Diabetes and immigration in Denmark

SDC
January 2016
http://BendixCarstensen.com/SDC/DMigr
Version 2.4

Compiled Saturday 23rd January, 2016, 16:17
from: /home/bendix/sdc/coll/maej/migrant/MiDM.tex

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

Contents

Chapter 1

Reading and setting up follow-up data

1.1 Data conversion

First we read the data from SAS format to xport format which is R-readable:

```
1                                    "Program: getdata.sas"     12:40 Thursday, November 14, 2013

NOTE: Copyright (c) 2002-2008 by SAS Institute Inc., Cary, NC, USA.
NOTE: SAS (r) Proprietary Software 9.2 (TS2M3)
      Licensed to NOVO NORDISK - BASIC PACKAGE, Site 50800704.
NOTE: This session is executing on the W32_VSPRO  platform.



NOTE: SAS initialization used:
      real time           2.73 seconds
      cpu time            0.48 seconds


NOTE: AUTOEXEC processing beginning; file is c:\stat\sas\autoexec.sas.

---------------------------------------------------------------
C:\Bendix\Steno\MaEJ\Migrant\sas\getdata.sas
---------------------------------------------------------------
NOTE: Libref HER was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\MaEJ\Migrant\sas
NOTE: Libref DATA was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\MaEJ\Migrant\data

NOTE: AUTOEXEC processing completed.

1          options fmtsearch = (data) ;
2          libname  data "../data" ;
NOTE: Libref DATA was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\MaEJ\Migrant\data
3
4          proc contents  data=data.dmigr ;
5          run ;

NOTE: PROCEDURE CONTENTS used (Total process time):
      real time           0.06 seconds
      cpu time            0.06 seconds

NOTE: The PROCEDURE CONTENTS printed page 1.

6
7          data dmigr ( drop = country ) ;
8            set data.dmigr ( rename = ( land=country ) ) ;
9            if region  eq "middle ea" then region = "MidEast" ;
10           if region  eq ""          then region = " DK" ;
11           if country eq "5525"      then region = "Africa" ;
12           land = input( country, 4. ) ;
13           lann = land ;
14         run ;

NOTE: There were 1203642 observations read from the data set DATA.DMIGR.
NOTE: The data set WORK.DMIGR has 1203642 observations and 9 variables.
NOTE: DATA statement used (Total process time):
      real time           0.56 seconds
      cpu time            0.56 seconds
```

```
15
16         proc tabulate  data = dmigr
17              noseps  missing  formchar="           " ;
18         class region land lann sex ;
19         var doBTH doDM doDTH doIND doUD ;
20         table region *
21               ( doBTH doDM doIND doUD doDTH ),
22               ( ( n nmiss ) * f=comma10.
23                 ( min max ) * f=ddmmyy10. )
24               / rts=25 ;
25         table region * lann * land,
26               ( all="Total" ( doBTH doDM doDTH ) * ( n nmiss ) ) * f=comma7.
27               / rts=50 indent=3 ;
28         format land Ncountry.
29                lann region. ;
30         run ;

NOTE: There were 1203642 observations read from the data set WORK.DMIGR.
NOTE: The PROCEDURE TABULATE printed pages 2-5.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.37 seconds
      cpu time            1.06 seconds


31
32         data dmigr ;
33           set dmigr ( drop = lann land ) ;
34         run ;

NOTE: There were 1203642 observations read from the data set WORK.DMIGR.
NOTE: The data set WORK.DMIGR has 1203642 observations and 7 variables.
NOTE: DATA statement used (Total process time):
      real time           1.66 seconds
      cpu time            0.48 seconds


35
36         libname xptout xport '../data/dmigr.xpt' ;
NOTE: Libref XPTOUT was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: C:\Bendix\Steno\MaEJ\Migrant\data\dmigr.xpt
37         proc copy  in = work  out = xptout  memtype = data ;
2 "Program: getdata.sas"                                           12:40 Thursday, November 14, 2013

38            select dmigr ;
39         run;

NOTE: Copying WORK.DMIGR to XPTOUT.DMIGR (memtype=DATA).
NOTE: There were 1203642 observations read from the data set WORK.DMIGR.
NOTE: The data set XPTOUT.DMIGR has 1203642 observations and 7 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           17.56 seconds
      cpu time            0.68 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           23.65 seconds
      cpu time            3.38 seconds

The SAS System                                                   12:40 Thursday, November 14, 2013   1

The CONTENTS Procedure

Data Set Name        DATA.DMIGR                             Observations          1203642
Member Type          DATA                                   Variables             8
Engine               V9                                     Indexes               0
Created              13. november 2013 onsdag 14:21:00     Observation Length    64
Last Modified        13. november 2013 onsdag 14:21:00     Deleted Observations  0
Protection                                                  Compressed            NO
Data Set Type                                               Sorted                NO
Label
Data Representation  WINDOWS_32
Encoding             wlatin1  Western (Windows)


                     Engine/Host Dependent Information

Data Set Page Size          8192
Number of Data Set Pages    9478
First Data Page             1
Max Obs per Page            127
Obs in First Data Page      99
Number of Data Set Repairs  0
Filename                    C:\Bendix\Steno\MaEJ\Migrant\data\dmigr.sas7bdat
Release Created             9.0202M3
Host Created                W32_VSPRO
```

```
Alphabetic List of Variables and Attributes

#    Variable    Type    Len

8    doBTH       Num       8
4    doDM        Num       8
5    doDTH       Num       8
6    doIND       Num       8
1    doUD        Num       8
2    land        Char      4
3    region      Char      9
7    sex         Num       8
The SAS System                                        12:40 Thursday, November 14, 2013   2

                             N        NMiss       Min        Max

 region
 DK          doBTH         414,494         0 06/04/1889 20/10/2008
             doDM          414,494         0 15/09/1941 31/12/2009
             doIND               0   414,494          .          .
             doUD                0   414,494          .          .
             doDTH         160,202   254,292 01/05/1971 31/12/2009
 Africa      doBTH          36,862         0 17/12/1899 16/11/2009
             doDM            1,301    35,561 13/01/1989 30/12/2009
             doIND          35,134     1,728 13/02/1965 22/08/2011
             doUD           12,549    24,313 31/07/1989 07/10/2011
             doDTH             502    36,360 12/06/1995 01/10/2011
 America     doBTH          63,451         0 14/07/1899 24/12/2009
             doDM              714    62,737 01/01/1987 29/12/2009
             doIND          56,587     6,864 29/11/1909 03/10/2011
             doUD           37,489    25,962 05/02/1966 12/10/2011
             doDTH           1,013    62,438 25/04/1995 20/09/2011
 Asia        doBTH          92,592         0 05/01/1902 31/12/2009
             doDM            2,927    89,665 01/05/1989 30/12/2009
             doIND          84,098     8,494 27/09/1968 13/10/2011
             doUD           30,988    61,604 11/03/1982 11/10/2011
             doDTH           1,194    91,398 14/03/1995 07/10/2011
 Europe      doBTH         474,534         0 24/07/1895 31/12/2009
             doDM            9,062   465,472 04/07/1978 31/12/2009
             doIND         340,278   134,256 26/09/1929 12/10/2011
             doUD          271,149   203,385 04/06/1966 13/10/2011
             doDTH          13,035   461,499 05/01/1995 11/10/2011
 MidEast     doBTH         120,846         0 02/06/1896 26/12/2009
             doDM            8,593   112,253 13/04/1984 30/12/2009
             doIND         109,540    11,306 08/11/1928 11/10/2011
             doUD           18,990   101,856 12/06/1974 11/10/2011
             doDTH           2,630   118,216 13/04/1994 10/10/2011
 Other       doBTH             863         0 06/02/1907 14/03/2009
             doDM               62       801 14/03/1990 17/02/2009
             doIND             657       206 26/01/1971 14/02/2011
             doUD              278       585 07/07/1996 01/10/2011
             doDTH              42       821 07/04/1997 07/10/2011

The SAS System                                        12:40 Thursday, November 14, 2013   3

                                                    Total      doBTH          doDM           doDTH

                                                      N       N    NMiss     N    NMiss     N    NMiss

 DK
    .
       .                                          414,494 414,494     0 414,494     0 160,202 254,292
 Africa
    Africa
       5204: Angola                                   241     241     0       3   238       0     241
       5207: Botswana                                 100     100     0       0   100       2      98
       5213: Burundi                                  535     535     0      18   517       7     528
       5214: Etiopien                               1,826   1,826     0      68 1,758      24   1,802
       5215: Comorerne                                 19      19     0       0    19       1      18
       5216: Eritrea                                  274     274     0      26   248       2     272
       5222: Gambia                                   723     723     0      41   682      14     709
       5228: Ghana                                  2,355   2,355     0      68 2,287      25   2,330
       5230: Ækvatorial Guinea                          2       2     0       0     2       0       2
       5231: Guinea Bissau                             74      74     0       1    73       0      74
       5232: Guinea                                   117     117     0       2   115       3     114
       5233: Cap Verde                                 23      23     0       0    23       1      22
       5234: Kenya                                  1,706   1,706     0      35 1,671      24   1,682
       5235: Lesotho                                   64      64     0       2    62       0      64
       5236: Liberia                                  177     177     0       4   173       6     171
       5240: Mozambique                               334     334     0       2   332       3     331
       5242: Madagascar                                65      65     0       0    65       0      65
       5243: Mali                                      52      52     0       1    51       1      51
       5245: Mauritius                                146     146     0      12   134       6     140
       5246: Nigeria                                1,439   1,439     0      37 1,402      11   1,428
       5247: Namibia                                   81      81     0       1    80       0      81
       5255: Sierra Leone                             388     388     0      10   378      11     377
```

5259: Swaziland	59	59	0	0	59	1	58
5262: Sydafrikanske Rep	2,301	2,301	0	24	2,277	35	2,266
5266: Tanzania	1,235	1,235	0	24	1,211	17	1,218
5269: Uganda	1,661	1,661	0	43	1,618	44	1,617
5276: Centr Afrikanske Rep	13	13	0	0	13	0	13
5277: Cameroun	750	750	0	11	739	1	749
5278: Den Demo.Rep.Congo	667	667	0	13	654	6	661
5279: Republikken Congo	665	665	0	11	654	5	660
5281: Benin	83	83	0	1	82	1	82
5282: Elfenbenskysten	545	545	0	12	533	7	538
5283: Gabon	10	10	0	0	10	0	10
5284: Mauretanien	34	34	0	0	34	0	34
5285: Niger	26	26	0	0	26	0	26
5287: Rwanda	337	337	0	12	325	6	331
5288: Senegal	164	164	0	3	161	2	162
5289: Somalia	15,671	15,671	0	792	14,879	208	15,463
5292: Tchad	14	14	0	0	14	0	14
5293: Togo	179	179	0	5	174	4	175
5294: Burkina Faso	87	87	0	1	86	0	87
5295: Zimbabwe	563	563	0	5	558	4	559
5296: Zambia	680	680	0	5	675	3	677
5297: Malawi	129	129	0	0	129	0	129
5298: Seychellerne	24	24	0	2	22	3	21
5299: Afrika	196	196	0	3	193	14	182
5525: Djibouti	27	27	0	3	24	0	27
5621: Sao Tome Og Principe	1	1	0	0	1	0	1
America							
Oceania							
5274: Kiribati	2	2	0	0	2	0	2
5275: Vanuatu	4	4	0	0	4	0	4
5502: Australien	6,050	6,050	0	30	6,020	32	6,018
5505: Tonga	24	24	0	0	24	0	24
5508: Fiji	30	30	0	0	30	1	29
5514: New Zealand	1,610	1,610	0	10	1,600	17	1,593
5522: Samoa	19	19	0	0	19	0	19
5534: Papua New Guinea	14	14	0	0	14	0	14
5599: Øer I Stillehavet	8	8	0	1	7	0	8
5623: Salomon-Øerne	3	3	0	0	3	0	3
5779: Cook Islands	1	1	0	0	1	0	1
America							
5302: Argentina	1,565	1,565	0	46	1,519	98	1,467
5303: Bahama-Øerne	20	20	0	0	20	0	20
5304: Bolivia	536	536	0	3	533	2	534
5305: Barbados	52	52	0	4	48	1	51
5306: Brasilien	4,614	4,614	0	47	4,567	31	4,583
5308: Guyana	102	102	0	7	95	4	98
5309: Antigua og Barbuda	3	3	0	0	3	0	3
5311: St Vincent Og Grenad	5	5	0	0	5	0	5
5314: Canada	6,682	6,682	0	90	6,592	139	6,543
5316: Chile	1,812	1,812	0	68	1,744	46	1,766
5318: Colombia	3,083	3,083	0	9	3,074	19	3,064
5322: Costa Rica	262	262	0	1	261	1	261
5324: Cuba	552	552	0	8	544	5	547
5326: Dominikanske Rep	169	169	0	1	168	1	168
5328: Ecuador	649	649	0	6	643	8	641
5338: Guatemala	262	262	0	2	260	1	261
5339: Grenada	11	11	0	0	11	0	11
5342: Haiti	151	151	0	0	151	4	147
5344: Suriname	27	27	0	2	25	3	24
5345: Dominica	14	14	0	0	14	0	14
5347: St.Lucia	9	9	0	0	9	0	9
5348: Honduras	181	181	0	0	181	2	179
5352: Jamaica	195	195	0	4	191	5	190
5354: Mexico	1,671	1,671	0	17	1,654	8	1,663
5356: Nicaragua	208	208	0	4	204	3	205
5358: Panama	92	92	0	4	88	4	88
5364: Paraguay	105	105	0	4	101	2	103
5366: Peru	1,154	1,154	0	13	1,141	7	1,147
5372: El Salvador	83	83	0	4	79	1	82
5374: Trinidad og Tobago	173	173	0	12	161	6	167
5376: Uruguay	276	276	0	13	263	8	268
5390: USA	30,099	30,099	0	290	29,809	536	29,563
5392: Venezuela	811	811	0	10	801	10	801
5395: Vestindiske Øer	26	26	0	3	23	6	20
5397: Nordamerika	6	6	0	0	6	1	5
5398: Syd Og Mell.Amerika	13	13	0	0	13	1	12
5526: Belize	12	12	0	1	11	0	12
5625: St.Christop.Og Nevis	1	1	0	0	1	0	1
Asia							
Asia							
5408: Bhutan	299	299	0	1	298	0	299
5410: Bangladesh	968	968	0	33	935	7	961
5412: Brunei Darussalem	8	8	0	1	7	0	8
5414: Myanmar (Burma)	1,184	1,184	0	24	1,160	6	1,178
5416: Cambodia	220	220	0	8	212	2	218
5418: Sri Lanka	8,279	8,279	0	1,198	7,081	215	8,064
5424: Taiwan	484	484	0	10	474	3	481
5432: Indien	12,154	12,154	0	346	11,808	148	12,006
5434: Indonesien	1,387	1,387	0	27	1,360	55	1,332

5444: Japan	4,211	4,211	0	43	4,168	47	4,164
5448: Kina	17,282	17,282	0	152	17,130	146	17,136
5454: Laos	77	77	0	2	75	1	76
5457: Maldiverne	12	12	0	0	12	0	12
5458: Malaysia	976	976	0	20	956	17	959
5459: Mongoliet	123	123	0	0	123	0	123
5464: Nepal	1,822	1,822	0	3	1,819	1	1,821
5466: Korea,Demo.Folkerep	458	458	0	3	455	1	457
5474: Filippinerne	11,224	11,224	0	207	11,017	70	11,154
5482: Singapore	1,112	1,112	0	12	1,100	12	1,100
5484: Korea,Republik	8,896	8,896	0	155	8,741	61	8,835
5488: Vietnam	9,745	9,745	0	512	9,233	288	9,457
5492: Thailand	11,651	11,651	0	166	11,485	113	11,538
5499: Asien	20	20	0	4	16	1	19
Europe							
Europe							
5104: Finland	7,965	7,965	0	210	7,755	403	7,562
5105: Island,Ligeret Dansk	33	33	0	0	33	0	33
5106: Island	21,796	21,796	0	152	21,644	229	21,567
5107: Liechtenstein	28	28	0	0	28	0	28
5108: Luxembourg	597	597	0	1	596	0	597
5109: Monaco	32	32	0	1	31	1	31
5110: Norge	33,577	33,577	0	794	32,783	1,658	31,919
5120: Sverige	49,667	49,667	0	748	48,919	1,686	47,981
5124: Andorra	2	2	0	0	2	0	2
5126: Belgien	3,735	3,735	0	36	3,699	65	3,670
5130: Frankrig	16,718	16,718	0	96	16,622	156	16,562
5134: Grækenland	2,596	2,596	0	46	2,550	42	2,554
5140: Nederlandene	10,452	10,452	0	154	10,298	198	10,254
5142: Irland	2,561	2,561	0	28	2,533	49	2,512
5150: Italien	11,221	11,221	0	130	11,091	138	11,083
5153: Malta	143	143	0	2	141	4	139
5156: Portugal	2,422	2,422	0	29	2,393	17	2,405
5159: San Marino	1	1	0	0	1	0	1
5160: Schweiz	3,855	3,855	0	49	3,806	116	3,739
5164: Spanien	12,707	12,707	0	68	12,639	78	12,629
5170: Storbritannien	27,017	27,017	0	462	26,555	722	26,295
5176: Vatikanstaten	1	1	0	0	1	0	1
5180: Tyskland,Forb.Rep	73,788	73,788	0	1,640	72,148	3,214	70,574
5182: Østrig	3,393	3,393	0	80	3,313	176	3,217
5199: Europa	29	29	0	3	26	6	23
5422: Cypern	170	170	0	7	163	5	165
5902: Færøerne	19,025	19,025	0	496	18,529	800	18,225
East Block							
5122: Albanien	391	391	0	6	385	0	391
5128: Bulgarien	4,173	4,173	0	26	4,147	22	4,151
5129: Tjekkoslovakiet	751	751	0	38	713	70	681
5151: Serbien & Montenegro	1,087	1,087	0	27	1,060	6	1,081
5152: Jugoslavien	9,556	9,556	0	816	8,740	521	9,035
5154: Polen	58,023	58,023	0	512	57,511	803	57,220
5158: Rumænien	10,692	10,692	0	104	10,588	74	10,618
5162: Sovjetunionen	1,209	1,209	0	93	1,116	232	977
5174: Ungarn	4,966	4,966	0	100	4,866	130	4,836
5607: Estland	2,907	2,907	0	12	2,895	14	2,893
5609: Letland	6,918	6,918	0	11	6,907	40	6,878
5611: Litauen	15,521	15,521	0	19	15,502	26	15,495
5700: Rusland	7,693	7,693	0	54	7,639	51	7,642
5704: Ukraine	11,492	11,492	0	25	11,467	32	11,460
5706: Belarus	841	841	0	5	836	4	837
5708: Armenien	712	712	0	27	685	15	697
5710: Azerbajdjan	296	296	0	7	289	6	290
5712: Moldova	442	442	0	0	442	1	441
5714: Uzbekistan	352	352	0	2	350	4	348
5716: Kazakhstan	351	351	0	2	349	0	351
5718: Turkmenistan	21	21	0	0	21	0	21
5720: Kirgizstan	58	58	0	1	57	0	58
5722: Tadzhikistan	32	32	0	0	32	0	32
5724: Georgien	292	292	0	4	288	2	290
5750: Kroatien	921	921	0	32	889	25	896
5752: Slovenien	635	635	0	2	633	3	632
5754: Bosnien-Herzegovina	19,564	19,564	0	1,696	17,868	1,099	18,465
5756: Makedonien	2,538	2,538	0	82	2,456	45	2,493
5757: Serbien	564	564	0	10	554	1	563
5758: Jugoslavien,Forb.Rep	1,695	1,695	0	75	1,620	37	1,658
5759: Montenegro	91	91	0	4	87	0	91
5761: Kosovo	751	751	0	28	723	4	747
5776: Tjekkiske Republik	2,753	2,753	0	7	2,746	3	2,750
5778: Slovakiet	2,715	2,715	0	3	2,712	2	2,713
MidEast							
Middle East							
5172: Tyrkiet	31,115	31,115	0	2,393	28,722	799	30,316
5202: Algeriet	1,007	1,007	0	66	941	30	977
5238: Libyen	254	254	0	13	241	6	248
5244: Marokko	5,048	5,048	0	402	4,646	122	4,926
5258: Sudan	737	737	0	38	699	10	727
5268: Tunesien	944	944	0	36	908	15	929
5272: Egypten	1,874	1,874	0	182	1,692	64	1,810
5406: Bahrain	114	114	0	1	113	0	114
5436: Irak	23,167	23,167	0	1,447	21,720	375	22,792

```
        5438: Iran                                13,800  13,800       0     768  13,032     313  13,487
        5442: Israel                               2,213   2,213       0      94   2,119      39   2,174
        5446: Jordan                               1,084   1,084       0     106     978      26   1,058
        5452: Kuwait                               1,505   1,505       0      62   1,443       9   1,496
        5462: Oman                                    27      27       0       0      27       1      26
        5478: Saudi Arabien                          349     349       0       0     349       0     349
        5486: Syrien                               2,163   2,163       0     139   2,024      35   2,128
        5487: Mellemøsten                            331     331       0      84     247      51     280
        5496: Qatar                                  118     118       0       0     118       0     118
    Asia
        5402: Yemen,Demo.Folkerep                    149     149       0      21     128       5     144
        5403: Foren.Arab.Emirater                    441     441       0       3     438       4     437
        5404: Afghanistan                         10,258  10,258       0     435   9,823     118  10,140
        5456: Libanon                             12,524  12,524       0     889  11,635     265  12,259
        5472: Pakistan                            11,624  11,624       0   1,414  10,210     343  11,281
 Other
    ?
        5102: Udlandet                               646     646       0      37     609      25     621
        5103: Statsløs                               185     185       0      21     164       8     177
        5800: Udlændingeservice                        4       4       0       0       4       1       3
        5999: Land Ukendt                             28      28       0       4      24       8      20
```

1.2 Data entry

The data in the just created SAS-xport file which contains records of all person who either

- are born outside of Denmark or
- has a diagnosis of DM

Thus the only persons not included here are persons born Denmark with no recorded DM.

First we read data and then groom the dataset a little:

```
> options( width=95 )
> memory.size(3500)
[1] 3500
> library(Epi)
> library(foreign)
> print( sessionInfo(), l=F )
R version 3.1.1 (2014-07-10)
Platform: i386-w64-mingw32/i386 (32-bit)

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

other attached packages:
[1] Epi_1.1.67     foreign_0.8-61

loaded via a namespace (and not attached):
[1] tools_3.1.1
> dmigr <- read.xport("./data/dmigr.xpt")
> names( dmigr ) <- tolower( names(dmigr) )
> levels( dmigr$region )[1] <- "DK"
> str( dmigr )
'data.frame':        1203642 obs. of  7 variables:
 $ doud  : num  NA NA NA NA NA ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 1 5 6 6 6 6 5 5 4 ...
 $ dodm  : num  10981 16700 NA NA NA ...
 $ dodth : num  11499 NA NA NA NA ...
 $ doind : num  NA NA 15183 15369 15519 ...
 $ sex   : num  2 2 1 1 1 1 1 1 2 1 ...
 $ dobth : num  -21914 14610 14610 14610 14610 ...
```

Sanity check:

```
> with( dmigr, ftable( region, Dead=!is.na(dodth),
+                                  DM=!is.na(dodm), col.vars=3:2 ) )
```

```
        DM    FALSE          TRUE
        Dead  FALSE   TRUE  FALSE   TRUE
region
DK                0      0 254292 160202
Africa        35144    417   1216     85
America       61878    859    560    154
Asia          88755    910   2643    284
Europe       454635  10837   6864   2198
MidEast      110452   1801   7764    829
Other           775     26     46     16
```

Then we transform dates to date-format, and subsequently transform all date variables in the data frame to `cal.yr` format:

```
> wh.var <-
+ function( dfr, pat )
+ {
+ print( names( dfr )[wh <- grep( pat, names(dfr) )] )
+ invisible( wh )
+ }
> dv <- wh.var( dmigr, "do" )
[1] "doud"  "dodm"  "dodth" "doind" "dobth"
> names( dmigr )[dv <- grep( "do", names(dmigr) )]
[1] "doud"  "dodm"  "dodth" "doind" "dobth"
> for( i in dv ) dmigr[,i] <- as.Date( dmigr[,i], origin="1960-01-01" )
> dmigr$sex <- factor( dmigr$sex, labels=c("M","F") )
> dmigr <- cal.yr( dmigr )
```

We then restrict the data by excluding persons that are dead or emigrated before 1.1.1995, have no date of birth or have date of DM on or after date of death:

```
> dim( dmigr )
[1] 1203642       7
> dmigr <- subset( dmigr, pmin( dodth, doud, 1995, na.rm=TRUE ) >= 1995 &
+                          !is.na(dobth) &
+                          ( is.na(dodm) | is.na(dodth) | (dodm < dodth) ) )
> dim( dmigr )
[1] 1176173       7
> zz <- formatC( with( dmigr, addmargins( table( Emigr=!is.na(doud),
+                                          Immigr=!is.na(doind) ) ) ),
+           format="f", big.mark=",", digits=0, pre="common" )
> print( zz, q=F )
       Immigr
Emigr   FALSE     TRUE      Sum
  FALSE   438,828   366,086   804,914
  TRUE    111,210   260,049   371,259
  Sum     550,038   626,135 1,176,173
```

Not all emigration dates are after immigration dates, so we assume that these are cases of re-immigration, and we decide just to follow these persons from the date of the immigration (`doind`), and ignore the earlier emigration date (`doud`) by setting the latter to `NA`:

```
> dmigr$doud <- with( dmigr, ifelse( doud>doind, doud, NA ) )
> with( dmigr, table( doud>doind, exclude=NULL ) )
  TRUE   <NA>
259935 916238
```

To get an overview of the material, we make scatter-plots of selected date variables, to check whether their joint ranges and distributions look sensible:

```
> dmigdk <- subset( dmigr , region != "DK" )
> nr <- nrow( dmigdk )
> dmigdk <- transform( dmigdk,
+                      doind = pmin(doind,2015-runif(nr,-0.6,0.2),na.rm=TRUE),
+                      doud  = pmin(doud ,2015-runif(nr,-0.6,0.2),na.rm=TRUE) )
> dmig10 <- subset( dmigdk, runif(nr)<0.1 )
> par( mfrow=c(2,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> with( dmig10, plot( doud , doind, pch=16, cex=0.3, ylim=c(1960,2015) ) )
> with( dmigdk, plot( dodth, doind, pch=16, cex=0.3, ylim=c(1960,2015) ) )
> with( dmig10, plot( dodm , doind, pch=16, cex=0.3, ylim=c(1980,2015), xlim=c(1980,2015) ) )
> with( dmigdk, plot( dodm , doud , pch=16, cex=0.3, ylim=c(1980,2015), xlim=c(1980,2015) ) )
```

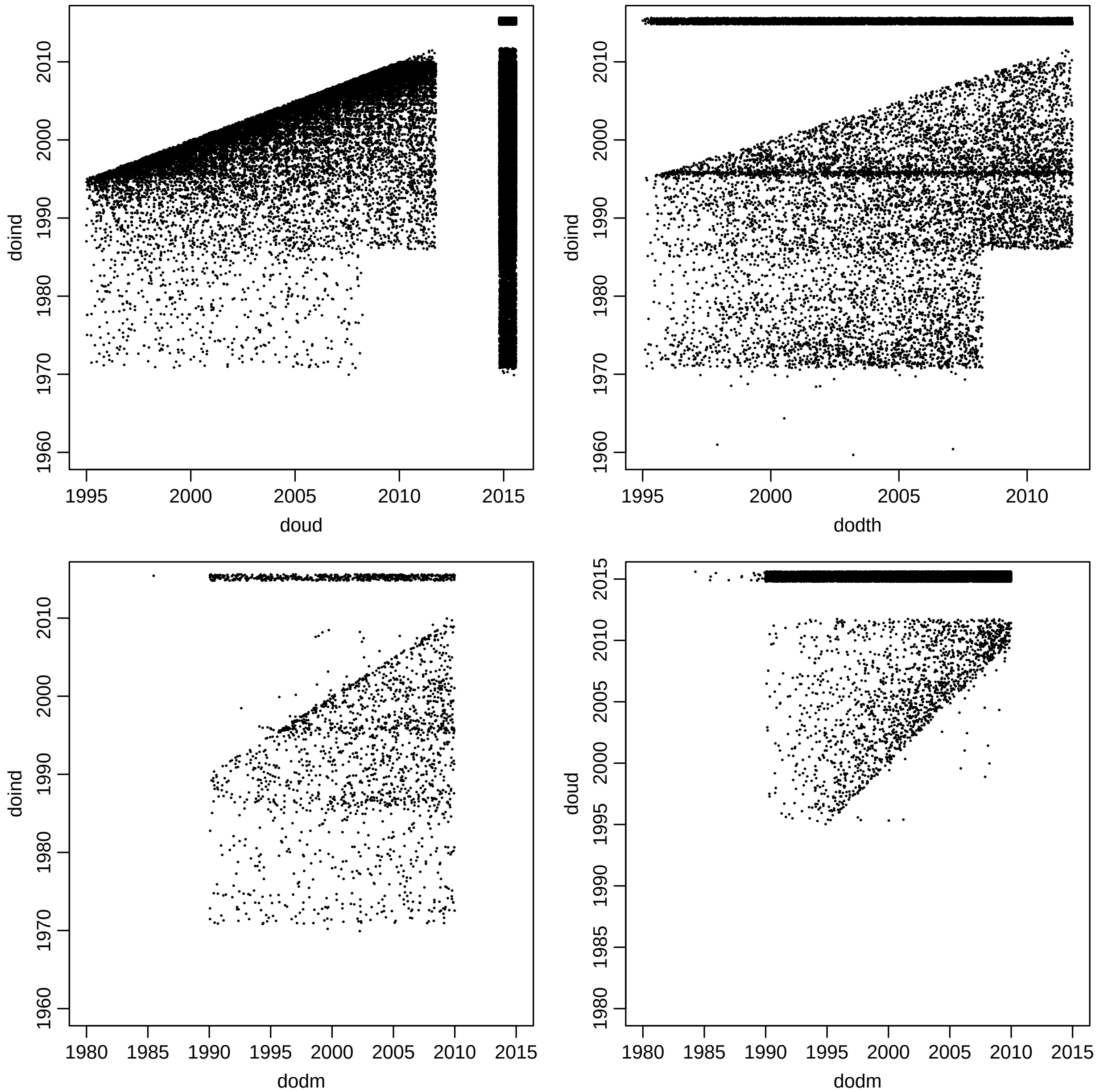

Figure 1.1: *Scatter plots of dates for foreign born persons, based on a 10% randomm sample. Missing values for* doind *and* doud *are set to 2015. It is obvious that something fishy is going on for persons with immigration before 1985.*

The plots in figure 1.1 show that there is a problem around 2008 for `doud` and `dodth` and 1985 for `doin`. The phenomenon is a consquence of an administrative cleanup in 2008, where persons immigrated before 1985 have had their immigration date deleted, but not their country of birth. Therefore this does not affect persons emigrated or dead before 2008.

```
> par( mfrow=c(3,2) )
> with( dmigr, hist(dobth,breaks=100, col=gray(0.2), border=gray(0.2) ) )
> with( dmigr, hist(dodth,breaks=100, col=gray(0.2), border=gray(0.2) ) )
> with( dmigr, hist(doind,breaks=100, col=gray(0.2), border=gray(0.2) ) )
> with( dmigr, hist(doud ,breaks=100, col=gray(0.2), border=gray(0.2) ) )
> with( dmigr, hist(dodm ,breaks=100, col=gray(0.2), border=gray(0.2) ) )
```

From the histograms in figure 1.2 it is clear that the death dates are incomplete beyond 01.01.2010, so we set the end of follow-up to 01.01.2010:

```
> ( end <- cal.yr( as.Date("2010-01-01") ) )
[1] 2010
attr(,"class")
[1] "cal.yr"  "numeric"
```

We can explore the apparent seasonality of the emigration date, by listing those dates that occur more the 500 times in the material:

```
> tt <- table( round(dmigr$doud,3) )
> cbind(tt[tt>400])
          [,1]
1998.496   416
1998.999   435
1999.495   455
2000.497   492
2001.496   478
2001.581   421
2002       551
2002.496   605
2003.492   539
2003.495   513
2003.996   530
2004.494   470
2004.497   680
2004.998   604
2005.414   411
2005.493   511
2005.496   840
2005.581   426
2005.997   646
2006.422   415
2006.493   588
2006.496   573
2006.58    437
2006.665   414
2006.997   507
2007.421   429
2007.492   663
2007.495   538
2007.58    501
2007.999   630
2008.015   492
2008.084   422
2008.409   548
2008.494   475
2008.497   741
2008.582   613
2008.667   470
```

```
2008.998   969
2009.001   670
2009.086   440
2009.417   625
2009.493   537
2009.496   815
2009.581   577
2009.666   536
2009.997  1060
2010       509
2010.085   504
2010.496   647
2010.58    417
2010.665   409
```

From this it is clear that dates 1 January and 1 July are clearly over-respresented in the material — this is likely due to the fact that some immigrants either do not know or do not want to disclose their true birthday. Therefore we should also be aware that age-information on immigrants is not always accurate.

In order to handle the follow-up properly, we define entry and exit dates. Note that we use the `na.rm=TRUE` argument to make sure that we get a valid date for all. Also note that we end follow up at `end` as defined above.

```
> dmigr <- transform( dmigr, entry = pmax( dobth, doind, 1995, na.rm=TRUE ),
+                             exit = pmin( dodth, doud ,  end, na.rm=TRUE ),
+                            dodm = pmin( dodm, na.rm=TRUE) )
> summary( dmigr )
      doud             region             dodm              dodth                doind
 Min.   :1995     DK      :387213   Min.   :1942     Min.   :1995       Min.   :1910
 1st Qu.:2001     Africa : 36856   1st Qu.:1995     1st Qu.:2000       1st Qu.:1996
 Median :2006     America: 63428   Median :2001     Median :2004       Median :2001
 Mean   :2005     Asia   : 92582   Mean   :2001     Mean   :2003       Mean   :2000
 3rd Qu.:2009     Europe :474424   3rd Qu.:2006     3rd Qu.:2007       3rd Qu.:2006
 Max.   :2012     MidEast:120807   Max.   :2010     Max.   :2012       Max.   :2012
 NA's   :916238   Other  :   863   NA's   :766304   NA's   :1024840    NA's   :550038
 sex            dobth             entry             exit
 M:619696   Min.   :1890   Min.   :1995   Min.   :1995
 F:556477   1st Qu.:1945   1st Qu.:1995   1st Qu.:2008
            Median :1968   Median :1995   Median :2010
            Mean   :1963   Mean   :1998   Mean   :2008
            3rd Qu.:1981   3rd Qu.:2002   3rd Qu.:2010
            Max.   :2010   Max.   :2012   Max.   :2010
```

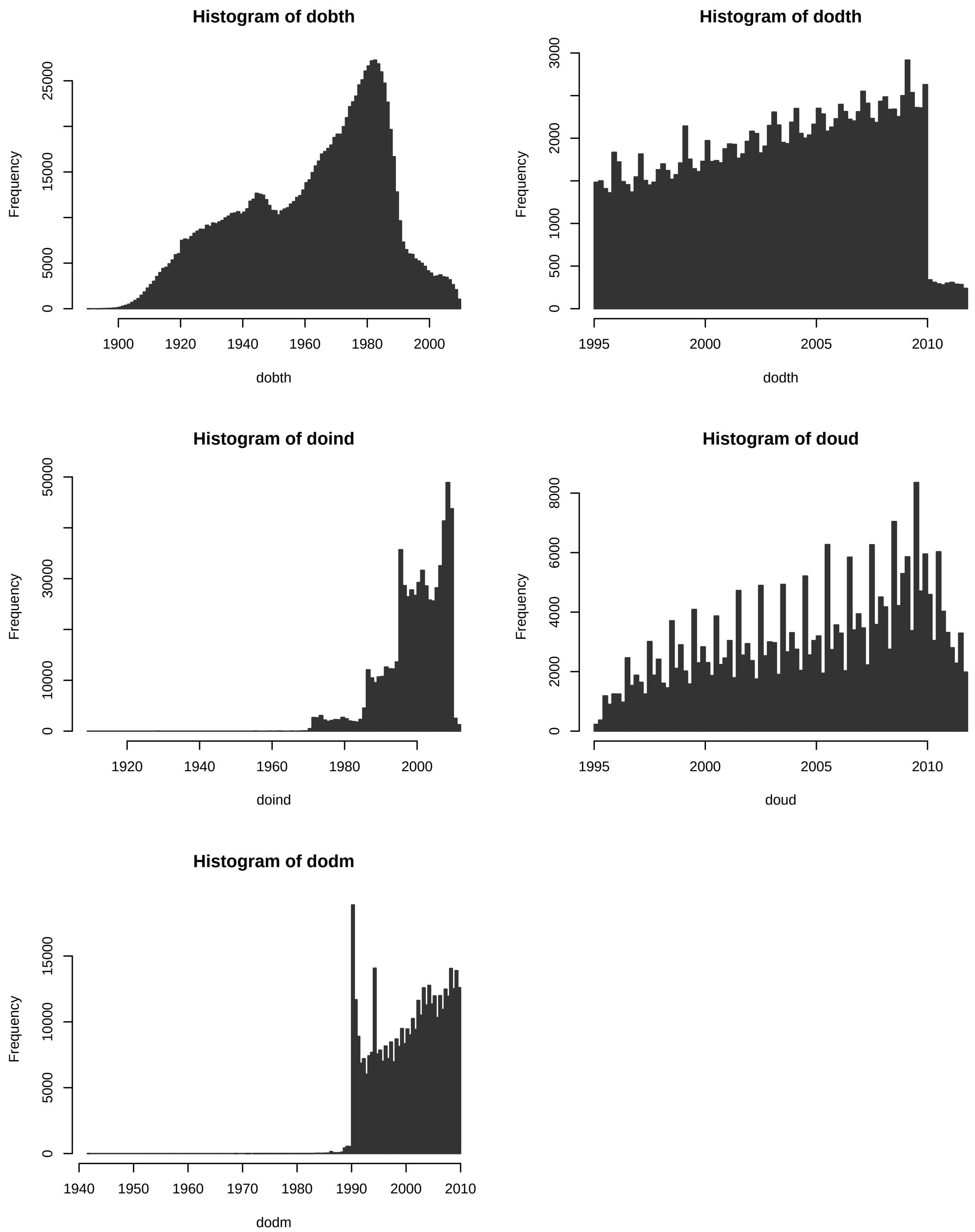

Figure 1.2: *Histograms of all the date variables. The very distinct seasonality of* `doud` *are from the massive over-representation of the dates 1 January and 1 July as seen below.*

1.3 Follow-up

We now set up a `Lexis` object to represent the follow-up; in the first instance just from start till end of follow-up (emigration, death, end of study):

```
> Lx <- Lexis( entry = list( date=entry,
+                            age=entry-dobth),
+               exit = list( date=exit ),
+        exit.status = factor( !is.na(dodth), labels=c("Well","Dead") ),
+               data = subset( dmigr, entry<exit ) )
NOTE: entry.status has been set to "Well" for all.
> summary( Lx )
Transitions:
     To
From      Well   Dead  Records:  Events: Risk time:  Persons:
  Well 1021005 151317   1172322   151317   11301352   1172322
```

For future reference we save the Lexis object:

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

We now cut the follow-up at the date of diagnosis of diabetes; however this is memory-demanding task, so we do this in chuncks:

```
> n.chunks <- 20
> lm <- round( seq(0,nrow(Lx),,n.chunks+1) )
> cLx <- NULL
> for( i in 1:n.chunks )
+    {
+    whr <- (lm[i]+1):(lm[i+1])
+    xL  <- Lx[whr,]
+ cat( "Start chunk ", i,
+      format(Sys.time(),format="%Y-%m-%d %H:%M:%S"), "\n" )
+ flush.console()
+ cLx <- rbind( cLx, cutLexis( xL, cut = xL$dodm,
+                                  pre = "Well",
+                               new.st = "DM",
+                               new.sc = "DMdur",
+                             split.st = TRUE ) )
+    }
Start chunk  1 2014-08-06 12:58:52
Start chunk  2 2014-08-06 12:59:05
Start chunk  3 2014-08-06 12:59:19
Start chunk  4 2014-08-06 12:59:35
Start chunk  5 2014-08-06 12:59:52
Start chunk  6 2014-08-06 13:00:09
Start chunk  7 2014-08-06 13:00:27
Start chunk  8 2014-08-06 13:00:45
Start chunk  9 2014-08-06 13:01:03
Start chunk  10 2014-08-06 13:01:22
Start chunk  11 2014-08-06 13:01:41
Start chunk  12 2014-08-06 13:01:59
Start chunk  13 2014-08-06 13:02:18
Start chunk  14 2014-08-06 13:02:37
Start chunk  15 2014-08-06 13:02:57
Start chunk  16 2014-08-06 13:03:17
Start chunk  17 2014-08-06 13:03:37
Start chunk  18 2014-08-06 13:03:59
Start chunk  19 2014-08-06 13:04:19
Start chunk  20 2014-08-06 13:04:41
```

Once we have cut the follow-up so that we have follow-up through the three states, we can show the amount of risk time and the transition rates between the states.

```
> summary( subset( cLx, region=="DK" ) )
Transitions:
     To
From    Well      DM Dead Dead(DM)  Records:  Events: Risk time:  Persons:
  Well     0 291281    0        0    291281   291281    2395675    291281
  DM       0 254292    0   132921    387213   132921    2488707    387213
  Sum      0 545573    0   132921    678494   424202    4884382    387213
> summary( subset( cLx, region!="DK" ) )
Transitions:
     To
From      Well     DM  Dead Dead(DM)  Records:  Events: Risk time:  Persons:
  Well 747677 19765 14836        0    782278    34601  6276435.9    782278
  DM        0 19036     0     3560     22596     3560   140534.1     22596
  Sum  747677 38801 14836     3560    804874    38161  6416970.0    785109
> par( mfrow=c(2,1) )
> boxes.Lexis( subset( cLx, region=="DK" ),
+                    boxpos=list( x=c(20,20,80,80),
+                                 y=c(80,15,80,15) ),
+                    show.BE=TRUE,
+                    hmult=1.2, wmult=1.1, scale.Y=1000, scale.R=1 )
> text( 10, 95, "DK", adj=c(0,1), cex=1.2, font=2 )
> boxes.Lexis( subset( cLx, region!="DK" ),
+                    boxpos=list( x=c(20,20,80,80),
+                                 y=c(80,15,80,15) ),
+                    show.BE=TRUE,
+                    hmult=1.2, wmult=1.1, scale.Y=1000, scale.R=1 )
> text( 10, 95, "not DK", adj=c(0,1), cex=1.2, font=2 )
```

The rates from the state ”Well” in figure 1.3 in the top panel are strongly misleading as the persons included here all contribute to the DM risk time, as is evident from the 0 persons ending in the “Well” state.

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

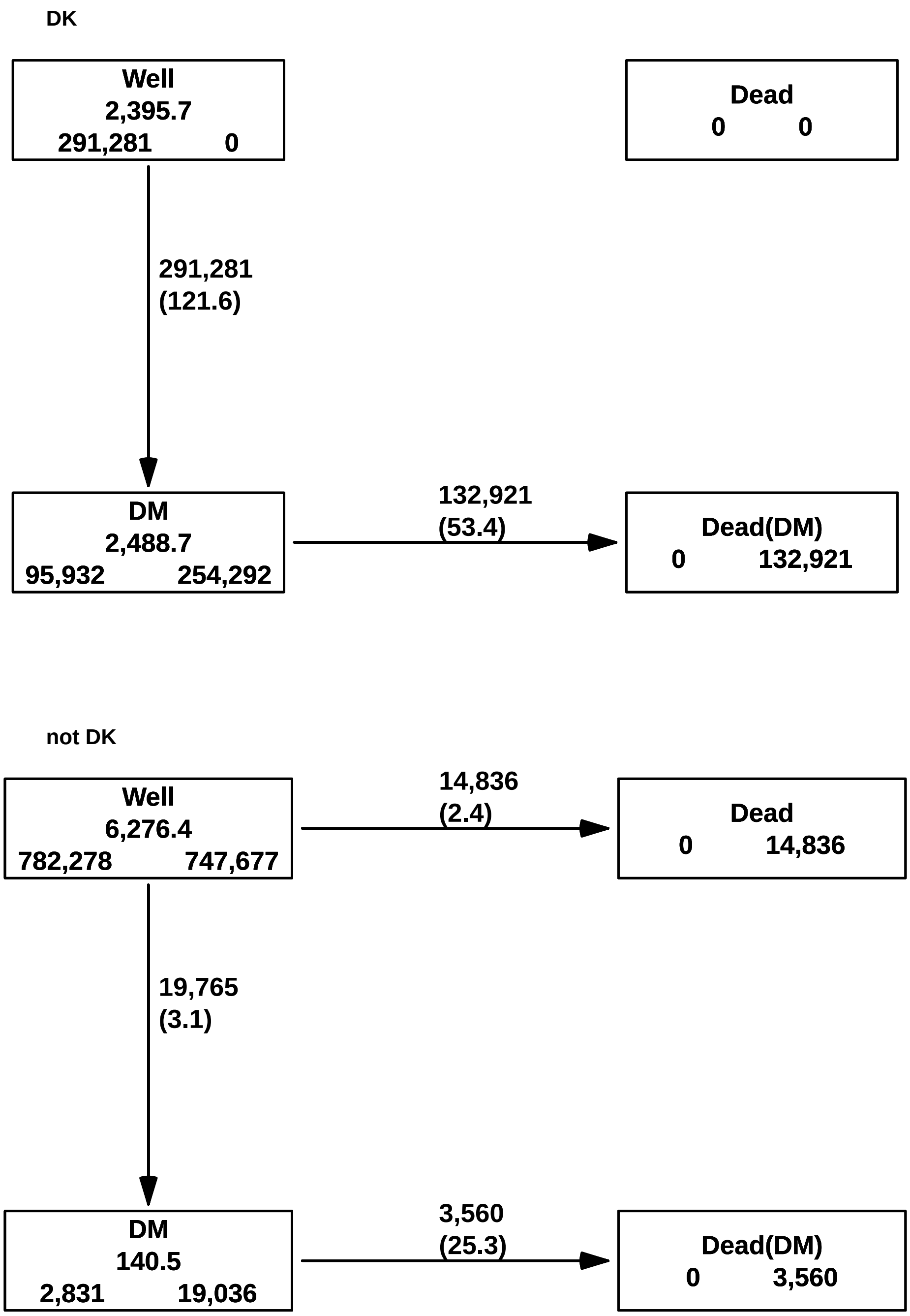

Figure 1.3: *States and transitions between them. Numbers in boxes are person-years in 1000s and number of persons starting, respectively ending in each state. Numbers on the arrows are number of transitions and transition rates per 1000 person-years. Top panel: Danish born persons, bottom panel: foreign born persons.*

Chapter 2

Splitting follow-up and adding population data

2.1 Splitting follow-up

We now split the follow-up data by age and calendar time in bands of 1 year in order to classify the risk time among those with diabetes and of foreign birth by sex, age and date of follow-up. We shall subsequently subtract the thus derived risk time from the overall population as obtained from Statistics Denmark, in order to obtain the correct risk time figures for the `Well` state for those born in Denmark.

In practice the time-splitting will produce some 30 intervals per person, so about 30 million intervals, which will not fit into this crap little office computer.

So we split the data for smaller chunks of `cLx` at at time, and aggregate the risk time and TB events into a dataset. This is then merged with and used to update the previous dataset, so we get a sequential updating of events and risk time (as well as a slowly increasing number of rows, as each chunk of the `Lexis` object contains a few combinations of the classifying factors that have not been encountered in previous chunks:

```
> Agg <- data.frame( A=0, P=0, U=0,
+                    cLx[1,c("sex","region","lex.Cst")],
+                    Y=0, D.dm=0, D.dd=0 )[NULL,]
> names( Agg )[6] <- "state"
> str( Agg )
'data.frame':        0 obs. of  9 variables:
 $ A     : num
 $ P     : num
 $ U     : num
 $ sex   : Factor w/ 2 levels "M","F":
 $ region: Factor w/ 7 levels "DK","Africa",..:
 $ state : Factor w/ 4 levels "Well","DM","Dead",..:
 $ Y     : num
 $ D.dm  : num
 $ D.dd  : num
> n.chunks <- 100
> lm <- round( seq(0,nrow(cLx),,n.chunks+1) )
> for( i in 1:n.chunks )
+ {
+ whr <- (lm[i]+1):(lm[i+1])
+ sLx <- splitLexis( cLx[whr,], 0:100, time.scale="age" )
+ sLx <- splitLexis( sLx,  1995:2012, time.scale="date" )
+ agg <- with( sLx, aggregate( cbind( y = lex.dur,
+                                     d.dm = ( lex.Xst == "DM" &
+                                              lex.Xst != lex.Cst )*1,
```

```
+                                  d.dd = ( lex.Xst %in% c("Dead","Dead(DM)") )*1 ),
+                              list( A = floor(age),
+                                    P = floor(date),
+                                    U = floor(date)-floor(age)-floor(dobth),
+                                  sex = sex,
+                               region = region,
+                                state = lex.Cst ),
+                              FUN = sum ) )
+ Agg <- merge( Agg, agg, by=names( Agg )[1:6], all=TRUE )
+ Agg <- transform( Agg, Y = pmax(Y   ,0,na.rm=TRUE) + pmax(y   ,0,na.rm=TRUE),
+                     D.dm = pmax(D.dm,0,na.rm=TRUE) + pmax(d.dm,0,na.rm=TRUE),
+                     D.dd = pmax(D.dd,0,na.rm=TRUE) + pmax(d.dd,0,na.rm=TRUE) )[,
+                   c("A","P","U","sex","region","state","Y","D.dm","D.dd")]
+ cat( "Merged in chunk", i, "now", nrow(Agg), "rows, at",
+      format(Sys.time(),format="%Y-%m-%d %H:%M:%S"), "\n" )
+ }

Merged in chunk 1 now 29503 rows, at 2013-11-14 13:00:56
Merged in chunk 2 now 37875 rows, at 2013-11-14 13:01:25
Merged in chunk 3 now 41388 rows, at 2013-11-14 13:01:54
Merged in chunk 4 now 43658 rows, at 2013-11-14 13:02:21
Merged in chunk 5 now 45754 rows, at 2013-11-14 13:02:49
Merged in chunk 6 now 47035 rows, at 2013-11-14 13:03:16
Merged in chunk 7 now 47995 rows, at 2013-11-14 13:03:44
Merged in chunk 8 now 48783 rows, at 2013-11-14 13:04:10
Merged in chunk 9 now 49408 rows, at 2013-11-14 13:04:37
Merged in chunk 10 now 50073 rows, at 2013-11-14 13:05:03
Merged in chunk 11 now 50729 rows, at 2013-11-14 13:05:29
Merged in chunk 12 now 51404 rows, at 2013-11-14 13:05:56
Merged in chunk 13 now 51924 rows, at 2013-11-14 13:06:22
Merged in chunk 14 now 52392 rows, at 2013-11-14 13:06:49
Merged in chunk 15 now 52976 rows, at 2013-11-14 13:07:16
Merged in chunk 16 now 53477 rows, at 2013-11-14 13:07:44
Merged in chunk 17 now 53862 rows, at 2013-11-14 13:08:10
Merged in chunk 18 now 54276 rows, at 2013-11-14 13:08:38
Merged in chunk 19 now 54568 rows, at 2013-11-14 13:09:06
Merged in chunk 20 now 54935 rows, at 2013-11-14 13:09:32
Merged in chunk 21 now 55297 rows, at 2013-11-14 13:09:59
Merged in chunk 22 now 55597 rows, at 2013-11-14 13:10:27
Merged in chunk 23 now 55879 rows, at 2013-11-14 13:10:53
Merged in chunk 24 now 56232 rows, at 2013-11-14 13:11:21
Merged in chunk 25 now 56446 rows, at 2013-11-14 13:11:47
Merged in chunk 26 now 56683 rows, at 2013-11-14 13:12:16
Merged in chunk 27 now 57045 rows, at 2013-11-14 13:12:42
Merged in chunk 28 now 57147 rows, at 2013-11-14 13:13:08
Merged in chunk 29 now 57319 rows, at 2013-11-14 13:13:35
Merged in chunk 30 now 57539 rows, at 2013-11-14 13:14:02
Merged in chunk 31 now 57854 rows, at 2013-11-14 13:14:28
Merged in chunk 32 now 57993 rows, at 2013-11-14 13:14:57
Merged in chunk 33 now 58185 rows, at 2013-11-14 13:15:25
Merged in chunk 34 now 58527 rows, at 2013-11-14 13:15:52
Merged in chunk 35 now 58758 rows, at 2013-11-14 13:16:19
Merged in chunk 36 now 58883 rows, at 2013-11-14 13:16:46
Merged in chunk 37 now 59115 rows, at 2013-11-14 13:17:13
Merged in chunk 38 now 59363 rows, at 2013-11-14 13:17:40
Merged in chunk 39 now 59504 rows, at 2013-11-14 13:18:08
Merged in chunk 40 now 59730 rows, at 2013-11-14 13:18:35
Merged in chunk 41 now 59983 rows, at 2013-11-14 13:19:02
Merged in chunk 42 now 60029 rows, at 2013-11-14 13:19:28
Merged in chunk 43 now 60217 rows, at 2013-11-14 13:19:55
Merged in chunk 44 now 60403 rows, at 2013-11-14 13:20:22
Merged in chunk 45 now 60559 rows, at 2013-11-14 13:20:49
Merged in chunk 46 now 60695 rows, at 2013-11-14 13:21:15
Merged in chunk 47 now 60883 rows, at 2013-11-14 13:21:42
Merged in chunk 48 now 61069 rows, at 2013-11-14 13:22:09
Merged in chunk 49 now 61236 rows, at 2013-11-14 13:22:36
Merged in chunk 50 now 61408 rows, at 2013-11-14 13:23:05
```

```
Merged in chunk 51 now 61595 rows, at 2013-11-14 13:23:33
Merged in chunk 52 now 61715 rows, at 2013-11-14 13:24:01
Merged in chunk 53 now 61837 rows, at 2013-11-14 13:24:27
Merged in chunk 54 now 61949 rows, at 2013-11-14 13:24:53
Merged in chunk 55 now 62070 rows, at 2013-11-14 13:25:20
Merged in chunk 56 now 62167 rows, at 2013-11-14 13:25:47
Merged in chunk 57 now 62252 rows, at 2013-11-14 13:26:14
Merged in chunk 58 now 62317 rows, at 2013-11-14 13:26:40
Merged in chunk 59 now 62404 rows, at 2013-11-14 13:27:07
Merged in chunk 60 now 62566 rows, at 2013-11-14 13:27:33
Merged in chunk 61 now 62592 rows, at 2013-11-14 13:28:00
Merged in chunk 62 now 62663 rows, at 2013-11-14 13:28:26
Merged in chunk 63 now 62746 rows, at 2013-11-14 13:28:52
Merged in chunk 64 now 62762 rows, at 2013-11-14 13:29:20
Merged in chunk 65 now 62841 rows, at 2013-11-14 13:29:46
Merged in chunk 66 now 63017 rows, at 2013-11-14 13:30:13
Merged in chunk 67 now 63092 rows, at 2013-11-14 13:30:41
Merged in chunk 68 now 63193 rows, at 2013-11-14 13:31:08
Merged in chunk 69 now 63251 rows, at 2013-11-14 13:31:35
Merged in chunk 70 now 63323 rows, at 2013-11-14 13:32:01
Merged in chunk 71 now 63416 rows, at 2013-11-14 13:32:28
Merged in chunk 72 now 63524 rows, at 2013-11-14 13:32:55
Merged in chunk 73 now 63633 rows, at 2013-11-14 13:33:22
Merged in chunk 74 now 63708 rows, at 2013-11-14 13:33:50
Merged in chunk 75 now 63756 rows, at 2013-11-14 13:34:16
Merged in chunk 76 now 63842 rows, at 2013-11-14 13:34:44
Merged in chunk 77 now 64000 rows, at 2013-11-14 13:35:10
Merged in chunk 78 now 64074 rows, at 2013-11-14 13:35:37
Merged in chunk 79 now 64171 rows, at 2013-11-14 13:36:04
Merged in chunk 80 now 64265 rows, at 2013-11-14 13:36:31
Merged in chunk 81 now 64396 rows, at 2013-11-14 13:36:59
Merged in chunk 82 now 64489 rows, at 2013-11-14 13:37:26
Merged in chunk 83 now 64547 rows, at 2013-11-14 13:37:52
Merged in chunk 84 now 64566 rows, at 2013-11-14 13:38:19
Merged in chunk 85 now 64653 rows, at 2013-11-14 13:38:46
Merged in chunk 86 now 64670 rows, at 2013-11-14 13:39:13
Merged in chunk 87 now 64778 rows, at 2013-11-14 13:39:42
Merged in chunk 88 now 64821 rows, at 2013-11-14 13:40:09
Merged in chunk 89 now 64918 rows, at 2013-11-14 13:40:35
Merged in chunk 90 now 65037 rows, at 2013-11-14 13:41:03
Merged in chunk 91 now 65088 rows, at 2013-11-14 13:41:29
Merged in chunk 92 now 65132 rows, at 2013-11-14 13:41:55
Merged in chunk 93 now 65144 rows, at 2013-11-14 13:42:22
Merged in chunk 94 now 65263 rows, at 2013-11-14 13:42:48
Merged in chunk 95 now 65304 rows, at 2013-11-14 13:43:15
Merged in chunk 96 now 65383 rows, at 2013-11-14 13:43:41
Merged in chunk 97 now 65459 rows, at 2013-11-14 13:44:07
Merged in chunk 98 now 65604 rows, at 2013-11-14 13:44:34
Merged in chunk 99 now 65632 rows, at 2013-11-14 13:45:01
Merged in chunk 100 now 65676 rows, at 2013-11-14 13:45:27
> summary( Agg )
       A                P              U             sex          region
 Min.   :  0.00   Min.   :1995   Min.   :0.0000   M:32473   DK     :12169
 1st Qu.: 27.00   1st Qu.:1999   1st Qu.:0.0000   F:33203   Africa : 8429
 Median : 49.00   Median :2002   Median :0.0000             America: 9112
 Mean   : 48.61   Mean   :2002   Mean   :0.4996             Asia   : 9737
 3rd Qu.: 70.00   3rd Qu.:2006   3rd Qu.:1.0000             Europe :11341
 Max.   :110.00   Max.   :2009   Max.   :1.0000             MidEast:10058
                                                            Other  : 4830
      state            Y                  D.dm             D.dd
 Well    :38387   Min.   :   0.0007   Min.   :  0.000   Min.   :  0.000
 DM      :27289   1st Qu.:   2.6295   1st Qu.:  0.000   1st Qu.:  0.000
 Dead    :    0   Median :  19.7546   Median :  0.000   Median :  0.000
 Dead(DM):    0   Mean   : 172.0773   Mean   :  4.736   Mean   :  2.304
                  3rd Qu.: 144.4735   3rd Qu.:  0.000   3rd Qu.:  0.000
                  Max.   :3016.9610   Max.   :259.000   Max.   :130.000
```

```
> round(
+ ftable( xtabs( Y/1000 ~ region + state,
+                data = Agg ),
+        row.vars=c(1) ) , 1 )
        state   Well     DM   Dead Dead(DM)
region
DK            2395.7 2488.7    0.0      0.0
Africa         301.7    7.3    0.0      0.0
America        335.3    4.1    0.0      0.0
Asia           669.8   19.1    0.0      0.0
Europe        3713.8   53.5    0.0      0.0
MidEast       1247.3   56.2    0.0      0.0
Other            8.5    0.4    0.0      0.0
> ftable( xtabs( cbind( D.dm, D.dd ) ~ region + state,
+                data = Agg ),
+        row.vars=c(3,1) )
             state   Well     DM   Dead Dead(DM)
     region
D.dm DK            291281      0      0        0
     Africa          1212      0      0        0
     America          606      0      0        0
     Asia            2584      0      0        0
     Europe          7831      0      0        0
     MidEast         7481      0      0        0
     Other             51      0      0        0
D.dd DK                 0 132921      0        0
     Africa           416     85      0        0
     America          858    154      0        0
     Asia             909    282      0        0
     Europe         10832   2194      0        0
     MidEast         1796    829      0        0
     Other             25     16      0        0
> save( Agg, file="./data/Agg.Rda" )
```

2.2 Splitting follow-up by duration

We will also be splitting the follow-up among those with diabetes by diabetes duration, however, only for diabetes patients diagnosed after 1.1.1995, since only dates of inclusion in the NDR after this are regraded to reflect the date of diagnosis in reasonable detail. This dataset will only be usable for analysis of mortality among persons with diabetes. Once constructed it should therefore be merged with the part of the mortality dataset for the non-diabetics.

```
> dLx <- subset( cLx, lex.Cst=="DM" & dodm>1995 )
> with( dLx, table( lex.Xst ) )
lex.Xst
    Well       DM     Dead Dead(DM)
       0   234451        0    76816
> nrow( dLx )
[1] 311267
```

The code to complete this task is almost the same as before, except that we only count detahs, we have included diabetes duration in fairly small intervals, and by that token made a shortcut in the splitting, as we only split by diabetes duration, and just classify follow-up according to where it belongs by age and date.

First we set up the empty dataframe for the aggregated data:

```
> Dgg <- data.frame( A=0, P=0, U=0, dur=0,
+                    cLx[1,c("sex","region","lex.Cst")],
+                    Y=0, D.dd=0 )[NULL,]
> names( Dgg )[7] <- "state"
> str( Dgg )
'data.frame':        0 obs. of  9 variables:
 $ A     : num
 $ P     : num
 $ U     : num
 $ dur   : num
 $ sex   : Factor w/ 2 levels "M","F":
 $ region: Factor w/ 7 levels "DK","Africa",..:
 $ state : Factor w/ 4 levels "Well","DM","Dead",..:
 $ Y     : num
 $ D.dd  : num
```

Then we can fill `Dgg` with deaths and person-years in the classes that are actually present in data:

```
> n.chunks <- 50
> lm <- round( seq(0,nrow(dLx),,n.chunks+1) )
> for( i in 1:n.chunks )
+ {
+ whr <- (lm[i]+1):(lm[i+1])
+ sLx <- splitLexis( dLx[whr,], breaks=seq(0,20,0.2), time.scale="DMdur" )
+ dgg <- with( sLx,
+              aggregate( cbind( y = lex.dur,
+                                d.dd = ( lex.Xst %in% c("Dead","Dead(DM)") )*1 ),
+                         list( A = floor(age+0.1),
+                                 P = floor(date+0.1),
+                                 U = floor(date+0.1)-floor(age+0.1)-floor(dobth),
+                               dur = timeBand( sLx, "DMdur", "left" ),
+                               sex = sex,
+                            region = region,
+                             state = lex.Cst ),
+                         FUN = sum ) )
+ Dgg <- merge( Dgg, dgg, by=names( Dgg )[1:7], all=TRUE )
+ Dgg <- transform( Dgg, Y = pmax(Y    ,0,na.rm=TRUE) + pmax(y    ,0,na.rm=TRUE),
+                      D.dd = pmax(D.dd,0,na.rm=TRUE) + pmax(d.dd,0,na.rm=TRUE) )[,
+                    c("A","P","U","dur","sex","region","state","Y","D.dd")]
+ cat( "Merged in chunk", i, "(", nrow(dgg), " rows), now total", nrow(Dgg), "rows, at",
+      format(Sys.time(),format="%Y-%m-%d %H:%M:%S"), "\n" )
+ }
Merged in chunk 1 ( 106085  rows), now total 106085 rows, at 2013-11-14 13:45:41
Merged in chunk 2 ( 110073  rows), now total 169234 rows, at 2013-11-14 13:45:54
Merged in chunk 3 ( 97948  rows), now total 199523 rows, at 2013-11-14 13:46:08
Merged in chunk 4 ( 99747  rows), now total 222529 rows, at 2013-11-14 13:46:22
Merged in chunk 5 ( 99114  rows), now total 240930 rows, at 2013-11-14 13:46:36
Merged in chunk 6 ( 98423  rows), now total 256650 rows, at 2013-11-14 13:46:51
Merged in chunk 7 ( 97741  rows), now total 271084 rows, at 2013-11-14 13:47:07
Merged in chunk 8 ( 99686  rows), now total 286348 rows, at 2013-11-14 13:47:22
Merged in chunk 9 ( 98875  rows), now total 297606 rows, at 2013-11-14 13:47:38
Merged in chunk 10 ( 99887  rows), now total 308226 rows, at 2013-11-14 13:47:54
Merged in chunk 11 ( 100753  rows), now total 318807 rows, at 2013-11-14 13:48:10
Merged in chunk 12 ( 97492  rows), now total 327885 rows, at 2013-11-14 13:48:27
Merged in chunk 13 ( 95970  rows), now total 336315 rows, at 2013-11-14 13:48:43
Merged in chunk 14 ( 98880  rows), now total 343858 rows, at 2013-11-14 13:49:00
Merged in chunk 15 ( 98421  rows), now total 351121 rows, at 2013-11-14 13:49:17
Merged in chunk 16 ( 98801  rows), now total 359903 rows, at 2013-11-14 13:49:34
Merged in chunk 17 ( 102297  rows), now total 370629 rows, at 2013-11-14 13:49:51
Merged in chunk 18 ( 99263  rows), now total 377565 rows, at 2013-11-14 13:50:08
Merged in chunk 19 ( 99270  rows), now total 384990 rows, at 2013-11-14 13:50:26
Merged in chunk 20 ( 101269  rows), now total 392818 rows, at 2013-11-14 13:50:44
Merged in chunk 21 ( 94442  rows), now total 398380 rows, at 2013-11-14 13:51:02
Merged in chunk 22 ( 98705  rows), now total 404101 rows, at 2013-11-14 13:51:19
Merged in chunk 23 ( 97258  rows), now total 408964 rows, at 2013-11-14 13:51:37
```

```
Merged in chunk 24 ( 102595  rows), now total 416252 rows, at 2013-11-14 13:51:55
Merged in chunk 25 ( 100154  rows), now total 423941 rows, at 2013-11-14 13:52:13
Merged in chunk 26 ( 96702  rows), now total 429577 rows, at 2013-11-14 13:52:32
Merged in chunk 27 ( 101844  rows), now total 436026 rows, at 2013-11-14 13:52:50
Merged in chunk 28 ( 97107  rows), now total 440747 rows, at 2013-11-14 13:53:09
Merged in chunk 29 ( 96651  rows), now total 445299 rows, at 2013-11-14 13:53:28
Merged in chunk 30 ( 96888  rows), now total 449969 rows, at 2013-11-14 13:53:46
Merged in chunk 31 ( 96348  rows), now total 454512 rows, at 2013-11-14 13:54:04
Merged in chunk 32 ( 102894  rows), now total 459967 rows, at 2013-11-14 13:54:23
Merged in chunk 33 ( 101593  rows), now total 465911 rows, at 2013-11-14 13:54:42
Merged in chunk 34 ( 99145  rows), now total 470494 rows, at 2013-11-14 13:55:03
Merged in chunk 35 ( 99329  rows), now total 474948 rows, at 2013-11-14 13:55:24
Merged in chunk 36 ( 96966  rows), now total 478571 rows, at 2013-11-14 13:55:44
Merged in chunk 37 ( 97258  rows), now total 482170 rows, at 2013-11-14 13:56:03
Merged in chunk 38 ( 100293  rows), now total 486106 rows, at 2013-11-14 13:56:23
Merged in chunk 39 ( 97353  rows), now total 490216 rows, at 2013-11-14 13:56:42
Merged in chunk 40 ( 99754  rows), now total 494560 rows, at 2013-11-14 13:57:02
Merged in chunk 41 ( 99972  rows), now total 498522 rows, at 2013-11-14 13:57:23
Merged in chunk 42 ( 98652  rows), now total 502116 rows, at 2013-11-14 13:57:42
Merged in chunk 43 ( 95570  rows), now total 505530 rows, at 2013-11-14 13:58:02
Merged in chunk 44 ( 98384  rows), now total 509278 rows, at 2013-11-14 13:58:21
Merged in chunk 45 ( 100559  rows), now total 513240 rows, at 2013-11-14 13:58:43
Merged in chunk 46 ( 97593  rows), now total 516511 rows, at 2013-11-14 13:59:03
Merged in chunk 47 ( 99432  rows), now total 519772 rows, at 2013-11-14 13:59:24
Merged in chunk 48 ( 96159  rows), now total 522897 rows, at 2013-11-14 13:59:44
Merged in chunk 49 ( 98075  rows), now total 525980 rows, at 2013-11-14 14:00:06
Merged in chunk 50 ( 100595  rows), now total 529426 rows, at 2013-11-14 14:00:26
```

```
> str( Dgg )
```

```
'data.frame':        529426 obs. of  9 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1996 ...
 $ U     : num  0 0 0 0 0 0 0 0 1 1 ...
 $ dur   : num  0 0.2 0.4 0.6 0 0.2 0.4 0.6 0 0.2 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 2 2 2 2 2 2 ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ state : Factor w/ 4 levels "Well","DM","Dead",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ Y     : num  0.2 0.2 0.2 0.2 0.4 0.4 0.4 0.2 0.6 0.4 ...
 $ D.dd  : num  0 0 0 0 0 0 0 0 0 0 ...
```

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

2.3 Acquiring the population risk time

So far we have only attended to persons who are either non-Danish or have a diagnosis of DM. So in the “Well” state we are missing the follow-up time from Danish persons without DM. But we actually have access to follow-up time in the object `Agg`, so if we take the risk time among non-Danish or DM and subtract this from the total risk time in the population, we get the the risk time among Danish in the state “Well”.

```
R version 3.0.2 (2013-09-25)
Platform: i386-w64-mingw32/i386 (32-bit)

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

other attached packages:
[1] Epi_1.1.59     foreign_0.8-55

loaded via a namespace (and not attached):
[1] tools_3.0.2
```

The data frame Agg contains all the risk time among the persons on whom we have follow-up in the various states.

```
> load( file="./data/Agg.Rda" )
> str( Agg )
'data.frame':        65676 obs. of  9 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1995 ...
 $ U     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 2 2 ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 1 2 3 4 5 6 7 1 2 ...
 $ state : Factor w/ 4 levels "Well","DM","Dead",..: 1 2 1 1 1 1 1 1 1 1 ...
 $ Y     : num  61.992 0.806 1.316 26.486 10.537 ...
 $ D.dm  : num  1 0 0 0 0 0 0 0 0 0 ...
 $ D.dd  : num  0 0 0 0 0 0 0 0 0 0 ...
> Agg$state <- factor( Agg$state )
> round(
+ ftable( xtabs( Y/1000 ~ region + state,
+                data = Agg ),
+         row.vars=c(1) ) , 1 )
        state   Well     DM
region
DK            2395.7 2488.7
Africa         301.7    7.3
America        335.3    4.1
Asia           669.8   19.1
Europe        3713.8   53.5
MidEast       1247.3   56.2
Other            8.5    0.4
> ftable( xtabs( cbind( D.dm, D.dd ) ~ region + state,
+                data = Agg ),
+         row.vars=c(3,1) )
             state   Well     DM
     region
D.dm DK            291281      0
     Africa          1212      0
     America          606      0
     Asia            2584      0
     Europe          7831      0
     MidEast         7481      0
     Other             51      0
D.dd DK                 0 132921
     Africa           416     85
     America          858    154
     Asia             909    282
     Europe         10832   2194
     MidEast         1796    829
     Other             25     16
```

The follow-up time and number of deaths for persons in region "DK" and state "Well" is wrong, because the dataset only includes persons who either are born outside DK or have a DM event recorded. Risk time and deaths in all other states is correct, and *all* transitions to DM are correct for all, as we included everyone with any of these events.

Thus the risk time computed in Agg for the DK should be replaced by the total population risk time *minus* the risk time accumulated by persons born outside of Denmark *or* by persons with a previous diagnosis of DM.

But the risk time to be subtracted is readily available in the dataframe of aggregated follow-up, we just sum over the state "DM" or among persons not in Denmark (region!="DK")

```
> system.time(
+ Cgg <- with( subset( Agg, A<100 & P>1994 & P<2010 &
+                         !(region=="DK" & state=="Well") ),
+               aggregate( cbind( X = Y ),
+                          list( A = A,
+                                P = P,
+                            upper = U,
+                              sex = sex ),
+                       FUN = sum ) ) )
   user  system elapsed
   0.56    0.00    0.56
> str( Cgg )
'data.frame':         6000 obs. of  5 variables:
 $ A    : num  0 1 2 3 4 5 6 7 8 9 ...
 $ P    : num  1995 1995 1995 1995 1995 ...
 $ upper: num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ X    : num  101 265 320 456 594 ...
> summary( Cgg )
       A               P              upper       sex            X
 Min.   : 0.00   Min.   :1995   Min.   :0.0   M:3000   Min.   :   0.546
 1st Qu.:24.75   1st Qu.:1998   1st Qu.:0.0   F:3000   1st Qu.: 715.677
 Median :49.50   Median :2002   Median :0.5            Median :1289.651
 Mean   :49.50   Mean   :2002   Mean   :0.5            Mean   :1484.152
 3rd Qu.:74.25   3rd Qu.:2006   3rd Qu.:1.0            3rd Qu.:2216.002
 Max.   :99.00   Max.   :2009   Max.   :1.0            Max.   :4444.710
```

`Cgg` now has the number of person-years lived by persons who are either non-Danish or who have a diagnosis of DM, classified by sex and Lexis triangles (age, period and cohort).

Then we get the population data from Denmark in Lexis triangles:

```
> data( Y.dk )
> Y.dk$sex <- factor( Y.dk$sex, labels=c("M","F") )
> Y.dk <- subset( Y.dk,
+                 A<100 & P>1994 & P<2010,
+                 select=c("sex","A","P","upper","Y") )
```

In `Y.dk` we now have the total person-years in the population (up to age 100), and can now subtract the person-years from the study in order to get the follow-up among the non-foreign, non-DM persons:

```
> Y.rev <- merge( Cgg, Y.dk, all.y=TRUE )
> summary( Y.rev )
       A               P              upper       sex            X
 Min.   : 0.00   Min.   :1995   Min.   :0.0   M:3000   Min.   :   0.546
 1st Qu.:24.75   1st Qu.:1998   1st Qu.:0.0   F:3000   1st Qu.: 715.677
 Median :49.50   Median :2002   Median :0.5            Median :1289.651
 Mean   :49.50   Mean   :2002   Mean   :0.5            Mean   :1484.152
 3rd Qu.:74.25   3rd Qu.:2006   3rd Qu.:1.0            3rd Qu.:2216.002
 Max.   :99.00   Max.   :2009   Max.   :1.0            Max.   :4444.710
       Y
 Min.   :   48.5
 1st Qu.: 9145.0
 Median :15918.9
 Mean   :13435.6
 3rd Qu.:18404.8
 Max.   :23096.3
> Y.rev <- transform( Y.rev, Y.pop = Y-pmax(X,0,na.rm=TRUE),
+                            state = "Well",
+                           region = "DK",
+                                U = upper )[,c("A","P","U","sex","state","region","Y.pop")]
> str( Y.rev )
```

```
'data.frame':        6000 obs. of  7 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1996 ...
 $ U     : num  0 0 1 1 0 0 1 1 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
 $ state : Factor w/ 1 level "Well": 1 1 1 1 1 1 1 1 1 1 ...
 $ region: Factor w/ 1 level "DK": 1 1 1 1 1 1 1 1 1 1 ...
 $ Y.pop : num  16923 17926 16949 17713 16374 ...
```

Thus `Y.rev` now contains the correct person-years in the "Well" state among persons born in DK (`region="DK"`), classified by sex, age, date of follow-up and date of birth.

2.3.1 Creating follow-up for all persons

The trick is now to merge the new population data into the data frame with the aggregate person-years. Note that we must merge the datasets, because we want to preserve the number of diabetes events (`D.dm`) from `Agg` among Danish born without diabetes.

```
> Afu <- merge( subset( Agg, A<100 & P>1994 & P<2010 ), Y.rev, all=TRUE )
> Afu <- transform( Afu, Y = pmax( Y,Y.pop,na.rm=TRUE),
+                     D.dm = pmax(D.dm,   0,na.rm=TRUE),
+                     D.dd = pmax(D.dd,   0,na.rm=TRUE) )[,
+                   c("sex","A","P","U","state","region","Y","D.dm","D.dd")]
> str( Afu )
'data.frame':        65348 obs. of  9 variables:
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 2 2 ...
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1995 ...
 $ U     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ state : Factor w/ 2 levels "Well","DM": 1 2 1 1 1 1 1 1 1 1 ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 1 2 3 4 5 6 7 1 2 ...
 $ Y     : num  1.79e+04 8.06e-01 1.32 2.65e+01 1.05e+01 ...
 $ D.dm  : num  1 0 0 0 0 0 0 0 0 0 ...
 $ D.dd  : num  0 0 0 0 0 0 0 0 0 0 ...
```

The data frame `Afu` now contains the correct number of person-years and transitions to DM, but not to death:

```
> round( addmargins( xtabs( Y    ~ region + state, data=Afu )/1000 ), 1 )
         state
region       Well      DM     Sum
  DK      71708.5  2488.1 74196.7
  Africa    301.7     7.3   309.0
  America   335.3     4.1   339.4
  Asia      669.8    19.1   688.9
  Europe   3713.7    53.5  3767.2
  MidEast  1247.3    56.2  1303.5
  Other       8.5     0.4     8.9
  Sum     77984.8  2628.6 80613.5
>         addmargins( xtabs( D.dm ~ region + state, data=Afu ) )

         state
region      Well     DM    Sum
  DK      291230      0 291230
  Africa    1212      0   1212
  America    606      0    606
  Asia      2584      0   2584
  Europe    7831      0   7831
  MidEast   7481      0   7481
  Other       51      0     51
  Sum     310995      0 310995
>         addmargins( xtabs( D.dd ~ region + state, data=Afu ) )
```

```
         state
region      Well     DM    Sum
  DK           0 132590 132590
  Africa     416     85    501
  America    854    154   1008
  Asia       907    282   1189
  Europe   10749   2182  12931
  MidEast   1792    828   2620
  Other       25     16     41
  Sum      14743 136137 150880
```

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

2.3.2 Creating deaths for all states

Along the same lines we can derive the number of deaths in the class ("Well","DK") by subtracting the number of deaths in all other classes from the total number of deaths in the population. To that end we first retrieve the total number of detahs from the human mortality database:

2.3.2.1 Getting mortality data from Human Mortality Database

In order to fetch mortality from the HMD in 1×1 Lexis triangles we need to provide a user id and a password, which is hidden in this output: We can now get the mortality data for Denmark, and reshpae them to our purpose:

```
> source( "C:/stat/R/BxC/Examples/HMD2R.r" )
> library( RCurl )
> DK <- HMD2R( "DNK",
+             wanted = "Deaths_lexis",
+           username = HMDBusr,
+           password = HMDBpwd )$Deaths_lexis[,1:5]
  *** Fetching... Deaths_lexis
```

```
> names( DK ) <- c("P","A","C","F","M")
> DK <- subset( DK, A < 100 & P>1994 & P<2010 )
> DK$upper <- with( DK,  P-A-C )
> M.dk <- reshape( DK, direction = "long",
+                      varying = c("M","F"),
+                      v.names = "D.pop",
+                      timevar = "sex" )[,-7]
> M.dk$sex <- factor( M.dk$sex, labels=c("M","F") )
> str( M.dk )
```

```
'data.frame':        6000 obs. of  6 variables:
 $ P     : int  1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
 $ A     : int  0 0 1 1 2 2 3 3 4 4 ...
 $ C     : int  1995 1994 1994 1993 1993 1992 1992 1991 1991 1990 ...
 $ upper: int  0 1 0 1 0 1 0 1 0 1 ...
 $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ D.pop: num  179 21 13 8 2 7 4 6 5 8 ...
```

```
> save( M.dk, file="./data/Mdk.Rda" )
```

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

2.3.3 Merging with the deaths

First we tabulate all the deaths that we *do* have recorded completely, namely those among diabetes patients and among immigrants, in order to subtract them from the total number of deaths:

```
> DMf <- with( subset( Afu, !(region=="DK" & state=="Well") ),
+              aggregate( cbind( X = D.dd ),
+                         list( A = A,
+                               P = P,
+                           upper = U,
+                             sex = sex ),
+                        FUN = sum ) )
> str( DMf )
'data.frame':        6000 obs. of  5 variables:
 $ A    : num  0 1 2 3 4 5 6 7 8 9 ...
 $ P    : num  1995 1995 1995 1995 1995 ...
 $ upper: num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex  : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ X    : num  0 0 0 0 0 0 0 0 0 0 ...
> summary( DMf )
       A               P             upper        sex             X
 Min.   : 0.00   Min.   :1995   Min.   :0.0   M:3000   Min.   :  0.00
 1st Qu.:24.75   1st Qu.:1998   1st Qu.:0.0   F:3000   1st Qu.:  1.00
 Median :49.50   Median :2002   Median :0.5            Median :  7.00
 Mean   :49.50   Mean   :2002   Mean   :0.5            Mean   : 25.15
 3rd Qu.:74.25   3rd Qu.:2006   3rd Qu.:1.0            3rd Qu.: 43.00
 Max.   :99.00   Max.   :2009   Max.   :1.0            Max.   :185.00
```

We now merge these data with the total population-deaths, and subtract in order to get the number of deaths among non-diabetics of Danish origin:

```
> D.rev <- merge( DMf, M.dk, all.y=TRUE )
> D.rev <- transform( D.rev, D.Wdk = pmax( D.pop - pmax(X,0,na.rm=TRUE),
+                                          0,
+                                          na.rm=TRUE ),
+                            state = "Well",
+                           region = "DK",
+                                U = upper )[,c("A","P","U","sex","state","region","D.Wdk")]
> str( D.rev )
'data.frame':        6000 obs. of  7 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1996 ...
 $ U     : num  0 0 1 1 0 0 1 1 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
 $ state : Factor w/ 1 level "Well": 1 1 1 1 1 1 1 1 1 1 ...
 $ region: Factor w/ 1 level "DK": 1 1 1 1 1 1 1 1 1 1 ...
 $ D.Wdk : num  137 179 16 21 134 189 23 30 152 171 ...
```

Thus `D.rev` now contains the correct no. of deaths in the “Well” state among persons born in DK (`region="DK"`), classified by sex, age, date of follow-up and date of birth in the variable `D.Wdk`. We then merge these numbers into the `Afu`

```
> Afu <- transform( merge( D.rev, Afu, all.y=TRUE ),
+                   D.dd = pmax( D.dd, D.Wdk, 0, na.rm=TRUE ) )
> round(
+ ftable( xtabs( Y/1000 ~ region + state,
+                data = Afu ),
+         row.vars=c(1) ) , 1 )
        state     Well      DM
region
DK             71708.5  2488.1
Africa           301.7     7.3
America          335.3     4.1
```

```
Asia            669.8    19.1
Europe         3713.7    53.5
MidEast        1247.3    56.2
Other             8.5     0.4
> ftable( xtabs( cbind( D.dm, D.dd ) ~ region + state,
+                data = Afu ),
+         row.vars=c(3,1) )
                state   Well     DM
     region
D.dm DK               291230      0
     Africa             1212      0
     America             606      0
     Asia               2584      0
     Europe             7831      0
     MidEast            7481      0
     Other                51      0
D.dd DK               710330 132590
     Africa              416     85
     America             854    154
     Asia                907    282
     Europe            10749   2182
     MidEast            1792    828
     Other                25     16
```

The data frame `Afu` now has person-years (`Y`), number of diabetes diagnoses (`D.dm`) and number of deaths (`D.dd`) for all combinations of sex, age, period, cohort, region and diabetes status. This table will thus form the basis for analysis of incidence and mortality rates by diabetes status and region of origin:

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

2.3.4 Merging with duration-classified data

If we want to analyse mortality by diabetes duration and compare with non-diabetics, then we must merge the mortality data for diabetes patients included after 1995, with the mortality among non-diabetics from the dataframe `Afu`:

```
> load( file="./data/Dgg.Rda" )
> round( addmargins( xtabs( cbind( PY=Y/1000, D.dd ) ~
+                          floor(dur)+region, data=Dgg ) ), 1 )
, ,  = PY

          region
floor(dur)     DK Africa America   Asia  Europe MidEast   Other     Sum
       0    269.3    1.1     0.6    2.4     7.3     7.1     0.1   287.8
       1    234.3    1.0     0.5    2.1     6.2     6.2     0.0   250.4
       2    203.5    0.8     0.4    1.9     5.3     5.5     0.0   217.5
       3    176.3    0.7     0.3    1.6     4.6     4.9     0.0   188.5
       4    151.6    0.6     0.3    1.4     3.9     4.2     0.0   162.2
       5    127.9    0.5     0.2    1.2     3.3     3.6     0.0   136.8
       6    105.1    0.4     0.2    1.0     2.7     3.0     0.0   112.5
       7     84.7    0.3     0.1    0.8     2.2     2.4     0.0    90.7
       8     67.5    0.3     0.1    0.7     1.8     2.0     0.0    72.2
       9     52.6    0.2     0.1    0.5     1.4     1.5     0.0    56.3
       10    39.8    0.1     0.1    0.4     1.1     1.2     0.0    42.7
       11    28.3    0.1     0.0    0.3     0.8     0.8     0.0    30.4
       12    18.9    0.1     0.0    0.2     0.6     0.6     0.0    20.3
       13    10.6    0.0     0.0    0.1     0.3     0.3     0.0    11.3
       14     3.3    0.0     0.0    0.0     0.1     0.1     0.0     3.5
       Sum 1573.7    6.3     3.0   14.8    41.6    43.4     0.3  1683.1

, ,  = D.dd
```

```
           region
floor(dur)      DK  Africa America    Asia  Europe MidEast   Other     Sum
       0   16951.0    13.0    14.0    22.0   263.0    64.0     1.0 17328.0
       1    9702.0     8.0    11.0    27.0   194.0    41.0     0.0  9983.0
       2    8249.0     8.0     9.0    19.0   170.0    47.0     1.0  8503.0
       3    7411.0     6.0    11.0    10.0   133.0    51.0     1.0  7623.0
       4    6574.0     8.0    14.0    19.0   138.0    41.0     3.0  6797.0
       5    5850.0     7.0    10.0    17.0   115.0    45.0     0.0  6044.0
       6    4731.0     3.0     6.0    22.0   111.0    44.0     1.0  4918.0
       7    3959.0     6.0     8.0    17.0    96.0    26.0     2.0  4114.0
       8    3231.0     4.0     6.0    14.0    92.0    45.0     0.0  3392.0
       9    2610.0     3.0     8.0    11.0    64.0    23.0     1.0  2720.0
       10   1941.0     5.0     3.0    10.0    72.0    29.0     0.0  2060.0
       11   1479.0     2.0     1.0     5.0    50.0    23.0     0.0  1560.0
       12    906.0     0.0     4.0     4.0    29.0    15.0     1.0   959.0
       13    551.0     0.0     1.0     5.0    42.0    14.0     0.0   613.0
       14    179.0     0.0     1.0     3.0    11.0     8.0     0.0   202.0
       Sum 74324.0    73.0   107.0   205.0  1580.0   516.0    11.0 76816.0

, ,  = Sum

           region
floor(dur)      DK  Africa America    Asia  Europe MidEast   Other     Sum
       0   17220.3    14.1    14.6    24.4   270.3    71.1     1.1 17615.8
       1    9936.3     9.0    11.5    29.1   200.2    47.2     0.0 10233.4
       2    8452.5     8.8     9.4    20.9   175.3    52.5     1.0  8720.5
       3    7587.3     6.7    11.3    11.6   137.6    55.9     1.0  7811.5
       4    6725.6     8.6    14.3    20.4   141.9    45.2     3.0  6959.2
       5    5977.9     7.5    10.2    18.2   118.3    48.6     0.0  6180.8
       6    4836.1     3.4     6.2    23.0   113.7    47.0     1.0  5030.5
       7    4043.7     6.3     8.1    17.8    98.2    28.4     2.0  4204.7
       8    3298.5     4.3     6.1    14.7    93.8    47.0     0.0  3464.2
       9    2662.6     3.2     8.1    11.5    65.4    24.5     1.0  2776.3
       10   1980.8     5.1     3.1    10.4    73.1    30.2     0.0  2102.7
       11   1507.3     2.1     1.0     5.3    50.8    23.8     0.0  1590.4
       12    924.9     0.1     4.0     4.2    29.6    15.6     1.0   979.3
       13    561.6     0.0     1.0     5.1    42.3    14.3     0.0   624.3
       14    182.3     0.0     1.0     3.0    11.1     8.1     0.0   205.5
       Sum 75897.7    79.3   110.0   219.8  1621.6   559.4    11.3 78499.1
```

Therefore, we must append the entire follow-up (both person-years and events) as constructed above:

```
> Dfu <- rbind( subset( Dgg, A<100 ),
+        cbind( subset( Afu[,intersect(names(Dgg),names(Afu))],
+                       state=="Well" ),
+               dur=NA ) )
> Dfu$state <- factor( Dfu$state )
> str( Dfu )
'data.frame':        565950 obs. of  9 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1996 ...
 $ U     : num  0 0 0 0 0 0 0 0 1 1 ...
 $ dur   : num  0 0.2 0.4 0.6 0 0.2 0.4 0.6 0 0.2 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 2 2 2 2 2 2 ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ state : Factor w/ 2 levels "Well","DM": 2 2 2 2 2 2 2 2 2 2 ...
 $ Y     : num  0.2 0.2 0.2 0.2 0.4 0.4 0.4 0.2 0.6 0.4 ...
 $ D.dd  : num  0 0 0 0 0 0 0 0 0 0 ...
```

The data frame `Dfu` now contains the correct number of person-years and deaths:

```
> round(  addmargins( xtabs( Y ~ region + state, data=Dfu ) )/1000, 1 )
```

```
         state
region       Well      DM     Sum
  DK      71708.5  1573.4 73281.9
  Africa    301.7     6.3   308.1
  America   335.3     3.0   338.3
  Asia      669.8    14.8   684.6
  Europe   3713.7    41.6  3755.3
  MidEast  1247.3    43.4  1290.6
  Other       8.5     0.3     8.8
  Sum     77984.8  1682.7 79667.6
> ftable( addmargins( xtabs( D.dd ~ region + state,
+                              data=Dfu ),
+                       margin = 1:2 ) )
        state   Well     DM    Sum
region
DK            710330  74116 784446
Africa           416     73    489
America          854    107    961
Asia             907    205   1112
Europe         10749   1573  12322
MidEast         1792    516   2308
Other             25     11     36
Sum           725073  76601 801674
> save( Dfu, file="./data/Dfu.Rda" )
```

Chapter 3

Basic tabulation

In this chapter we provide an overview of the follow-up, so we first load the tabulated data:

```
> options( width=90 )
> library( Epi )
> load( file="./data/Afu.Rda" )
> str( Afu )
'data.frame':        65348 obs. of  10 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1995 ...
 $ U     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 2 2 ...
 $ state : Factor w/ 2 levels "Well","DM": 1 1 1 1 1 1 1 2 1 1 ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 2 3 4 5 6 7 1 1 2 ...
 $ D.Wdk : num  179 NA NA NA NA NA NA NA 137 NA ...
 $ Y     : num  17926.24 1.32 26.49 10.54 51.68 ...
 $ D.dm  : num  1 0 0 0 0 0 0 0 0 0 ...
 $ D.dd  : num  179 0 0 0 0 0 0 0 137 0 ...
```

We want to show the number of persons, the amount of follow-up time, the number of events (DM, death) and the event rates by sex and place of birth. Except for the number of persons, we can derive it all from the `Afu` data frame of aggregated follow-up:

```
> tt <- xtabs( cbind(Y/1000,D.dm,D.dd) ~ sex + region + state, data=Afu )
> str( tt )
 xtabs [1:2, 1:7, 1:2, 1:3] 35059 36649 158 143 172 ...
 - attr(*, "dimnames")=List of 4
  ..$ sex   : chr [1:2] "M" "F"
  ..$ region: chr [1:7] "DK" "Africa" "America" "Asia" ...
  ..$ state : chr [1:2] "Well" "DM"
  ..$       : chr [1:3] "V1" "D.dm" "D.dd"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = cbind(Y/1000, D.dm, D.dd) ~ sex + region + state, data
> round( ftable( tt, col.vars=c(3,4) ), 1 )
            state     Well                          DM
                        V1     D.dm     D.dd        V1     D.dm     D.dd
sex region
M   DK             35059.2 152024.0 345365.0    1269.4      0.0  69501.0
    Africa           158.3    670.0    252.0       4.1      0.0     57.0
    America          171.9    287.0    406.0       1.9      0.0     70.0
    Asia             264.3   1241.0    460.0       8.9      0.0    152.0
    Europe          2159.9   3851.0   4795.0      24.5      0.0    963.0
    MidEast          709.0   4057.0   1235.0      29.8      0.0    480.0
    Other              5.4     22.0     16.0       0.2      0.0      5.0
F   DK             36649.4 139206.0 364965.0    1218.7      0.0  63089.0
    Africa           143.4    542.0    164.0       3.1      0.0     28.0
    America          163.4    319.0    448.0       2.2      0.0     84.0
```

```
    Asia               405.5   1343.0    447.0     10.1      0.0    130.0
    Europe            1553.8   3980.0   5954.0     28.9      0.0   1219.0
    MidEast            538.3   3424.0    557.0     26.5      0.0    348.0
    Other                3.1     29.0      9.0      0.2      0.0     11.0
> tm <-
+ cbind( round( tt["M",,"Well","V1"  ], 1 ),
+        round( tt["M",,"Well","D.dm"], 0 ),
+        round( tt["M",,"Well","D.dm"]/tt["M",,"Well","V1"], 1 ),
+        round( tt["M",,"Well","D.dd"], 0 ),
+        round( tt["M",,"Well","D.dd"]/tt["M",,"Well","V1"], 1 ),
+        round( tt["M",,"DM"  ,"V1"  ], 1 ),
+        round( tt["M",,"DM"  ,"D.dd"], 0 ),
+        round( tt["M",,"DM"  ,"D.dd"]/tt["M",,"DM"  ,"V1"], 1 ) )
> tf <-
+ cbind( round( tt["F",,"Well","V1"  ], 1 ),
+        round( tt["F",,"Well","D.dm"], 0 ),
+        round( tt["F",,"Well","D.dm"]/tt["F",,"Well","V1"], 1 ),
+        round( tt["F",,"Well","D.dd"], 0 ),
+        round( tt["F",,"Well","D.dd"]/tt["F",,"Well","V1"], 1 ),
+        round( tt["F",,"DM"  ,"V1"  ], 1 ),
+        round( tt["F",,"DM"  ,"D.dd"], 0 ),
+        round( tt["F",,"DM"  ,"D.dd"]/tt["F",,"DM"  ,"V1"], 1 ) )
> cc <- c("PY(1000)","DM ca","DM inc","Deaths","Mort",
+         "  DM PY(1000)","DM death","DM mort")
> colnames(tm) <-
+ colnames(tf) <- cc
> rord <- c(1,5,3,6,2,4)
> aa <- rbind(NA,tm[rord,],NA,tf[rord,] )
> rownames( aa )[c(1,8)] <- c("Men","Women")
> print( aa, na.print=" " )
        PY(1000)  DM ca DM inc Deaths Mort   DM PY(1000) DM death DM mort
Men
DK       35059.2 152024    4.3 345365  9.9        1269.4    69501    54.8
Europe    2159.9   3851    1.8   4795  2.2          24.5      963    39.2
America    171.9    287    1.7    406  2.4           1.9       70    37.5
MidEast    709.0   4057    5.7   1235  1.7          29.8      480    16.1
Africa     158.3    670    4.2    252  1.6           4.1       57    13.9
Asia       264.3   1241    4.7    460  1.7           8.9      152    17.0
Women
DK       36649.4 139206    3.8 364965 10.0        1218.7    63089    51.8
Europe    1553.8   3980    2.6   5954  3.8          28.9     1219    42.1
America    163.4    319    2.0    448  2.7           2.2       84    37.8
MidEast    538.3   3424    6.4    557  1.0          26.5      348    13.2
Africa     143.4    542    3.8    164  1.1           3.1       28     8.9
Asia       405.5   1343    3.3    447  1.1          10.1      130    12.8
```

For the sake of the argument we also compute the total number of persons in the study. We have the total number of persons by ethnicity and diabetes status in the `Lx` dataset:

```
> load( file="./data/cLx.Rda" )
> Narr <- NArray( list( sex = levels( cLx$sex ),
+                    region = levels( cLx$region ),
+                     entry = levels( cLx$lex.Cst ),
+                      exit = levels( cLx$lex.Xst ) ) )
> for( sx in levels(cLx$sex) )
+ for( rg in levels(cLx$region) )
+    {
+    slx <- subset( cLx, sex==sx & region==rg )
+    Narr[sx,rg,,] <- table( status(slx,at="entry",by.id=TRUE),
+                            status(slx,at="exit" ,by.id=TRUE) )
+   }
> ftable( Narr[,,1:2,1:2], col.vars=3:4 )
           entry   Well            DM
           exit    Well     DM   Well     DM
sex region
M   DK                0 112013      0  18654
```

```
    Africa        18472    621      0     45
    America       30198    240      0     25
    Asia          35330   1125      0    127
    Europe       251998   3139      0    267
    MidEast       61957   3755      0    394
    Other           472     19      0      3
F   DK                0 104944      0  18681
    Africa        16612    521      0     27
    America       31511    260      0     30
    Asia          53222   1255      0    131
    Europe       199405   3141      0    293
    MidEast       48199   3216      0    378
    Other           301     21      0      3
```

However we are only interested in the distribution of persons by sex and region:

```
> ( NN <- apply( Narr, 2:1, sum ) )
         sex
region         M      F
  DK      200227 186986
  Africa   19447  17352
  America  30939  32337
  Asia     37195  55186
  Europe  261176 210093
  MidEast  67823  52701
  Other      515    345
```

The total number of persons possibly involved in follow-up from 1.1.1995 till 1.1.2010 is the entire population at 1.1.1995 *plus* all born in the period. These figures are available by sex in the `Epi` package:

```
> data( N.dk )
> str( N.dk )
'data.frame':         8600 obs. of  4 variables:
 $ sex: num  1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P  : num  1971 1971 1971 1971 1971 ...
 $ N  : num  35839 34108 36302 34153 37855 ...
 - attr(*, "Contents")= chr "Population size as of 1 January in Denmark"
> ( pp <- xtabs( N ~ sex, data=subset(N.dk,P==1995) ) )
sex
      1       2
2573379 2642452
> data( B.dk )
> str( B.dk )
'data.frame':         1308 obs. of  4 variables:
 $ year : int  1901 1901 1901 1901 1901 1901 1901 1901 1901 1901 ...
 $ month: int  1 2 3 4 5 6 7 8 9 10 ...
 $ m    : num  2948 2851 3476 3326 3324 ...
 $ f    : num  2786 2695 3284 3143 3141 ...
 - attr(*, "Contents")= chr "Number of births by month in Denmark"
> ( bp <- with( subset(B.dk,year>1994 & year<2010), c( sum(m), sum(f) ) ) )
[1] 505759 478863
> NN[1,] <- bp+pp - apply( NN[-1,], 2, sum )
> NN
         sex
region          M       F
  DK      2662043 2753301
  Africa    19447   17352
  America   30939   32337
  Asia      37195   55186
  Europe   261176  210093
  MidEast   67823   52701
  Other       515     345
```

We can now append this to the tables `tm` and `tf` above:

```
> tm <- cbind( NN[,"M"], tm )
> tf <- cbind( NN[,"F"], tf )
> colnames( tm )[1] <-
+ colnames( tf )[1] <- "N"
> rbind(NA,tm[rord,],NA,tf[rord,] )
              N PY(1000)  DM ca DM inc Deaths Mort   DM PY(1000) DM death DM mort
             NA       NA     NA     NA     NA   NA            NA       NA      NA
DK      2662043  35059.2 152024    4.3 345365  9.9        1269.4    69501    54.8
Europe   261176   2159.9   3851    1.8   4795  2.2          24.5      963    39.2
America   30939    171.9    287    1.7    406  2.4           1.9       70    37.5
MidEast   67823    709.0   4057    5.7   1235  1.7          29.8      480    16.1
Africa    19447    158.3    670    4.2    252  1.6           4.1       57    13.9
Asia      37195    264.3   1241    4.7    460  1.7           8.9      152    17.0
             NA       NA     NA     NA     NA   NA            NA       NA      NA
DK      2753301  36649.4 139206    3.8 364965 10.0        1218.7    63089    51.8
Europe   210093   1553.8   3980    2.6   5954  3.8          28.9     1219    42.1
America   32337    163.4    319    2.0    448  2.7           2.2       84    37.8
MidEast   52701    538.3   3424    6.4    557  1.0          26.5      348    13.2
Africa    17352    143.4    542    3.8    164  1.1           3.1       28     8.9
Asia      55186    405.5   1343    3.3    447  1.1          10.1      130    12.8
> write.csv2( rbind(tm[rord,],NA,tf[rord,]), file="table1.csv" )
```

Chapter 4

DM prevalence by country of origin

In this chapter we provide the age-specific prevalences by sex, date and ethnicity.

```
> options( width=90 )
> library( Epi )
> clear()
```

4.1 Migrant and diabetes data

To this end we load the cLx Lexis data frame from which we derive the number of prevalent cases and

```
> load( file="./data/Lx.Rda" )
> Lx <- transform( subset(Lx,region!="Other"),
+                  region = Relevel( factor( region ),
+                                    list("Danish born" = 1,
+                                              "Europe" = 5,
+                                  "Sub Saharan Africa" = 2,
+                          "Middle East & North Africa" = 6,
+                                                "Asia" = 4,
+                                   "America & Oceania" = 3 ) ) )
> head( Lx )
      date      age   lex.dur lex.Cst lex.Xst lex.id doud                     region
2 1999.999 0.000000 10.001369    Well    Well      1   NA                Danish born
3 2001.567 1.568789  8.432580    Well    Well      2   NA                     Europe
4 2002.077 2.078029  7.923340    Well    Well      3   NA Middle East & North Africa
5 2002.487 2.488706  7.512663    Well    Well      4   NA Middle East & North Africa
6 2004.385 4.386037  5.615332    Well    Well      5   NA Middle East & North Africa
7 2004.760 4.761123  5.240246    Well    Well      6   NA Middle East & North Africa
      dodm dodth    doind sex    dobth    entry exit
2 2005.721    NA       NA   F 1999.999 1999.999 2010
3       NA    NA 2001.567   M 1999.999 2001.567 2010
4       NA    NA 2002.077   M 1999.999 2002.077 2010
5       NA    NA 2002.487   M 1999.999 2002.487 2010
6       NA    NA 2004.385   M 1999.999 2004.385 2010
7       NA    NA 2004.760   M 1999.999 2004.760 2010
> table( Lx$region )

               Danish born                     Europe         Sub Saharan Africa
                    387213                     471269                      36799
Middle East & North Africa                       Asia          America & Oceania
                    120524                      92381                      63276
```

We set up an array to hold the resulting no of cases

```
> Parr <- NArray( list( date = 1995:2010,
+                        sex = levels( Lx$sex ),
+                     region = levels( Lx$region ),
+                        age = 0:99,
+                       what = c("D","N") ) )
> str( Parr )
 logi [1:16, 1:2, 1:6, 1:100, 1:2] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ date  : chr [1:16] "1995" "1996" "1997" "1998" ...
  ..$ sex   : chr [1:2] "M" "F"
  ..$ region: chr [1:6] "Danish born" "Europe" "Sub Saharan Africa" "Middle East & North Africa" ...
  ..$ age   : chr [1:100] "0" "1" "2" "3" ...
  ..$ what  : chr [1:2] "D" "N"
```

We fill in the array by year:

```
> prv <- NULL
> for( yy in 1995:2010 )
+ prv <- rbind( prv, cbind( P=yy,
+ as.data.frame( with( subset( Lx, entry<=yy & exit>=yy ),
+                      table( sex,
+                             region,
+                         A = floor(yy-dobth),
+                    status = factor( (!is.na(dodm) & dodm<yy),
+                                     labels=c("Well","DM") ) ) ) ) ) )
> str( prv )
'data.frame':        40704 obs. of  6 variables:
 $ P     : int  1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 2 1 2 1 2 1 2 1 2 ...
 $ region: Factor w/ 6 levels "Danish born",..: 1 1 2 2 3 3 4 4 5 5 ...
 $ A     : Factor w/ 111 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ status: Factor w/ 2 levels "Well","DM": 1 1 1 1 1 1 1 1 1 1 ...
 $ Freq  : int  170 148 109 109 3 3 15 22 41 27 ...
> tt <- xtabs( Freq ~ sex + factor(A) + factor(P) + region + status,
+              data = subset( prv, as.numeric(as.character(A))<100 ) )
> str( tt )
 int [1:2, 1:100, 1:16, 1:6, 1:2] 170 148 157 179 185 147 152 175 156 164 ...
 - attr(*, "dimnames")=List of 5
  ..$ sex      : chr [1:2] "M" "F"
  ..$ factor(A): chr [1:100] "0" "1" "2" "3" ...
  ..$ factor(P): chr [1:16] "1995" "1996" "1997" "1998" ...
  ..$ region   : chr [1:6] "Danish born" "Europe" "Sub Saharan Africa" "Middle East & North Africa"
  ..$ status   : chr [1:2] "Well" "DM"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = Freq ~ sex + factor(A) + factor(P) + region + status, d
> ftable( apply( tt, c(1,3:5), sum ), row.vars=c(2,3), col.vars=c(4,1) )
                                   status   Well            DM
                                   sex         M      F      M      F
factor(P) region
1995      Danish born                     151229 138395  48197  47722
          Europe                          113157  72574    441    629
          Sub Saharan Africa                5656   4416     45     25
          Middle East & North Africa       33929  23206    508    486
          Asia                             12686  16788    155    164
          America & Oceania                 7776   7123     43     51
1996      Danish born                     143927 131761  51882  50866
          Europe                          124904  83681    564    776
          Sub Saharan Africa                6702   5346     63     33
          Middle East & North Africa       35364  24483    611    602
          Asia                             13382  17883    185    210
          America & Oceania                 8632   7906     56     59
1997      Danish born                     136450 125000  55788  54312
          Europe                          130120  88808    735   1013
          Sub Saharan Africa                8195   6729     76     54
          Middle East & North Africa       37169  26277    752    743
```

	Asia	13817	19080	223	257
	America & Oceania	9060	8308	63	69
1998	Danish born	129075	118029	59492	57902
	Europe	133593	92563	863	1159
	Sub Saharan Africa	9182	7729	113	78
	Middle East & North Africa	38812	27922	893	873
	Asia	14290	20308	275	327
	America & Oceania	9544	8775	73	87
1999	Danish born	120810	110600	63952	61885
	Europe	136051	95223	1014	1339
	Sub Saharan Africa	9832	8545	141	95
	Middle East & North Africa	41338	30248	1055	1014
	Asia	14839	21564	332	395
	America & Oceania	10055	9242	79	95
2000	Danish born	112113	102651	68541	66156
	Europe	137853	97465	1128	1510
	Sub Saharan Africa	10308	9234	183	119
	Middle East & North Africa	43425	32272	1247	1163
	Asia	15374	22934	403	455
	America & Oceania	10472	9700	84	106
2001	Danish born	103151	94370	73356	70601
	Europe	139899	100028	1280	1621
	Sub Saharan Africa	10880	9940	214	144
	Middle East & North Africa	46545	35017	1445	1339
	Asia	15985	24414	476	522
	America & Oceania	10869	10149	99	114
2002	Danish born	93624	85566	78626	75447
	Europe	141615	102373	1430	1759
	Sub Saharan Africa	11278	10497	250	177
	Middle East & North Africa	50526	38301	1677	1534
	Asia	16612	25964	524	588
	America & Oceania	11234	10545	108	129
2003	Danish born	83128	75368	84442	81461
	Europe	143241	104435	1572	1920
	Sub Saharan Africa	11549	10864	295	208
	Middle East & North Africa	52478	40391	1948	1767
	Asia	17373	27398	590	662
	America & Oceania	11645	11035	126	149
2004	Danish born	71561	64720	91193	87696
	Europe	144699	106276	1780	2131
	Sub Saharan Africa	11598	10808	322	236
	Middle East & North Africa	53268	41321	2243	2005
	Asia	18497	28745	674	762
	America & Oceania	11868	11337	134	159
2005	Danish born	59896	53882	97927	94174
	Europe	146782	108605	2021	2318
	Sub Saharan Africa	11695	10833	351	284
	Middle East & North Africa	53321	41688	2529	2214
	Asia	19242	30209	760	861
	America & Oceania	12293	11781	157	178
2006	Danish born	49095	44078	103662	99333
	Europe	149752	111438	2228	2514
	Sub Saharan Africa	11668	10817	395	314
	Middle East & North Africa	53202	41729	2836	2460
	Asia	20014	31685	837	941
	America & Oceania	12712	12247	168	201
2007	Danish born	37828	33936	109715	104627
	Europe	154700	114818	2416	2724
	Sub Saharan Africa	11822	10994	423	342
	Middle East & North Africa	53150	41770	3101	2657
	Asia	20519	33140	917	1030
	America & Oceania	13207	12815	177	220
2008	Danish born	26008	22991	116143	110663
	Europe	162108	120156	2652	2907
	Sub Saharan Africa	12012	11231	467	384
	Middle East & North Africa	53455	42150	3367	2887

```
         Asia                                22036  35481    989   1124
         America & Oceania                   13878  13582    194    239
2009     Danish born                         13047  11327 123598 117381
         Europe                             171346 127280   2952   3179
         Sub Saharan Africa                  12348  11595    512    434
         Middle East & North Africa          53698  42365   3700   3185
         Asia                                23642  38030   1074   1239
         America & Oceania                   14475  14219    219    249
2010     Danish born                             0      0 130655 123545
         Europe                             174475 132101   3261   3390
         Sub Saharan Africa                  12623  11883    569    475
         Middle East & North Africa          54127  42709   4085   3528
         Asia                                24265  40382   1200   1340
         America & Oceania                   14826  14807    224    269
```

We know that entries with `region=DK` and `N=FALSE` should represent non-diabetic persons born in DK, but they do not, it only comprises those that at a *later time*, contracts DM, as is seen from the previous table where the number is 0 in 2010. So first we set the wrong entries to 0:

```
> tt[,,,"Danish born","Well"] <- 0
> ftable( apply( tt, c(1,3:5), sum ), row.vars=c(2,3), col.vars=c(4,1) )
                                    status   Well            DM
                                    sex         M      F      M      F
factor(P) region
1995      Danish born                           0      0  48197  47722
          Europe                           113157  72574    441    629
          Sub Saharan Africa                 5656   4416     45     25
          Middle East & North Africa        33929  23206    508    486
          Asia                              12686  16788    155    164
          America & Oceania                  7776   7123     43     51
1996      Danish born                           0      0  51882  50866
          Europe                           124904  83681    564    776
          Sub Saharan Africa                 6702   5346     63     33
          Middle East & North Africa        35364  24483    611    602
          Asia                              13382  17883    185    210
          America & Oceania                  8632   7906     56     59
1997      Danish born                           0      0  55788  54312
          Europe                           130120  88808    735   1013
          Sub Saharan Africa                 8195   6729     76     54
          Middle East & North Africa        37169  26277    752    743
          Asia                              13817  19080    223    257
          America & Oceania                  9060   8308     63     69
1998      Danish born                           0      0  59492  57902
          Europe                           133593  92563    863   1159
          Sub Saharan Africa                 9182   7729    113     78
          Middle East & North Africa        38812  27922    893    873
          Asia                              14290  20308    275    327
          America & Oceania                  9544   8775     73     87
1999      Danish born                           0      0  63952  61885
          Europe                           136051  95223   1014   1339
          Sub Saharan Africa                 9832   8545    141     95
          Middle East & North Africa        41338  30248   1055   1014
          Asia                              14839  21564    332    395
          America & Oceania                 10055   9242     79     95
2000      Danish born                           0      0  68541  66156
          Europe                           137853  97465   1128   1510
          Sub Saharan Africa                10308   9234    183    119
          Middle East & North Africa        43425  32272   1247   1163
          Asia                              15374  22934    403    455
          America & Oceania                 10472   9700     84    106
2001      Danish born                           0      0  73356  70601
          Europe                           139899 100028   1280   1621
          Sub Saharan Africa                10880   9940    214    144
          Middle East & North Africa        46545  35017   1445   1339
```

	Asia	15985	24414	476	522
	America & Oceania	10869	10149	99	114
2002	Danish born	0	0	78626	75447
	Europe	141615	102373	1430	1759
	Sub Saharan Africa	11278	10497	250	177
	Middle East & North Africa	50526	38301	1677	1534
	Asia	16612	25964	524	588
	America & Oceania	11234	10545	108	129
2003	Danish born	0	0	84442	81461
	Europe	143241	104435	1572	1920
	Sub Saharan Africa	11549	10864	295	208
	Middle East & North Africa	52478	40391	1948	1767
	Asia	17373	27398	590	662
	America & Oceania	11645	11035	126	149
2004	Danish born	0	0	91193	87696
	Europe	144699	106276	1780	2131
	Sub Saharan Africa	11598	10808	322	236
	Middle East & North Africa	53268	41321	2243	2005
	Asia	18497	28745	674	762
	America & Oceania	11868	11337	134	159
2005	Danish born	0	0	97927	94174
	Europe	146782	108605	2021	2318
	Sub Saharan Africa	11695	10833	351	284
	Middle East & North Africa	53321	41688	2529	2214
	Asia	19242	30209	760	861
	America & Oceania	12293	11781	157	178
2006	Danish born	0	0	103662	99333
	Europe	149752	111438	2228	2514
	Sub Saharan Africa	11668	10817	395	314
	Middle East & North Africa	53202	41729	2836	2460
	Asia	20014	31685	837	941
	America & Oceania	12712	12247	168	201
2007	Danish born	0	0	109715	104627
	Europe	154700	114818	2416	2724
	Sub Saharan Africa	11822	10994	423	342
	Middle East & North Africa	53150	41770	3101	2657
	Asia	20519	33140	917	1030
	America & Oceania	13207	12815	177	220
2008	Danish born	0	0	116143	110663
	Europe	162108	120156	2652	2907
	Sub Saharan Africa	12012	11231	467	384
	Middle East & North Africa	53455	42150	3367	2887
	Asia	22036	35481	989	1124
	America & Oceania	13878	13582	194	239
2009	Danish born	0	0	123598	117381
	Europe	171346	127280	2952	3179
	Sub Saharan Africa	12348	11595	512	434
	Middle East & North Africa	53698	42365	3700	3185
	Asia	23642	38030	1074	1239
	America & Oceania	14475	14219	219	249
2010	Danish born	0	0	130655	123545
	Europe	174475	132101	3261	3390
	Sub Saharan Africa	12623	11883	569	475
	Middle East & North Africa	54127	42709	4085	3528
	Asia	24265	40382	1200	1340
	America & Oceania	14826	14807	224	269

4.2 Population data

Hence we use `N.dk` to get the total population at each time point by sex and age, and the subtract foreign born and/or persons with diabetes, to get the Danish born non-diabetics.

```
> data( N.dk )
> str( N.dk )
'data.frame':        8600 obs. of  4 variables:
 $ sex: num  1 2 1 2 1 2 1 2 1 2 ...
 $ A  : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P  : num  1971 1971 1971 1971 1971 ...
 $ N  : num  35839 34108 36302 34153 37855 ...
 - attr(*, "Contents")= chr "Population size as of 1 January in Denmark"
> pp <- xtabs( N ~ sex + factor(A) + factor(P),
+               data = subset( N.dk, A<100 & P>1994 & P<2011 ) )
```

4.3 Total prevalence data

Then we must subtract those remaining in `tt`:

```
> str( pp )
 xtabs [1:2, 1:100, 1:16] 35612 34094 34747 32967 35082 ...
 - attr(*, "dimnames")=List of 3
  ..$ sex      : chr [1:2] "1" "2"
  ..$ factor(A): chr [1:100] "0" "1" "2" "3" ...
  ..$ factor(P): chr [1:16] "1995" "1996" "1997" "1998" ...
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = N ~ sex + factor(A) + factor(P), data = subset(N.dk, A
> str( apply(tt,1:3,sum) )
 num [1:2, 1:100, 1:16] 214 196 476 405 645 ...
 - attr(*, "dimnames")=List of 3
  ..$ sex      : chr [1:2] "M" "F"
  ..$ factor(A): chr [1:100] "0" "1" "2" "3" ...
  ..$ factor(P): chr [1:16] "1995" "1996" "1997" "1998" ...
> str( tt[,,,"Danish born","Well"] )
 num [1:2, 1:100, 1:16] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 3
  ..$ sex      : chr [1:2] "M" "F"
  ..$ factor(A): chr [1:100] "0" "1" "2" "3" ...
  ..$ factor(P): chr [1:16] "1995" "1996" "1997" "1998" ...
> tt[,,,"Danish born","Well"] <-  pp - apply(tt,1:3,sum)
> ftable( apply( tt, c(1,3:5), sum ), row.vars=c(2,3), col.vars=c(4,1) )
                                     status    Well             DM
                                     sex          M       F       M       F
factor(P) region
1995      Danish born                       2350786 2469268   48197   47722
          Europe                             113157   72574     441     629
          Sub Saharan Africa                   5656    4416      45      25
          Middle East & North Africa          33929   23206     508     486
          Asia                                12686   16788     155     164
          America & Oceania                    7776    7123      43      51
1996      Danish born                       2349932 2467015   51882   50866
          Europe                             124904   83681     564     776
          Sub Saharan Africa                   6702    5346      63      33
          Middle East & North Africa          35364   24483     611     602
          Asia                                13382   17883     185     210
          America & Oceania                    8632    7906      56      59
1997      Danish born                       2348984 2464583   55788   54312
          Europe                             130120   88808     735    1013
          Sub Saharan Africa                   8195    6729      76      54
          Middle East & North Africa          37169   26277     752     743
          Asia                                13817   19080     223     257
          America & Oceania                    9060    8308      63      69
1998      Danish born                       2348586 2461519   59492   57902
          Europe                             133593   92563     863    1159
          Sub Saharan Africa                   9182    7729     113      78
          Middle East & North Africa          38812   27922     893     873
```

	Asia	14290	20308	275	327
	America & Oceania	9544	8775	73	87
1999	Danish born	2346782	2458565	63952	61885
	Europe	136051	95223	1014	1339
	Sub Saharan Africa	9832	8545	141	95
	Middle East & North Africa	41338	30248	1055	1014
	Asia	14839	21564	332	395
	America & Oceania	10055	9242	79	95
2000	Danish born	2345151	2454836	68541	66156
	Europe	137853	97465	1128	1510
	Sub Saharan Africa	10308	9234	183	119
	Middle East & North Africa	43425	32272	1247	1163
	Asia	15374	22934	403	455
	America & Oceania	10472	9700	84	106
2001	Danish born	2343321	2451056	73356	70601
	Europe	139899	100028	1280	1621
	Sub Saharan Africa	10880	9940	214	144
	Middle East & North Africa	46545	35017	1445	1339
	Asia	15985	24414	476	522
	America & Oceania	10869	10149	99	114
2002	Danish born	2340318	2446948	78626	75447
	Europe	141615	102373	1430	1759
	Sub Saharan Africa	11278	10497	250	177
	Middle East & North Africa	50526	38301	1677	1534
	Asia	16612	25964	524	588
	America & Oceania	11234	10545	108	129
2003	Danish born	2337164	2440794	84442	81461
	Europe	143241	104435	1572	1920
	Sub Saharan Africa	11549	10864	295	208
	Middle East & North Africa	52478	40391	1948	1767
	Asia	17373	27398	590	662
	America & Oceania	11645	11035	126	149
2004	Danish born	2333859	2436029	91193	87696
	Europe	144699	106276	1780	2131
	Sub Saharan Africa	11598	10808	322	236
	Middle East & North Africa	53268	41321	2243	2005
	Asia	18497	28745	674	762
	America & Oceania	11868	11337	134	159
2005	Danish born	2330214	2430968	97927	94174
	Europe	146782	108605	2021	2318
	Sub Saharan Africa	11695	10833	351	284
	Middle East & North Africa	53321	41688	2529	2214
	Asia	19242	30209	760	861
	America & Oceania	12293	11781	157	178
2006	Danish born	2328372	2427934	103662	99333
	Europe	149752	111438	2228	2514
	Sub Saharan Africa	11668	10817	395	314
	Middle East & North Africa	53202	41729	2836	2460
	Asia	20014	31685	837	941
	America & Oceania	12712	12247	168	201
2007	Danish born	2326515	2425285	109715	104627
	Europe	154700	114818	2416	2724
	Sub Saharan Africa	11822	10994	423	342
	Middle East & North Africa	53150	41770	3101	2657
	Asia	20519	33140	917	1030
	America & Oceania	13207	12815	177	220
2008	Danish born	2325365	2422321	116143	110663
	Europe	162108	120156	2652	2907
	Sub Saharan Africa	12012	11231	467	384
	Middle East & North Africa	53455	42150	3367	2887
	Asia	22036	35481	989	1124
	America & Oceania	13878	13582	194	239
2009	Danish born	2324456	2420275	123598	117381
	Europe	171346	127280	2952	3179
	Sub Saharan Africa	12348	11595	512	434
	Middle East & North Africa	53698	42365	3700	3185

```
         Asia                              23642   38030    1074    1239
         America & Oceania                 14475   14219     219     249
2010     Danish born                     2322976 2417023  130655  123545
         Europe                           174475  132101    3261    3390
         Sub Saharan Africa                12623   11883     569     475
         Middle East & North Africa        54127   42709    4085    3528
         Asia                              24265   40382    1200    1340
         America & Oceania                 14826   14807     224     269
```

Thus tt contains for each date 1.1.1995 – 1.1.2010 the entire population classified by sex, age, region and diabetes status.

4.4 Modelling prevalences

Based on this we model the age-specific prevalences; basically we just smooth the empirical rates and the result is simply an array of smoothed prevalences with confidence intervals.

4.4.1 Separate modeling

We shall put these in an array of almost the same structure as tt - the only difference is that the last dimension is not the number of no DM / DM persons but the estimate and the 95% c.i.:

```
> pr <- NArray( c( dimnames(tt)[-5], list( c("Est","lo","hi") ) ) )
> str( pr )
 logi [1:2, 1:100, 1:16, 1:6, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ sex       : chr [1:2] "M" "F"
  ..$ factor(A): chr [1:100] "0" "1" "2" "3" ...
  ..$ factor(P): chr [1:16] "1995" "1996" "1997" "1998" ...
  ..$ region    : chr [1:6] "Danish born" "Europe" "Sub Saharan Africa" "Middle East & North Africa"
  ..$           : chr [1:3] "Est" "lo" "hi"
```

In order to fill this we model the prevalences by a logistic regression:

```
> library( splines )
> ( a.kn <- seq( 10,95,,5 ) )
[1] 10.00 31.25 52.50 73.75 95.00
> a.pr <- 0:99 + 0.5
> for( sx in dimnames(tt)[[1]] )
+ for( ip in dimnames(tt)[[3]] )
+ for( ir in dimnames(tt)[[4]][-7] )
+ pr[sx,,ip,ir,] <- ci.pred( glm( tt[sx,,ip,ir,2:1] ~ Ns( a.pr, knots=a.kn ),
+                                  family = binomial ),
+                             newdata = data.frame(a.pr=a.pr) )
> pr <- pr / (1+pr)
> pr <- pr * 100
```

We can now plot these in nice little film showing how prevalences develop over time

```
> c.ord <- 1:6 #c(1,5,3,6,2,4)
> cbind( dimnames(tt)[[4]][-7],
+        rcol <- c("red","blue","limegreen","black","orange","magenta") )
     [,1]                         [,2]
[1,] "Danish born"                "red"
[2,] "Europe"                     "blue"
[3,] "Sub Saharan Africa"         "limegreen"
[4,] "Middle East & North Africa" "black"
[5,] "Asia"                       "orange"
[6,] "America & Oceania"          "magenta"
```

```
> prpl <-
+ function(pr)
+ {
+ 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.pr, as.matrix(ftable(pr["M",,ip,,],col.vars=2:3)),
+          type="l", lty=c(1,0,0), lwd=c(4,1,1), col=rep(rcol,each=3),
+          ylim=c(0,40), xlab="", ylab="" )
+ text( 10, 35, ip )
+ text( 80, 5, "M" )
+ matplot( a.pr, as.matrix(ftable(pr["F",,ip,,],col.vars=2:3)),
+          type="l", lty=c(1,0,0), lwd=c(4,1,1), col=rep(rcol,each=3),
+          ylim=c(0,40), xlab="", ylab="", yaxt="n" )
+ text( rep(0,6), par("usr")[3:4] %*% rbind(xx<-seq(0.05,0.2,,6),1-xx),
+       dimnames(pr)[[4]][-7][c.ord], col=rcol[c.ord], font=2, adj=0 )
+ text( 80, 5, "F" )
+ mtext( "DM prevalence (%)", side=2, line=3, outer=T, las=0 )
+ mtext( "Age", side=1, line=2, outer=T, las=0 )
+    }
> pdf( "./graph/Prev-film-raw.pdf", height=8, width=12, pointsize=15 )
> for( ip in dimnames(tt)[[3]] ) prpl(pr)
> dev.off()
null device
          1
```

Here is the prevalences as they look in 2009

```
> ip <- "2009"
> prpl(pr)
> postscript("./graph/2paper/Fig2.eps", height=8, width=12, pointsize=15 )
> prpl(pr)
> dev.off()
pdf
  2
>        pdf("./graph/2paper/Fig2.pdf", height=8, width=12, pointsize=15 )
> prpl(pr)
> dev.off()
pdf
  2
```

4.4.2 Modeling time-trends in prevalence

In order to model time-trends in the prevalence, and also use information on the age-specific prevalences across periods we construct an analysis dataset:

```
> str( tt )
 xtabs [1:2, 1:100, 1:16, 1:6, 1:2] 35398 33898 34271 32562 34437 ...
 - attr(*, "dimnames")=List of 5
  ..$ sex      : chr [1:2] "M" "F"
  ..$ factor(A): chr [1:100] "0" "1" "2" "3" ...
  ..$ factor(P): chr [1:16] "1995" "1996" "1997" "1998" ...
  ..$ region   : chr [1:6] "Danish born" "Europe" "Sub Saharan Africa" "Middle East & North Africa" 
  ..$ status   : chr [1:2] "Well" "DM"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = Freq ~ sex + factor(A) + factor(P) + region + status, d
> dm <- data.frame( as.table( tt[,,,-7,"DM"] ) )
> wl <- data.frame( as.table( tt[,,,-7,"Well"] ) )
> names( dm )[c(2,3,5)] <- c("A","P","X")
> names( wl )[c(2,3,5)] <- c("A","P","N")
> pana <- transform( merge( dm, wl ),
+                    A = as.numeric( as.character(A) ) + 0.5,
+                    P = as.numeric( as.character(P) ) )
> str( pana )
```

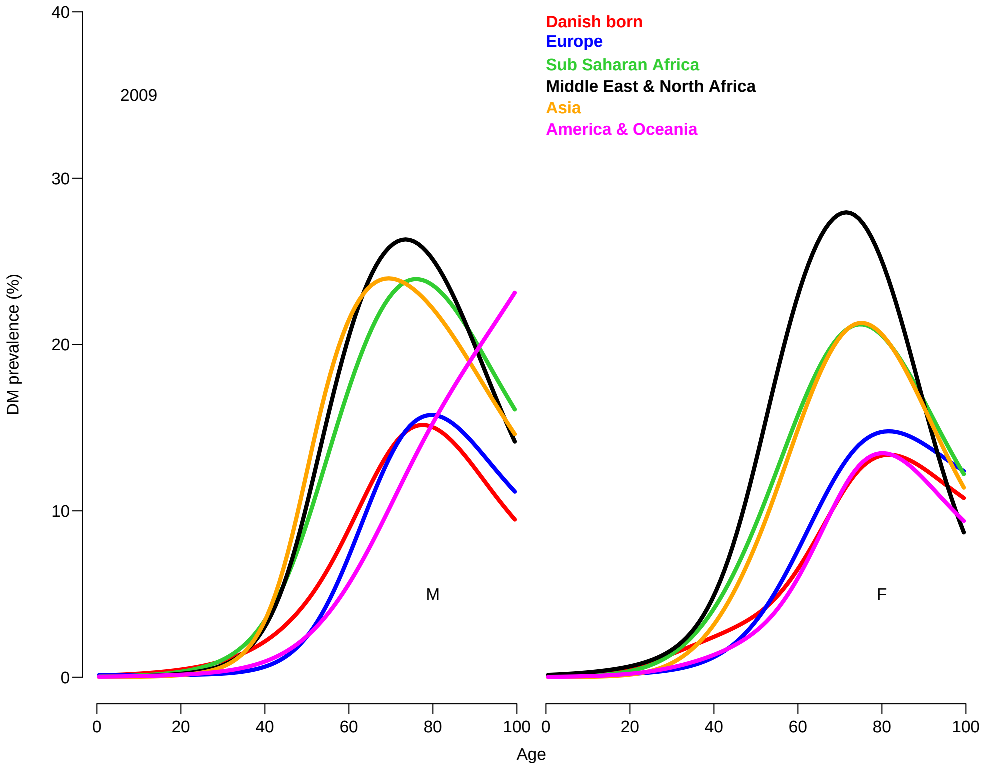

Figure 4.1: *Prevalence of Diabetes in immigrants from different regions as of 1.1.2009.*

```
'data.frame':        19200 obs. of  6 variables:
 $ sex   : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 2 2 2 2 ...
 $ A     : num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
 $ P     : num  1995 1995 1995 1995 1995 ...
 $ region: Factor w/ 6 levels "Danish born",..: 6 5 1 2 4 3 6 5 1 2 ...
 $ X     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ N     : num  35 27 33898 109 22 ...
```

The model for the prevalences that we shall use to describe the trends will be one with a fairly detailed age-specific prevalence overall, and regional-specific deviations from this with few parameters, and on top of this a linear trend in the prevalence.

Since we obtain more stability in a joint model we will use a log-link binomial, enabling interpretation as average annual relative increases:

```
> a.kn <- seq( 5,95,,10)
> r.kn <- seq(10,90,,4)
> mtot <- glm( cbind(X,N) ~ Ns( A, knots=a.kn ) +
+                          region +
+                          region:Ns(A,knots=r.kn) +
+                          region:I(P-2005),
+              family = binomial(link="log"),
+                data = subset( pana, sex=="M" ) )
> ftot <- update( mtot, data = subset( pana, sex=="F" ) )
```

Because we have an age×region interaction there is no such thing as a relative prevalence between regions, we will have to visualize the estimated age-specific prevalences at a given date, say 1.1.2009, but we *can* show the region specific average trends in prevalence over the period, here computed as percent per year sine we fitted a binomial model with log link:

```
> chg <- cbind( (ci.exp(mtot,subset="P")-1)*100,
+               (ci.exp(ftot,subset="P")-1)*100 )
> rownames( chg ) <- levels( pana$region )
> round( chg[c.ord,], 1 )
                           exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
Danish born                      5.9  5.8   5.9       6.2  6.1   6.2
Europe                           4.6  4.3   4.9       6.0  5.7   6.3
Sub Saharan Africa               4.6  3.9   5.4       6.9  6.0   7.9
Middle East & North Africa       4.4  4.2   4.7       4.0  3.7   4.3
Asia                             5.2  4.7   5.7       4.1  3.7   4.6
America & Oceania                4.5  3.5   5.6       6.5  5.6   7.5
```

In order to visualize the resulting estimated prevalences, we make predictions exactly as before:

```
> ps <- pr
> for( ip in 1995:2010 )
+ for( ir in levels(pana$region ) )
+    {
+    nd <- data.frame( A = 0:99+0.5,
+                      P = ip,
+                 region = factor( ir, levels=levels(pana$region) ) )
+    ps["M",,paste(ip),ir,] <- ci.pred( mtot, newdata = nd )
+    ps["F",,paste(ip),ir,] <- ci.pred( ftot, newdata = nd )
+    }
> ps <- ps * 100
```

With these predictions in place we can make the same type of film as before:

```
> pdf( "./graph/Prev-film-smooth.pdf", height=8, width=10 )
> for( ip in dimnames(tt)[[3]] ) prpl(ps)
> dev.off()
null device
          1
```

For the report we show the estimated prevalences in 2010

```
> ip <- "2009"
> prpl(ps)
> postscript("./graph/2paper/Fig2-s.eps", height=8, width=12, pointsize=15 )
> prpl(ps)
> dev.off()
pdf
  2
>         pdf("./graph/2paper/Fig2-s.pdf", height=8, width=12, pointsize=15 )
> prpl(ps)
> dev.off()
pdf
  2
```

For illustration in the report we also make a display with the estimated prevalences in 1995, 2002, and 2009:

```
> par( mfcol=c(2,3), mar=c(0,0,0,0), oma=c(3,4,1,1), mgp=c(3,1,0)/1.6,
+      las=1, bty="n" )
> for( ip in c("1995","2002","2009") )
+    {
+ matplot( a.pr, as.matrix(ftable(ps["M",,ip,-7,],col.vars=2:3)),
+          type="l", lty=c(1,0,0), lwd=c(4,1,1), col=rep(rcol,each=3),
+          ylim=c(0,45), xlab="", ylab="",
+          yaxt= if( ip!="1995") "n" else"s" )
+ text( 10, 40, paste( ip, ", M" ) )
+ if( ip=="1995" )
+ text( rep(0,6), par("usr")[3:4] %*% rbind(xx<-seq(0.25,0.4,,6),1-xx),
+       dimnames(ps)[[4]][-7][c.ord], col=rcol[c.ord], font=2, adj=0 )
```

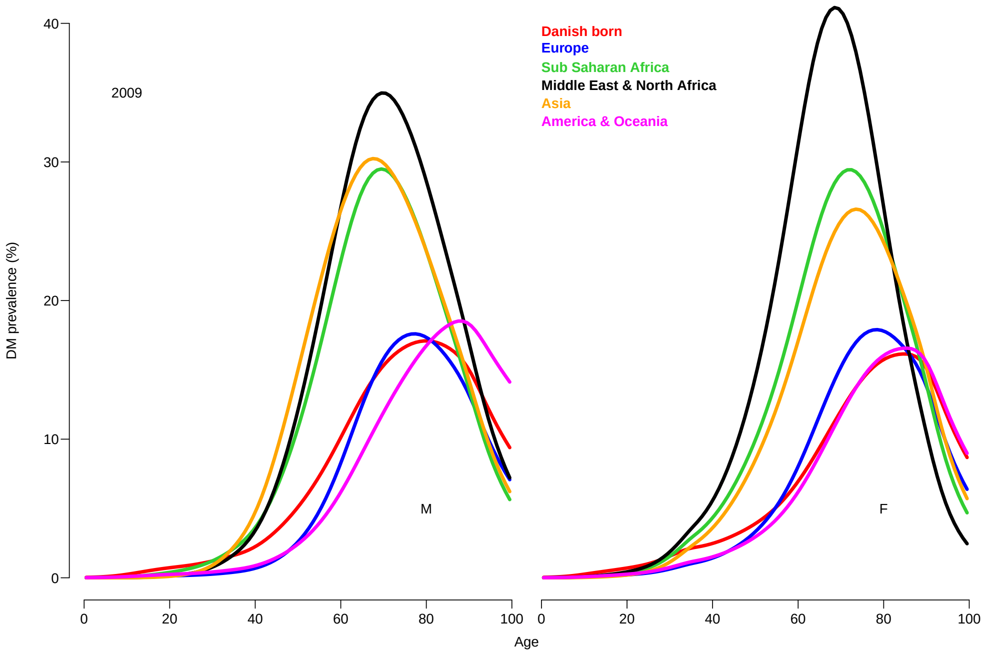

Figure 4.2: *Estimated prevalences of DM in immigrants from different regions as of 1.1.2009. Estimates are from a model with constant relative increase within each region, allowing different age-specific prevalence curves between regions, however constrained to be constant over time.*

```
+ matplot( a.pr, as.matrix(ftable(ps["F",,ip,-7,],col.vars=2:3)),
+          type="l", lty=c(1,0,0), lwd=c(4,1,1), col=rep(rcol,each=3),
+          ylim=c(0,45), xlab="", ylab="",
+          yaxt= if( ip!="1995") "n" else"s" )
+ text( 10, 40, paste( ip, ", F" ) )
+    }
> mtext( "DM prevalence (%)", side=2, line=3, outer=T, las=0 )
> mtext( "Age", side=1, line=2, outer=T, las=0 )
```

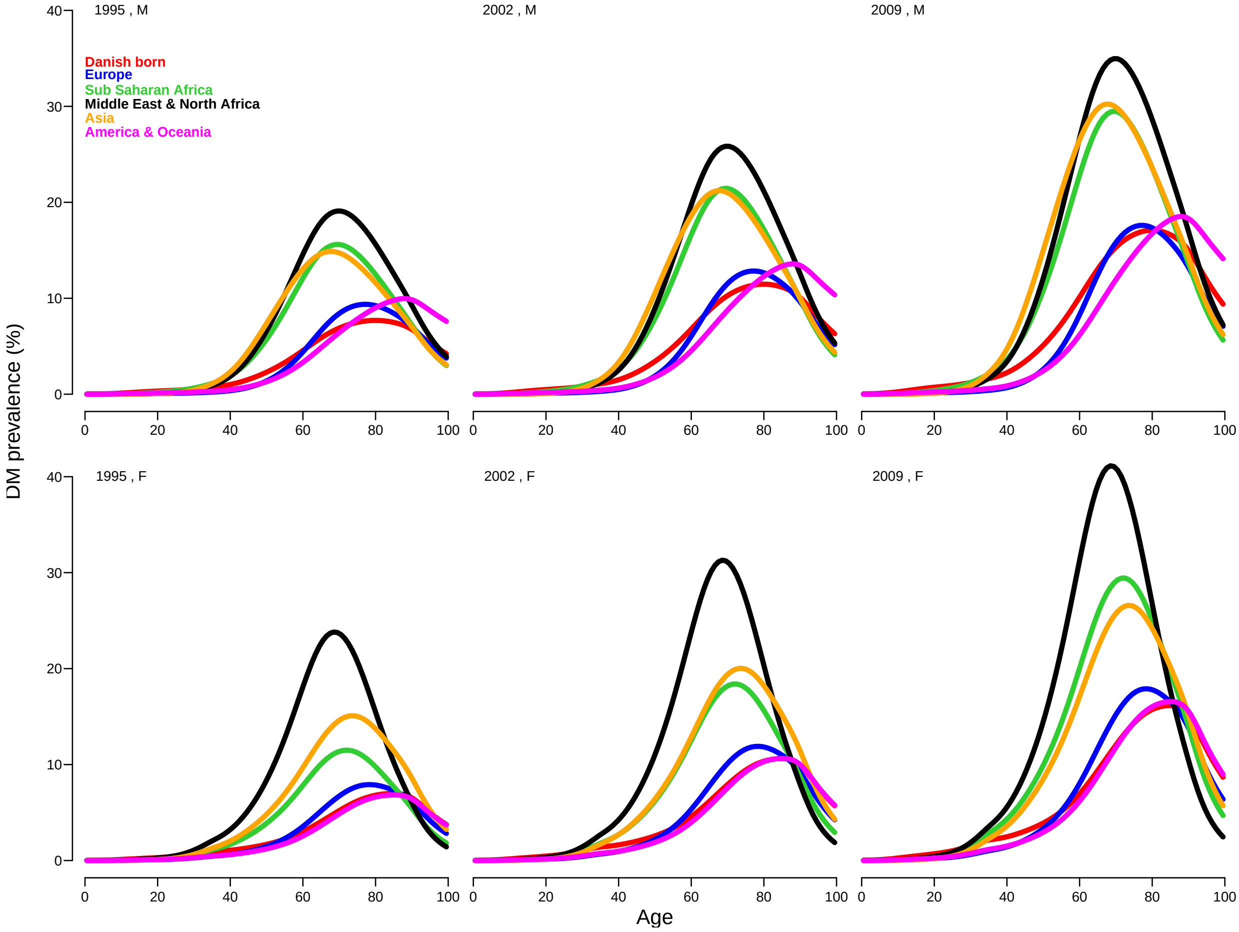

Figure 4.3: *Estimated prevalences from a model with constant relative increase within each region, allowing different age-specific prevalence curves between regions, however constrained to be constant over time.*

Chapter 5

DM incidence by country of origin

In this chapter we analyse the incidence rates of DM as recorded in the registers, so we first load the tabulated data:

```
> options( width=90 )
> library( Epi )
> load( file="./data/Afu.Rda" )
> str( Afu )
'data.frame':        65348 obs. of  10 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1995 ...
 $ U     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 2 2 ...
 $ state : Factor w/ 2 levels "Well","DM": 1 1 1 1 1 1 1 2 1 1 ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 2 3 4 5 6 7 1 1 2 ...
 $ D.Wdk : num  179 NA NA NA NA NA NA NA 137 NA ...
 $ Y     : num  17926.24 1.32 26.49 10.54 51.68 ...
 $ D.dm  : num  1 0 0 0 0 0 0 0 0 0 ...
 $ D.dd  : num  179 0 0 0 0 0 0 0 137 0 ...
```

Since we are analyzing diabetes incidence we do not need the part of the follow-up among the diabetes patients, and neither do we need the number of deaths:

```
> idat <- subset( Afu, state=="Well" & region != "Other",
+                      select=-c(state,D.Wdk,D.dd) )
> addmargins( xtabs( D.dm ~ region + sex, data=idat ) )
         sex
region         M      F    Sum
  DK      152024 139206 291230
  Africa     670    542   1212
  America    287    319    606
  Asia      1241   1343   2584
  Europe    3851   3980   7831
  MidEast   4057   3424   7481
  Other        0      0      0
  Sum     162130 148814 310944
```

Further we need to define the midpoint of follow-up in the Lexis triangles, shrink the region to the actually occurring levels and also define a grouping of regions in western (DK,America, Europe) and other, south-east (Africa, Middle East and Asia):

```
> idat <- transform( idat, A = A + (1+U)/3,
+                          P = P + (2-U)/3,
+                          D = D.dm,
+                          Y = Y/1000,
+                     region = Relevel( factor( region ),
+                                  list("Danish born" = 1,
+                                            "Europe" = 5,
```

```
+                                   "Sub Saharan Africa" = 2,
+                            "Middle East & North Africa" = 6,
+                                                "Asia" = 4,
+                                  "America & Oceania" = 3 ) ),
+                    Region = Relevel( factor(region),
+                                      list(DK=1,
+                                            W=c(3,5),
+                                           SE=c(2,4,6) ) ) )
> str( idat )
'data.frame':        34111 obs. of  9 variables:
 $ A     : num  0.333 0.333 0.333 0.333 0.333 ...
 $ P     : num  1996 1996 1996 1996 1996 ...
 $ U     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 2 2 2 2 ...
 $ region: Factor w/ 6 levels "Danish born",..: 1 3 6 5 2 4 1 3 6 5 ...
 $ Y     : num  17.92624 0.00132 0.02649 0.01054 0.05168 ...
 $ D.dm  : num  1 0 0 0 0 0 0 0 0 0 ...
 $ D     : num  1 0 0 0 0 0 0 0 0 0 ...
 $ Region: Factor w/ 3 levels "DK","W","SE": 1 3 2 3 2 3 1 3 2 3 ...
> with( idat, table( region, Region ) )
                            Region
region                         DK    W   SE
  Danish born                6000    0    0
  Europe                        0 5978    0
  Sub Saharan Africa            0    0 5065
  Middle East & North Africa    0    0 5648
  Asia                          0    0 5671
  America & Oceania             0 5749    0
```

5.1 Incidence of diabetes

First a brief overview of events (DM diagnoses) and person-years:

```
> round( ff <-ftable( addmargins( xtabs( cbind( D, Y ) ~ region + sex,
+                                        data=idat ),
+                                 margin = 1:2 ),
+                   col.vars=3:2 ), 1 )
                                   D                          Y
                           sex     M        F      Sum        M        F      Sum
region
Danish born                 152024.0 139206.0 291230.0  35059.2  36649.4  71708.5
Europe                        3851.0   3980.0   7831.0   2159.9   1553.8   3713.7
Sub Saharan Africa             670.0    542.0   1212.0    158.3    143.4    301.7
Middle East & North Africa    4057.0   3424.0   7481.0    709.0    538.3   1247.3
Asia                          1241.0   1343.0   2584.0    264.3    405.5    669.8
America & Oceania              287.0    319.0    606.0    171.9    163.4    335.3
Sum                         162130.0 148814.0 310944.0  38522.5  39453.8  77976.3
```

It is seen that America, Africa are quite thin on events, together even smaller than Asia. But Europe and Middle East are quite well represented, hence also the variable `Region`:


```
> with( idat, table( region, Region ) )
                            Region
region                         DK    W   SE
  Danish born                6000    0    0
  Europe                        0 5978    0
  Sub Saharan Africa            0    0 5065
  Middle East & North Africa    0    0 5648
  Asia                          0    0 5671
  America & Oceania             0 5749    0
```

```
> round( ff <-ftable( addmargins( xtabs( cbind( D, Y ) ~ Region + sex,
+                                    data=idat ),
+                               margin = 1:2 ),
+                   col.vars=3:2 ), 1 )
                  D                             Y
       sex        M        F      Sum         M        F      Sum
Region
DK         152024.0 139206.0 291230.0   35059.2  36649.4  71708.5
W            4138.0   4299.0   8437.0    2331.7   1717.2   4049.0
SE           5968.0   5309.0  11277.0    1131.6   1087.2   2218.8
Sum        162130.0 148814.0 310944.0   38522.5  39453.8  77976.3
```

5.1.1 Age-Period-Cohort models for each region

For a start we show results from separate age-period-cohort models for each region. For display we collect the results in a matrix of lists with the results from each model:

```
> dnam <- list( region = levels(idat$region),
+                 sex = levels(idat$sex) )
> effs <- c("Age","Per","Coh","Drift")
> res <- list( NULL )
>   length( res ) <- prod( sapply( dnam, length ) )
>      dim( res ) <-       sapply( dnam, length )
> dimnames( res ) <-              dnam
> for( rg in levels(idat$region) )
+    {
+    cat( "\n---------------------------------------------------\n", rg )
+ for( sx in levels(idat$sex) )
+    {
+    cat( "\n", sx, "\n" )
+    res[[rg,sx]] <- apc.fit( subset( idat, region==rg & sex==sx ),
+                              npar=c(5,3,5),
+                              parm = "ACP",
+                             ref.c = 1950 )
+    }
+    }
---------------------------------------------------
 Danish born
 M
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
Age                    2994     8446.7
Age-drift              2993     4564.1  1   3882.6 < 2.2e-16
Age-Cohort             2989     4503.7  4     60.4 2.336e-12
Age-Period-Cohort      2987     4424.3  2     79.3 < 2.2e-16
Age-Period             2991     4482.1 -4    -57.8 8.587e-12
Age-drift              2993     4564.1 -2    -82.0 < 2.2e-16

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
Age                    2994     8691.4
Age-drift              2993     4769.1  1   3922.3 < 2.2e-16
Age-Cohort             2989     4673.1  4     96.0 < 2.2e-16
Age-Period-Cohort      2987     4416.3  2    256.8 < 2.2e-16
Age-Period             2991     4533.8 -4   -117.5 < 2.2e-16
Age-drift              2993     4769.1 -2   -235.3 < 2.2e-16
```

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
Age                    2972     2727.1
Age-drift              2971     2714.4  1  12.6320 0.0003792
Age-Cohort             2967     2693.6  4  20.8312 0.0003420
Age-Period-Cohort      2965     2692.8  2   0.8609 0.6502271
Age-Period             2969     2713.6 -4 -20.8147 0.0003446
Age-drift              2971     2714.4 -2  -0.8774 0.6448765

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
Age                    2994     2988.7
Age-drift              2993     2978.7  1  10.0028 0.0015630
Age-Cohort             2989     2957.5  4  21.1164 0.0003003
Age-Period-Cohort      2987     2952.4  2   5.1321 0.0768398
Age-Period             2991     2974.3 -4 -21.9249 0.0002074
Age-drift              2993     2978.7 -2  -4.3236 0.1151190

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
Age                    2531     1460.2
Age-drift              2530     1458.3  1   1.8989   0.1682
Age-Cohort             2526     1452.9  4   5.3329   0.2548
Age-Period-Cohort      2524     1449.4  2   3.5272   0.1714
Age-Period             2528     1455.0 -4  -5.5559   0.2349
Age-drift              2530     1458.3 -2  -3.3043   0.1916

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
Age                    2522     1381.4
Age-drift              2521     1381.1  1   0.2404   0.6239
Age-Cohort             2517     1376.1  4   5.0094   0.2863
Age-Period-Cohort      2515     1372.8  2   3.2683   0.1951
Age-Period             2519     1378.6 -4  -5.7670   0.2172
Age-drift              2521     1381.1 -2  -2.5107   0.2850

---------------------------------------------------
 Middle East & North Africa
 M
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
Age                    2793     2054.8
Age-drift              2792     2052.4  1   2.4166 0.1200551
```

```
Age-Cohort             2788     2042.1  4  10.2694 0.0361267
Age-Period-Cohort      2786     2024.4  2  17.7275 0.0001414
Age-Period             2790     2034.5 -4 -10.0908 0.0389251
Age-drift              2792     2052.4 -2 -17.9060 0.0001293

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
Age                    2843     2337.9
Age-drift              2842     2337.1  1   0.7583 0.383863
Age-Cohort             2838     2328.9  4   8.2706 0.082153
Age-Period-Cohort      2836     2315.5  2  13.3388 0.001269
Age-Period             2840     2323.9 -4  -8.3829 0.078517
Age-drift              2842     2337.1 -2 -13.2265 0.001342

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
Age                    2806     1913.1
Age-drift              2805     1908.2  1   4.9043  0.02679
Age-Cohort             2801     1902.5  4   5.7634  0.21753
Age-Period-Cohort      2799     1897.6  2   4.8841  0.08698
Age-Period             2803     1902.9 -4  -5.2989  0.25798
Age-drift              2805     1908.2 -2  -5.3485  0.06896

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
Age                    2853     2024.5
Age-drift              2852     2017.7  1   6.7036 0.009622
Age-Cohort             2848     2011.6  4   6.1650 0.187158
Age-Period-Cohort      2846     2009.8  2   1.7917 0.408265
Age-Period             2850     2015.3 -4  -5.5340 0.236754
Age-drift              2852     2017.7 -2  -2.4227 0.297800

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

Analysis of deviance for Age-Period-Cohort model

                  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
Age                    2808     1169.8
Age-drift              2807     1169.0  1   0.8385 0.359833
Age-Cohort             2803     1153.2  4  15.8088 0.003287
Age-Period-Cohort      2801     1150.0  2   3.1208 0.210056
Age-Period             2805     1167.6 -4 -17.5370 0.001520
Age-drift              2807     1169.0 -2  -1.3925 0.498455

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

Analysis of deviance for Age-Period-Cohort model
```

```
                  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
Age                    2929     1296.5
Age-drift              2928     1294.2  1   2.3366  0.12636
Age-Cohort             2924     1287.5  4   6.7129  0.15186
Age-Period-Cohort      2922     1281.1  2   6.4389  0.03998
Age-Period             2926     1288.7 -4  -7.6528  0.10516
Age-drift              2928     1294.2 -2  -5.4990  0.06396
```

A preliminary overview can be created by plotting age-period-cohort effects on diabetes incidence separately for the different regions:

```
> cbind( levels(idat$region),
+        rcol <- c("red","blue","limegreen","black","orange","magenta") )
     [,1]                          [,2]
[1,] "Danish born"                 "red"
[2,] "Europe"                      "blue"
[3,] "Sub Saharan Africa"          "limegreen"
[4,] "Middle East & North Africa"  "black"
[5,] "Asia"                        "orange"
[6,] "America & Oceania"           "magenta"
> par( mfcol=c(2,1), mar=c(0,0,0,0), oma=c(3,4,1,4), mgp=c(3,1,0)/1.6, las=1 )
> for( sx in c("M","F") )
+    {
+ apc.frame( a.lab = seq(10,90,20),
+           cp.lab = seq(1920,2010,30),
+            r.lab = c(c(if(sx=="F")2,5,10)/100,
+                      c(2,5,10)/10,
+                      c(2,5,10),
+                      c(2,5,10)*10),
+            a.tic = seq(10,100,10),
+           cp.tic = seq(1900,2010,10),
+            r.tic = outer(c(2:10),10^c(-2:1)),
+           rr.ref = 1,
+            a.txt = "Age",
+           cp.txt = "Calendar time (birth / follow-up)",
+            r.txt = "",
+           rr.txt = "",
+         ref.line = TRUE,
+         col.grid = gray(0.85),
+            sides = if( sx=="M" ) c(2,4) else c(1,2,4) )
+ text( 1920 - getOption( "apc.frame.par" )[1], 93, paste(sx,"DM cases"), adj=c(0.5,1), font=1 )
+ for( i in 1:nrow(res) )
+    {
+ text( 1921 - getOption( "apc.frame.par" )[1], 93*(0.7^i),
+       rownames(res)[i], col=rcol[i], adj=c(0,1) )
+ text( 1920 - getOption( "apc.frame.par" )[1], 93*(0.7^i),
+       formatC( sum( subset( idat, region==levels(region)[i] & sex==sx )$D ),
+                format="f", digits=0, big.mark="," ),
+       col=rcol[i], adj=c(1,1) )
+ apc.lines( res[[i,sx]], col=rcol[i] )
+    }
+ mtext( "Rate of DM per 1000 PY", side=2, line=3, outer=TRUE, las=0)
+ mtext( "Rate ratio", side=4, line=3, outer=TRUE, las=0)
+    }
```

From the figure 5.1 it is clear that the age-incidence curves come in two groups: DK/Europe/America with a late peak and AFrica/Asia/MidEast with higher incidences in younger ages and an earlier peak, and — at least for women, a tendency to decline after age 60.

5.1.2 Common incidence models

The simplest possible model is of course a model that ignores any differences between ethnic groups. This is hardly tenable in the wiew of the results from the separate

APC-analyses shown in figure 5.1.

We therefore fit 3 different extensions to the simplified APC-model (as previously, everything still separately for the two sexes):

1. A separate overall RR for each ethnicity (`m0,f0`).

   This is just reported as a table of RRs relative to the Danish population.

2. An additional separate trend (drift) for each ethnic group (`mi,fi`).

   In principle this could be reported as the RR at say 2009 and the average annual change in DM incidence. If we took a cohort perspective, we might instead report this as the RRs in the 1945 cohort (which would be different from the RRs at 2009), and the average annual change in incidence from cohort to cohort, which would be the same as before.

   We will estimate the trend in two different guises; one with allowance for non-linear effects of period and cohort (assumed identical across ethnic groups) and one where only a linear effect of period/cohort is included.

3. Allowing separate age-specific incidence rates for the three groups of ethninc groups in addition to the separate trends (`mia,fia`).

   This would naturally be reported in the same way as the the model above, it basically just corresponds to allowing different shapes between ethnicities, in both cases we must refer the age-specific incidneces to a particular period or cohort.

Wehen reporting the differences between the ethnic groups we can either do it by showing the age-specific rates as cross-sectional (period) or longitudinal (cohort) rates. It is only the reporting that differs; the model fit to the data is the same, but for comparability with most other reports we have chosen to report from the period models, that is cross-sectional age-rates.

```
>  a.kn <- seq(  10,   95,,9)
>  p.kn <- seq(1995,2009,,5)
>  c.kn <- seq(1890,1990,,8)
> m0 <- glm( D ~ Ns( A, knots=a.kn, intercept=TRUE ) - 1 +
+               detrend( Ns( P   , knots=p.kn ), P    ) +
+               detrend( Ns( P-A, knots=c.kn ), P-A ) +
+               I( P-2009 ) + Relevel(region,c(2:6,1)),
+           offset = log( Y ),
+           family = poisson,
+             data = subset(idat,sex=="M") )
> mi  <- update( m0, . ~ . + region:I(P-2009) - I(P-2009) )
> mil <- update( m0, . ~ Ns( A, knots=a.kn, intercept=TRUE ) - 1 +
+                        Relevel(region,c(2:6,1)) + region:I(P-2009) )
> mia <- update( m0, . ~ Ns( A, knots=a.kn ):Region + region + region:I(P-2009) )
> f0  <- update( m0, data = subset(idat,sex=="F") )
> fi  <- update( mi, data = subset(idat,sex=="F") )
> fil <- update( mil,data = subset(idat,sex=="F") )
> fia <- update( mia,data = subset(idat,sex=="F") )
> round( cbind( ci.exp( m0, subset=c("region","I") ),
+               ci.exp( f0, subset=c("region","I") ) ), 3 )
                                                 exp(Est.)  2.5% 97.5% exp(Est.)
Relevel(region, c(2:6, 1))Europe                     0.810 0.784 0.836     1.095
Relevel(region, c(2:6, 1))Sub Saharan Africa         2.601 2.410 2.807     2.676
Relevel(region, c(2:6, 1))Middle East & North Africa 2.526 2.447 2.607     3.633
Relevel(region, c(2:6, 1))Asia                       2.610 2.468 2.761     2.032
Relevel(region, c(2:6, 1))America & Oceania          0.751 0.668 0.843     0.959
```

```
Relevel(region, c(2:6, 1))Danish born                1.000 1.000 1.000      1.000
I(P - 2009)                                          1.034 1.032 1.036      1.036
                                                   2.5% 97.5%
Relevel(region, c(2:6, 1))Europe                   1.061 1.130
Relevel(region, c(2:6, 1))Sub Saharan Africa       2.458 2.913
Relevel(region, c(2:6, 1))Middle East & North Africa 3.509 3.761
Relevel(region, c(2:6, 1))Asia                     1.925 2.145
Relevel(region, c(2:6, 1))America & Oceania        0.859 1.070
Relevel(region, c(2:6, 1))Danish born              1.000 1.000
I(P - 2009)                                        1.034 1.038
> round( rr.lin <-
+        cbind( ci.exp( mi, subset=c("region") ),
+               ci.exp( fi, subset=c("region") ) ), 3 )
                                                   exp(Est.)  2.5% 97.5% exp(Est.)
Relevel(region, c(2:6, 1))Europe                       0.748 0.711 0.787     0.956
Relevel(region, c(2:6, 1))Sub Saharan Africa           2.093 1.849 2.370     2.491
Relevel(region, c(2:6, 1))Middle East & North Africa   2.248 2.138 2.364     3.142
Relevel(region, c(2:6, 1))Asia                         2.438 2.231 2.664     1.635
Relevel(region, c(2:6, 1))America & Oceania            0.664 0.551 0.800     0.875
Relevel(region, c(2:6, 1))Danish born                  1.000 1.000 1.000     1.000
I(P - 2009):regionDanish born                          1.036 1.034 1.038     1.039
I(P - 2009):regionEurope                               1.020 1.012 1.028     1.014
I(P - 2009):regionSub Saharan Africa                   0.991 0.973 1.010     1.024
I(P - 2009):regionMiddle East & North Africa           1.012 1.004 1.020     1.010
I(P - 2009):regionAsia                                 1.022 1.009 1.036     0.995
I(P - 2009):regionAmerica & Oceania                    1.012 0.984 1.040     1.021
                                                   2.5% 97.5%
Relevel(region, c(2:6, 1))Europe                   0.907 1.008
Relevel(region, c(2:6, 1))Sub Saharan Africa       2.176 2.853
Relevel(region, c(2:6, 1))Middle East & North Africa 2.971 3.324
Relevel(region, c(2:6, 1))Asia                     1.497 1.786
Relevel(region, c(2:6, 1))America & Oceania        0.731 1.046
Relevel(region, c(2:6, 1))Danish born              1.000 1.000
I(P - 2009):regionDanish born                      1.037 1.041
I(P - 2009):regionEurope                           1.006 1.021
I(P - 2009):regionSub Saharan Africa               1.003 1.047
I(P - 2009):regionMiddle East & North Africa       1.001 1.018
I(P - 2009):regionAsia                             0.982 1.008
I(P - 2009):regionAmerica & Oceania                0.995 1.049
> round( RR.lin <-
+        cbind( ci.exp( mil, subset=c("region") ),
+               ci.exp( fil, subset=c("region") ) ), 3 )
                                                   exp(Est.)  2.5% 97.5% exp(Est.)
Relevel(region, c(2:6, 1))Europe                       0.739 0.703 0.778     0.948
Relevel(region, c(2:6, 1))Sub Saharan Africa           2.064 1.825 2.335     2.419
Relevel(region, c(2:6, 1))Middle East & North Africa   2.221 2.114 2.334     3.087
Relevel(region, c(2:6, 1))Asia                         2.405 2.202 2.625     1.600
Relevel(region, c(2:6, 1))America & Oceania            0.657 0.546 0.791     0.865
Relevel(region, c(2:6, 1))Danish born                  1.000 1.000 1.000     1.000
regionDanish born:I(P - 2009)                          1.038 1.037 1.039     1.040
regionEurope:I(P - 2009)                               1.019 1.011 1.027     1.012
regionSub Saharan Africa:I(P - 2009)                   0.990 0.972 1.009     1.018
regionMiddle East & North Africa:I(P - 2009)           1.011 1.003 1.019     1.006
regionAsia:I(P - 2009)                                 1.021 1.008 1.035     0.992
regionAmerica & Oceania:I(P - 2009)                    1.011 0.984 1.039     1.019
                                                   2.5% 97.5%
Relevel(region, c(2:6, 1))Europe                   0.899 0.998
Relevel(region, c(2:6, 1))Sub Saharan Africa       2.117 2.764
Relevel(region, c(2:6, 1))Middle East & North Africa 2.922 3.262
Relevel(region, c(2:6, 1))Asia                     1.467 1.745
Relevel(region, c(2:6, 1))America & Oceania        0.725 1.032
Relevel(region, c(2:6, 1))Danish born              1.000 1.000
regionDanish born:I(P - 2009)                      1.038 1.041
regionEurope:I(P - 2009)                           1.005 1.020
regionSub Saharan Africa:I(P - 2009)               0.997 1.040
regionMiddle East & North Africa:I(P - 2009)       0.998 1.014
```

```
regionAsia:I(P - 2009)                                0.979 1.005
regionAmerica & Oceania:I(P - 2009)                   0.993 1.046
> anova( m0, mi, mil, mia, test="Chisq" )
Analysis of Deviance Table

Model 1: D ~ Ns(A, knots = a.kn, intercept = TRUE) - 1 + detrend(Ns(P,
    knots = p.kn), P) + detrend(Ns(P - A, knots = c.kn), P -
    A) + I(P - 2009) + Relevel(region, c(2:6, 1))
Model 2: D ~ Ns(A, knots = a.kn, intercept = TRUE) + detrend(Ns(P, knots = p.kn),
    P) + detrend(Ns(P - A, knots = c.kn), P - A) + Relevel(region,
    c(2:6, 1)) + I(P - 2009):region - 1
Model 3: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Relevel(region, c(2:6,
    1)) + region:I(P - 2009) - 1
Model 4: D ~ region + Ns(A, knots = a.kn):Region + region:I(P - 2009)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     16916      13841
2     16911      13768  5    73.17 2.245e-14
3     16920      14017 -9  -249.80 < 2.2e-16
4     16904      13191 16   826.16 < 2.2e-16
> anova( f0, fi, fil, fia, test="Chisq" )
Analysis of Deviance Table

Model 1: D ~ Ns(A, knots = a.kn, intercept = TRUE) - 1 + detrend(Ns(P,
    knots = p.kn), P) + detrend(Ns(P - A, knots = c.kn), P -
    A) + I(P - 2009) + Relevel(region, c(2:6, 1))
Model 2: D ~ Ns(A, knots = a.kn, intercept = TRUE) + detrend(Ns(P, knots = p.kn),
    P) + detrend(Ns(P - A, knots = c.kn), P - A) + Relevel(region,
    c(2:6, 1)) + I(P - 2009):region - 1
Model 3: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Relevel(region, c(2:6,
    1)) + region:I(P - 2009) - 1
Model 4: D ~ region + Ns(A, knots = a.kn):Region + region:I(P - 2009)
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     17147      15075
2     17142      14957  5   118.72 < 2.2e-16
3     17151      15484 -9  -527.12 < 2.2e-16
4     17135      14478 16  1005.65 < 2.2e-16
```

From the tables of estimates it is seen that if we restrict to the model with common age-structure, the immmigrants from Africa, Asia and Middle East have DM incidence rates about 2.5 times that of the Danish men and less for Asian women (2.0) and much more for East women (3.6). Among immigrants from Europe and America (incl. Oceania) men have 20% lower rates in men, and in women about the same rates.

When allowing for a separate annual change per year, we see that is is 3.5% in Denmark, and about 3% in all other regions, except for Europe where the increase is only 1.5% among men, but 3% among women.

Since this is a register study, the extensions of the models with separate trends of incidence by calendar time as well as with region-specific age-effects of course represent significant improvements, but we will restrict the graphical reporting to the model with age-specific incidence rates in the three regions, and linear trends by date of birth. Thus, we are extracting estimates from the models `mia` and `fia`, so reporting the RR between ethnicities at 1.1.2009, showing them as three different age-specific incidence curves, and on top of this reporting the average annual trend in incicences for each of the 5 ethnic groups.

To this end we need the contrast matrix for the age-spline used in the model:

```
> a.pt <- 20:90
> CA <- Ns( a.pt, knots=a.kn )
> effs <- NArray( list( sex = c("M","F"),
+                      region = levels( idat$region ),
+                           A = a.pt,
+                        what = c("Est","lo","hi") ) )
> str( effs )
```

```
 logi [1:2, 1:6, 1:71, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ sex   : chr [1:2] "M" "F"
  ..$ region: chr [1:6] "Danish born" "Europe" "Sub Saharan Africa" "Middle East & North Africa" ...
  ..$ A     : chr [1:71] "20" "21" "22" "23" ...
  ..$ what  : chr [1:3] "Est" "lo" "hi"
> effs["M",1,,] <- ci.exp( mia, ctr.mat=cbind(1,0,0,0,0,0, CA*1,CA*0,CA*0, 0,0,0,0,0,0) )
> effs["M",2,,] <- ci.exp( mia, ctr.mat=cbind(1,1,0,0,0,0, CA*0,CA*1,CA*0, 0,0,0,0,0,0) )
> effs["M",3,,] <- ci.exp( mia, ctr.mat=cbind(1,0,1,0,0,0, CA*0,CA*0,CA*1, 0,0,0,0,0,0) )
> effs["M",4,,] <- ci.exp( mia, ctr.mat=cbind(1,0,0,1,0,0, CA*0,CA*0,CA*1, 0,0,0,0,0,0) )
> effs["M",5,,] <- ci.exp( mia, ctr.mat=cbind(1,0,0,0,1,0, CA*0,CA*0,CA*1, 0,0,0,0,0,0) )
> effs["M",6,,] <- ci.exp( mia, ctr.mat=cbind(1,0,0,0,0,1, CA*0,CA*1,CA*0, 0,0,0,0,0,0) )
> effs["F",1,,] <- ci.exp( fia, ctr.mat=cbind(1,0,0,0,0,0, CA*1,CA*0,CA*0, 0,0,0,0,0,0) )
> effs["F",2,,] <- ci.exp( fia, ctr.mat=cbind(1,1,0,0,0,0, CA*0,CA*1,CA*0, 0,0,0,0,0,0) )
> effs["F",3,,] <- ci.exp( fia, ctr.mat=cbind(1,0,1,0,0,0, CA*0,CA*0,CA*1, 0,0,0,0,0,0) )
> effs["F",4,,] <- ci.exp( fia, ctr.mat=cbind(1,0,0,1,0,0, CA*0,CA*0,CA*1, 0,0,0,0,0,0) )
> effs["F",5,,] <- ci.exp( fia, ctr.mat=cbind(1,0,0,0,1,0, CA*0,CA*0,CA*1, 0,0,0,0,0,0) )
> effs["F",6,,] <- ci.exp( fia, ctr.mat=cbind(1,0,0,0,0,1, CA*0,CA*1,CA*0, 0,0,0,0,0,0) )
```

On top of these estimated age-specific rates (cross-sectional from 2009) we also want to report the average annual change in DM rates in each of the ethnic groups:

```
> RR.ann <- NArray( list( region = levels( idat$region ),
+                            sex = levels( idat$sex ),
+                                  c("RR","up","lo") ) )
> str( RR.ann )
 logi [1:6, 1:2, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ region: chr [1:6] "Danish born" "Europe" "Sub Saharan Africa" "Middle East & North Africa" ...
  ..$ sex   : chr [1:2] "M" "F"
  ..$       : chr [1:3] "RR" "up" "lo"
> RR.ann[,"M",] <- ( ci.exp( mia, subset="P" ) - 1)*100
> RR.ann[,"F",] <- ( ci.exp( fia, subset="P" ) - 1)*100
> round( ftable( RR.ann, row.vars=1 ), 2 )
                           sex     M                   F
                                  RR    up    lo    RR    up    lo
region
Danish born                     3.79  3.66  3.91  3.99  3.86  4.12
Europe                          1.48  0.70  2.26  1.20  0.46  1.94
Sub Saharan Africa             -1.76 -3.58  0.09  0.24 -1.85  2.38
Middle East & North Africa      0.71 -0.05  1.48 -0.31 -1.12  0.51
Asia                            1.60  0.25  2.97 -1.72 -2.96 -0.45
America & Oceania               0.91 -1.82  3.73  1.72 -0.88  4.39
```

With this table in place we can plot the three age-specific incidences as they look in 2009 and annotate the plot with the percentwise annual changes:

```
> ap.lin <-
+ function( mfac=0.9, wh=0.8, ci=FALSE, a0=75, ax=seq(-5,5,5), as=3, nt=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 )
+ # limits of secondary axis (on the age-scale)
+ axl <- range( a0+ax*as )
+ tpos <- seq(a0+min(ax*as),a0+max(ax*as),,nt)
+ # order of regions
+ c.ord <- 1:6 # c(1,5,3,6,2,4)
+ # plot for males
+ matplot( a.pt, t(effs["M",,,1]), bty="n", ylim=yl<-c(0.2,50),
+          log="y", type="l", lty=1, col=rcol, lwd=5 )
+ text( rep(20,6), (10^par("usr")[4])*((mfac^1.5)^c(1:6)),
+       levels( idat$region )[c.ord], col=rcol[c.ord], font=2, adj=0 )
+ # confidence limits
+ if( ci ) for( j in 2:3 )
+ matlines( a.pt, t(effs["M",,,j]), type="l", lty=1, col=rcol, lwd=2 )
+ text( a0, wh, "Men", font=2 )
```

```
+ segments( tpos, wh*(mfac^7),
+           tpos, wh*(mfac^1), col=gray(0.8) )
+ segments( a0, wh*(mfac^7), a0, wh*(mfac^1) )
+ axis( side=1, at=a0+ax*as, labels=ax, pos=wh*(mfac^7), cex=0.9 )
+ axis( side=1, at=tpos, labels=FALSE, pos=wh*(mfac^7), tcl=-0.3 )
+ linesEst( RR.ann[c.ord,"M",]*as+a0, y=wh*(mfac^(1:6)), col=rcol[c.ord], lwd=3 )
+ text( a0, wh*(mfac^10), "Annual change(%)", cex=0.9 )
+
+ matplot( a.pt, t(effs["F",,,1]), yaxt="n", bty="n", ylim=yl,
+          log="y", type="l", lty=1, col=rcol, lwd=5 )
+ if( ci ) for( j in 2:3 )
+ matlines( a.pt, t(effs["F",,,j]), type="l", lty=1, col=rcol, lwd=2 )
+ text( a0, wh, "Women", font=2 )
+ segments( tpos, wh*(mfac^7),
+           tpos, wh*(mfac^1), col=gray(0.8) )
+ segments( a0, wh*(mfac^7), a0, wh*(mfac^1) )
+ axis( side=1, at=a0+ax*as, labels=ax, pos=wh*(mfac^7), cex=0.9 )
+ axis( side=1, at=tpos, labels=FALSE, pos=wh*(mfac^7), tcl=-0.3 )
+ linesEst( RR.ann[c.ord,"F",]*as+a0, y=wh*(mfac^(1:6)), col=rcol[c.ord], lwd=3 )
+ text( a0, wh*(mfac^10), "Annual change(%)", cex=0.9 )
+
+ mtext( "DM incidence rate 2009 (per 1000 PY)", side=2, line=3, outer=T, las=0 )
+ mtext( "Age", side=1, line=2, outer=T, las=0 )
+    }
> ap.lin()


> ap.lin(ci=TRUE)


>           pdf( "./graph/2paper/Fig1.pdf", height=8, width=12, pointsize=15 ) ; ap.lin(0.88,1.0,a0=
null device
          1
> #win.metafile( "./graph/2paper/Fig1.emf", height=8, width=12, pointsize=15 ) ; ap.lin(0.88,1.0) ;
>    postscript( "./graph/2paper/Fig1.eps", height=8, width=12, pointsize=15 ) ; ap.lin(0.88,1.0) ;
null device
          1
>           pdf( "./graph/2paper/Fig1-ci.pdf", height=8, width=12, pointsize=15 ) ; ap.lin(0.88,1.0,
null device
          1
> #win.metafile( "./graph/2paper/Fig1-ci.emf", height=8, width=12, pointsize=15 ) ; ap.lin(0.88,1.0,
>    postscript( "./graph/2paper/Fig1-ci.eps", height=8, width=12, pointsize=15 ) ; ap.lin(0.88,1.0,
null device
          1
```

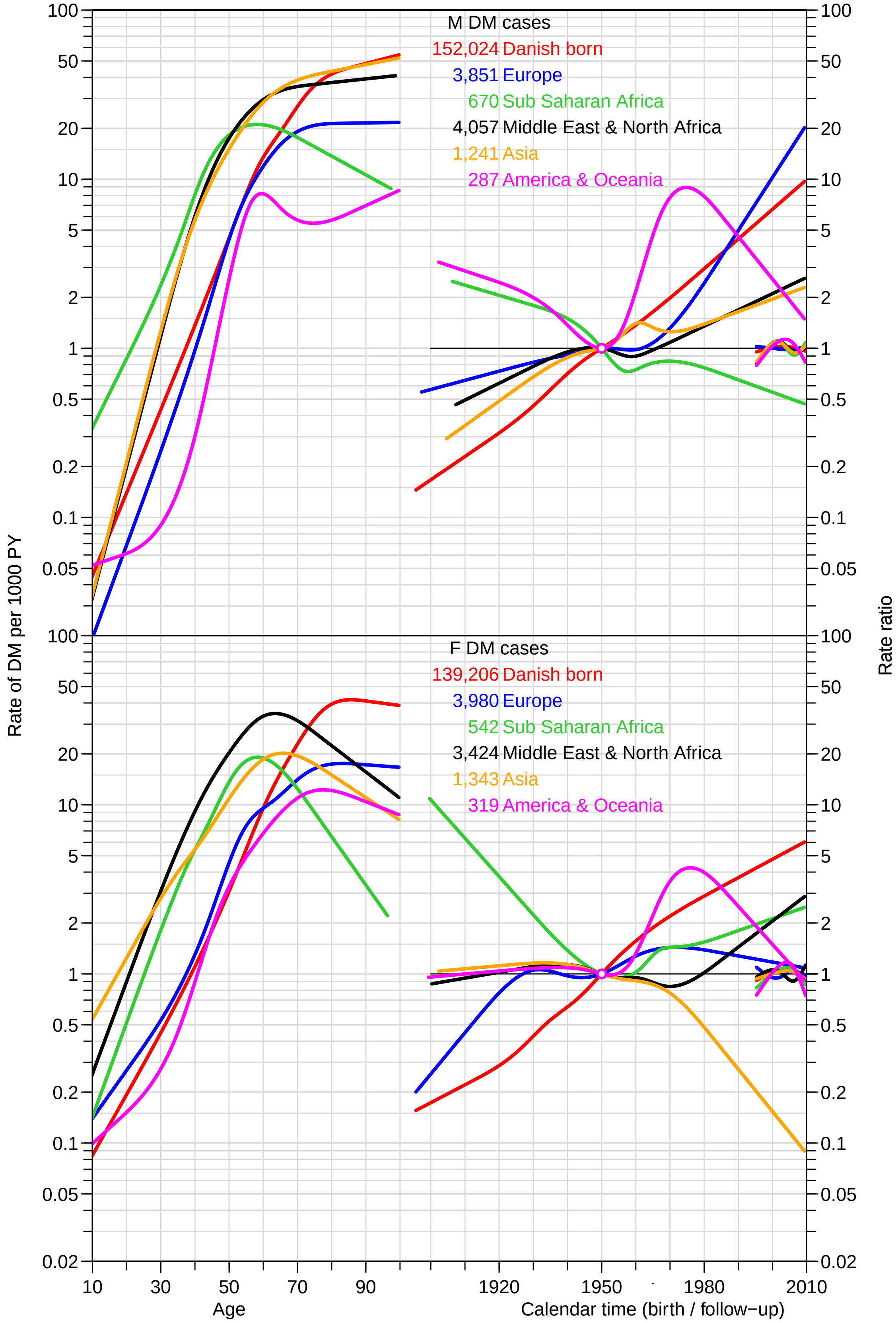

Figure 5.1: *Estimates from separate age-period-cohort models for each region and sex; models are models with period effects fixed to be 0 on average and 1950 as the reference birth cohort, and hence the age-effects interpretable as longitudinal (cohort) effects. The numbers are the total number of DM cases in each region.*

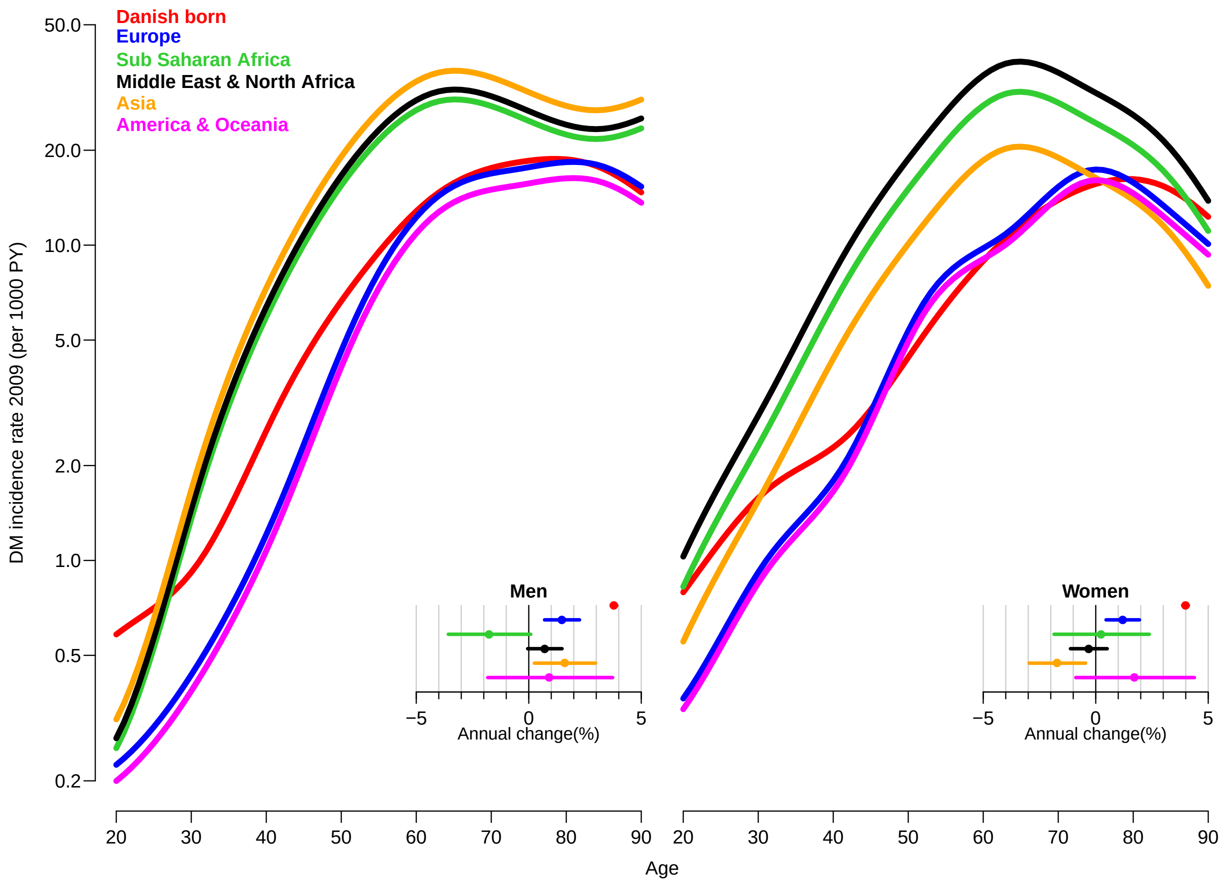

Figure 5.2: *Incidence rates of diabetes in different ethnic groups. From a model with a linear trend in calendar time (=constant relative annual change).*

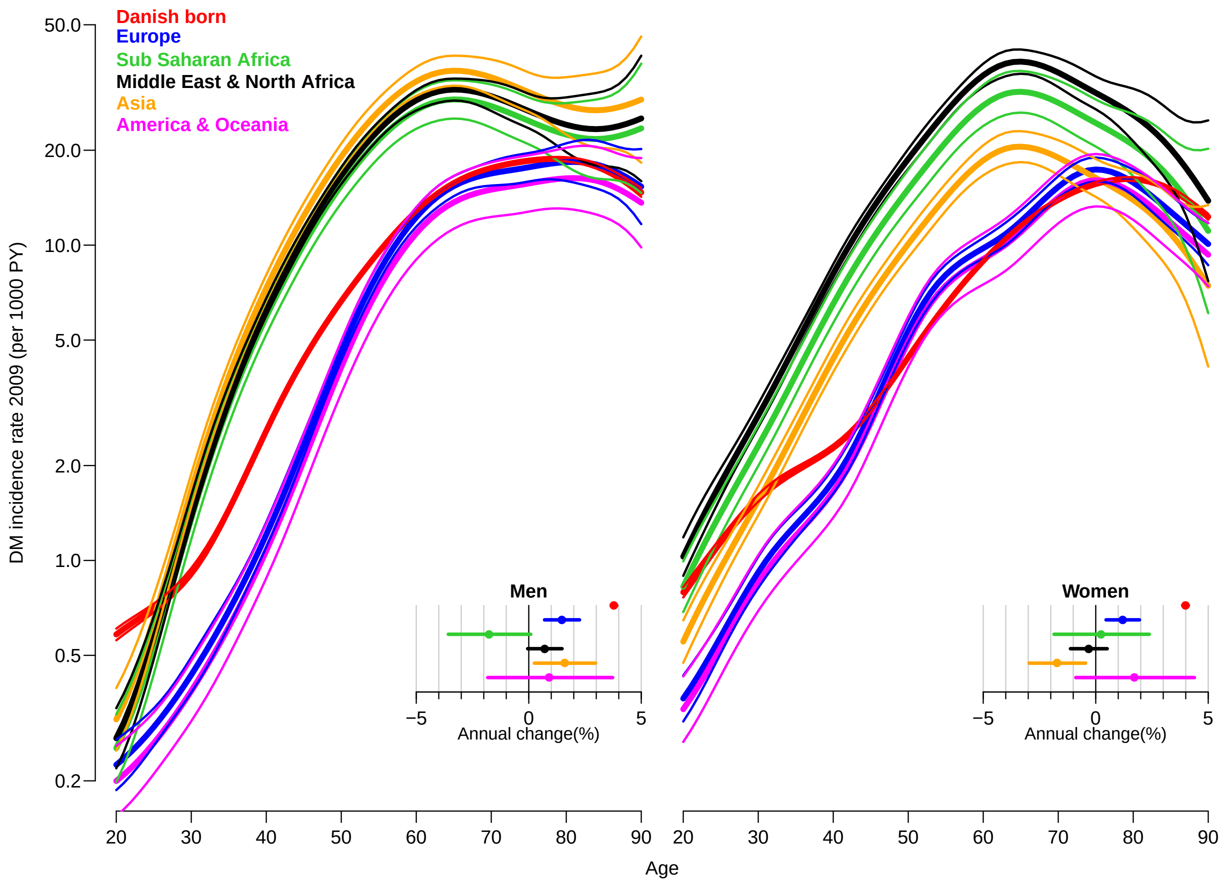

Figure 5.3: *Incidence rates of diabetes in different ethnic groups with 95% confidence limits. From a model with a linear trend in calendar time (=constant relative annual change).*

Chapter 6

Mortality and SMR by country of origin

In this chapter we analyze the mortality rates as reported in the registers, so we load the tabulated data, and the outcome variable of interest will be D.dd and Y.

```
> options( width=90 )
> library( Epi )
> library( splines )
> load( file="./data/Afu.Rda" )
> str( Afu )
'data.frame':        65348 obs. of  10 variables:
 $ A     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ P     : num  1995 1995 1995 1995 1995 ...
 $ U     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 2 2 ...
 $ state : Factor w/ 2 levels "Well","DM": 1 1 1 1 1 1 1 2 1 1 ...
 $ region: Factor w/ 7 levels "DK","Africa",..: 1 2 3 4 5 6 7 1 1 2 ...
 $ D.Wdk : num  179 NA NA NA NA NA NA NA 137 NA ...
 $ Y     : num  17926.24 1.32 26.49 10.54 51.68 ...
 $ D.dm  : num  1 0 0 0 0 0 0 0 0 0 ...
 $ D.dd  : num  179 0 0 0 0 0 0 0 137 0 ...
```

For the analysis of age, period and cohort effects on mortality, we need (as for the incidence analyses) to define the midpoint of follow-up in the Lexis triangles, and the event variable as D.dd

```
> mdat <- transform( subset( Afu, region != "Other" ),
+                     A = A + (1+U)/3,
+                     P = P + (2-U)/3,
+                     Y = Y/1000,
+                     D = D.dd,
+                region = Relevel( factor(region),
+                                  list("Danish born" = 1,
+                                             "Europe" = 5,
+                                "Sub Saharan Africa" = 2,
+                         "Middle East & North Africa" = 6,
+                                               "Asia" = 4,
+                                 "America & Oceania" = 3 ) ) )
> str( mdat )
'data.frame':        60518 obs. of  11 variables:
 $ A     : num  0.333 0.333 0.333 0.333 0.333 ...
 $ P     : num  1996 1996 1996 1996 1996 ...
 $ U     : num  0 0 0 0 0 0 0 0 0 0 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 2 2 2 ...
 $ state : Factor w/ 2 levels "Well","DM": 1 1 1 1 1 1 2 1 1 1 ...
 $ region: Factor w/ 6 levels "Danish born",..: 1 3 6 5 2 4 1 1 3 6 ...
```

```
 $ D.Wdk : num  179 NA NA NA NA NA NA 137 NA NA ...
 $ Y     : num  17.92624 0.00132 0.02649 0.01054 0.05168 ...
 $ D.dm  : num  1 0 0 0 0 0 0 0 0 0 ...
 $ D.dd  : num  179 0 0 0 0 0 0 137 0 0 ...
 $ D     : num  179 0 0 0 0 0 0 137 0 0 ...
```

6.1 SMR: Age-Period-Cohort models for mortality relative to the non-DM population

We fit a common mortality model for the entire material with a diabetes-specific RR depending on age. This is presumably not a tenable model, so therefore we fit 3 different extensions to the simplified APC-model (and, as previously, everything still separately for the two sexes).

Note that we shall make the implicit assumption here that the age-period-cohort shape of mortality among non-diabetics for all ethnic groups is the same — essentially as determined by the Danish population, and only the *level* of mortality differ between ethnicites. It is on top of these (proportional) ethnic-group specific mortality rates that we estimate the DM-related SMR in three different guises. In the following we use “SMR” as the mortality HR between DM pateintsns and non-DM persons (within each ethnicity).

1. A separate overall SMR for each ethnicity (`m0,f0`), on top of the age-specific mortality rates described above.

   This is just reported as a table of relative SMRs as well as the ratio of the SMRs to the SMR in the Danish population.

2. An additional separate secular trend (drift) in SMR for each ethnic group (`mi,fi`).

   This is reported as the SMR at 2009 and the average annual change in SMR for each ethnicity.

   In a cohort perspective, we might instead have reported this as the SMRs in the 1945 cohort (which would be different from the SMRs at 2009), and the average annual change in SMR by date of birth (cohort). Since age is not included in the mdel fro the SMR, the cohort trend will not necessarily be the same as the period trend, so we have not fitted the model with cohort trend in SMR.

   Thus, this model is not of particular interest.

3. What is of interest is to allow separate age-specific incidence rates for the three groups of ethnic groups in addition to the separate trends (`mia,fia`).

   This would naturally be reported in the same way as the the model above, it basically just corresponds to allowing different shapes of the age-specific SMRs between ethnicities, in both cases we must refer the age-specific incidences to a particular period or cohort.

   In this model the period and cohort trend *will* be the same (because the linear effect of age on SMR is in the model), and is reported as an annual “drift” in SMR.

   Thus the SMR-model is really a model with a separate age-drift effect for each ethnicity. Thus the reporting of the SMR must be as an age-specific SMR at a given date - corresponding to cross-sectional SMRs at a given date. Alternatively we might report the *longitudinal* age-specific SMRs referring to a specific birth cohort.

4. For the last model (SMR age-drift) we also report the SMR-ratios relative to the Danish population. These will also be age-drift models and as above they may either be reported as cross-sectional or longitudinal SMR-ratios.

We shall however stick to the cross-sectional reporting of the SMRs and SMR-ratios, because the follow-up period is rather small, and hence substantial extrapolations are needed to justify longitudinal SMR-age curves.

```
> a.kn <- seq(  10,   95,,4)
> r.kn <- seq(  10,   95,,4)
> p.kn <- seq(1995,2009,,4)
> c.kn <- seq(1890,1990,,4)
> m0 <- glm( D ~ Ns(   A, knots=a.kn, intercept=TRUE ) - 1 +
+                Ns( P   , knots=p.kn ) +
+                Ns( P-A, knots=c.kn ) +
+                I( P-2009 ) + Relevel(region,c(2:6,1)),
+            offset = log( Y ),
+            family = poisson,
+              data = subset(mdat,sex=="M") )
> mi  <- update( m0 , . ~ . + region:I((state=="DM")*1) )
> mil <- update( mi , . ~ . + region:I((state=="DM")*(P-2009)) )
> mia <- update( mil, . ~ . + region:I((state=="DM")*1):Ns( A, knots=r.kn ) )
> f0  <- update( m0 ,data = subset(mdat,sex=="F") )
> fi  <- update( mi ,data = subset(mdat,sex=="F") )
> fil <- update( mil,data = subset(mdat,sex=="F") )
> fia <- update( mia,data = subset(mdat,sex=="F") )
> anova( m0, mi, mil, mia, test="Chisq" )
Analysis of Deviance Table

Model 1: D ~ Ns(A, knots = a.kn, intercept = TRUE) - 1 + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
    1))
Model 2: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
    1)) + region:I((state == "DM") * 1) - 1
Model 3: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
    1)) + region:I((state == "DM") * 1) + region:I((state ==
    "DM") * (P - 2009)) - 1
Model 4: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
    1)) + region:I((state == "DM") * 1) + region:I((state ==
    "DM") * (P - 2009)) + region:I((state == "DM") * 1):Ns(A,
    knots = r.kn) - 1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     29865      62731
2     29859      41490  6  21241.5 < 2.2e-16
3     29853      40984  6    505.4 < 2.2e-16
4     29835      36642 18   4341.8 < 2.2e-16
> anova( f0, fi, fil, fia, test="Chisq" )
Analysis of Deviance Table

Model 1: D ~ Ns(A, knots = a.kn, intercept = TRUE) - 1 + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
    1))
Model 2: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
    1)) + region:I((state == "DM") * 1) - 1
Model 3: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
```

```
    1)) + region:I((state == "DM") * 1) + region:I((state ==
    "DM") * (P - 2009)) - 1
Model 4: D ~ Ns(A, knots = a.kn, intercept = TRUE) + Ns(P, knots = p.kn) +
    Ns(P - A, knots = c.kn) + I(P - 2009) + Relevel(region, c(2:6,
    1)) + region:I((state == "DM") * 1) + region:I((state ==
    "DM") * (P - 2009)) + region:I((state == "DM") * 1):Ns(A,
    knots = r.kn) - 1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     30625      55054
2     30619      39737  6  15317.5 < 2.2e-16
3     30613      39147  6    589.5 < 2.2e-16
4     30595      36066 18   3081.4 < 2.2e-16
```

Thus we see that the last model provides a substantially better fit then the three simpler ones. Particularly we see that the (first) model ignoring the effect of DM om mortality provides a really bad fit compared to the oher models, both for men and women.

First we set up an array to collect the said SMRs as well as SMR-ratios, the mean SMR and slope, for different regions and the two sexes.

```
> a.pt <- 20:90
> CA <- Ns( a.pt, knots=r.kn )
> pnam <- names(coef(mia))
> SMRr <- NArray( list( sex = c("M","F"),
+                    region = levels( mdat$region ),
+                      type = c("SMR","SMR vs DK"),
+                         A = a.pt,
+                      what = c("Est","lo","hi") ) )
> for( rg in levels(mdat$region) )
+    {
+ # All the interacion parameters
+ stn <- grep( "state", pnam )
+ # The Danish ones among these
+ dgn <- grep( "Danish born", pnam[stn] )
+ # The regional ones among these
+ rgn <- grep( rg, pnam[stn] )
+ # The set wanted
+ pnam[ c(stn[rgn],stn[dgn]) ]
+ # Extract the SMRs
+ SMRr["M",rg,"SMR",,] <- ci.exp( mia, subset=stn[rgn], ctr.mat=cbind(1,0,CA) )
+ SMRr["F",rg,"SMR",,] <- ci.exp( fia, subset=stn[rgn], ctr.mat=cbind(1,0,CA) )
+ # Extract the SMR ratios
+ SMRr["M",rg,"SMR vs DK",,] <- ci.exp( mia, subset=c(stn[rgn],stn[dgn]), ctr.mat=cbind(1,0,CA,-1,0,
+ SMRr["F",rg,"SMR vs DK",,] <- ci.exp( fia, subset=c(stn[rgn],stn[dgn]), ctr.mat=cbind(1,0,CA,-1,0,
+    }
```

On top of these estimated age-specific SMRs (cross-sectional from 2009) we also want to report the average annual change in SMR DM in each of the ethnic groups, as well as the ratio of change between each ethnic group and the Danish born.

```
> RR.ann <- NArray( list( model = c("const","age","vs DK"),
+                        region = levels( mdat$region ),
+                           sex = levels( mdat$sex ),
+                                 c("RR","up","lo") ) )
> str( RR.ann )
 logi [1:3, 1:6, 1:2, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ model : chr [1:3] "const" "age" "vs DK"
  ..$ region: chr [1:6] "Danish born" "Europe" "Sub Saharan Africa" "Middle East & North Africa" ...
  ..$ sex   : chr [1:2] "M" "F"
  ..$       : chr [1:3] "RR" "up" "lo"
> RR.ann["const",,"M",] <- ( ci.exp( mil, subint=c("state","P") ) - 1)*100
> RR.ann["const",,"F",] <- ( ci.exp( fil, subint=c("state","P") ) - 1)*100
> RR.ann["age",,"M",]   <- ( ci.exp( mia, subint=c("state","P") ) - 1)*100
> RR.ann["age",,"F",]   <- ( ci.exp( fia, subint=c("state","P") ) - 1)*100
```

```
> Cd <- diag(nlevels(mdat$region))
> Cd[,1] <- Cd[,1]-1
> RR.ann["vs DK",,"M",]   <- ( ci.exp( mia, subint=c("state","P"), ctr.mat=Cd ) - 1)*100
> RR.ann["vs DK",,"F",]   <- ( ci.exp( fia, subint=c("state","P"), ctr.mat=Cd ) - 1)*100
> round( ftable( RR.ann, row.vars=2:1 ), 1 )
                                  sex     M                 F
                                         RR   up   lo   RR   up   lo
region                     model
Danish born                const       -1.4 -1.6 -1.3 -1.9 -2.1 -1.7
                           age         -1.3 -1.5 -1.1 -1.8 -2.0 -1.6
                           vs DK        0.0  0.0  0.0  0.0  0.0  0.0
Europe                     const       10.7  8.7 12.7  9.8  8.1 11.4
                           age         10.6  8.6 12.6  9.7  8.1 11.4
                           vs DK       12.1 10.1 14.1 11.8 10.1 13.5
Sub Saharan Africa         const        3.7 -3.7 11.8 10.6 -2.0 24.8
                           age          5.3 -2.4 13.5 11.8 -1.0 26.2
                           vs DK        6.7 -1.0 15.1 13.9  0.8 28.6
Middle East & North Africa const       12.2  9.1 15.5  8.7  5.4 12.1
                           age         12.8  9.6 16.0  9.3  5.9 12.7
                           vs DK       14.3 11.1 17.6 11.3  7.9 14.8
Asia                       const       20.2 13.5 27.3 10.5  4.9 16.4
                           age         21.0 14.2 28.1 10.9  5.2 16.8
                           vs DK       22.6 15.8 29.9 12.9  7.2 19.0
America & Oceania          const       10.4  3.5 17.8 12.6  6.1 19.5
                           age         10.0  3.1 17.4 12.7  6.1 19.7
                           vs DK       11.5  4.5 19.0 14.8  8.0 21.9
```

This is the annual change (in %) of the SMR of death within each ethnic group (for the entries "age" and "). From the table it is seen that there is not much difference in the estimated annual change in SMR between the model with constant SMR and the model with age-varying SMR. It is seen that the SMR is decreasing for Danish born, but increasing for all other ethnic groups.

With this table in place we can plot the age-specific SMRs as they look in 2009 and annotate the plot with the percentwise annual changes:

```
> cbind( levels(mdat$region),
+        rcol <- c("red","blue","limegreen","black","orange","magenta") )
     [,1]                         [,2]
[1,] "Danish born"                "red"
[2,] "Europe"                     "blue"
[3,] "Sub Saharan Africa"         "limegreen"
[4,] "Middle East & North Africa" "black"
[5,] "Asia"                       "orange"
[6,] "America & Oceania"          "magenta"
```

```
> ap.smr <-
+ function( mfac=0.9, wh=0.8, ci=FALSE, a0=60, ax=seq(-5,25,5), tp="SMR" )
+   {
+ 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 )
+ # plot SMR for men
+ matplot( a.pt, t(SMRr["M",,tp,,1]), bty="n", ylim=yl<-c(0.2,50)/((tp!="SMR")+1),
+          log="y", type="l", lty=1, col=rcol, lwd=5 )
+ # confidence intervals ?
+ if( ci ) for( j in 2:3 )
+ matlines( a.pt, t(SMRr["M",,tp,,j]), type="l", lty=1, col=rcol, lwd=2 )
+ # reference line
+ abline( h=1 )
+ if(tp!="SMR") abline( h=1, lwd=5, col=rcol[1] )
+ # order of ethnic groups
+ c.ord <- 1:6 # c(1,5,3,6,2,4)
+ text( rep(90,6), (10^par("usr")[4])*((mfac^1.2)^c(1:6)),
+       levels( mdat$region )[c.ord], col=rcol[c.ord], font=2, adj=1 )
+ axis( side=1, at=a0+ax, labels=ax, pos=wh*(mfac^7), cex=0.8 )
+ text( a0, wh, "Men, ", adj=1 )
```

```
+ text( a0, wh, " % annual chg. in SMR", adj=0 )
+ # gridlines
+ segments( a0+ax, wh*(mfac^7),
+           a0+ax, wh*(mfac^1), col=gray(0.8) )
+ segments( a0, wh*(mfac^7), a0, wh*(mfac^1) )
+ # annual changes
+ linesEst( RR.ann["age",c.ord,"M",]+a0, y=wh*(mfac^(1:6)), col=rcol[c.ord], lwd=3 )
+ # text( a0+mean(range(ax)), wh*(mfac^11), "Annual SMR change(%)", cex=0.9 )
+
+ # plot for women
+ matplot( a.pt, t(SMRr["F",,tp,,1]), yaxt="n", bty="n", ylim=yl,
+          log="y", type="l", lty=1, col=rcol, lwd=5 )
+ if( ci ) for( j in 2:3 )
+ matlines( a.pt, t(SMRr["F",,tp,,j]), type="l", lty=1, col=rcol, lwd=2 )
+ abline( h=1 )
+ if(tp!="SMR") abline( h=1, lwd=5, col=rcol[1] )
+ axis( side=1, at=a0+ax, labels=ax, pos=wh*(mfac^7), cex=0.8 )
+ text( a0, wh, "Women, ", adj=1 )
+ text( a0, wh, " % annual chg. in SMR", adj=0 )
+ # gridlines and reference line
+ segments( a0+ax, wh*(mfac^7),
+           a0+ax, wh*(mfac^1), col=gray(0.8) )
+ segments( a0, wh*(mfac^7), a0, wh*(mfac^1) )
+ # annual changes
+ linesEst( RR.ann["age",c.ord,"F",]+a0, y=wh*(mfac^(1:6)), col=rcol[c.ord], lwd=3 )
+ # text( a0+mean(range(ax)), wh*(mfac^11), "Annual SMR change(%)", cex=0.9 )
+
+ mtext( if(tp=="SMR") "SMR in 2009" else "SMR ratio vs Danish born (2009)" ,
+        side=2, line=3, outer=T, las=0 )
+ mtext( "Age", side=1, line=2, outer=T, las=0 )
+    }
> ap.smr(wh=0.6)
```

```
> ap.smr(wh=0.6,ci=TRUE)
```

```
> ap.smr(wh=0.3,tp="SMR vs DK")
```

```
> ap.smr(wh=0.3,ci=TRUE,tp="SMR vs DK")
```

```
>        pdf( "./graph/2paper/Fig3.pdf"    , height=8, width=12, pointsize=15 )
>        ap.smr(0.85,0.8) ; dev.off()
null device
          1
> postscript( "./graph/2paper/Fig3.eps"    , height=8, width=12, pointsize=15 )
>        ap.smr(0.85,0.8) ; dev.off()
null device
          1
>        pdf( "./graph/2paper/Fig3-ci.pdf" , height=8, width=12, pointsize=15 )
>        ap.smr(0.85,0.8,ci=TRUE) ; dev.off()
null device
          1
> postscript( "./graph/2paper/Fig3-ci.eps" , height=8, width=12, pointsize=15 )
>        ap.smr(0.85,0.8,ci=TRUE) ; dev.off()
null device
          1
>        pdf( "./graph/2paper/Fig3r.pdf"   , height=8, width=12, pointsize=15 )
>        ap.smr(0.85,0.5,tp="SMR vs DK") ; dev.off()
null device
          1
```

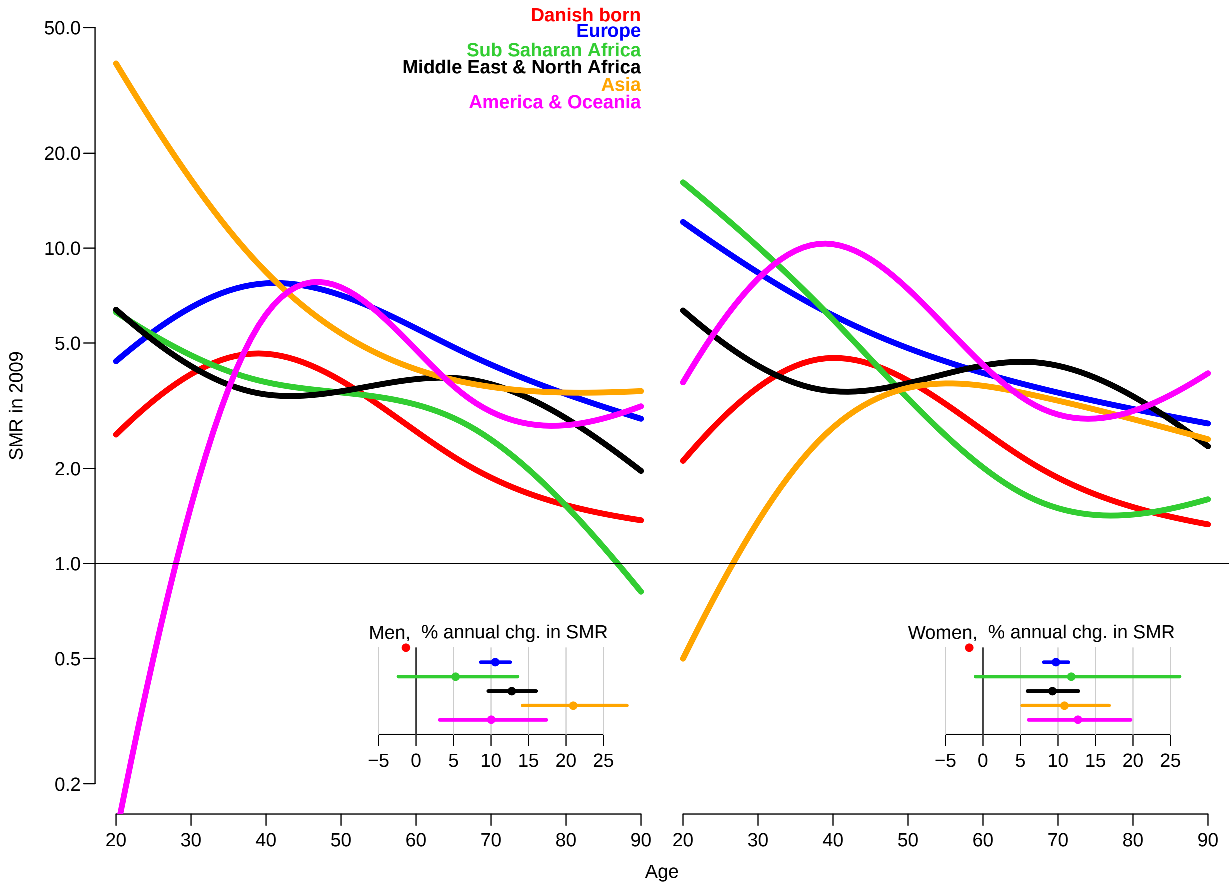

Figure 6.1: *Mortality rate ratio between DM and no DM (SMR) in different ethnic groups. From a model with a linear trend in calendar time and a 3-parameter natural spline in age for each ethnic group.*

```
> postscript( "./graph/2paper/Fig3r.eps"    , height=8, width=12, pointsize=15 )
>         ap.smr(0.85,0.5,tp="SMR vs DK") ; dev.off()
null device
          1
>         pdf( "./graph/2paper/Fig3r-ci.pdf", height=8, width=12, pointsize=15 )
>         ap.smr(0.85,0.5,tp="SMR vs DK",ci=TRUE) ; dev.off()
null device
          1
> postscript( "./graph/2paper/Fig3r-ci.eps", height=8, width=12, pointsize=15 )
>         ap.smr(0.85,0.5,tp="SMR vs DK",ci=TRUE) ; dev.off()
null device
          1
```

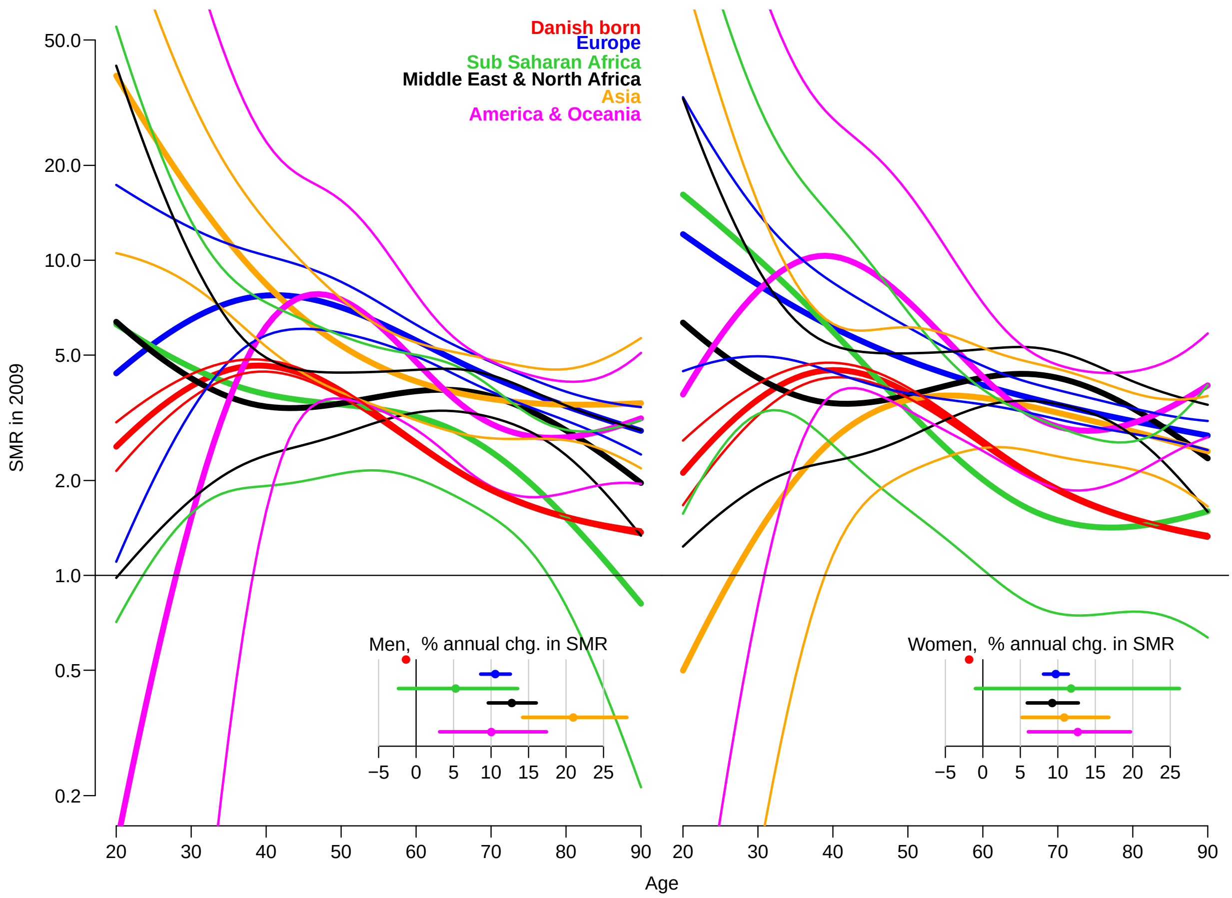

Figure 6.2: *Mortality rate ratio between DM and no DM (SMR) in different ethnic groups. with 95% confidence limits. From a model with a linear trend in calendar time and a 3-parameter natural spline in age separately for each ethnic group.*

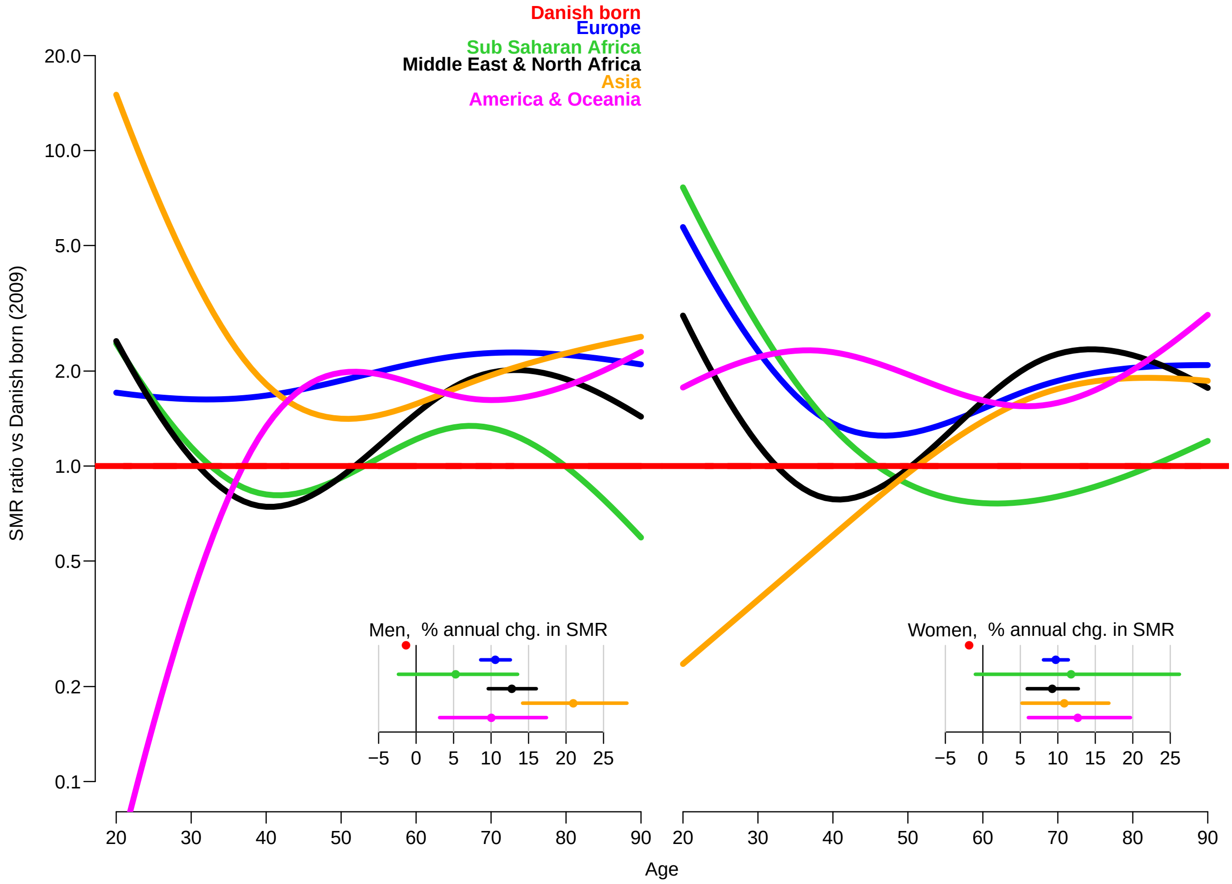

Figure 6.3: *Ratio of SMRs in different ethnic groups realtive to Danish born. From a model with a linear trend in calendar time and a 3-parameter natural spline in age for each ethnic group.*

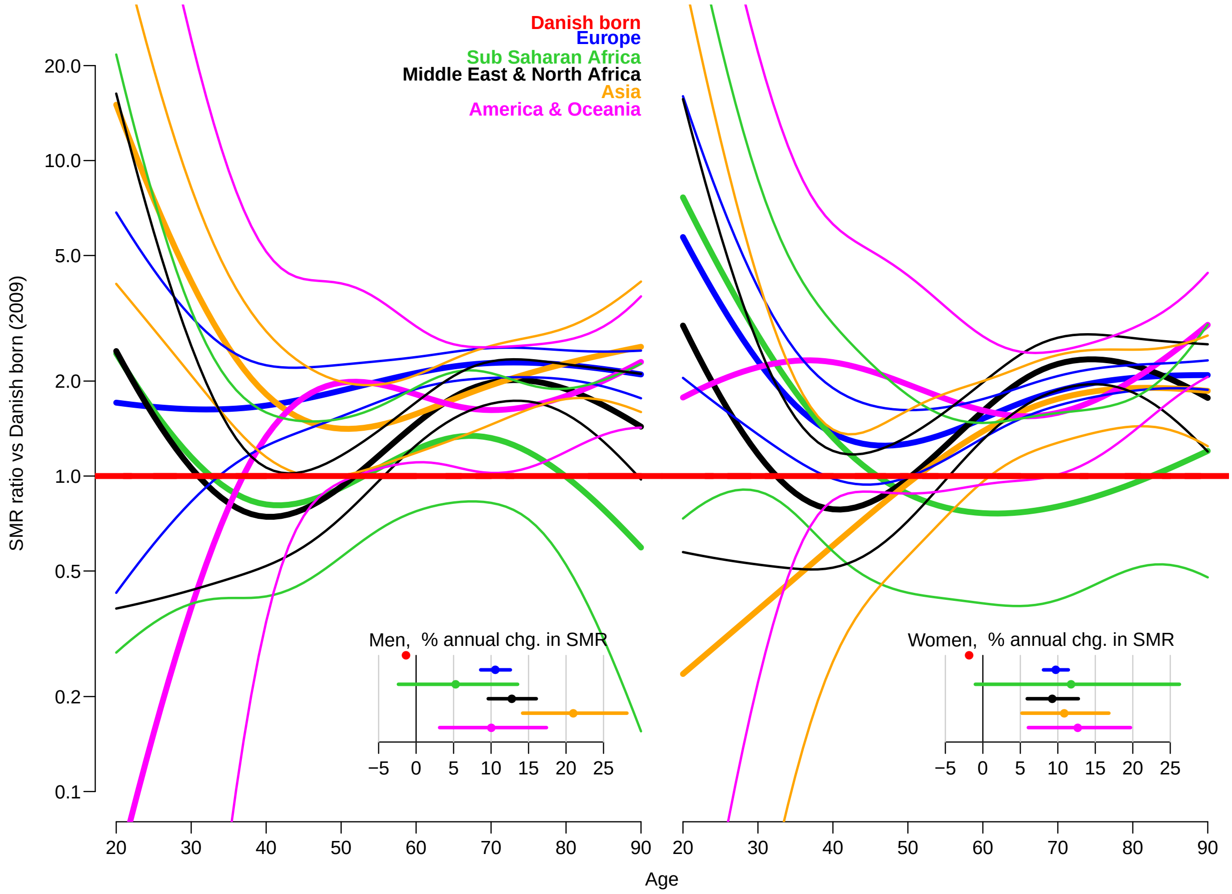

Figure 6.4: *Ratio of SMRs in different ethnic groups realtive to Danish born with 95% confidence limits. From a model with a linear trend in calendar time and a 3-parameter natural spline in age separately for each ethnic group.*