" & Y.nD>0 ) )
> m.mw.aPC <- update( m.mw.ap, . ~ . - Ns(P  ,kn=kmw.p) + I(P) +
+                        detrend( Ns(P  ,kn=kmw.p), P  , D.nD ) +
+                        detrend( Ns(P-A,kn=kmw.c), P-A, D.nD ) )
> m.mw.apc <- update( m.mw.ap, . ~ . + Ns(P-A,kn=kmw.c) )
> f.mw.ap  <- update( m.mw.ap , data = subset( TT, sex=="F" & Y.nD>0 ) )
> f.mw.apc <- update( m.mw.apc, data = subset( TT, sex=="F" & Y.nD>0 ) )
> f.mw.aPC <- update( m.mw.aPC, data = subset( TT, sex=="F" & Y.nD>0 ) )
> 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 )
```

2.5 Secular 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       3.9  3.8   4.0
                      Women       4.0  3.9   4.1
Mortality change,   DM: Men      -4.0 -4.2  -3.9
                      Women      -3.8 -4.0  -3.7
Mortality change, Well: Men      -2.7 -2.8  -2.6
                      Women      -2.3 -2.4  -2.2
```

Thus it appears that the incidence rates of diabetes are increasing by some 4% per year, while mortality rates are decreasing 4% per your for persons with diabetes, but only 2–2.5% per year for perons without.

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="./data/Ests.Rda" )
```

Chapter 3

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( file="./data/Ests.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 empirical age-specific prevalences at 1.1.1995–1.1.2012, 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 1995–2011 or assuming they were constant at the 1995 level. These refer to intervals of length `int` and are therefore labeled on the period dimension by the midpoint of these, a total of 17/`int`.

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

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

3.1 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 [1:1200, 1:204, 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 [1:204] "1995.04166666667" "1995.125" "1995.20833333333" "1995.29166666667" ...
  ..$ sex  : chr [1:2] "M" "F"
  ..$ mod  : 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 0.083 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.

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) for either incidence, mortality or both:

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

```
  p
204
```

```
> 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 15 mil. numbers:

```
> str( TR )
 num [1:1200, 1:204, 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" "0.291666666666667" ...
  ..$ p    : chr [1:204] "1995.04166666667" "1995.125" "1995.20833333333" "1995.29166666667" ...
  ..$ sex  : chr [1:2] "M" "F"
  ..$ mod  : 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] 15667200
```

3.2 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( "/home/bendix/sdc/DMreg/NDR-demo/2012/data/prev-o.Rda" )
> summary( pr )
 sex           A               P              X                N        
 M:1800   Min.   : 0.00   Min.   :1995   Min.   :   0.0   Min.   :   77  
 F:1800   1st Qu.:24.75   1st Qu.:1999   1st Qu.: 139.0   1st Qu.:18405  
          Median :49.50   Median :2004   Median : 593.0   Median :32166  
          Mean   :49.50   Mean   :2004   Mean   : 950.7   Mean   :26980  
          3rd Qu.:74.25   3rd Qu.:2008   3rd Qu.:1463.2   3rd Qu.:36725  
          Max.   :99.00   Max.   :2012   Max.   :5982.0   Max.   :46208  
```

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

These are empirical prevalences (`X`—no. of cases of DM, `N`—population size) for each of the 18 dates 1.1.1995 – 1.1.2010 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 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 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%       50% 56.66667%
       10        28        40        47        52        56        59        62        64        67
63.33333%       70% 76.66667% 83.33333%       90% 96.66667%
       69        72        75        78        82        87
```

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

```
 logi [1:1200, 1:2, 1:18] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ a  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ t  : chr [1:18] "1995" "1996" "1997" "1998" ...
```

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

```
[1] 43200
```

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.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=as.numeric(dimnames(pr.fit)[["a"]]) ),
+                            type = "response" )
+    }
> round( ftable( pr.fit[c(1:2,NA,floor(dim(pr.fit)[1]*3/4)+1:5),,],
+                row.vars=2:1 )*100, 1 )
```

```
                       t 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
sex a
M   0.0416666666666667    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
    0.125                 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
    75.041666666667       7.2  7.6  7.9  8.4  8.9  9.6 10.3 11.2 12.0 12.8 13.7 14.5 15.5 16.7 18.0
    75.125                7.2  7.6  8.0  8.4  9.0  9.6 10.3 11.2 12.0 12.8 13.7 14.5 15.5 16.7 18.0
    75.2083333333333      7.2  7.7  8.0  8.4  9.0  9.6 10.3 11.2 12.0 12.8 13.7 14.5 15.5 16.7 18.0
    75.291666666667       7.2  7.7  8.0  8.4  9.0  9.6 10.3 11.2 12.0 12.8 13.7 14.5 15.5 16.7 18.0
    75.375                7.2  7.7  8.0  8.4  9.0  9.6 10.4 11.2 12.0 12.9 13.7 14.5 15.5 16.7 18.0
F   0.0416666666666667    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
    0.125                 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
    NA                     NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
    75.041666666667       6.5  6.9  7.3  7.6  8.0  8.3  8.8  9.5 10.3 11.2 12.1 12.8 13.5 14.3 15.1
    75.125                6.5  6.9  7.3  7.7  8.0  8.3  8.8  9.5 10.3 11.2 12.1 12.8 13.5 14.3 15.1
    75.2083333333333      6.5  6.9  7.3  7.7  8.0  8.4  8.8  9.5 10.3 11.2 12.1 12.8 13.5 14.3 15.1
    75.291666666667       6.5  6.9  7.3  7.7  8.0  8.4  8.8  9.5 10.4 11.2 12.1 12.8 13.6 14.3 15.2
    75.375                6.5  6.9  7.3  7.7  8.0  8.4  8.8  9.5 10.4 11.2 12.1 12.8 13.6 14.3 15.2
```

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(0,22), 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)[[1]] )[1]
+ a80 <- grep( "80.", dimnames(pr.fit)[[1]] )[1]
+ text( 89, pr.fit[a89,"M","1995"]* 99, "1995", col="blue", adj=c(1,1) )
+ text( 80, pr.fit[a80,"M","2012"]*101, "2012", col="blue", adj=c(0,0) )
+ axis( side=1 )
+ axis( side=2 )
```

```
+ axis( side=2, at=0:22, labels=NA )
+ matplot( a.pt, pr.fit[,"F",]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", lty=1, col="red", lwd=c(1,2) )
+ 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[a89,"F","1995"]* 99, "1995", col="red", adj=c(1,1) )
+ text( 80, pr.fit[a80,"F","2012"]*101, "2012", col="red", adj=c(1,0) )
+ axis( side=1 )
+ mtext( "Age", side=1, line=1, outer=T )
+ mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
+ }
> plp()
```

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

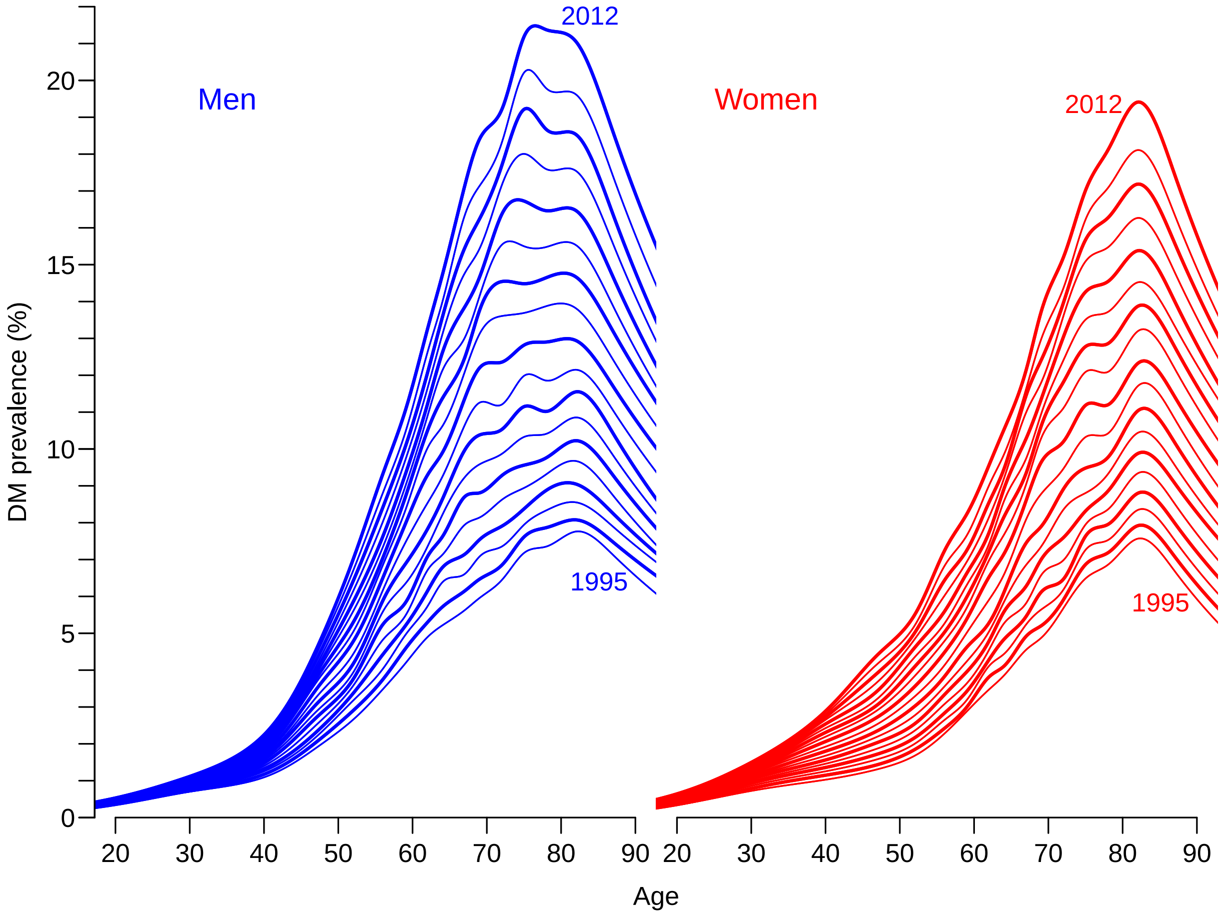

Figure 3.1: *Smoothed age-specific prevalences for the years 1995–2010. Blue is men, red is women.*

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

3.3 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 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 [1:1200, 1:205, 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" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

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

```
[1] 6888000
```

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-born) 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,floor(dim(prv)[1]/1.5)+1:3),1:2,,,],
+                col.vars=4:5 )*100, 1 )
```

```
                                        mod  apc                                   ap
                                        what obs m-fix i-fix all-f mort inc const obs m-fix i-fix al
a                  t                sex
0.0416666666666667 1995             M        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
                   1995.08333333333 M        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
0.125              1995             M        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
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
                                    F         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
66.7083333333333   1995             M        5.6   5.6   5.6   5.6  5.6 5.6   5.6 5.6   5.6   5.6
                                    F        4.5   4.5   4.5   4.5  4.5 4.5   4.5 4.5   4.5   4.5
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
                                    F         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
66.7916666666667   1995             M        5.6   5.6   5.6   5.6  5.6 5.6   5.6 5.6   5.6   5.6
                                    F        4.5   4.5   4.5   4.5  4.5 4.5   4.5 4.5   4.5   4.5
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
```

```
                                    F         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
66.875             1995             M        5.6   5.6   5.6   5.6  5.6 5.6   5.6 5.6   5.6   5.6
                                    F        4.5   4.5   4.5   4.5  4.5 4.5   4.5 4.5   4.5   4.5
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
                                    F         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,,,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
  8.944   0.000   8.940
```

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.

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

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

                    t
a                      1995 1995.08333333333
  0.0416666666666667 0.000            0.000
  0.125              0.041            0.001
  0.208333333333333  0.042            0.042
  0.291666666666667  0.042            0.042

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

3.3.1 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 1.1.2012.

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:

```
> nd <- c( grep("2000",dimnames(prv)[[2]])[1],
+          grep("2006",dimnames(prv)[[2]])[1],
+          grep("2012",dimnames(prv)[[2]])[1] )
> ( wh <- dimnames(prv)[[2]][nd] )
[1] "2000" "2006" "2012"
```

```
> 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(0,22), 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(0,22), 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(0,22), yaxs="i",
+          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 )
> 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(0,22), yaxs="i",
+          xlab="Age", ylab="", yaxt="n",
+          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 )
> 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 (see figure **??**), we will use these in the reporting of the different scenarios.

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

```
> str( prv )
 num [1:1200, 1:205, 1:2, 1:2, 1:7] 0 0.000411 0.000415 0.000419 0.000423 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

```
> prv[floor(dim(prv)[1]/1.5)+1:5,np,"M","apc",]*100
```

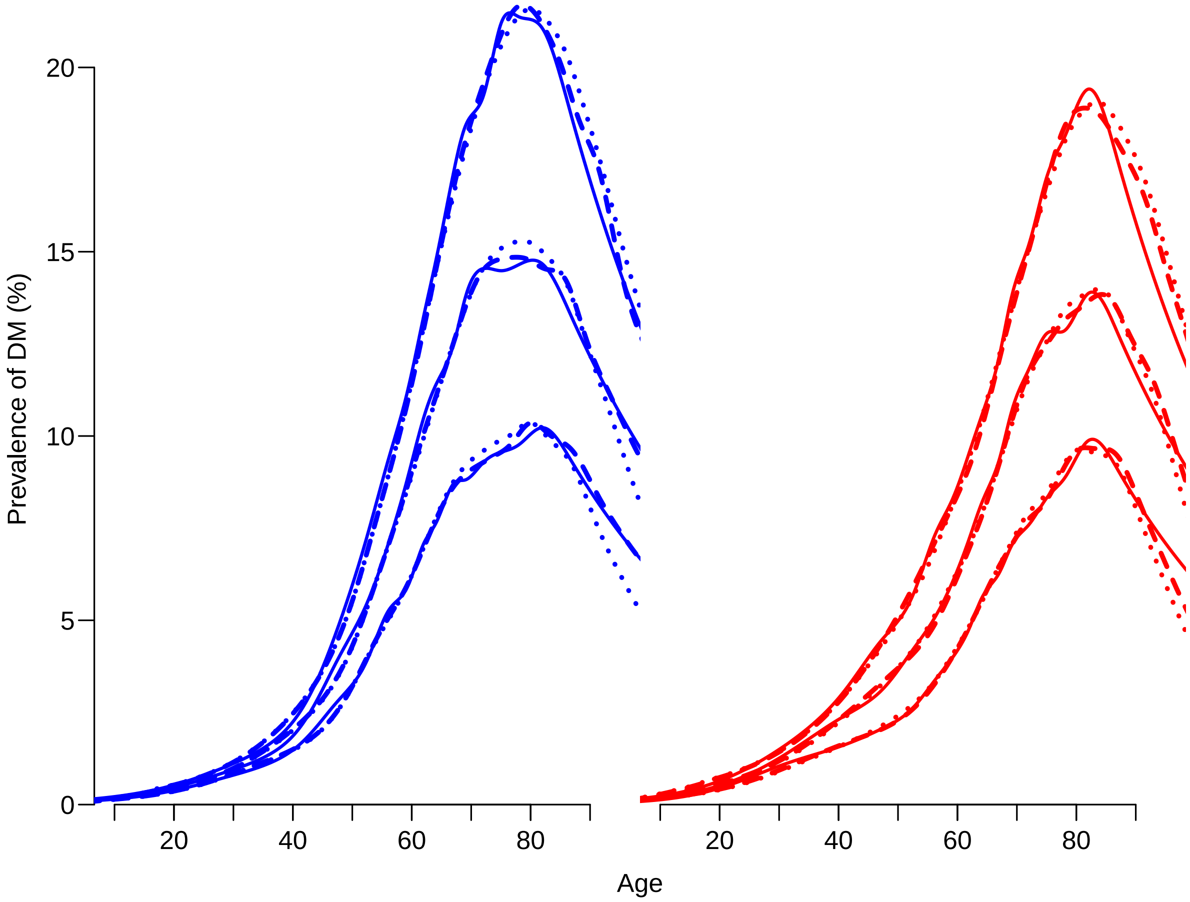

Figure 3.2: *Plot of observed (full lines) and predicted prevalences in 2000, 2006 and 2012, using simple age-period-models (dotted lines) or age-period-cohort models (broken lines). Clearly the broken lines gives a better approximation to the smoothed empirical rates.*

```
                    what
a                         obs    m-fix    i-fix    all-f mort inc const
  66.7083333333333 16.43350 15.08703 11.30686 10.22518   NA  NA    NA
  66.7916666666667 16.49216 15.13425 11.34443 10.25344   NA  NA    NA
  66.875           16.55047 15.18107 11.38187 10.28152   NA  NA    NA
  66.9583333333333 16.60841 15.22747 11.41918 10.30941   NA  NA    NA
  67.0416666666667 16.66599 15.27343 11.45636 10.33711   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" )
> matplot( a.pt, cbind(prv[,np,"M","apc",],prv[,1,"M","apc",1])*100,
+          xlim=c(20,90), ylim=c(0,22), yaxs="i",
+          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" )
> axis( side=2, at=0:22, labels=NA )
> matplot( a.pt, cbind(prv[,np,"F","apc",],prv[,1,"F","apc",1])*100,
+          xlim=c(20,90), ylim=c(0,22), yaxs="i",
+          xlab="Age", ylab="", yaxt="n",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
> matlines( a.pt, prv[,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" )
```

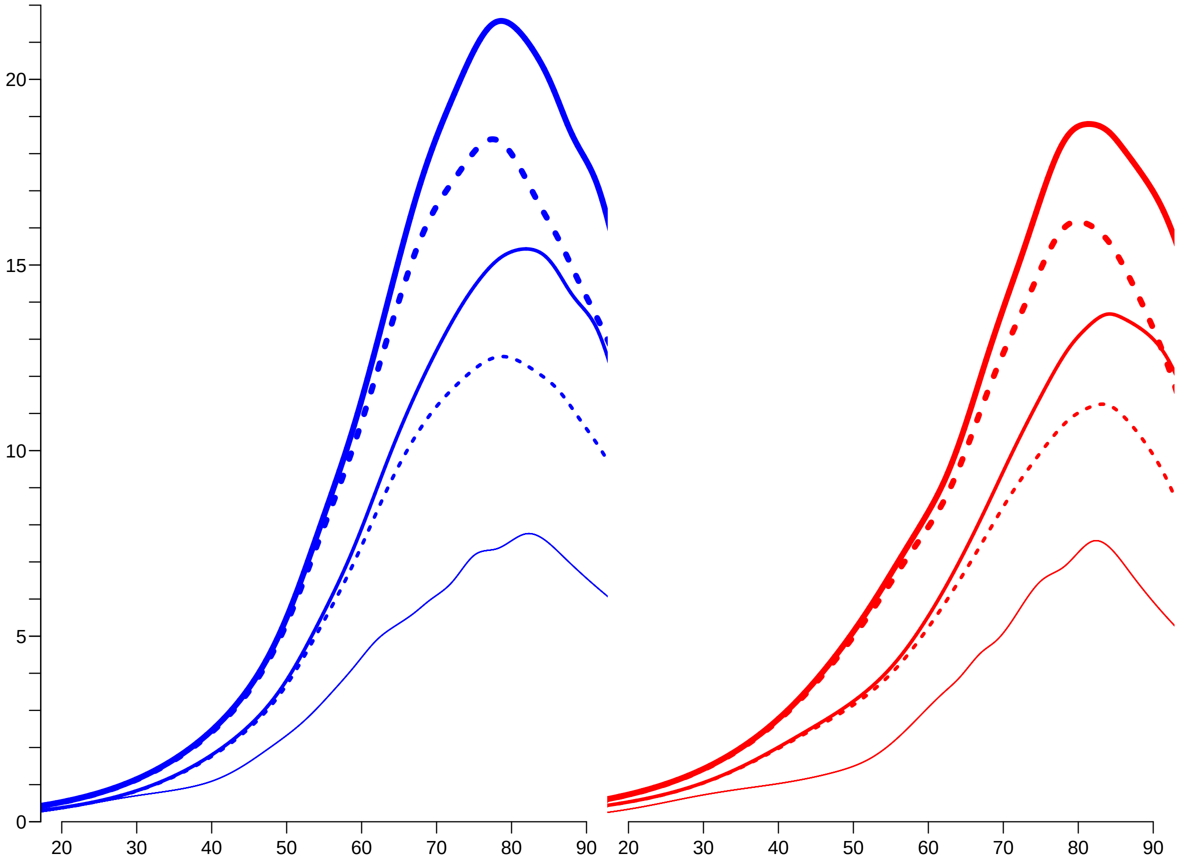

Figure 3.3: *The predicted prevalences under different scenarios:*
Full lines: Mortality rates evolve as observed, Broken lines: Mortality rates remain as 1995.
Thick lines: Incidence rates evolve as observed, Thin lines: Incidence rates remain as in 1995.
The very thin lines lowest in the two displays are the observed prevalences in 1995.

```
> scen <- c("Mort obs, Inc obs",
+           "Mort 1995, Inc obs",
+           "Mort obs, Inc 1995",
+           "Mort 1995, Inc 1995")
> 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( "./graph/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(0,22), 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=0:22, labels=NA )
+ 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(0,22), 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)
```

```
X11cairo
       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.

3.4 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.

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(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","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(0,7), 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 )
> 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(0,7), 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 4.4 we see that the two possible ways of computing the contribution give pretty much the same results — the differences never exceed some 0.5%. 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 1995 (“`all-f`”). This is obtained by taking the average of the two curves in each scenario.

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

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 — the “epidemiological disequilibrium” 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 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.

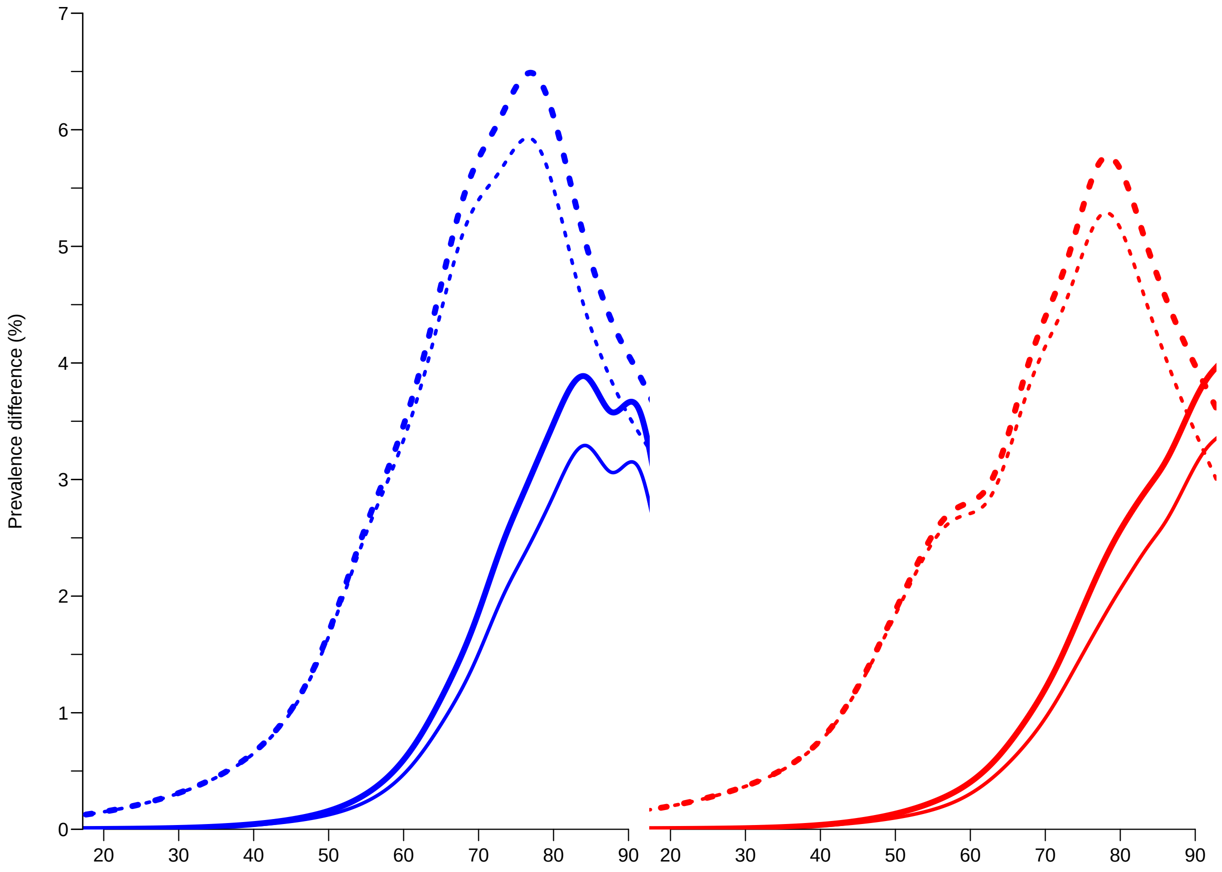

Figure 3.4: *Suggested contributions to age-specific prevalences from increasing mortalities over the period 1995–2012; the thick lines are obtained by subtracting the prediction based on fixing one rate from the one using the observed rates; thin lines based on subtracting the prediction based on fixing both rates from that where one is fixed. Full lines are for differences attributable to changes in mortality rates, broken lines are for changes attributable to changes in incidence rates.*

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(0,22),
+             txt=dimnames(prv)[[2]][yn[j]] )
+ abline(h=0)
+ if( j==1 ) text( rep(25,3), c(13,15,17),
+                  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(0,22) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> axis( side=2 )
> axis( side=2, at=0:22, labels=NA )
> text( rep(25,3), 17:19+2.5,
+       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(0,22) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> # axis( side=2 )
> text( rep(25,3), 17:19+2.5,
+       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 )
```

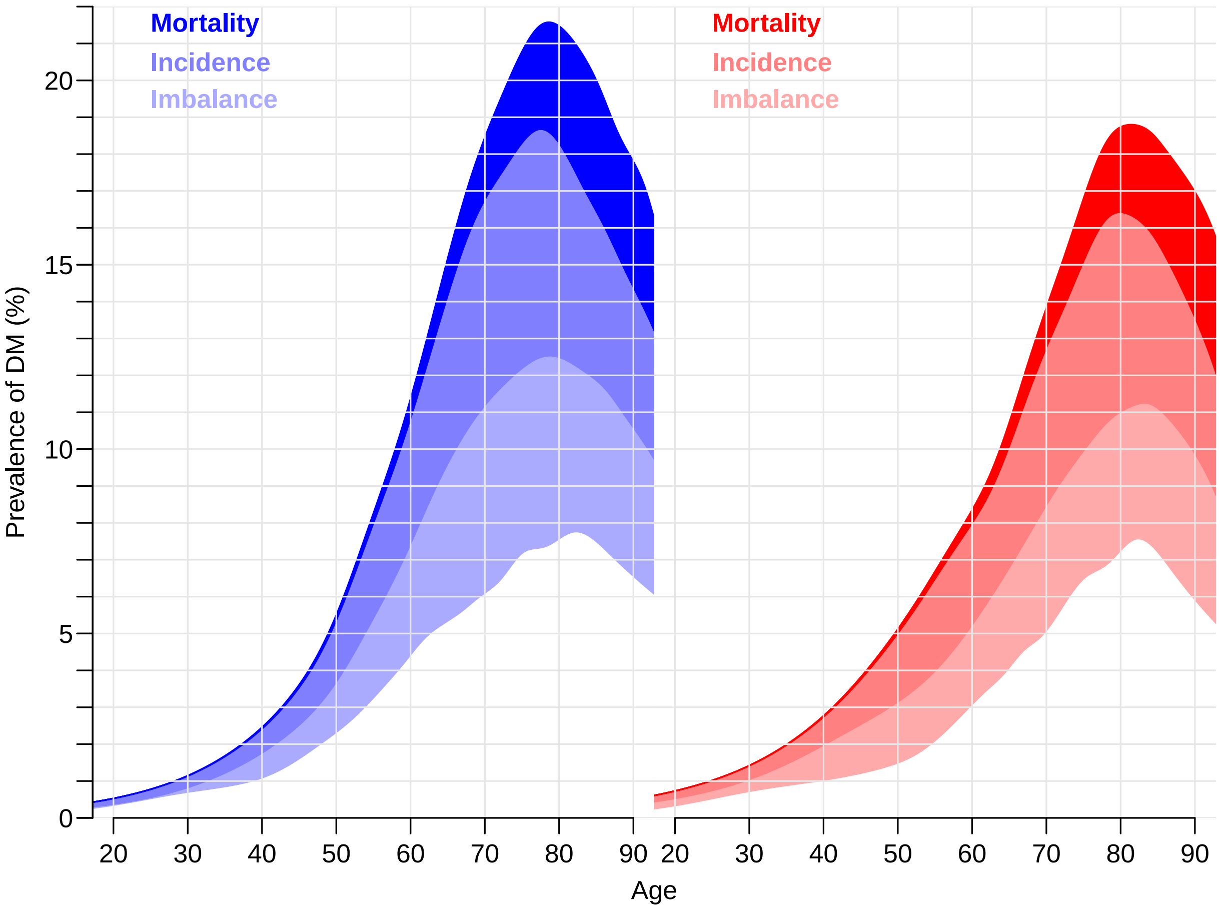

Figure 3.5: *Predicted age-specific prevalences of DM in Denmark 2012 among men (blue) and women (red), partitioned by the contribution from rates as they were in 1995 ("Imbalance"), increases in incidence and decrease in mortality, respectively.*

3.5 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` (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]] <- 1995:2012
> dn[[5]] <- dn[[5]][c(5:7,1)]
> prn <- NArray( dn )
> length(prv) ; str( prv ) ; table( prv<0 )
[1] 6888000

 num [1:1200, 1:205, 1:2, 1:2, 1:7] 0 0.000411 0.000415 0.000419 0.000423 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...

  FALSE    TRUE 
6792623   95377 

> hist( prv, breaks=200, col="black", ylim=c(0,10000) ) ; abline(v=0,col="red")
> length(prn) ; str( prn )
[1] 28800

 logi [1:100, 1:18, 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 [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:4] "mort" "inc" "const" "obs"
```

In order to fill in the numbers we use the estimates age-specific prevalences at 1st January each year, that is at the dates 1.1.1995,...,1.1.2012 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, 1:18, 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 [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ 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
```

```
> 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 )
+    }
> range( prv )
```

```
[1] 0.0000000 0.2169337
```

```
> range( prn )
```

```
[1] 0.00000000 0.07759244
```

Now `prn` contains the prevalences (as fractions) for 100 age classes and 18 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   M 0 1995 3 35612
2   M 0 1996 1 36055
3   M 0 1997 0 34853
4   M 0 1998 1 34774
5   M 0 1999 2 34076
6   M 0 2000 1 33906
```

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

```
 xtabs [1:100, 1:18, 1:2] 35612 34747 35082 33330 32974 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = N ~ A + P + sex, data = pr)
```

```
> str( dmp )

 xtabs [1:100, 1:18, 1:2] 3 4 6 5 12 21 22 34 29 29 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = X ~ A + P + sex, data = pr)

> str( prn )

 num [1:100, 1:18, 1:2, 1:2, 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 [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : 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(=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 <- "2012"
> 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) )
+ 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(1996:2012) ) draw.dmp(pp)
> dev.off()

pdf
  3
```

```
> for( pp in paste(1995:2012) )
+ {
+ 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" "#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) )
+ 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=2 )
+ 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( "./graph/DMpr-film.pdf", width=9, height=6 )
> for( pp in paste(1996:2012) ) draw.pyr(pp)
> dev.off()
```

```
pdf
  3
```

```
> for( pp in paste(1996:2012) )
+ {
+ pdf( paste("./graph/DMpr-", pp, ".pdf", sep=""), width=8, height=6 )
+ draw.pyr(pp)
+ dev.off()
+ }
```

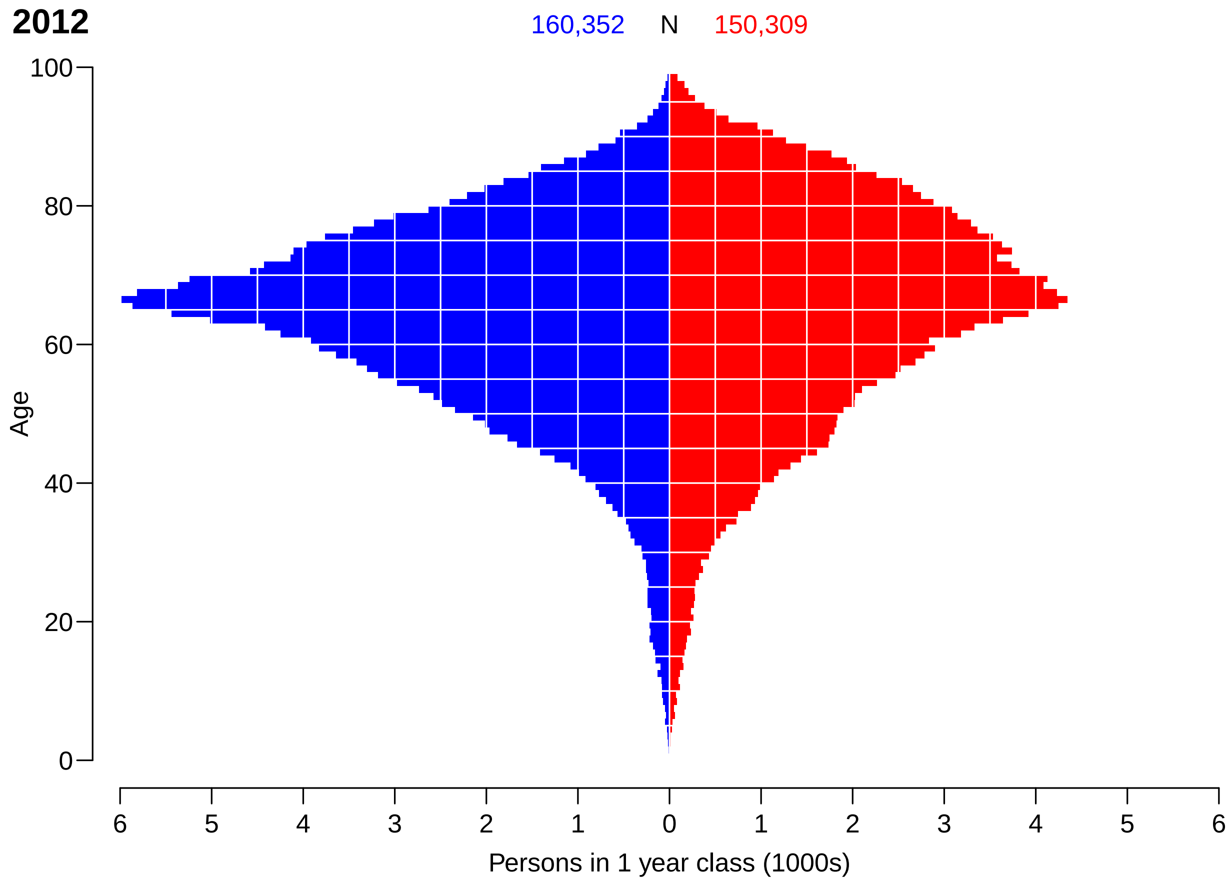

Figure 3.6: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2012.*

3.6 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.

```
> lls()
   name       mode      class       size
1  a.pt       numeric   numeric     1200
2  c.a        character character   1
3  cau.exp    function  function    1
4  clr        character character   8
5  comp       character character   3
6  dmp        numeric   xtabs table 100 18 2
7  dn         list      list        5
8  dpr        list      list        5
9  draw.dmp   function  function    1
```

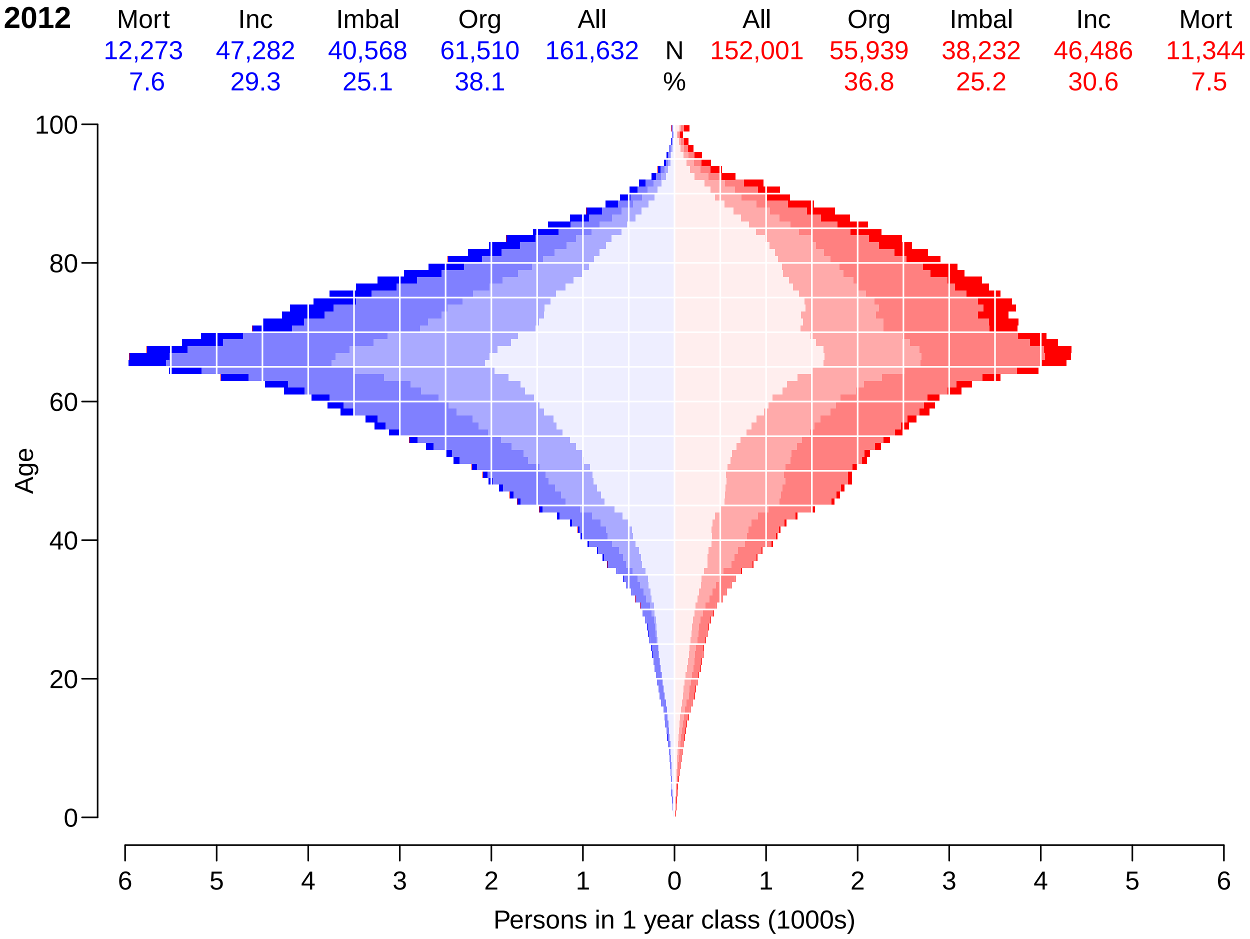

Figure 3.7: *Age-distribution of the predicted no. of diabetes cases in Denmark as of 1.1.2012, subdivided by the components of disease prevalence: Mort: decrease in mortality, Inc: increase in incidence, Imbal: constant rates from 1995 (non-steady-state imbalance between incidence and mortality in 1995, Org: age-specific prevalence in 1995.*

```
10 draw.pyr   function  function    1
11 dt         character character   1
12 een        numeric   numeric     4
13 ff         logical   logical     1
14 hts        numeric   numeric     4
15 i          numeric   integer     1
16 ia         numeric   integer     1
17 int        numeric   numeric     1
18 ip         numeric   integer     1
19 j          character character   1
20 kp.a       numeric   numeric     16
21 Lambda     numeric   array       1200 204 2 2
22 lim        numeric   numeric     1
23 Mu.DM      numeric   array       1200 204 2 2
24 Mu.W       numeric   array       1200 204 2 2
25 n.a        numeric   numeric     1
26 nd         numeric   integer     3
27 np         numeric   integer     1
28 one.comp   function  function    1
29 oo         character character   4
30 plp        function  function    1
31 poly.parts function  function    1
32 pop        numeric   xtabs table 100 18 2
```

```
33 pp         character character   1
34 p.pt       numeric   numeric     204
35 pr         list      data.frame  3600 5
36 pr.fit     numeric   array       1200 2 18
37 pr.mod     list      glm lm      30
38 prn        numeric   array       100 18 2 2 4
39 prn.a      numeric   numeric     100
40 prn.p      numeric   numeric     18
41 prv        numeric   array       1200 205 2 2 7
42 prv.a      numeric   numeric     1200
43 prv.p      numeric   numeric     205
44 qn         function  function    1
45 scen       character character   4
46 shd        numeric   numeric     4
47 states     character character   2
48 sx         character character   1
49 t.pt       numeric   numeric     205
50 TR         numeric   array       1200 204 2 2 2 2 4
51 wh         character character   3
52 wh.a       numeric   integer     12
53 wh.p       numeric   integer     18
```

```
> str( prv )
```

```
 num [1:1200, 1:205, 1:2, 1:2, 1:7] 0 0.000411 0.000415 0.000419 0.000423 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ 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, 1:205, 1:2, 1:4] 0.0443 0.061 0.0757 0.0445 0.0614 ...
 - attr(*, "dimnames")=List of 4
  ..$ a   : chr [1:3] "60.0416666666667" "70.0416666666667" "80.0416666666667"
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ 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(1995,2012), 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+1995, labels=NA ) }
+ 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(2000,2010,5), 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( 1996, 70, paste( if(sx=="F") "W" else sx,"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 )
```

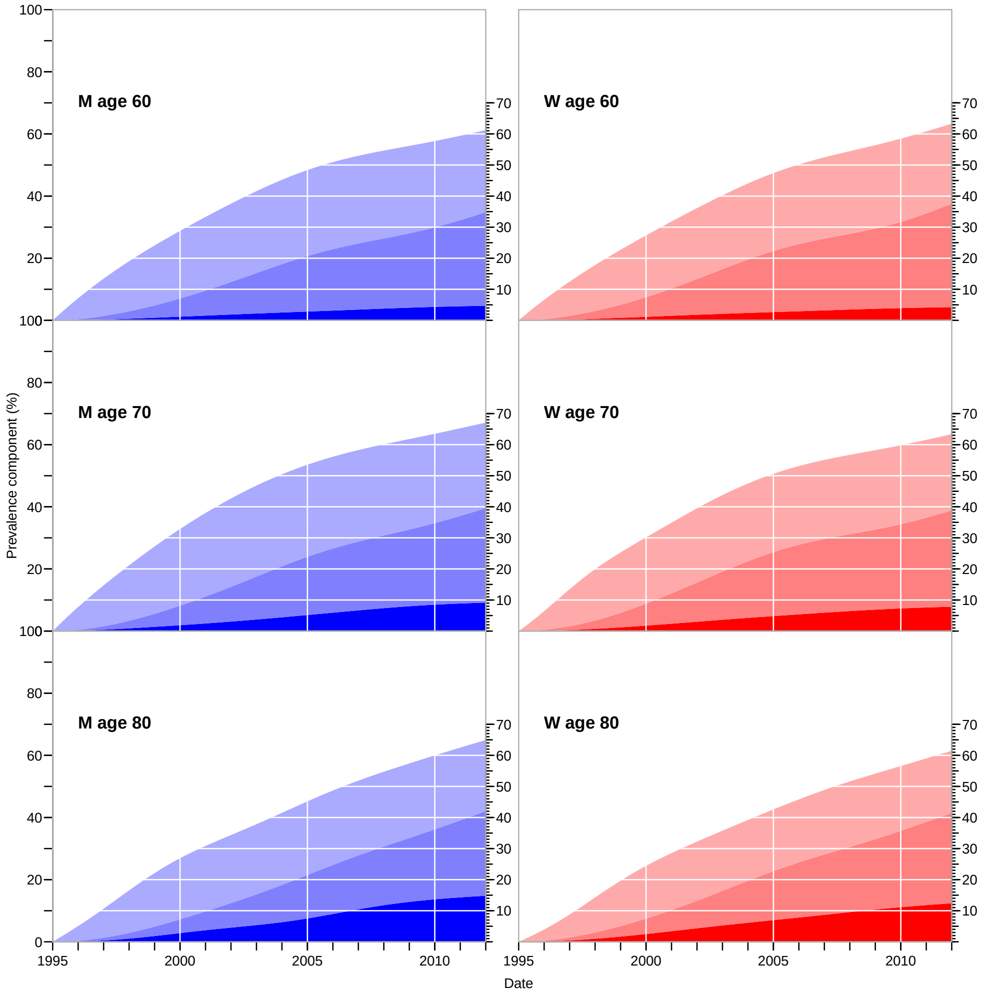

Figure 3.8: *Fraction of the prevalent cases at different times attributable to a) declining mortality (bottom, full color), b) increasing incidence (middle, pale color) and c) prevalence/mortality imbalance 1995 (top, weak color). The white area above the curves corresponds to the fraction of the cases that would have been around if incidence and mortality rates had remained as in 1995.*

Chapter 4

Calculations using modified definition

This chapter contains a repeat of the calculations in the previous two chapters, but building all calculations on a revised version of the diabetes register where the blood glucose criteria are not used as inclusion criterion.

4.1 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 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`

```
> int <- 1/12
> a.pt <- seq(0,100,int)[-1] - int/2
> t.pt <- seq(1995,2012,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 0.7666667 0.8733333
[10] 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, so that the outer intervals are $1/(\texttt{nk} \times \texttt{bd}$.

4.2 Rates from registers

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( "/home/bendix/sdc/DMreg/NDR-demo/2012/data/FU-m.Rda" )
> lls()
   name mode     class            size
1  a.pt numeric  numeric          1200
2  int  numeric  numeric          1
3  Lx   list     Lexis data.frame 382873 23
4  nk.a numeric  numeric          1
5  nk.c numeric  numeric          1
6  nk.p numeric  numeric          1
7  p.pt numeric  numeric          204
8  qn   function function         1
9  t.pt numeric  numeric          205
10 TT   list     data.frame       6732 9
```

```
> str(TT)
'data.frame':        6732 obs. of  9 variables:
 $ sex : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
 $ A   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P   : num  1995 1995 1996 1996 1997 ...
 $ U   : num  0 1 0 1 0 1 0 1 0 1 ...
 $ Y.nD: num  17026 17101 16468 17067 16434 ...
 $ Y.DM: num  0 0.13 1.44 1.86 0 ...
 $ D.DM: num  0 0 0 0 0 0 0 0 0 0 ...
 $ D.nD: num  137 16 134 23 152 14 132 16 95 17 ...
 $ X   : int  0 2 4 4 0 2 0 0 0 1 ...
 - attr(*, "Variables")= chr [1:9, 1] "Sex" "1-year age class" "1-year period" "Indicator of upper L
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr  "sex" "A" "P" "U" ...
  .. ..$ : chr "Data frame using the modified definintion of DM from NDR"
```

```
> attr(TT,"Variables")
     Data frame using the modified definintion of DM from NDR
sex  "Sex"
A    "1-year age class"
P    "1-year period"
U    "Indicator of upper Lexis triangle"
Y.nD "P-Y among non-diabetics"
Y.DM "P-Y among diabetes patients"
D.DM "Deaths among non-diabetics"
D.nD "Deaths among diabetes patients"
X    "Diabetes diagnoses among non-diabetics"
```

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

```
> summary( TT )
 sex            A             P              U              Y.nD              Y.DM
 M:3366   Min.   : 0   Min.   :1995   Min.   :0.0   Min.   :   46.38   Min.   :   0.00
 F:3366   1st Qu.:24   1st Qu.:1999   1st Qu.:0.0   1st Qu.: 8607.75   1st Qu.:  66.54
          Median :49   Median :2003   Median :0.5   Median :15972.72   Median : 243.43
          Mean   :49   Mean   :2003   Mean   :0.5   Mean   :13235.67   Mean   : 387.91
          3rd Qu.:74   3rd Qu.:2007   3rd Qu.:1.0   3rd Qu.:18011.50   3rd Qu.: 610.31
          Max.   :98   Max.   :2011   Max.   :1.0   Max.   :22944.81   Max.   :2499.02
```

```
      D.DM             D.nD             X
 Min.   :  0.00   Min.   : -1.0   Min.   :  0.00
 1st Qu.:  0.00   1st Qu.:  8.0   1st Qu.:  7.00
 Median :  5.00   Median : 56.0   Median : 26.00
 Mean   : 20.27   Mean   :122.9   Mean   : 44.27
 3rd Qu.: 36.00   3rd Qu.:206.0   3rd Qu.: 74.00
 Max.   :117.00   Max.   :582.0   Max.   :319.00
```

```
> TT[TT$D.nD<0,]
```

```
    sex  A    P U     Y.nD     Y.DM D.DM D.nD X
393   F 19 2004 0 14088.26 59.57769    1   -1 6
```

When fitting models we need the mean age and period, and we need to remove the negative number of non-DM deaths. We see from the tabulation that we truly have data in Lexis-triangles:

```
> TT <- transform( TT, A = A + (1+U)/3,
+                      P = P + (2-U)/3,
+                   D.nD = pmax(D.nD,0) )
> with( subset( TT, A<5 & P<1999 ),
+       print( table( Age=round(A,2),
+                     Per=round(P,2) ),
+              zero.print="." ) )
```

```
      Per
Age    1995.33 1995.67 1996.33 1996.67 1997.33 1997.67 1998.33 1998.67
  0.33       .       2       .       2       .       2       .       2
  0.67       2       .       2       .       2       .       2       .
  1.33       .       2       .       2       .       2       .       2
  1.67       2       .       2       .       2       .       2       .
  2.33       .       2       .       2       .       2       .       2
  2.67       2       .       2       .       2       .       2       .
  3.33       .       2       .       2       .       2       .       2
  3.67       2       .       2       .       2       .       2       .
  4.33       .       2       .       2       .       2       .       2
  4.67       2       .       2       .       2       .       2       .
```

A brief overview of the number of events and PY:

```
> tt <- xtabs( cbind(X,D.nD,D.DM,Y.nD/1000,Y.DM/1000) ~ sex + floor(P),
+              data = TT )
> round( ftable( addmargins(tt,1:2), row.vars=1:2 ), 1 )
```

```
                     X     D.nD     D.DM       V4       V5
sex floor(P)
M   1995        6924.0  27788.0   3393.0   2537.5     45.2
    1996        7052.0  27003.0   3310.0   2549.8     48.7
    1997        6846.0  26020.0   3439.0   2557.9     52.3
    1998        7742.0  25431.0   3486.0   2564.4     56.1
    1999        7929.0  24937.0   3780.0   2569.4     60.3
    2000        7988.0  24431.0   3754.0   2574.6     64.5
    2001        8308.0  24404.0   3919.0   2580.3     68.9
    2002        9465.0  24045.0   4203.0   2584.5     73.7
    2003       10434.0  23043.0   4357.0   2586.8     79.3
    2004       10539.0  23720.0   4350.0   2588.1     85.4
    2005        9603.0  22220.0   4500.0   2590.3     91.2
    2006        9707.0  22492.0   4646.0   2594.9     96.1
    2007       10165.0  22182.0   4765.0   2603.2    101.3
    2008       11015.0  21934.0   4829.0   2615.0    107.2
    2009       11720.0  21530.0   5312.0   2624.1    113.4
    2010       12599.0  21326.0   5294.0   2629.4    120.3
    2011       15545.0  20414.0   5423.0   2632.6    128.8
    Sum       163581.0 402920.0  72760.0  43982.8   1392.6
F   1995        6079.0  28210.0   3339.0   2607.0     43.1
    1996        6051.0  27233.0   3109.0   2618.1     45.9
    1997        5919.0  26926.0   3194.0   2625.5     48.7
```

```
    1998        6251.0  25997.0   3182.0   2631.6     51.5
    1999        6696.0  26713.0   3322.0   2636.7     54.8
    2000        6685.0  25927.0   3451.0   2641.7     58.1
    2001        6794.0  26023.0   3561.0   2647.6     61.3
    2002        8326.0  26254.0   3639.0   2651.6     65.3
    2003        8842.0  24108.0   3814.0   2653.5     70.0
    2004        8993.0  25275.0   3737.0   2654.8     75.2
    2005        7832.0  23741.0   3976.0   2657.2     79.8
    2006        7471.0  23778.0   4049.0   2661.8     83.3
    2007        8243.0  23926.0   4165.0   2668.9     86.9
    2008        8662.0  23182.0   4070.0   2679.0     91.3
    2009        8905.0  23133.0   4280.0   2688.5     95.8
    2010        9561.0  22535.0   4481.0   2696.0    100.6
    2011       13142.0  21660.0   4318.0   2700.4    107.3
    Sum       134452.0 424621.0  63687.0  45119.7   1218.8
Sum 1995       13003.0  55998.0   6732.0   5144.5     88.3
    1996       13103.0  54236.0   6419.0   5167.8     94.6
    1997       12765.0  52946.0   6633.0   5183.4    101.0
    1998       13993.0  51428.0   6668.0   5195.9    107.6
    1999       14625.0  51650.0   7102.0   5206.1    115.0
    2000       14673.0  50358.0   7205.0   5216.4    122.6
    2001       15102.0  50427.0   7480.0   5227.9    130.2
    2002       17791.0  50299.0   7842.0   5236.1    139.0
    2003       19276.0  47151.0   8171.0   5240.3    149.3
    2004       19532.0  48995.0   8087.0   5242.9    160.6
    2005       17435.0  45961.0   8476.0   5247.4    171.0
    2006       17178.0  46270.0   8695.0   5256.8    179.4
    2007       18408.0  46108.0   8930.0   5272.1    188.2
    2008       19677.0  45116.0   8899.0   5293.9    198.5
    2009       20625.0  44663.0   9592.0   5312.6    209.2
    2010       22160.0  43861.0   9775.0   5325.4    220.9
    2011       28687.0  42074.0   9741.0   5333.1    236.1
    Sum       298033.0 827541.0 136447.0  89102.5   2611.4
```

```
> tt <- xtabs( cbind(X,D.nD,D.DM,Y.nD/1000,Y.DM/1000) ~ sex + gP,
+              data = transform( TT,
+                                gP = factor( (P>2000)+(P>2006),
+                                             labels=c("1995-2000",
+                                                      "2001-2006",
+                                                      "2006-2011") ) ) )
> round( ftable( addmargins(tt,1:2), row.vars=1:2 ), 1 )
```

```
                     X     D.nD     D.DM       V4       V5
sex gP
M   1995-2000   36493.0 131179.0  17408.0  12778.9    262.6
    2001-2006   56337.0 141863.0  25083.0  15504.6    462.9
    2006-2011   70751.0 129878.0  30269.0  15699.2    667.1
    Sum        163581.0 402920.0  72760.0  43982.8   1392.6
F   1995-2000   30996.0 135079.0  16146.0  13118.8    244.0
    2001-2006   47472.0 151328.0  22178.0  15906.3    409.7
    2006-2011   55984.0 138214.0  25363.0  16094.6    565.1
    Sum        134452.0 424621.0  63687.0  45119.7   1218.8
Sum 1995-2000   67489.0 266258.0  33554.0  25897.7    506.6
    2001-2006  103809.0 293191.0  47261.0  31410.9    872.6
    2006-2011  126735.0 268092.0  55632.0  31793.8   1232.2
    Sum        298033.0 827541.0 136447.0  89102.5   2611.4
```

Then 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("apc","ap") ) )
> str( Lambda )
 logi [1:1200, 1:204, 1:2, 1:2] 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 [1:204] "1995.04166666667" "1995.125" "1995.20833333333" "1995.29166666667" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ mod: chr [1:2] "apc" "ap"

> prod( dim(Lambda) )

[1] 979200
```

4.3 Incidence rates

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

```
> ( ki.a <- with( TT, quantile( rep(  A,X), qn(nk.a) ) ) )
      5%      15%      25%      35%      45%      55%      65%      75%      85%      95% 
32.66667 45.66667 52.33333 56.66667 60.66667 64.33333 68.33333 72.66667 77.33333 84.66667 

> ( ki.p <- with( TT, quantile( rep(P   ,X), qn(nk.p) ) ) )

     10%      30%      50%      70%      90% 
1997.333 2001.333 2004.667 2008.333 2010.667 

> ( ki.c <- with( TT, quantile( rep(P-A,X), qn(nk.c) ) ) )

   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75% 
1919.333 1928.333 1934.667 1940.333 1944.667 1949.667 1956.667 1969.333 
```

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. However we want to extract the avearge trend from the APC-model, so we also fit the model with the parametrization that allows us to extract the trend by using the information (in this case `X`) as weights:

```
> m.inc.ap <- glm( X ~ Ns(A,knots=ki.a) + Ns(P,knots=ki.p),
+                      offset = log(Y.nD),
+                      family = poisson,
+                        data = subset(TT,sex=="M") )
> m.inc.aPC <- update( m.inc.ap, . ~ . - Ns(P   ,kn=ki.p) + I(P) +
+                               detrend( Ns(P   ,kn=ki.p), P   , X ) +
+                               detrend( Ns(P-A,kn=ki.c), P-A, X ) )
> m.inc.apc <- update( m.inc.ap, . ~ . + Ns(P-A,kn=ki.c) )
> f.inc.ap  <- update( m.inc.ap , data = subset(TT,sex=="F") )
> f.inc.apc <- update( m.inc.apc, data = subset(TT,sex=="F") )
> f.inc.aPC <- update( m.inc.aPC, data = subset(TT,sex=="F") )
```

The average annual trends in incidence:

```
> 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       3.7  3.5   3.8
                      Women       3.6  3.5   3.7
```

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.nD = int, Y.DM = int )
> head( nd )
           A        P       Y.nD       Y.DM
1 0.04166667 1995.042 0.08333333 0.08333333
2 0.12500000 1995.042 0.08333333 0.08333333
3 0.20833333 1995.042 0.08333333 0.08333333
4 0.29166667 1995.042 0.08333333 0.08333333
5 0.37500000 1995.042 0.08333333 0.08333333
6 0.45833333 1995.042 0.08333333 0.08333333

> dim( nd )

[1] 244800      4
```

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.

4.4 Mortality rates

We have the mortality-rates in `TT` too, by using `D.nD` and `Y.nD` for mortality in the population without diabetes and `D.DM` and `Y.DM` for the mortality in persons with diabetes.

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( TT, quantile( rep(A  ,D.DM), qn(nk.a) ) ) )
      5%      15%      25%      35%      45%      55%      65%      75%      85%      95% 
54.66667 64.33333 69.33333 73.33333 76.66667 79.33333 82.33333 85.33333 88.33333 92.33333 

> ( kmd.p <- with( TT, quantile( rep(P  ,D.DM), qn(nk.p) ) ) )

     10%      30%      50%      70%      90% 
1997.333 2001.333 2004.333 2007.667 2010.667 

> ( kmd.c <- with( TT, quantile( rep(P-A,D.DM), qn(nk.c) ) ) )

   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75% 
1910.667 1916.667 1920.667 1924.333 1928.333 1932.667 1938.667 1948.667 
```

```
> m.md.ap <- glm( D.DM ~ Ns(A,knots=kmd.a) + Ns(P,knots=kmd.p),
+                        offset = log(Y.DM),
+                        family = poisson,
+                          data = subset( TT, sex=="M" & Y.DM>0 ) )
> m.md.aPC <- update( m.md.ap, . ~ . - Ns(P   ,kn=kmd.p) + I(P) +
+                             detrend( Ns(P   ,kn=kmd.p), P   , D.DM ) +
+                             detrend( Ns(P-A,kn=kmd.c), P-A, D.DM ) )
> m.md.apc <- update( m.md.ap, . ~ . + Ns(P-A,kn=kmd.c) )
> f.md.ap  <- update( m.md.ap , data = subset( TT, sex=="F" & Y.DM>0 ) )
> f.md.apc <- update( m.md.apc, data = subset( TT, sex=="F" & Y.DM>0 ) )
> f.md.aPC <- update( m.md.aPC, data = subset( TT, sex=="F" & Y.DM>0 ) )
> 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 )
```

4.4.2 Persons without diabetes

The mortality in the population without diabetes is modeled in exactly the same way:

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

```
      5%      15%      25%      35%      45%      55%      65%      75%      85%      95% 
46.33333 60.33333 67.66667 72.66667 77.33333 80.33333 83.33333 86.33333 89.66667 93.66667 
```

```
> ( kmw.p <- with( TT, quantile( rep(P   ,D.nD), qn(nk.p) ) ) )
```

```
     10%      30%      50%      70%      90% 
1996.333 1999.667 2002.667 2006.333 2010.333 
```

```
> ( kmw.c <- with( TT, quantile( rep(P-A,D.nD), qn(nk.c) ) ) )
```

```
   6.25%   18.75%   31.25%   43.75%   56.25%   68.75%   81.25%   93.75% 
1908.333 1914.333 1918.667 1922.333 1926.667 1932.333 1940.667 1954.667 
```

```
> m.mw.ap <- glm( D.nD ~ Ns(A,knots=kmw.a) + Ns(P,knots=kmw.p),
+                        offset = log(Y.nD),
+                        family = poisson,
+                          data = subset( TT, sex=="M" & Y.nD>0 ) )
> m.mw.aPC <- update( m.mw.ap, . ~ . - Ns(P   ,kn=kmw.p) + I(P) +
+                             detrend( Ns(P   ,kn=kmw.p), P   , D.nD ) +
+                             detrend( Ns(P-A,kn=kmw.c), P-A, D.nD ) )
> m.mw.apc <- update( m.mw.ap, . ~ . + Ns(P-A,kn=kmw.c) )
> f.mw.ap  <- update( m.mw.ap , data = subset( TT, sex=="F" & Y.nD>0 ) )
> f.mw.apc <- update( m.mw.apc, data = subset( TT, sex=="F" & Y.nD>0 ) )
> f.mw.aPC <- update( m.mw.aPC, data = subset( TT, sex=="F" & Y.nD>0 ) )
> 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 )
```

4.5 Secular 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       3.7  3.5   3.8
                      Women       3.6  3.5   3.7
Mortality change,   DM: Men      -3.8 -4.0  -3.7
                      Women      -3.5 -3.7  -3.4
Mortality change, Well: Men      -2.7 -2.7  -2.6
                      Women      -2.3 -2.4  -2.2
```

Thus it appears that the incidence rates of diabetes are increasing by some 4% per year, while mortality rates are decreasing 4% per your for persons with diabetes, but only 2–2.5% per year for perons without.

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="./data/Ests-m.Rda" )
```

4.6 Prevalence predictions

The purpose of the remainings sections 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/Ests-m.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 empirical age-specific prevalences at 1.1.1995–1.1.2012, 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 1995–2011 or assuming they were constant at the 1995 level. These refer to intervals of length `int` and are therefore labeled on the period dimension by the midpoint of these, a total of 17/`int`.

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

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

4.7 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 [1:1200, 1:204, 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 [1:204] "1995.04166666667" "1995.125" "1995.20833333333" "1995.29166666667" ...
  ..$ sex  : chr [1:2] "M" "F"
  ..$ mod  : 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 0.083 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.

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) for either incidence, mortality or both:

```
> ( np <- dim(Lambda)[2] )
  p
204
> 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 15 mil. numbers:

```
> str( TR )
 num [1:1200, 1:204, 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" "0.291666666666667" ...
  ..$ p    : chr [1:204] "1995.04166666667" "1995.125" "1995.20833333333" "1995.29166666667" ...
  ..$ sex  : chr [1:2] "M" "F"
  ..$ mod  : 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] 15667200
```

4.8 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( "/home/bendix/sdc/DMreg/NDR-demo/2012/data/prev-m.Rda" )
> summary( pr )
 sex            A               P              X                N
 M:1800   Min.   : 0.00   Min.   :1995   Min.   :   0.0   Min.   :   77
 F:1800   1st Qu.:24.75   1st Qu.:1999   1st Qu.: 128.8   1st Qu.:18405
          Median :49.50   Median :2004   Median : 473.0   Median :32166
          Mean   :49.50   Mean   :2004   Mean   : 770.8   Mean   :26980
          3rd Qu.:74.25   3rd Qu.:2008   3rd Qu.:1196.5   3rd Qu.:36725
          Max.   :99.00   Max.   :2012   Max.   :4997.0   Max.   :46208
> head( pr )
  sex A    P X     N
1   M 0 1995 3 35612
2   M 0 1996 1 36055
3   M 0 1997 0 34853
4   M 0 1998 1 34774
5   M 0 1999 2 34076
6   M 0 2000 1 33906
```

These are empirical prevalences (`X`—no. of cases of DM, `N`—population size) for each of the 18 dates 1.1.1995 – 1.1.2010 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 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 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%       50% 56.66667%
       10        27        40        47        52        55        58        61        64        66
63.33333%       70% 76.66667% 83.33333%       90% 96.66667%
       69        72        74        78        81        87
> pr.fit <- NArray( c( dimnames(Lambda)[c(1,3)],
+                      list( t = sort(unique(pr$P)) ) ) )
> str( pr.fit )
 logi [1:1200, 1:2, 1:18] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ a  : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ sex: chr [1:2] "M" "F"
  ..$ t  : chr [1:18] "1995" "1996" "1997" "1998" ...
```

```
> prod( dim(pr.fit) )
[1] 43200
```

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.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=as.numeric(dimnames(pr.fit)[["a"]]) ),
+                            type = "response" )
+    }
> round( ftable( pr.fit[c(1:2,NA,floor(dim(pr.fit)[1]*3/4)+1:5),,],
+                row.vars=2:1 )*100, 1 )
                       t 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
sex a
M   0.0416666666666667    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
    0.125                 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
    NA                     NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
    75.041666666667       6.0  6.4  6.7  7.0  7.4  8.0  8.5  9.2  9.8 10.5 11.3 11.8 12.5 13.2 14.0
    75.125                6.0  6.4  6.7  7.0  7.4  8.0  8.5  9.2  9.8 10.5 11.3 11.8 12.5 13.2 14.0
    75.2083333333333      6.0  6.4  6.7  7.0  7.4  8.0  8.5  9.1  9.8 10.5 11.3 11.8 12.5 13.2 14.0
    75.2916666666667      6.0  6.4  6.7  7.1  7.4  8.0  8.5  9.1  9.8 10.5 11.3 11.8 12.5 13.2 14.0
    75.375                6.0  6.5  6.7  7.1  7.4  8.0  8.5  9.1  9.8 10.5 11.3 11.8 12.5 13.2 14.0
F   0.0416666666666667    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  0.1  0.1
    0.125                 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  0.1  0.1
    NA                     NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA   NA
    75.041666666667       5.4  5.7  6.1  6.3  6.5  6.7  7.0  7.4  8.2  8.9  9.4  9.8 10.1 10.6 11.0
    75.125                5.4  5.7  6.1  6.3  6.5  6.7  7.0  7.4  8.2  8.9  9.4  9.8 10.1 10.6 11.0
    75.2083333333333      5.4  5.8  6.1  6.3  6.5  6.7  7.0  7.4  8.2  8.9  9.4  9.8 10.1 10.6 11.1
    75.2916666666667      5.5  5.8  6.1  6.3  6.5  6.7  7.0  7.4  8.1  8.9  9.4  9.8 10.1 10.6 11.1
    75.375                5.5  5.8  6.1  6.3  6.5  6.7  7.0  7.4  8.1  8.9  9.4  9.8 10.1 10.6 11.1
```

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(0,22), 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)[[1]] )[1]
+ a80 <- grep( "80.", dimnames(pr.fit)[[1]] )[1]
+ text( 89, pr.fit[a89,"M","1995"]* 99, "1995", col="blue", adj=c(1,1) )
+ text( 80, pr.fit[a80,"M","2012"]*101, "2012", col="blue", adj=c(0,0) )
+ axis( side=1 )
+ axis( side=2 )
+ axis( side=2, at=0:22, labels=NA )
+ matplot( a.pt, pr.fit[,"F",]*100,
+          ylim=c(0,22), xlim=c(20,90), yaxs="i", xaxt="n", yaxt="n", xlab="", ylab="",
+          type="l", lty=1, col="red", lwd=c(1,2) )
+ 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[a89,"F","1995"]* 99, "1995", col="red", adj=c(1,1) )
```

```
+ text( 80, pr.fit[a80,"F","2012"]*101, "2012", col="red", adj=c(1,0) )
+ axis( side=1 )
+ mtext( "Age", side=1, line=1, outer=T )
+ mtext( "DM prevalence (%)", side=2, line=2, outer=T, las=0 )
+ }
> plp()
```

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

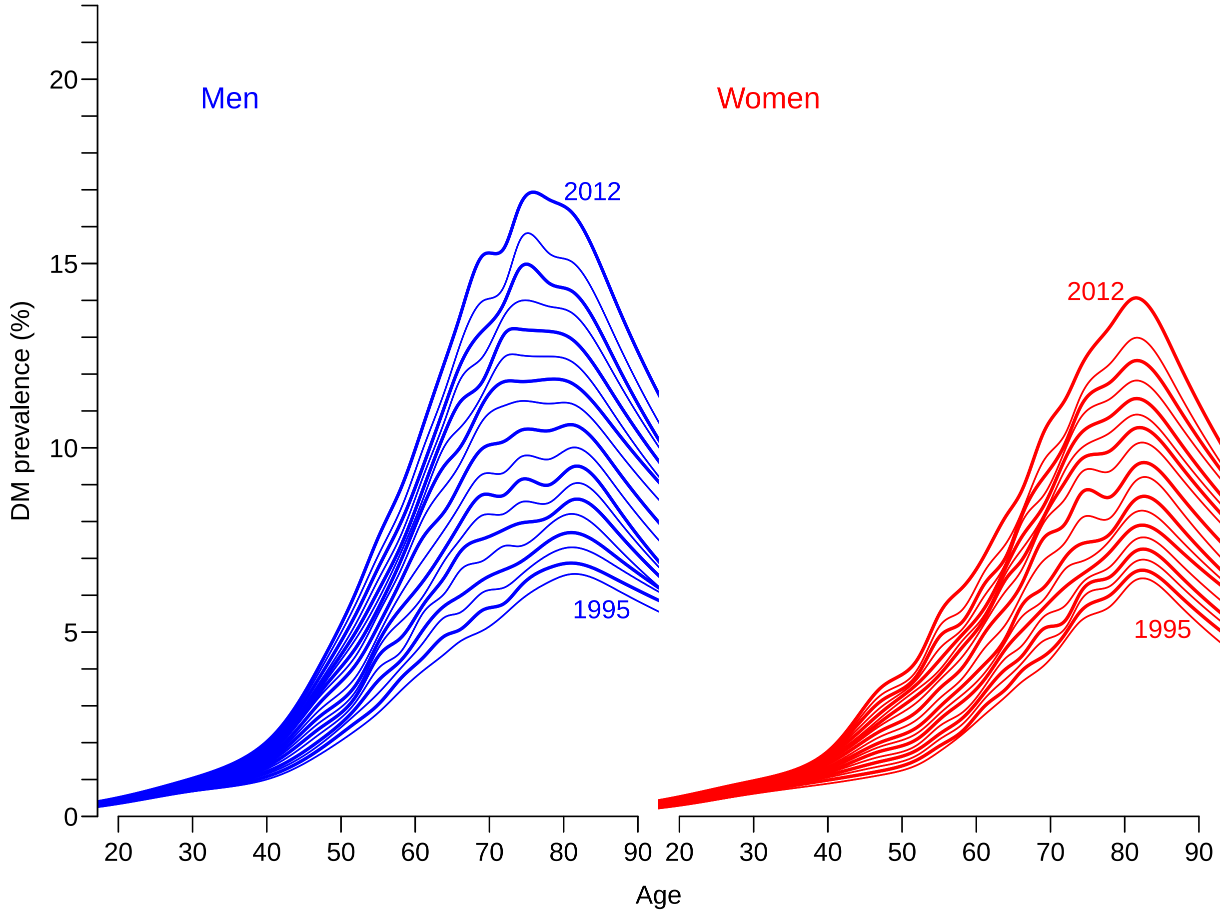

Figure 4.1: *Smoothed age-specific prevalences for the years 1995–2010. Blue is men, red is women.*

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

4.9 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 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 [1:1200, 1:205, 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" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

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

```
[1] 6888000
```

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-born) 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,floor(dim(prv)[1]/1.5)+1:3),1:2,,,],
+                col.vars=4:5 )*100, 1 )
```

```
                                        mod  apc                                   ap
                                        what obs m-fix i-fix all-f mort inc const obs m-fix i-fix al
a                  t                sex
0.0416666666666667 1995             M        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
                   1995.08333333333 M        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
0.125              1995             M        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
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
                                    F         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
66.7083333333333   1995             M        4.8   4.8   4.8   4.8  4.8 4.8   4.8 4.8   4.8   4.8
                                    F        3.7   3.7   3.7   3.7  3.7 3.7   3.7 3.7   3.7   3.7
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
                                    F         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
66.7916666666667   1995             M        4.8   4.8   4.8   4.8  4.8 4.8   4.8 4.8   4.8   4.8
                                    F        3.7   3.7   3.7   3.7  3.7 3.7   3.7 3.7   3.7   3.7
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
                                    F         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
66.875             1995             M        4.8   4.8   4.8   4.8  4.8 4.8   4.8 4.8   4.8   4.8
                                    F        3.8   3.8   3.8   3.8  3.8 3.8   3.8 3.8   3.8   3.8
                   1995.08333333333 M         NA    NA    NA    NA   NA  NA    NA  NA    NA    NA
                                    F         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,,,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
  8.923   0.000   8.920
```

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.

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

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

                    t
a                     1995 1995.08333333333
  0.0416666666666667 0.000            0.000
  0.125              0.040            0.001
  0.208333333333333  0.040            0.041
  0.291666666666667  0.041            0.041

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

4.9.1 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 1.1.2012.

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:

```
> nd <- c( grep("2000",dimnames(prv)[[2]])[1],
+          grep("2006",dimnames(prv)[[2]])[1],
+          grep("2012",dimnames(prv)[[2]])[1] )
> ( wh <- dimnames(prv)[[2]][nd] )
[1] "2000" "2006" "2012"

> 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(0,22), 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(0,22), 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(0,22), yaxs="i",
+          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 )
> 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(0,22), yaxs="i",
+          xlab="Age", ylab="", yaxt="n",
+          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 )
> 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 (see figure **??**), we will use these in the reporting of the different scenarios.

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

```
> str( prv )
```

```
 num [1:1200, 1:205, 1:2, 1:2, 1:7] 0 0.000399 0.000403 0.000407 0.000411 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...
```

```
> 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 13.69592 12.51609  9.953617 8.968855   NA  NA    NA
  66.7916666666667 13.74097 12.55212  9.984925 8.992574   NA  NA    NA
  66.875           13.78554 12.58763 10.016075 9.016096   NA  NA    NA
  66.9583333333333 13.82961 12.62261 10.047070 9.039421   NA  NA    NA
  67.0416666666667 13.87316 12.65704 10.077908 9.062544   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" )
> matplot( a.pt, cbind(prv[,np,"M","apc",],prv[,1,"M","apc",1])*100,
+          xlim=c(20,90), ylim=c(0,22), yaxs="i",
+          xlab="Age", ylab="Prevalence (%)",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="blue" )
```

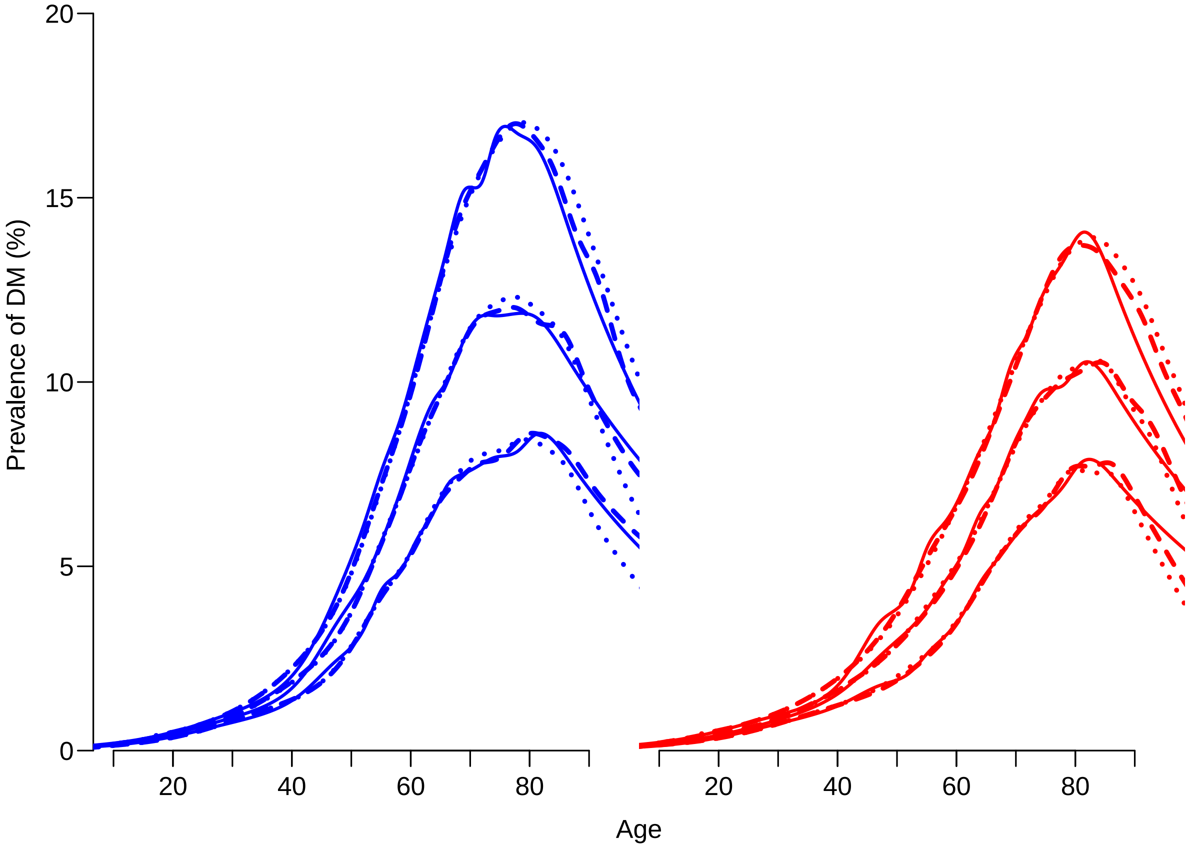

Figure 4.2: *Plot of observed (full lines) and predicted prevalences in 2000, 2006 and 2012, using simple age-period-models (dotted lines) or age-period-cohort models (broken lines). Clearly the broken lines gives a better approximation to the smoothed empirical rates.*

```
> 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" )
> axis( side=2, at=0:22, labels=NA )
> matplot( a.pt, cbind(prv[,np,"F","apc",],prv[,1,"F","apc",1])*100,
+          xlim=c(20,90), ylim=c(0,22), yaxs="i",
+          xlab="Age", ylab="", yaxt="n",
+          type="l", lty=rep(c(1,0),2), lwd=c(4,4,2,2,0)+1, col="red" )
> matlines( a.pt, prv[,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")
> 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 )
+ {
```

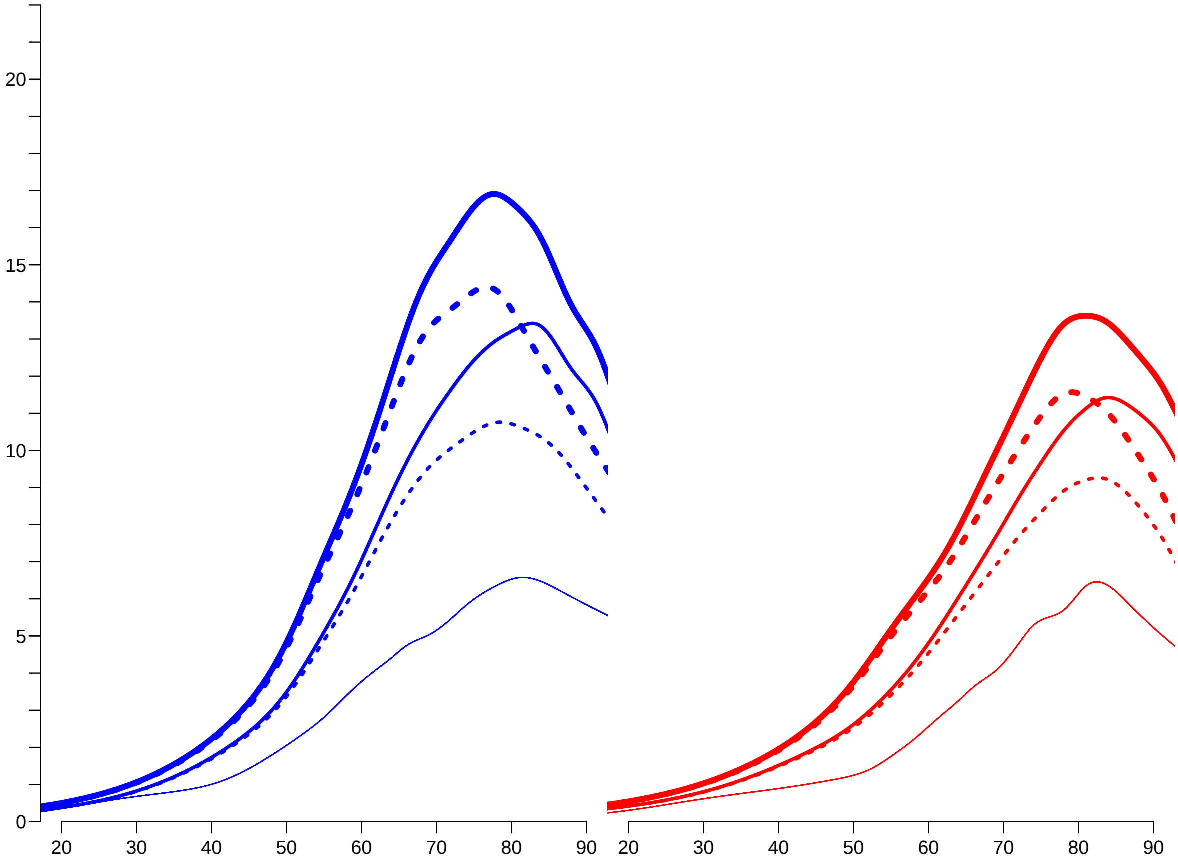

Figure 4.3: *The predicted prevalences under different scenarios:*
Full lines: Mortality rates evolve as observed, Broken lines: Mortality rates remain as 1995. Thick lines: Incidence rates evolve as observed, Thin lines: Incidence rates remain as in 1995.
The very thin lines lowest in the two displays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

```
+ pdf( paste( "./graph/mDMpr-", 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(0,22), 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=0:22, labels=NA )
+ 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(0,22), 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 4.3 shows the predicted prevalences under 4 different scenarios compared to the observed prevalences as of 1.1.1995.

4.10 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(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","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(0,7), 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 )
> 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(0,7), 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" )
```

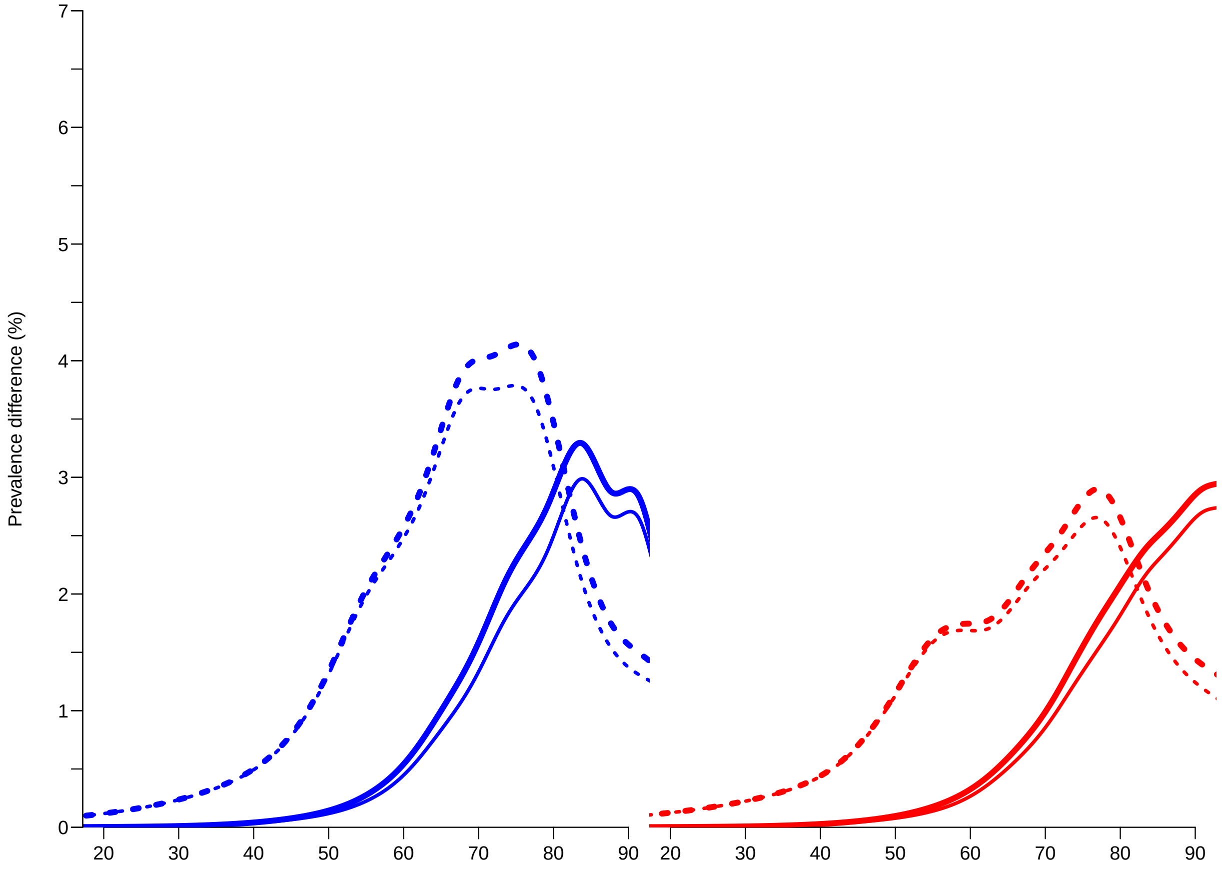

Figure 4.4: *Suggested contributions to age-specific prevalences from increasing mortalities over the period 1995–2012; the thick lines are obtained by subtracting the prediction based on fixing one rate from the one using the observed rates; thin lines based on subtracting the prediction based on fixing both rates from that where one is fixed. Full lines are for differences attributable to changes in mortality rates, broken lines are for changes attributable to changes in incidence rates.*

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 0.5%. 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 1995 (“`all-f`”). This is obtained by taking the average of the two curves in each scenario.

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

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 — the "epidemiological disequilibrium" 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 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(0,22),
+             txt=dimnames(prv)[[2]][yn[j]] )
+ abline(h=0)
+ if( j==1 ) text( rep(25,3), c(13,15,17),
+                  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(0,22) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> axis( side=2 )
> axis( side=2, at=0:22, labels=NA )
> text( rep(25,3), 17:19+2.5,
+       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(0,22) )
> abline(h=0:22,v=2:9*10,col=gray(0.9))
> axis( side=1 )
> # axis( side=2 )
> text( rep(25,3), 17:19+2.5,
+       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 )
```

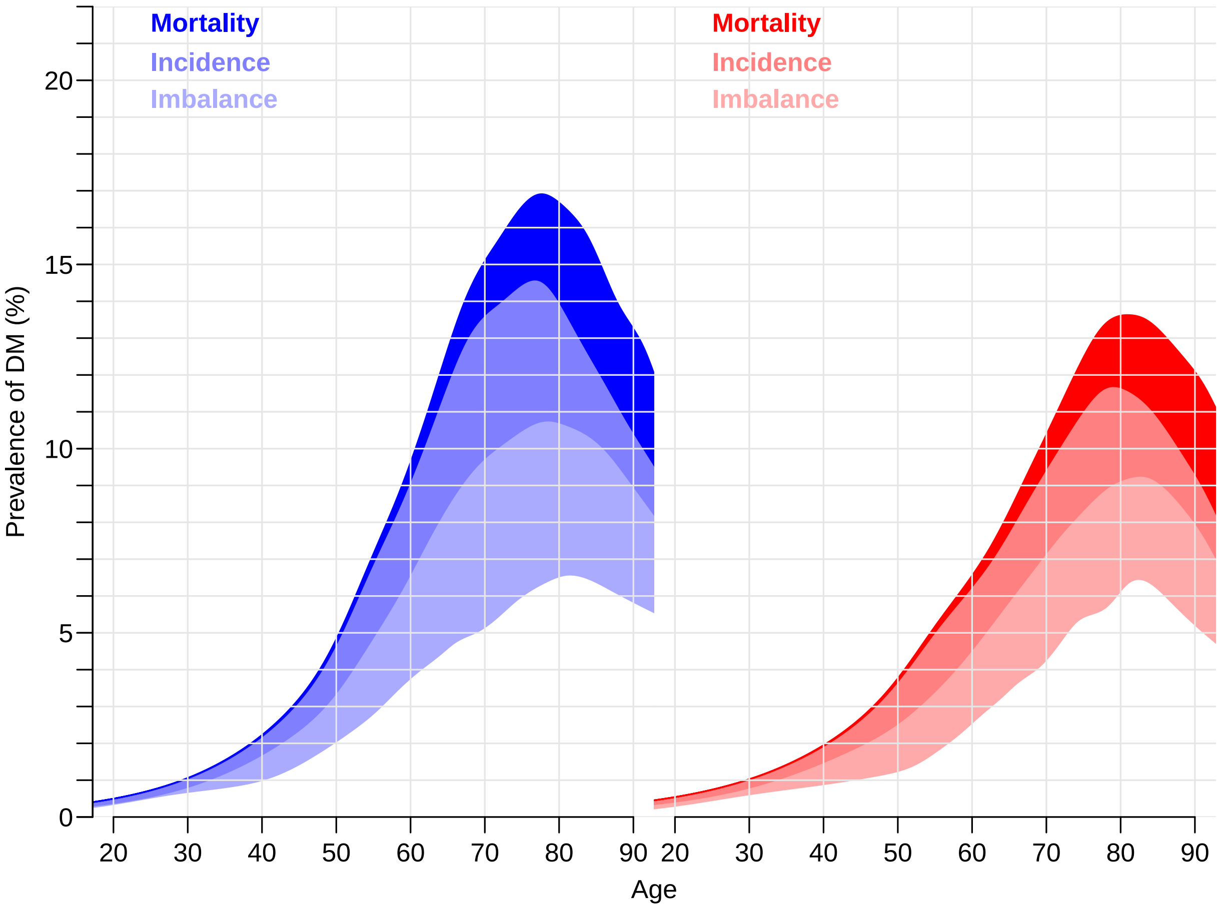

Figure 4.5: *Predicted age-specific prevalences of DM in Denmark 2012 among men (blue) and women (red), partitioned by the contribution from rates as they were in 1995 ("Imbalance"), increases in incidence and decrease in mortality, respectively.*

4.11 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` (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]] <- 1995:2012
> dn[[5]] <- dn[[5]][c(5:7,1)]
> prn <- NArray( dn )
> length(prv) ; str( prv ) ; table( prv<0 )
[1] 6888000

 num [1:1200, 1:205, 1:2, 1:2, 1:7] 0 0.000399 0.000403 0.000407 0.000411 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:7] "obs" "m-fix" "i-fix" "all-f" ...

  FALSE    TRUE
6780695  107305

> hist( prv, breaks=200, col="black", ylim=c(0,10000) ) ; abline(v=0,col="red")
> length(prn) ; str( prn )
[1] 28800

 logi [1:100, 1:18, 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 [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ what: chr [1:4] "mort" "inc" "const" "obs"
```

In order to fill in the numbers we use the estimates age-specific prevalences at 1st January each year, that is at the dates 1.1.1995,…,1.1.2012 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, 1:18, 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 [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ 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
```

```
> 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 )
+    }
> range( prv )
```

```
[1] 0.0000000 0.1703914
```

```
> range( prn )
```

```
[1] 0.00000000 0.06575705
```

Now `prn` contains the prevalences (as fractions) for 100 age classes and 18 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   M 0 1995 3 35612
2   M 0 1996 1 36055
3   M 0 1997 0 34853
4   M 0 1998 1 34774
5   M 0 1999 2 34076
6   M 0 2000 1 33906
```

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

```
 xtabs [1:100, 1:18, 1:2] 35612 34747 35082 33330 32974 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = N ~ A + P + sex, data = pr)
```

```
> str( dmp )
```

```
 xtabs [1:100, 1:18, 1:2] 3 4 6 5 12 21 22 34 29 29 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = X ~ A + P + sex, data = pr)
```

```
> str( prn )
```

```
 num [1:100, 1:18, 1:2, 1:2, 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 [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : 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(=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 <- "2012"
> 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) )
+ 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/mDMpr-obs-film.pdf", width=8, height=6 )
> for( pp in paste(1996:2012) ) draw.dmp(pp)
> dev.off()
```

```
pdf
  3
```

```
> for( pp in paste(1995:2012) )
+ {
+ pdf( paste("./graph/mDMpr-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" "#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) )
+ 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=2 )
+ 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( "./graph/mDMpr-film.pdf", width=9, height=6 )
> for( pp in paste(1996:2012) ) draw.pyr(pp)
> dev.off()

pdf
  3

> for( pp in paste(1996:2012) )
+ {
+ pdf( paste("./graph/mDMpr-", pp, ".pdf", sep=""), width=8, height=6 )
+ draw.pyr(pp)
+ dev.off()
+ }
```

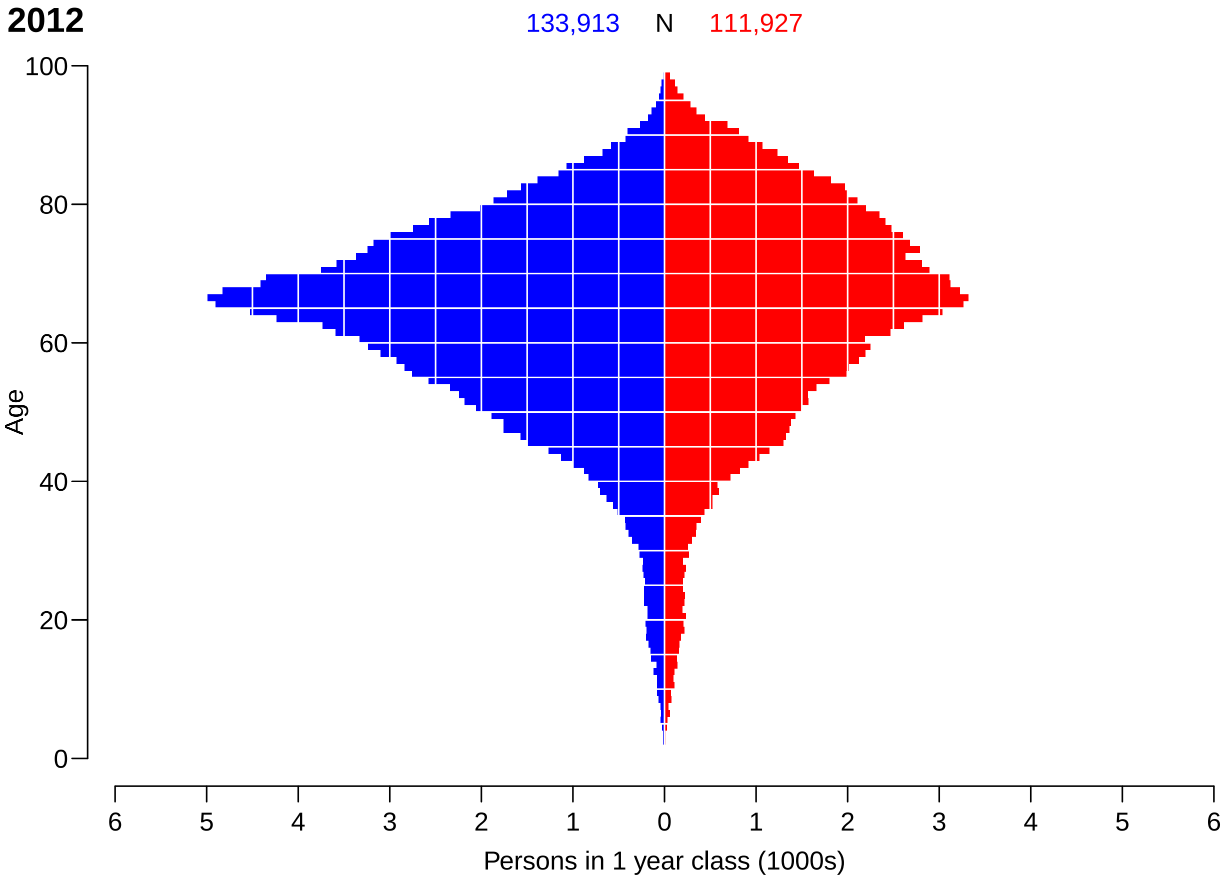

Figure 4.6: *Empirical age-distribution of the diabetes cases in Denmark as of 1.1.2012.*

4.12 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.

```
> lls()
   name       mode      class       size
1  a.pt       numeric   numeric     1200
2  c.a        character character   1
3  cau.exp    function  function    1
4  clr        character character   8
5  comp       character character   3
6  dmp        numeric   xtabs table 100 18 2
7  dn         list      list        5
8  dpr        list      list        5
9  draw.dmp   function  function    1
10 draw.pyr   function  function    1
11 dt         character character   1
12 een        numeric   numeric     4
13 ff         logical   logical     1
14 hts        numeric   numeric     4
15 i          numeric   integer     1
16 ia         numeric   integer     1
17 int        numeric   numeric     1
18 ip         numeric   integer     1
19 j          character character   1
20 kp.a       numeric   numeric     16
```

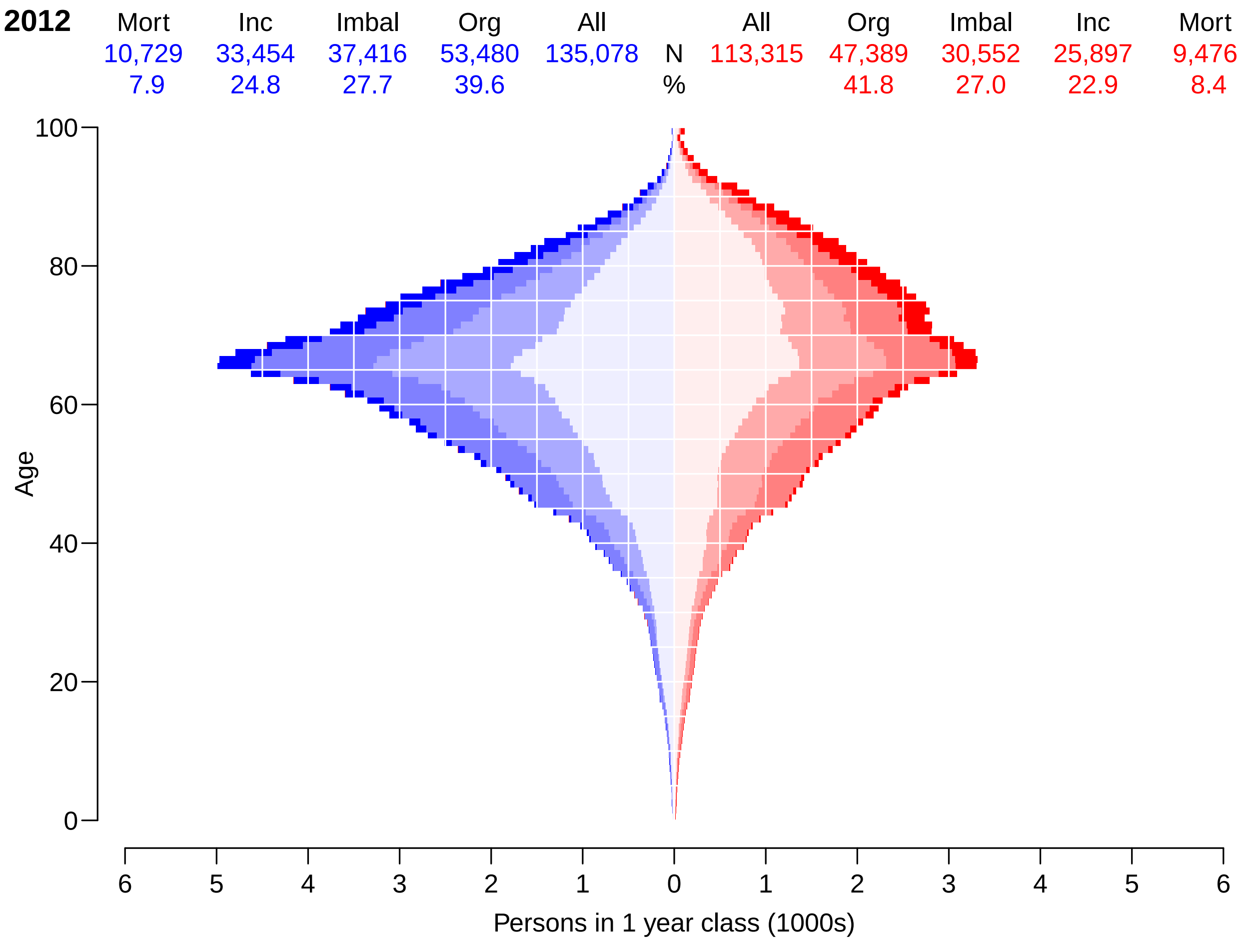

Figure 4.7: *Age-distribution of the predicted no. of diabetes cases in Denmark as of 1.1.2012, subdivided by the components of disease prevalence: Mort: decrease in mortality, Inc: increase in incidence, Const: constant rates from 1995 (non-steady-state between incidence and mortality in 1995, Org: age-specific prevalence in 1995.*

```
21 Lambda     numeric   array       1200 204 2 2
22 lim        numeric   numeric     1
23 Mu.DM      numeric   array       1200 204 2 2
24 Mu.W       numeric   array       1200 204 2 2
25 n.a        numeric   numeric     1
26 nd         numeric   integer     3
27 np         numeric   integer     1
28 one.comp   function  function    1
29 oo         character character   4
30 plp        function  function    1
31 poly.parts function  function    1
32 pop        numeric   xtabs table 100 18 2
33 pp         character character   1
34 p.pt       numeric   numeric     204
35 pr         list      data.frame  3600 5
36 pr.fit     numeric   array       1200 2 18
37 pr.mod     list      glm lm      30
38 prn        numeric   array       100 18 2 2 4
39 prn.a      numeric   numeric     100
40 prn.p      numeric   numeric     18
41 prv        numeric   array       1200 205 2 2 7
42 prv.a      numeric   numeric     1200
43 prv.p      numeric   numeric     205
```

```
44 qn         function  function    1
45 scen       character character   4
46 shd        numeric   numeric     4
47 states     character character   2
48 sx         character character   1
49 t.pt       numeric   numeric     205
50 TR         numeric   array       1200 204 2 2 2 2 4
51 wh         character character   3
52 wh.a       numeric   integer     12
53 wh.p       numeric   integer     18
```

```
> str( prv )
```

```
 num [1:1200, 1:205, 1:2, 1:2, 1:7] 0 0.000399 0.000403 0.000407 0.000411 ...
 - attr(*, "dimnames")=List of 5
  ..$ a   : chr [1:1200] "0.0416666666666667" "0.125" "0.208333333333333" "0.291666666666667" ...
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ mod : chr [1:2] "apc" "ap"
  ..$ 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, 1:205, 1:2, 1:4] 0.0377 0.0515 0.0653 0.038 0.0519 ...
 - attr(*, "dimnames")=List of 4
  ..$ a   : chr [1:3] "60.0416666666667" "70.0416666666667" "80.0416666666667"
  ..$ t   : chr [1:205] "1995" "1995.08333333333" "1995.16666666667" "1995.25" ...
  ..$ 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 fro 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(1995,2012), 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+1995, labels=NA ) }
+ 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(2000,2010,5), 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( 1996, 70, paste( if(sx=="F") "W" else sx,"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 )
```

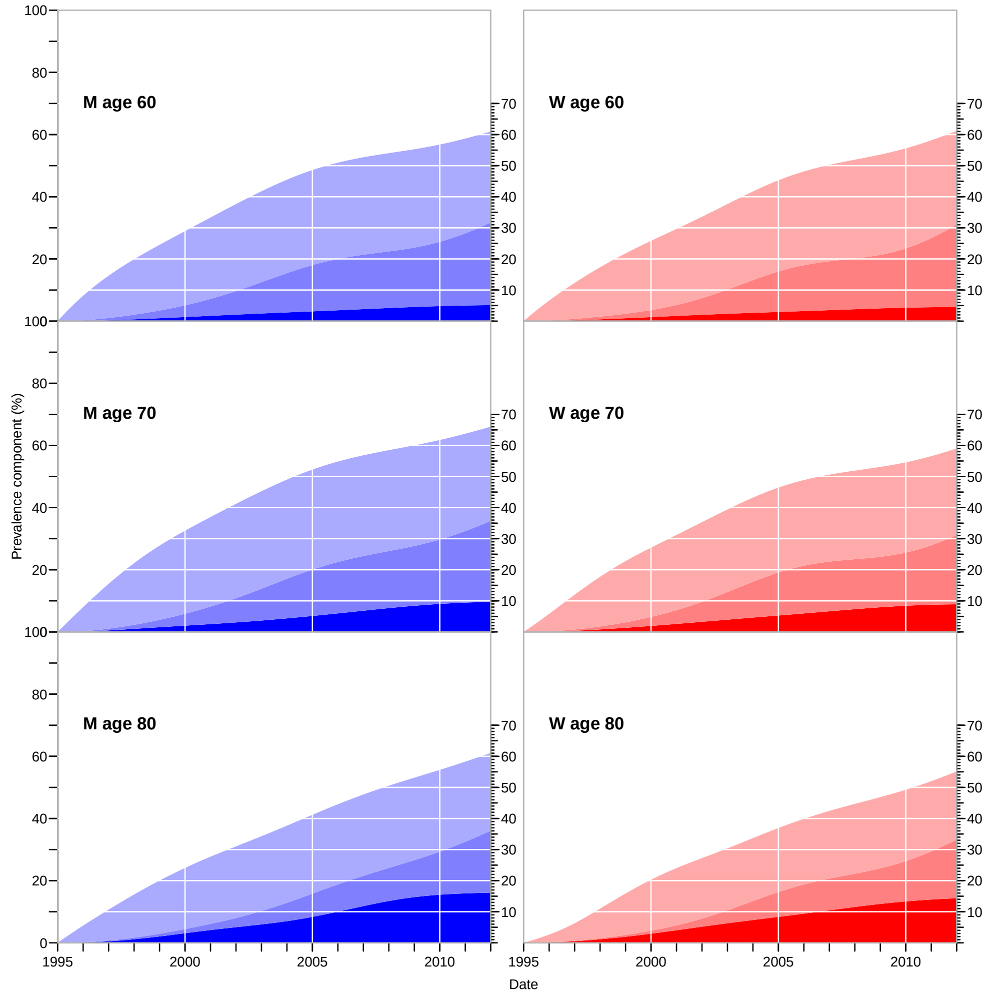

Figure 4.8: *Fraction of the prevalent cases at different times attributable to a) declining mortality (bottom, full color), b) increasing incidence (middle, pale color) and c) prevalence/mortality imbalance 1995 (top, weak color). The white area above the curves corresponds to the fraction of the cases that would have been around if incidence and mortality rates had remained as in 1995.*