Cancer in T1D patients: Protocol & Data analysis

SDC
January 2016
http://BendixCarstensen.com/DMCa/T1D/T1D-Ca.pdf
Version 6.1

Compiled Thursday 21st January, 2016, 19:31
from: /home/bendix/sdc/DMreg/T1DCa/r/T1D-Ca.tex

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

Data collaboration:
Søren Friis KB, Copenhagen, Denmark
Reijo Sund KTL, Helsinki, Finland
Rickard Ljung KI, Stockholm, Sweden
Jan J Kerssens NHS National Services, Edinburgh, Scotland
Jessica Harding BakerIDI, Melbourne, Australia

Contents

Chapter 1

Analysis protocol

This chapter is an outline of the analysis procedure, and hence the data layout for the joint analysis of cancer incidence among T1D patients.

The main aim of the analysis is to estimate a single cancer incidence RR between T1D patients and non-T1D persons.

Analyses will be conducted separately for each cancer site and sex, and also for all cancers combined.

Sensitivity analyses wil be carried out for all cancers combined by restricting analyses to the subgroups defined by age at diagnosis of DM and by including time since diagnosis in the analysis. We consider data too thin for doing these sensitivity analyses for any of the specific sites of cancer.

1.1 Analysis dataset

Since the T1D patients constitute only a small part of the population, we will use the *entire population* as comparison group, instead of the non-T1D persons or non-DM in general.

Thus we will need a combination of two datasets, one with the number of cases and follow-up time for the T1D patients, and one with the number of cases and follow-up for the entire population.

The combined (stacked dataset) will then be classified by the following variables:

- sex
- age at follow-up — preferably in 1-year classes, scored as the *mean* of the age-class.
- date of follow-up — preferably in 1-year classes, scored as the *mean* of the period.
- T1D status — Population, 0–30, 30–35, 35–40 (T1D diagnosis age interval)
- T1D duration — Population, 0, 1, 2, 5, 10, 15 (left endpoints of intervals)

In addition to these five classification variables, the data set should contain the following outcome variables:

- `Y` — person-years
- `D0` — no. of cases of any cancer (counting only the first for each person).

- `D1` — no. of cases of cancer 1.
- `D2` — no. of cases of cancer 2.
- ...

When counting the cancers at specific sites we do not exclude cancers occurring after cancers at a different site.

Note that in the combined dataset the rows corresponding to the T1D patients will have quite few cases and person-years, whereas the rows corresponding to the population will have large number of cases and person-years.

Moreover there is no need that the age and calendar time classification be the same in the T1D part and the population part of the dataset, but of course, the age and calendar time range of the population data should include that of the T1D data.

Once the dataset has been constructed, rows (observations) with 0 person-years can of course be deleted.

1.2 Analysis models

For each site we fit a model for the temporal dependence of cancer incidence rates as a function of age and date of follow-up and date of birth (age-period-cohort model). These dependencies are largely going to be determined by the rates in the general population. On top of this we add an effect of being T1D, so a single RR parameter describing the cancer occurrence among T1D patients as depending on age, period and cohort in the same way as in the general population, but just at a different level.

For all cancers the model will be amended by estimating separate RRs for T1D patients diagnosed in different ages (< 30, 30–35, 35–40), and a test for the equality of these will be made[1]. Similarly we will include duration of diabetes in the model and test wheter this has an effect.

The dependencies on age, period and cohort are going to be modelled by smooth functions of the variables (splines). This has the implication that for each observational unit in the dataset, the age and date of follow-up must be coded as the *mean* of age, resp. date of follow up for the unit.

This way of modelling the RR between T1D and the general population is very similar to old-fashioned calculation of the SMR, but with two distinct advantages:

1. The method here takes the uncertainty of the population rates into account, which in this study might be of some importance because the observation time among T1D patients is in quite young ages.

2. It is not necessary to have the same classification of age and date of observation among T1D cases and in the general population. It is preferable to have both as detailed as possible, though.

[1] This is really a shortcut, since we should have taken age at diagnosis of T1D into account as a continuous variable; in practice in 1-year intervals.

1.3 Pooling data from different countries

These analyses will be done for each country separately. This corresponds to a joint analysis of all data with an interaction between age, period, cohort and country, as well as an interaction between T1D status and country.

The relevant pooling of data would simply be reducing this model to one where the RR associated with T1D does not depend on country, but is the same across countries. This will also make it possible to test whether the T1D effect is the same between counties by a simple likelihood ratio test.

The outlined structure of data will facilitate this analysis since there will be no requirements about the age and period grouping being the same between counties; the central point is the coding of each age/period group by the *mean* age/period in the group, and the use of these as continuous variables in the analysis.

1.4 Reporting

1.4.1 Country-wise

The reporting will be a separate RR for men and women for each cancer site analyzed, the number of cancer cases among the T1D patients and for background information, the age-distribution of the follow-up among the T1D patients.

The RRs should preferably be reported as forest plots to allow for comparison across sites, sexes and countries. These results should be seen as secondary and appear in supplementary material.

1.4.2 Pooled analysis

The pooled analysis will for each cancer site consist of the no. of cancer cases among T1D persons, the pooled RR as well as the test for equality of RRs across countries.

The effects will be reported in a forest plot for easy comparison.

Chapter 2

Tumour site coding

The tumour classification numbering is different between the coding used by the Danish CRG and NORDCAN. The Danish coding is here:

```
> tcl <- read.csv2( "../data/DKtumorcls.csv",
+                   header=TRUE, skip=2, as.is=TRUE)[1:118,1:3]
> tcl <- tcl[tcl[,1] != "",]
> rownames(tcl) <- 1:nrow(tcl)
> tcl
                               Tumorgruppering DIAGGRPa   DIAGb
1                      Buccal cavity and pharynx       10   11-15
2                                            Lip       NA      11
3                                         Tongue       NA      12
4                                          Mouth       NA      13
5                               Salivary glands       NA      14
6                                        Pharynx       NA      15
7                               Digestive organs       20   21-29
8                                     Oesophagus       NA      21
9                                        Stomach       NA      22
10                              Small intestine       NA      23
11                   Colon incl. rectosigmoideum       NA      24
12                               Rectum and anus       NA      25
13                                         Liver       NA      26
14                  Gallbladder and biliary tract       NA      27
15                                      Pancreas       NA      28
16           Other and unspecified digestive organs       NA      29
17 Respiratory system and intrathoracic organs       30   31-36
18         Nasal cavities, middle ear and sinuses       NA      31
19                                        Larynx       NA      32
20                      Lung, bronchus and trachea       NA      33
21                                        Thymus       NA      34
22                          Heart and mediastinum       NA      35
23                                        Pleura       NA      36
24         Bones, joints and articular cartilage       40      40
25                                          Skin       50   51-52
26                              Melanoma of skin       NA      51
27                                    Other skin       NA      52
28            Mesothelium and connective tissue       60   61-64
29                      Mesothelium, non-pleural       NA      61
30              Peripheral nerves and autonomic        NA      62
31                                nervous system       NA
32              Peritoneum and retroperitoneum       NA      63
33                       Other connective tissue       NA      64
34                                      Breast e       70      70
35              Female genital organs incl. skin       80   81-85
36    External female genital organs and vagina       NA      81
37                                  Cervix uteri       NA      82
38                                  Corpus uteri       NA      83
39    Ovary, fallopian tube and broad ligament       NA      84
40 Other and unspecified female genital organs       NA      85
```

```
41              Male genital organs incl. skin       90   91-93
42                                    Prostate       NA      91
43                                      Testis       NA      92
44   Other and unspecified male genital organs       NA      93
45                           Urinary tract f,g      100 101-104
46                                      Kidney       NA     101
47                   Renal pelvis and ureter f,g     NA     102
48                         Urinary bladder f,g       NA     103
49    Other and unspecified urinary organs f,g       NA     104
50              Eye, brain and other parts of       110 111-114
51                      central nervous system       NA
52                                         Eye       NA     111
53                                    Meninges       NA     112
54                                       Brain       NA     113
55  Spinal cord, cranial nerves and other and        NA     114
56 unspecified parts of central nervous system       NA
57                             Endocrine glands     120 121-123
58                                Thyroid gland      NA     121
59                                Adrenal gland      NA     122
60                      Other endocrine glands       NA     123
61         Lymphatic and haematopoietic tissue      130 131-138
62                             Hodgkin lymphoma      NA     131
63         Non-Hodgkin lymphoma and malignant        NA     132
64                 immunoproliferative disease       NA
65                             Multiple myeloma      NA     133
66                          Lymphatic leukaemia      NA     134
67                            Myeloid leukaemia      NA     135
68                          Monocytic leukaemia      NA     136
69             Other and unspecified leukaemia       NA     137
70  Other and unspecified cancer in lymphatic        NA     138
71                 and haematopoietic tissue h       NA
72           Ill-defined and unspecified cancer     140     140
73                                                   NA
74                         Not counted as cancer i  150     150
```

Likewise we load the naming of tumours which is used in NORDCAN (and hence population rate data:

```
> load( "../data/n.ana.Rda" )
> n.ana
   ln code                          name
1   1   d0                     All sites
2   2   d1                           Lip
3   3   d2                        Tongue
4   4   d3               Salivary glands
5   5   d4                         Mouth
6   6   d5                       Pharynx
7   7   d6                    Oesophagus
8   8   d7                       Stomach
9   9   d8               Small intestine
10 10   d9                         Colon
11 11  d10               Rectum and anus
12 12  d11                         Liver
13 13  d12                  Gallbladder
14 14  d13                      Pancreas
15 15  d14                Nose, sinuses
16 16  d15                        Larynx
17 17  d16                          Lung
18 18  d17                        Pleura
19 19  d18                        Breast
20 20  d19                 Cervix uteri
21 21  d20                  Corpus uteri
22 22  d21                Uterus, other
23 23  d22                    Ovary etc.
24 24  d23       Other female genital organs
25 25  d24                      Prostate
```

```
26 26  d25                                   Testis
27 27  d26                               Penis etc.
28 28  d27                                   Kidney
29 29  d28                             Bladder etc.
30 30  d29                         Melanoma of skin
32 31  d31                                      Eye
33 32  d32           Brain, central nervous system
34 33  d33                                  Thyroid
35 34  d34                                     Bone
36 35  d35                             Soft tissues
37 36  d36                     Non-Hodgkin lymphoma
38 37  d37                         Hodgkin lymphoma
39 38  d38                         Multiple myeloma
40 39  d40                                Leukaemia
48 40  d48            Other and unspecified cancers
50 41  d50 All sites but non-melanoma skin cancer
51 42  d51            Lip, oral cavity and pharynx
52 43  d52                               Colorectal
```

A bit of hand-coding gives the conversion, so that we can rename the columns in the incidence data to match the ones one in the population data (and eventually also the Finnish data) — or vice versa.

```
> o.tcl <- c(  NA, 2, 3, 5, 4,
+               6, 8, 9,10,11,
+              12,13,14,15,18,
+              19,20,23,34,37,
+              38,NA,39,40,42,
+              43,44,46,48,26,
+              52,54,58,24,28,
+              63,62,65,NA,72,
+              rep(NA,2),66:69 )
> zz <- cbind( tcl[o.tcl,], dnam <- n.ana[c(1:40,42,43,rep(NA,4)),])
> ww <- zz
> for( i in c(1,6) ) ww[,i] <- substr(ww[,i],1,15)
> cbind( ww[,c(6,1:5)], 1:nrow(ww) )
                name Tumorgruppering DIAGGRPa DIAGb ln code 1:nrow(ww)
NA         All sites            <NA>       NA  <NA>  1   d0          1
2                Lip             Lip       NA    11  2   d1          2
3             Tongue          Tongue       NA    12  3   d2          3
5    Salivary glands Salivary glands       NA    14  4   d3          4
4              Mouth           Mouth       NA    13  5   d4          5
6            Pharynx         Pharynx       NA    15  6   d5          6
8         Oesophagus      Oesophagus       NA    21  7   d6          7
9            Stomach         Stomach       NA    22  8   d7          8
10   Small intestine Small intestine       NA    23  9   d8          9
11             Colon Colon incl. rec       NA    24 10   d9         10
12   Rectum and anus Rectum and anus       NA    25 11  d10         11
13             Liver           Liver       NA    26 12  d11         12
14       Gallbladder Gallbladder and       NA    27 13  d12         13
15          Pancreas        Pancreas       NA    28 14  d13         14
18     Nose, sinuses Nasal cavities,       NA    31 15  d14         15
19            Larynx          Larynx       NA    32 16  d15         16
20              Lung Lung, bronchus        NA    33 17  d16         17
23            Pleura          Pleura       NA    36 18  d17         18
34            Breast        Breast e       70    70 19  d18         19
37      Cervix uteri    Cervix uteri       NA    82 20  d19         20
38      Corpus uteri    Corpus uteri       NA    83 21  d20         21
NA.1   Uterus, other            <NA>       NA  <NA> 22  d21         22
39        Ovary etc. Ovary, fallopia       NA    84 23  d22         23
40   Other female ge Other and unspe       NA    85 24  d23         24
42          Prostate        Prostate       NA    91 25  d24         25
43            Testis          Testis       NA    92 26  d25         26
44        Penis etc. Other and unspe       NA    93 27  d26         27
46            Kidney          Kidney       NA   101 28  d27         28
48      Bladder etc. Urinary bladder       NA   103 29  d28         29
```

```
26   Melanoma of ski Melanoma of ski       NA    51 30  d29           30
52              Eye             Eye       NA   111 31  d31           31
54   Brain, central           Brain       NA   113 32  d32           32
58          Thyroid   Thyroid gland       NA   121 33  d33           33
24             Bone Bones, joints a       40    40 34  d34           34
28      Soft tissues Mesothelium and       60 61-64 35  d35           35
63   Non-Hodgkin lym Non-Hodgkin lym       NA   132 36  d36           36
62   Hodgkin lymphom Hodgkin lymphom       NA   131 37  d37           37
65   Multiple myelom Multiple myelom       NA   133 38  d38           38
NA.2       Leukaemia            <NA>       NA  <NA> 39  d40           39
72   Other and unspe Ill-defined and      140   140 40  d48           40
NA.3 Lip, oral cavit            <NA>       NA  <NA> 42  d51           41
NA.4      Colorectal            <NA>       NA  <NA> 43  d52           42
66             <NA> Lymphatic leuka       NA   134 NA <NA>           43
67             <NA> Myeloid leukaem       NA   135 NA <NA>           44
68             <NA> Monocytic leuka       NA   136 NA <NA>           45
69             <NA> Other and unspe       NA   137 NA <NA>           46
```

```
> DKnam <- paste("d",zz$DIAGb,sep="")
> DKnam <- gsub( "61-64", "63", DKnam )
> DKnam[c(1,39,41,42)] <- c("d0","d139","d151","d251")
> wh <- c(11,23,29,32,41)
> nn <- c("Rectum","Ovary","Bladder","Brain, CNS","Oral etc.")
> cbind( dnam$name[wh],   nn )
                                     nn
[1,] "Rectum and anus"               "Rectum"
[2,] "Ovary etc."                    "Ovary"
[3,] "Bladder etc."                  "Bladder"
[4,] "Brain, central nervous system" "Brain, CNS"
[5,] "Lip, oral cavity and pharynx"  "Oral etc."
>        dnam$name[wh] <- nn
> ( conv <- ( cnv <- data.frame( DKnam,
+                                NCnam=dnam$code,
+                                Clab=dnam$name,
+                                stringsAsFactors=FALSE ) )[wh.c <- c(1:21,23:42),] )
   DKnam NCnam                        Clab
1     d0    d0                   All sites
2    d11    d1                         Lip
3    d12    d2                      Tongue
4    d14    d3             Salivary glands
5    d13    d4                       Mouth
6    d15    d5                     Pharynx
7    d21    d6                  Oesophagus
8    d22    d7                     Stomach
9    d23    d8             Small intestine
10   d24    d9                       Colon
11   d25   d10                      Rectum
12   d26   d11                       Liver
13   d27   d12                 Gallbladder
14   d28   d13                    Pancreas
15   d31   d14              Nose, sinuses
16   d32   d15                      Larynx
17   d33   d16                        Lung
18   d36   d17                      Pleura
19   d70   d18                      Breast
20   d82   d19                Cervix uteri
21   d83   d20                Corpus uteri
23   d84   d22                       Ovary
24   d85   d23   Other female genital organs
25   d91   d24                    Prostate
26   d92   d25                      Testis
27   d93   d26                  Penis etc.
28  d101   d27                      Kidney
29  d103   d28                     Bladder
30   d51   d29            Melanoma of skin
```

```
31  d111   d31                              Eye
32  d113   d32                      Brain, CNS
33  d121   d33                          Thyroid
34   d40   d34                             Bone
35   d63   d35                     Soft tissues
36  d132   d36             Non-Hodgkin lymphoma
37  d131   d37                 Hodgkin lymphoma
38  d133   d38                 Multiple myeloma
39  d139   d40                        Leukaemia
40  d140   d48 Other and unspecified cancers
41  d151   d51                        Oral etc.
42  d251   d52                       Colorectal
> cnv[-wh.c,]
   DKnam NCnam          Clab
22   dNA   d21 Uterus, other
43  d134  <NA>          <NA>
44  d135  <NA>          <NA>
45  d136  <NA>          <NA>
46  d137  <NA>          <NA>
> rownames( conv ) <- 1:nrow(conv)
```

We then save the conversion data frame:

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

Chapter 3

Danish T1D data

The following is a detailed technical account of the approach to construction and analysis of data from Denmark. The precise data lay-out is very closely tied to the intended construction of the statistical analysis, hence this is briefly detailed here.

The analysis will be done in parallel for the different types of cancers; the outcome will be (d, y), events and person-years classified by sex, age and calendar time of observation as well as by diabetes status (yes/no) and diabetes duration. The model will be an age-period-cohort model for the rates as function of age and time, with a constant diabetes effect estimated, and in some analyses an effect of time since diagnosis (diabetes duration). This corresponds to a standardized mortality ratio analysis, except that this analysis also takes the uncertainty in the population rates into account.

Since this is analysis of rates in T1D patients a fair bit of the follow-up is in younger ages where the population cancer rates, particularly for the rarer cancers are not terribly high, and so is better modelled by smoothing than averaging in intervals of age and date of diagnosis. This is the background for the chosen approach.

Analyses will be done separately for men and women, and in some circumstances a pooled analysis will be done too. Some analyses will also take age at diagnosis into account, to explore how this influences the rate-ratio.

3.1 The Register data

We use SAS to merge the cancer register and the diabetes register to provide a datafile with one record per cancer diagnosed or diabetes diagnosed:

```
1                                    "Program: DMCR.sas"        13:01 Friday, September 5, 2014

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           1.91 seconds
      cpu time            0.46 seconds


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

---------------------------------------------------------------
C:\Bendix\Steno\DM-register\NDR\projects\Cancer\T1D\sas\DMCR.sas
---------------------------------------------------------------
NOTE: Libref HER was successfully assigned as follows:
      Engine:        V9
```

```
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\T1D\sas
NOTE: Libref DATA was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\T1D\data

NOTE: AUTOEXEC processing completed.

1          ***************************************************************
2          NOTE: This version of the program takes only T1DM patients defined
3                in different ways, and also splits follow-up by DM-duration
4                Datasets are produced for the entire available follow-up
5                too, NOT classified by the duration variables the names of
6                these are preceded with "x" (for eXtended).
7          *************************************************************** ;
8
9          * The date from which we trust the inclusion date to be the first ;
10         %let  validdate = '01JAN1995'd ;
11         * Set the entry and exit dates for the entire follow-up endeavour ;
12         %let  truncdate = '01JAN1995'd ;
13         %let   censdate = '31DEC2012'd ;
14         * Just to check it all wemt well ;
15         %put validdate = &validdate.
16              truncdate = &truncdate.
17               censdate = &censdate. ;
validdate = '01JAN1995'd      truncdate = '01JAN1995'd       censdate = '31DEC2012'd
18         * Set the selector of subgroups to analyse ;
19         %let dgrp = 21,22,241,242,243,249,251,26,28,
20                      33,
21                      51,
22                      70,
23                      82,83,84,
24                      91,92,
25                     101,103,
26                     113,
27                     121,
28                     131,132,133,139 ;
29         %let diagselect = diag in (&dgrp.) ;
30         * Variable names for tabulation purposes, note DX and D259 here ;
31         %let dvars = D0 D999
32                      D21 D22 D241 D242 D243 D249 D251 D259 D26 D28
33                      D33
34                      D51
35                      D70
36                      D82 D83 D84
37                      D91 D92
38                     D101 D103
39                     D113
40                     D121
41                     D131 D132 D133 D139 ;
42
43         * Get the relevant formats ;
44         options nosource2 ;
45         %inc "..\..\sas\CRG-fmts.sas" ;
NOTE: Format SEX has been output.
NOTE: Format DIAG has been output.

NOTE: PROCEDURE FORMAT used (Total process time):
      real time           0.12 seconds
      cpu time            0.03 seconds


130
131        * Where we read and out the data ;
132        libname DMCA "..\..\data" ;
NOTE: Libref DMCA was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\data
133
134        *---------------------------------------------------------------------;
135        * Preprocessing of the cancer register to primary tumours ;
136
137        * First take the cancer registry, remove all non-cancers ;
138        data cancer ;
139          set DMCA.crg2012 ; * DMCA.cancer ;
140          doca = d_diagnosedato ;
141        * Remove 'not counted as cancer' and non-melanoma skin cancer ;
142          if ( diag in (52,150) ) then delete ;
143        * Recode the leukaemias to one group (139 is a not used value in formats) ;
144          if diag in (134,135,136,137) then diag = 139 ;
145        * Recode the colon cancers to the three separate subsites and the rest ;
146          * 24.1 Ascending colon C18.0, C18.1, C18.2
147          * 24.2 Transverse colon C18.3, C18.4, C18.5
148          * 24.3 Descending and sigmoid colon C18.6, C18.7, C19, C19.9
149          * 24.9 Other colon (unspec. or multiple)
150          * 25.1 Rectum (excl. anus) C20, C209
151          * This means that colorectal cancers are to be taken as the sum of these
152          * 5 groups, but also that the group 24.9 is NOT of interest per se ;
153          if( diag eq 24 )                         then diag = 249 ;
154          if( icdpyrs in ("C180","C181","C182") )  then diag = 241 ;
```

```
155          if( icdpyrs in ("C183","C184","C185") )       then diag = 242 ;
156          if( icdpyrs in ("C186","C187","C19","C199") ) then diag = 243 ;
157          if( icdpyrs in ("C20","C209") )               then diag = 251 ;
158          * Finally make a single code for the sites not among those to be anlysed ;
159          if not ( diag in ( &dgrp. ) ) then diag = 999 ;
160        run ;

NOTE: There were 1929170 observations read from the data set DMCA.CRG2012.
NOTE: The data set WORK.CANCER has 1397464 observations and 34 variables.
NOTE: DATA statement used (Total process time):
      real time           47.71 seconds
      cpu time            2.09 seconds


161
162        * Sort by id and date of diagnosis ;
163        proc sort data = cancer ;
164          by id doCA ;
165        run ;

NOTE: There were 1397464 observations read from the data set WORK.CANCER.
NOTE: The data set WORK.CANCER has 1397464 observations and 34 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           7.55 seconds
      cpu time            3.24 seconds


166
167        * Sort by id ;
168        proc sort data = DMCA.dmr2012  out = diabetes ;
169          by id ;
170        run ;

NOTE: There were 497232 observations read from the data set DMCA.DMR2012.
NOTE: The data set WORK.DIABETES has 497232 observations and 12 variables.
NOTE: PROCEDURE SORT used (Total process time):
      real time           10.98 seconds
      cpu time            0.73 seconds


171
172        * Then merge with the diabetes register ;
173        data DMCR;
174          merge cancer diabetes ;
175          by id ;
176          keep id sex diag diaggrp DMprev T1D
177                  doBT doDM doI doCA doX doDD ;
178          format sex sex.
179                 doBT doDM doI doCA doX doDD ddmmyy10. ;
180          sex = 1 + ( C_SEX in ("K","2") ) ;
181        * Demografic dates collected from CRG and NDR ;
182          doBT = min( D_foddto  , D_fdsdato ) ;
183          doDD = min( D_statdato, D_dodsdto ) ;
184          doX  = min( D_statdato, D_dodsdto, &censdate. ) ;
185        * Event-dates ;
186          doDM = D_inkldto ;
187          doI  = D_ins ;
188          doCA = D_diagnosedato ;
189        * If included before 1.1.1995 set DMprev to 1 otherwise 0 ;
190          DMprev = ( doDM lt &validdate. ) + doDM - doDM ;
191        * Define T1D solely by date of diagnosis ;
192        * Coded as 30 if dx<30, 35 if btw 30 and 35 and 40 if btw 35 and 40
193          and as 42 if over 40, and 0 if no diagnosis of DM ;
194          T1D = min( max( floor((doDM-doBT)/(365.25*5))*5 + 5, 30 ), 42 ) ;
195          if doDM le .z then T1D = 0 ;
196        * Only persons alive on 1.1.1995 ;
197          if doDD gt '31DEC94'd or doDD le .z ;
198        run ;

NOTE: Missing values were generated as a result of performing an operation on missing values.
      Each place is given by: (Number of times) at (Line):(Column).
      543533 at 183:10    1290991 at 190:36   1290991 at 194:19   1290991 at 194:30   1290991 at 194:36
      1290991 at 194:48   1290991 at 194:51
NOTE: There were 1397464 observations read from the data set WORK.CANCER.
NOTE: There were 497232 observations read from the data set WORK.DIABETES.
NOTE: The data set WORK.DMCR has 1066038 observations and 12 variables.
NOTE: DATA statement used (Total process time):
      real time           1.72 seconds
      cpu time            1.71 seconds


199
200        * The dataset DMCR now has a record for each person who has either a
201        * a diabetes diagnosis or a cancer diagnosis. Persons with more than
202        * one recorded tumour are represented by a record for each tumour ;
203
204        title "All cancers (tumours) diagnosed 1995 ff." ;
205        &tab. DMCR ;
206          where ( doca ge &truncdate. ) ;
```

```
        class sex diag doca DMprev T1D ;
        table ( sex all ) * ( all diag = "Dia" ),
              doca * f = comma6.
               all * f = comma7.
              / rts = 9 indent=2 ;
        format doca year4. ;
        keylabel n = " " ;
      run ;

NOTE: There were 539063 observations read from the data set WORK.DMCR.
      WHERE doca>='01JAN1995'D;
NOTE: The PROCEDURE TABULATE printed pages 1-2.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.35 seconds
      cpu time            0.81 seconds



      title "All cancers diagnosed 1995 ff. among T1D patients" ;
      &tab. DMCR ;
        where ( doca ge &truncdate. and
                 T1D in (30,35,40)  and
                doca gt dodm ) ;
        class sex diag doca DMprev T1D ;
        table ( sex all ) * ( DMprev all ) * T1D all,
              doca * f = 4.
               all * f = comma5.
              / rts = 9 indent=2 ;
        table ( sex all ) * ( all diag = "Dia" ),
              doca * f = 4.
               all * f = comma5.
              / rts = 9 indent=2 ;
        format doca year4. ;
        keylabel n = " " ;
      run ;

NOTE: There were 1113 observations read from the data set WORK.DMCR.
      WHERE (doca>='01JAN1995'D) and T1D in (30, 35, 40) and (doca>dodm);
NOTE: The PROCEDURE TABULATE printed pages 3-4.
NOTE: PROCEDURE TABULATE used (Total process time):
      real time           0.19 seconds
      cpu time            0.17 seconds



      * Finally export the relevant data for analysis in XPT format ;
      libname xptcrg xport '../data/DMCR.xpt';
NOTE: Libref XPTCRG was successfully assigned as follows:
      Engine:        XPORT
      Physical Name: C:\Bendix\Steno\DM-register\NDR\projects\Cancer\T1D\data\DMCR.xpt
      proc copy  in = work
                out = xptcrg memtype = data ;
        select DMCR ;
      run;

NOTE: Copying WORK.DMCR to XPTCRG.DMCR (memtype=DATA).
NOTE: There were 1066038 observations read from the data set WORK.DMCR.
NOTE: The data set XPTCRG.DMCR has 1066038 observations and 12 variables.
NOTE: PROCEDURE COPY used (Total process time):
      real time           7.22 seconds
      cpu time            0.98 seconds


NOTE: SAS Institute Inc., SAS Campus Drive, Cary, NC USA 27513-2414
NOTE: The SAS System used:
      real time           1:18.36
      cpu time            10.32 seconds
```

	doca														
	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Males															
All	11,625	11,798	12,076	12,424	12,660	12,860	12,951	13,258	13,741	15,069	15,528	16,077	16,718	17,439	18,076
21	230	257	266	260	254	254	290	308	296	306	274	278	276	324	323
22	364	354	331	311	345	307	330	338	341	330	358	373	376	382	403
26	168	143	154	158	179	181	181	200	197	194	181	196	180	218	221
28	289	291	344	348	347	361	358	366	397	418	445	435	447	463	481
33	2,079	2,002	2,067	1,980	2,042	2,174	2,132	2,141	2,085	2,169	2,186	2,203	2,314	2,247	2,298
51	434	407	410	409	454	508	466	491	549	475	609	631	716	705	912
70	20	13	28	28	30	16	35	26	31	26	20	26	20	32	16
91	1,417	1,554	1,651	1,812	1,902	2,023	2,083	2,306	2,494	3,380	3,599	3,733	4,189	4,567	4,773
92	301	294	302	289	300	280	249	279	280	303	275	295	320	286	320
101	262	308	293	313	308	326	332	304	331	367	374	408	396	424	444
103	1,130	1,148	1,224	1,280	1,255	1,208	1,178	1,208	1,323	1,307	1,310	1,297	1,274	1,298	1,346
113	368	347	370	365	361	363	354	345	373	386	399	427	413	437	430
121	44	45	34	44	37	53	39	40	46	45	43	52	60	61	60
131	67	68	80	80	75	58	63	82	86	64	78	88	74	74	70

132	412	402	409	392	421	399	399	413	470	476	460	484	527	508	521
133	168	184	174	161	179	181	197	178	202	208	226	235	266	246	260
139	396	372	395	407	460	434	440	401	411	499	482	527	526	489	507
241	275	253	271	316	322	253	333	288	337	345	333	370	359	390	400
242	118	135	126	150	146	138	184	157	178	190	177	161	213	214	216
243	514	552	537	599	572	620	623	631	594	593	644	713	626	633	672
249	74	94	77	90	83	106	87	94	105	116	101	116	113	147	162
251	576	610	661	594	631	677	666	693	713	793	804	823	826	884	904
999	1,919	1,965	1,872	2,038	1,957	1,940	1,932	1,969	1,902	2,079	2,150	2,206	2,207	2,410	2,337
Females															
All	12,869	12,966	13,268	13,745	13,831	13,902	14,119	14,355	14,298	14,884	15,115	15,943	15,891	16,972	18,193
21	85	100	117	93	104	105	120	110	102	132	122	141	121	127	118
22	217	215	203	223	204	156	211	167	205	195	182	213	189	191	178
26	103	98	86	108	119	99	106	116	108	86	63	78	89	102	90
28	332	322	357	399	376	377	437	407	392	441	416	450	493	458	521
33	1,357	1,354	1,470	1,469	1,567	1,580	1,625	1,726	1,689	1,798	1,872	1,944	2,047	2,041	2,075
51	572	490	510	483	573	605	591	577	688	662	745	817	821	903	1,042
70	3,320	3,526	3,490	3,626	3,719	3,821	3,920	4,153	3,987	4,032	4,050	4,218	4,225	4,856	5,843
82	507	495	438	440	444	394	417	372	417	396	414	392	371	367	400
83	648	629	629	629	643	645	689	641	687	683	677	701	682	755	781
84	612	579	598	626	634	630	624	624	577	535	596	576	575	559	595
101	195	197	182	254	202	230	213	177	201	227	200	222	229	249	265
103	362	386	409	425	411	412	439	441	436	436	455	475	475	477	481
113	297	306	281	371	330	356	320	332	302	349	324	342	374	449	404
121	83	110	86	104	108	120	102	98	114	156	117	122	130	142	156
131	51	50	49	51	39	56	64	54	55	44	47	65	73	46	67
132	348	344	365	361	352	385	375	370	382	388	423	419	444	432	431
133	147	132	137	140	136	163	144	131	131	147	160	210	185	185	226
139	296	304	312	353	334	308	283	378	315	367	356	374	371	376	379
241	400	387	454	447	394	413	458	412	421	458	504	532	516	531	519
242	144	158	184	197	210	173	191	185	183	206	200	200	219	226	247
243	571	507	593	551	555	550	525	603	548	530	582	557	534	567	528
249	95	96	99	99	108	96	99	77	109	130	109	119	141	154	182
251	460	453	501	509	474	461	496	490	475	555	496	630	559	584	560
999	1,667	1,728	1,718	1,787	1,795	1,767	1,670	1,714	1,774	1,931	2,005	2,146	2,028	2,195	2,105
All															
All	24,494	24,764	25,344	26,169	26,491	26,762	27,070	27,613	28,039	29,953	30,643	32,020	32,609	34,411	36,269
21	315	357	383	353	358	359	410	418	398	438	396	419	397	451	441
22	581	569	534	534	549	463	541	505	546	525	540	586	565	573	581
26	271	241	240	266	298	280	287	316	305	280	244	274	269	320	311
28	621	613	701	747	723	738	795	773	789	859	861	885	940	921	1,002
33	3,436	3,356	3,537	3,449	3,609	3,754	3,757	3,867	3,774	3,967	4,058	4,147	4,361	4,288	4,373
51	1,006	897	920	892	1,027	1,113	1,057	1,068	1,237	1,137	1,354	1,448	1,537	1,608	1,954
70	3,340	3,539	3,518	3,654	3,749	3,837	3,955	4,179	4,018	4,058	4,070	4,244	4,245	4,888	5,859
82	507	495	438	440	444	394	417	372	417	396	414	392	371	367	400
83	648	629	629	629	643	645	689	641	687	683	677	701	682	755	781
84	612	579	598	626	634	630	624	624	577	535	596	576	575	559	595
91	1,417	1,554	1,651	1,812	1,902	2,023	2,083	2,306	2,494	3,380	3,599	3,733	4,189	4,567	4,773
92	301	294	302	289	300	280	249	279	280	303	275	295	320	286	320
101	457	505	475	567	510	556	545	481	532	594	574	630	625	673	709
103	1,492	1,534	1,633	1,705	1,666	1,620	1,617	1,649	1,759	1,743	1,765	1,772	1,749	1,775	1,827
113	665	653	651	736	691	719	674	677	675	735	723	769	787	886	834
121	127	155	120	148	145	173	141	138	160	201	160	174	190	203	216
131	118	118	129	131	114	114	127	136	141	108	125	153	147	120	137
132	760	746	774	753	773	784	774	783	852	864	883	903	971	940	952
133	315	316	311	301	315	344	341	309	333	355	386	445	451	431	486
139	692	676	707	760	794	742	723	779	726	866	838	901	897	865	886
241	675	640	725	763	716	666	791	700	758	803	837	902	875	921	919
242	262	293	310	347	356	311	375	342	361	396	377	361	432	440	463
243	1,085	1,059	1,130	1,150	1,127	1,170	1,148	1,234	1,142	1,123	1,226	1,270	1,160	1,200	1,200
249	169	190	176	189	191	202	186	171	214	246	210	235	254	301	344
251	1,036	1,063	1,162	1,103	1,105	1,138	1,162	1,183	1,188	1,348	1,300	1,453	1,385	1,468	1,464
999	3,586	3,693	3,590	3,825	3,752	3,707	3,602	3,683	3,676	4,010	4,155	4,352	4,235	4,605	4,442

(Continued)

	doca			
	2010	2011	2012	All
Males				
All	17,752	18,213	18,139	266,404
21	360	383	431	5,370
22	390	363	326	6,322
26	231	289	268	3,539
28	504	529	515	7,338
33	2,326	2,324	2,368	39,137
51	883	999	995	11,053
70	34	30	42	473
91	4,149	4,348	4,328	54,308
92	264	289	317	5,243
101	498	495	466	6,649
103	1,302	1,378	1,270	22,736
113	537	479	492	7,246
121	52	68	78	901
131	71	81	91	1,350
132	573	583	598	8,447
133	237	279	252	3,833

139	511	504	488	8,249
241	453	437	426	6,161
242	233	216	221	3,173
243	633	638	646	11,040
249	193	115	140	2,013
251	893	865	932	13,545
999	2,425	2,521	2,449	38,278
Females				
All	17,546	17,639	17,123	272,659
21	126	141	148	2,112
22	185	195	182	3,511
26	118	105	133	1,807
28	468	477	465	7,588
33	2,236	2,234	2,201	32,285
51	995	1,169	1,071	13,314
70	5,149	4,689	4,565	75,189
82	356	409	361	7,390
83	732	817	793	12,461
84	570	573	559	10,642
101	235	248	264	3,990
103	474	474	446	7,914
113	423	383	434	6,377
121	175	208	198	2,329
131	61	63	60	995
132	449	454	432	7,154
133	180	178	170	2,902
139	364	374	345	6,189
241	555	600	606	8,607
242	257	226	268	3,674
243	532	544	541	9,918
249	175	194	135	2,217
251	550	622	573	9,448
999	2,181	2,262	2,173	34,646
All				
All	35,298	35,852	35,262	539,063
21	486	524	579	7,482
22	575	558	508	9,833
26	349	394	401	5,346
28	972	1,006	980	14,926
33	4,562	4,558	4,569	71,422
51	1,878	2,168	2,066	24,367
70	5,183	4,719	4,607	75,662
82	356	409	361	7,390
83	732	817	793	12,461
84	570	573	559	10,642
91	4,149	4,348	4,328	54,308
92	264	289	317	5,243
101	733	743	730	10,639
103	1,776	1,852	1,716	30,650
113	960	862	926	13,623
121	227	276	276	3,230
131	132	144	151	2,345
132	1,022	1,037	1,030	15,601
133	417	457	422	6,735
139	875	878	833	14,438
241	1,008	1,037	1,032	14,768
242	490	442	489	6,847
243	1,165	1,182	1,187	20,958
249	368	309	275	4,230
251	1,443	1,487	1,505	22,993
999	4,606	4,783	4,622	72,924

doca

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	All
Males																			
0																			
30	1	.	.	.	1	1	.	1	4	4	.	3	3	2	2	4	6	4	36
35	.	.	.	.	.	1	1	.	.	1	2	5	5	2	6	3	4	12	42
40	.	1	1	1	.	3	5	3	3	6	8	10	8	6	6	11	15	20	107
1																			
30	.	2	.	.	5	2	.	1	5	2	2	8	5	3	6	2	7	11	61
35	1	1	2	1	1	.	4	1	1	1	1	5	3	5	4	5	6	7	49
40	2	4	5	6	4	5	6	7	8	5	5	5	9	12	14	10	17	11	135
All																			
30	1	2	.	.	6	3	.	2	9	6	2	11	8	5	8	6	13	15	97
35	1	1	2	1	1	1	5	1	1	2	3	10	8	7	10	8	10	19	91
40	2	5	6	7	4	8	11	10	11	11	13	15	17	18	20	21	32	31	242
Females																			
0																			
30	.	.	.	2	1	1	3	5	1	4	.	5	4	6	13	9	12	18	84
35	1	.	.	1	3	1	2	2	1	6	3	4	4	12	11	15	12	13	91
40	.	.	.	5	2	2	6	9	6	5	13	9	11	18	25	17	14	22	164
1																			
30	3	5	.	5	3	3	7	5	7	2	13	5	7	6	12	11	7	10	111
35	2	4	2	2	5	5	6	3	3	15	2	6	5	8	7	5	7	7	94
40	2	7	6	2	5	2	5	9	5	8	4	13	8	14	13	19	5	12	139

All																			
30	3	5	.	7	4	4	10	10	8	6	13	10	11	12	25	20	19	28	195
35	3	4	2	3	8	6	8	5	4	21	5	10	9	20	18	20	19	20	185
40	2	7	6	7	7	4	11	18	11	13	17	22	19	32	38	36	19	34	303
All																			
0																			
30	1	.	.	2	2	2	3	6	5	8	.	8	7	8	15	13	18	22	120
35	1	.	.	1	3	2	3	2	1	7	5	9	9	14	17	18	16	25	133
40	.	1	1	6	2	5	11	12	9	11	21	19	19	24	31	28	29	42	271
1																			
30	3	7	.	5	8	5	7	6	12	4	15	13	12	9	18	13	14	21	172
35	3	5	4	3	6	5	10	4	4	16	3	11	8	13	11	10	13	14	143
40	4	11	11	8	9	7	11	16	13	13	9	18	17	26	27	29	22	23	274
All																			
30	4	7	.	7	10	7	10	12	17	12	15	21	19	17	33	26	32	43	292
35	4	5	4	4	9	7	13	6	5	23	8	20	17	27	28	28	29	39	276
40	4	12	12	14	11	12	22	28	22	24	30	37	36	50	58	57	51	65	545
All	12	24	16	25	30	26	45	46	44	59	53	78	72	94	119	111	112	147	1,113

doca

	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	All
Males																			
All	4	8	8	8	11	12	16	13	21	19	18	36	33	30	38	35	55	65	430
21	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	1	.	1	4
22	.	.	.	.	.	.	.	.	1	.	.	.	.	.	2	2	1	.	6
26	.	.	.	1	1	.	.	1	.	.	.	2	1	.	.	.	.	3	9
28	.	.	.	.	1	1	.	1	2	.	.	1	4	1	1	.	2	1	15
33	1	.	.	1	.	.	3	1	1	4	2	3	5	2	4	4	3	5	39
51	.	1	.	.	.	1	1	.	.	.	.	3	4	.	2	.	4	5	21
91	.	.	.	.	.	.	.	.	1	.	.	.	.	3	2	1	2	3	12
92	.	1	1	1	3	2	1	.	4	3	1	1	3	1	3	2	5	5	37
101	.	.	2	.	.	1	.	1	1	1	.	4	1	2	3	1	4	2	23
103	.	.	1	1	.	.	.	2	.	.	1	3	.	2	2	.	1	4	17
113	2	1	1	.	2	1	.	1	1	2	1	2	1	4	2	3	2	5	31
121	.	.	.	.	.	1	.	1	.	.	1	.	1	.	.	.	.	2	6
131	.	1	.	.	1	1	.	.	3	1	.	1	.	.	.	.	2	1	11
132	.	.	.	.	.	3	.	.	1	2	2	4	2	3	2	3	3	1	26
133	.	.	.	.	.	.	.	.	1	.	1	.	1	.	1	.	.	.	4
139	.	.	.	.	.	.	1	1	.	.	.	3	.	2	1	3	1	2	14
241	1	.	.	.	.	.	1	.	.	.	.	.	.	.	1	3	1	2	9
242	.	1	.	.	.	.	.	.	.	.	.	1	.	1	.	1	1	2	7
243	.	.	.	.	1	.	.	.	.	1	.	.	.	1	2	.	2	1	8
249	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	2
251	.	.	.	.	.	.	1	1	1	.	.	.	3	.	1	.	3	3	13
999	.	3	2	4	2	1	8	3	4	5	9	8	6	8	8	11	18	16	116
Females																			
All	8	16	8	17	19	14	29	33	23	40	35	42	39	64	81	76	57	82	683
21	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	1
22	.	.	.	.	.	.	1	.	1	2	2	2	.	4	1	1	.	.	14
26	.	.	.	.	.	.	.	.	1	.	.	1	2	.	.	.	1	1	6
28	.	.	.	1	.	1	.	.	1	.	.	1	.	1	1	1	1	1	9
33	.	1	.	1	.	1	1	1	2	3	2	1	4	5	3	2	6	5	38
51	.	1	.	.	2	.	1	2	3	5	3	.	3	6	7	8	6	12	59
70	1	4	3	4	7	3	9	12	6	6	6	10	13	18	24	20	13	25	184
82	1	3	2	3	1	1	2	2	3	6	2	3	2	3	5	4	1	4	48
83	.	.	2	.	1	.	2	2	1	2	1	2	3	4	7	4	7	2	40
84	.	.	.	.	2	.	1	1	1	.	2	3	.	3	2	2	1	7	25
101	.	.	.	.	.	.	.	.	.	1	3	.	.	2	1	1	2	5	15
103	.	.	.	.	.	.	.	2	.	.	.	1	.	.	1	1	1	1	7
113	3	1	1	1	2	1	5	4	2	1	2	3	2	1	4	2	4	3	42
121	.	.	.	1	1	2	1	.	.	2	3	.	2	1	5	3	4	4	29
131	.	.	.	.	.	.	.	1	.	.	.	.	1	1	.	1	.	.	4
132	.	1	.	.	.	.	.	.	1	1	1	.	.	3	3	3	1	.	14
133	.	.	.	.	.	1	.	.	.	.	.	.	.	.	1	1	1	1	5
139	.	1	.	1	.	1	.	.	1	1	.	2	.	2	2	3	.	2	16
241	.	.	.	.	.	.	.	1	.	.	1	.	.	1	.	1	.	.	4
243	.	.	.	.	2	.	.	1	.	.	1	2	1	.	2	.	.	1	10
249	.	.	.	.	.	.	.	.	.	.	.	1	.	.	1	.	.	.	2
251	.	1	.	.	.	.	.	1	.	.	1	2	.	.	.	3	.	2	10
999	3	3	.	5	1	3	5	3	.	10	5	8	6	9	11	15	8	6	101
All																			
All	12	24	16	25	30	26	45	46	44	59	53	78	72	94	119	111	112	147	1,113
21	.	.	.	.	.	.	1	.	.	.	.	.	1	.	1	1	.	1	5
22	.	.	.	.	.	.	1	.	2	2	2	2	.	4	3	3	1	.	20
26	.	.	.	1	1	.	.	1	1	.	.	3	3	.	.	.	1	4	15
28	.	.	.	1	1	2	.	1	3	.	.	2	4	2	2	1	3	2	24
33	1	1	.	2	.	1	4	2	3	7	4	4	9	7	7	6	9	10	77
51	.	2	.	.	2	1	2	2	3	5	3	3	7	6	9	8	10	17	80
70	1	4	3	4	7	3	9	12	6	6	6	10	13	18	24	20	13	25	184
82	1	3	2	3	1	1	2	2	3	6	2	3	2	3	5	4	1	4	48
83	.	.	2	.	1	.	2	2	1	2	1	2	3	4	7	4	7	2	40
84	.	.	.	.	2	.	1	1	1	.	2	3	.	3	2	2	1	7	25
91	.	.	.	.	.	.	.	.	1	.	.	.	.	3	2	1	2	3	12
92	.	1	1	1	3	2	1	.	4	3	1	1	3	1	3	2	5	5	37
101	.	.	2	.	.	1	.	1	1	2	3	4	1	4	4	2	6	7	38
103	.	.	1	1	.	.	.	4	.	.	1	4	.	2	3	1	2	5	24

```
 113      5    2    2    1    4    2    5    5    3    3    3    5    3    5    6    5    6    8   73
 121      .    .    .    1    1    3    1    1    .    2    4    .    3    1    5    3    4    6   35
 131      .    1    .    .    1    1    .    1    3    1    .    1    1    1    .    1    2    1   15
 132      .    1    .    .    .    3    .    .    2    3    3    4    2    6    5    6    4    1   40
 133      .    .    .    .    .    1    .    .    1    .    1    .    1    .    2    1    1    1    9
 139      .    1    .    1    .    1    1    1    1    1    .    5    .    4    3    6    1    4   30
 241      1    .    .    .    .    .    1    1    .    .    1    .    .    1    1    4    1    2   13
 242      .    1    .    .    .    .    .    .    .    .    .    1    .    1    .    1    1    2    7
 243      .    .    .    .    3    .    .    1    .    1    1    2    1    1    4    .    2    2   18
 249      .    .    1    .    .    .    .    .    .    .    .    1    .    .    1    .    .    1    4
 251      .    1    .    .    .    .    1    2    1    .    1    2    3    .    1    3    3    5   23
 999      3    6    2    9    3    4   13    6    4   15   14   16   12   17   19   26   26   22  217
```

The central point in this SAS-program is that we construct a variable, `T1D`, which takes the value 40 if the person is included in the diabetes register in ages 35–39 (*i.e.* between his 35th and 40th birthdays), 35 if the person is included in the register at ages 30–34 (*i.e.* between his 30th and 35th birthdays), and 30 if the person is included in the register at age 29 or younger (*i.e.* before his 30th birthday). The variable is coded 42 if the patient is included in the register in age over 40, and 0 if the person does not have a diagnosis of diabetes.

The variable `DMprev` is the indicator of being diagnosed before 1.1.1995, that is the indicator of being a prevalent case as of 1.1.1995. Note we only collect some of the persons diagnosed below age 40 before 1995. Some persons may have been diagnosed before 1995 in age under 40, but not picked up by the registers till after their 40th birthday, and some may have died before 1.1.1995. Hence we can only follow persons for cancer occurrence after 1.1.1995. By the age of the Danish Cancer Registry we are pretty sure that all tumours diagnosed in these person's lives are recorded.

We now read the constructed SAS-dataset:

```
> options( width=90 )
> crg <- read.xport( "../data/DMCR.xpt" )
> names( crg ) <- tolower( names(crg) )
> crg <- transform( crg, sex = factor( sex, labels=c("M","F") ),
+                        t1d = factor( t1d ),
+                     dmprev = factor( dmprev, levels=0:2, labels=c("Inc","Prv","Pop") ) )
> levels( crg$t1d )[c(1,5)] <- c("NoDM","T2D")
```

We tabulate the number of cancers and diabetes cases in it. Note however that the number of diabetes patients is slightly exaggerated, since some of them are represented by more than one record (namely those who are recorded with more than one cancer):

```
> with( crg, table(dmprev,t1d,exclude=NULL) )
      t1d
dmprev   NoDM     30     35     40    T2D   <NA>
  Inc       0  15715   8114  11340 345142      0
  Prv       0   7332   3212   4006  86460      0
  Pop       0      0      0      0      0      0
  <NA> 584717      0      0      0      0      0
> with( subset(crg, is.na(doca)), table( table(id) ) )

     1
381938
> with( subset(crg,!is.na(doca)), table( table(id) ) )

     1      2      3      4      5      6
552566  57451   4956    409     22      3
```

The last two tables here illustrate that persons without a cancer diagnosis only have one entry in the dataset, whereas persons with a cancer diagnosis may have more entries, that is be registered with more than one cancer.

3.1.1 Dates in `cal.yr` format

We have got the dates as SAS-dates, that is as number of days since 1 January 1960, so we convert them to `cal.yr` objects, that is years since A.D., using 365.25 as the length of a year.

```
> library(Epi)
> names( crg )[wh<-grep("do",names(crg))]
[1] "doca" "dobt" "dodm" "doi"  "dox"  "dodd"
> for( i in wh ) crg[,i] <- cal.yr( as.Date( crg[,i], origin="1960-01-01" ) )
```

Finally we also define the date of entry into the study (only relevant for the diabetes patients), which is the latest of date of diabetes diagnosis and 1.1.1995:

```
> crg$doe <- with( crg, pmax( 1995, dodm, na.rm=TRUE ) )
```

Finally we make a few sanity checks:

```
> with( crg, ftable( "bt<dm"= dobt <= dodm | is.na(dodm),
+                    "bt<ca"= dobt <= doca | is.na(doca),
+                    "dm<dd"= dodm <= dodd | is.na(dodm) | is.na(dodd),
+                    "ca<dd"= doca <= dodd | is.na(doca) | is.na(dodd) ) )
                  ca<dd   FALSE    TRUE
bt<dm bt<ca dm<dd
FALSE FALSE FALSE             0       0
            TRUE              0       0
      TRUE  FALSE             0       0
            TRUE              0       1
TRUE  FALSE FALSE             0       0
            TRUE              0      40
      TRUE  FALSE             5      86
            TRUE            187 1065719
```

Finally we exclude persons with dates that does not meet all of these normal sanity checks:

```
> nrow( crg )
[1] 1066038
> crg <- subset( crg, ( dobt <= dodm | is.na(dodm) ) &
+                     ( dobt <= doca | is.na(doca) ) &
+                     ( dodm <= dodd | is.na(dodm) | is.na(dodd) ) &
+                     ( doca <= dodd | is.na(doca) | is.na(dodd) ) )
> nrow( crg )
[1] 1065719
```

3.1.2 Overview of data

For completeness we first reproduce the table from the SAS-program, where we note that the allocation of the dates of diagnosis to years is slightly different, because this program uses years of equal length, whereas SAS uses the traditional years of unequal length. But the distribution by

```
> options( width=91 )
> with( subset( crg, dodm<doca & doca>1995 & t1d %in% c("30","35","40") ),
+       addmargins( table( diag, floor(doca), useNA="ifany" ) ) )
```

```
diag 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011
  21    0    0    0    0    0    0    1    0    0    0    0    0    1    0    1    1    0
  22    0    0    0    0    0    0    1    0    2    2    2    2    0    4    3    3    1
  26    0    0    0    1    1    0    0    1    1    0    0    3    3    0    0    0    1
  28    0    0    0    1    1    2    0    1    3    0    0    2    4    2    2    1    3
  33    1    1    0    2    0    1    4    2    3    7    4    4    9    7    7    6    9
  51    1    1    0    0    2    1    2    2    3    5    3    3    7    6    9    8   10
  70    1    4    3    4    8    2    9   13    5    6    6   10   13   18   24   20   13
  82    1    3    2    3    1    1    2    2    3    6    2    3    2    3    5    4    1
  83    0    0    2    0    1    0    2    2    1    2    1    2    3    4    7    4    7
  84    0    0    0    0    2    0    1    1    1    0    2    3    0    3    2    2    1
  91    0    0    0    0    0    0    0    0    1    0    0    0    0    3    2    1    2
  92    0    1    1    1    4    1    1    0    4    3    1    1    3    1    3    2    5
 101    0    0    2    0    0    1    0    1    1    2    3    4    1    4    4    2    6
 103    0    0    1    1    0    0    0    4    0    0    1    4    0    2    3    1    2
 113    6    1    2    1    5    1    5    5    3    3    3    5    3    5    6    5    6
 121    0    0    0    1    2    2    1    1    0    2    4    0    3    1    5    3    4
 131    0    1    0    0    1    1    0    1    3    1    0    1    1    1    0    1    2
 132    0    1    0    0    0    3    0    0    2    3    3    4    2    6    5    6    4
 133    0    0    0    0    0    1    0    0    1    0    1    0    1    0    2    1    1
 139    0    1    0    1    0    1    1    1    1    1    0    5    0    4    3    6    1
 241    1    0    0    0    0    0    1    1    0    0    1    0    0    1    1    4    1
 242    1    0    0    0    0    0    0    0    0    0    0    1    0    1    0    1    1
 243    0    0    0    0    3    0    0    1    0    1    1    2    1    1    4    0    2
 249    0    0    1    0    0    0    0    0    0    0    0    1    0    0    1    0    0
 251    0    1    0    0    0    0    1    2    1    0    1    2    3    0    1    3    3
 999    4    5    2    9    4    3   13    6    4   15   14   16   12   17   19   26   26
 Sum   16   20   16   25   35   21   45   47   43   59   53   78   72   94  119  111  112

diag 2012  Sum
  21    1    5
  22    0   20
  26    4   15
  28    2   24
  33   10   77
  51   17   80
  70   25  184
  82    4   48
  83    2   40
  84    7   25
  91    3   12
  92    5   37
 101    7   38
 103    5   24
 113    8   73
 121    6   35
 131    1   15
 132    1   40
 133    1    9
 139    4   30
 241    2   13
 242    2    7
 243    2   18
 249    1    4
 251    5   23
 999   22  217
 Sum  147 1113
> tt <- with( crg, table( t1d,
+             "St.1995" = dmprev,
+                has.ca = doca>dodm & doca>1995,
+                         useNA="ifany" ) )
> ftable( addmargins( tt, margin=1 ), row.vars=1 )
     St.1995   Inc                   Prv                   Pop                   NA
     has.ca  FALSE   TRUE      NA  FALSE   TRUE      NA  FALSE   TRUE      NA  FALSE   TRUE      NA
t1d
NoDM              0       0       0       0       0       0       0       0       0 126979       0 457517
```

```
30              115    120  15480    31    172   7128      0      0      0      0      0      0
35               92    133   7887    28    143   3041      0      0      0      0      0      0
40              182    271  10887    50    274   3682      0      0      0      0      0      0
T2D           34750  41126 269171  7584  14270  64606      0      0      0      0      0      0
Sum           35139  41650 303425  7693  14859  78457      0      0      0 126979      0 457517
> sum( tt[c("30","35","40"), ,2] )
[1] 1113
> sum( tt[  "30"          , ,2] )
[1] 292
> sum( tt[c("30","35","40"),1,2] )
[1] 524
> sum( tt[c("30","35","40"), , ] )
[1] 49716
```

So we see that we have 1113 recorded tumours among (verified) T1 patients after 1995, but only 292 if we use the "normal" strict definition of T1D (diagnosis before age 30). If we further restrict to T1D patients diagnosed after 1995, we only have 524, by the broad definition, and a mere 120 by the narrow definition.

We also see that the maximal number of persons that we will be following is 49,716 (somewhat less because these are *records* in the dataset, not persons), less than 1% of the population (but presumably a bit more in the age-bracket we are looking at), so when doing the comparison with the rest of the population, we can to a very close approximation use the rates of all the cancers in the entire population. Formally the T1D patients are then counted twice in follow-up but the error will be tiny.

3.2 T1D analysis dataset

On the basis of this total dataset with all cancers and diabetes diagnoses, we now set up an analysis dataset, which for the tabulated follow-up of the T1D patients will have the following variables:

- Response variables

  `dxx` no. of events in diagnosis group `xx` (cancer type)

  `Y` person-years (corresponding to each diagnosis group)

- Classification (explanatory) variables:

  `sex` sex

  `age` age at follow-up, 1-year classes

  `per` date of follow-up, 1 year classes

  `dur` time since diagnosis, left end point of intervals $0, 1, 2, 5, 10, 15$

  `t1d` T1 diabetes status: 30 — DM diagnosis < 30, 35 — DM diagnosis aged 30–34, 40 — DM diagnosis aged 35–39

  `DMprev` T1D present at 1.1.1995: Yes/No

The dataset will be constructed from multi-way tables of cases and person-years derived from the dataset of cases.

For the comparison population (which will be the entire population) we will construct a similar dataset. In the population dataset there will of course be separte values for the variables `dur`, `t1d` and `DMprev` — these variables will enter the analysis models, and must therefore also be present in the population dataset, but with a special value that allows for a special (reference) level of their effect for the population part of data. Stacking the two datasets appropriately will then enable a joint analysis, similar to an SMR-analysis.

3.2.1 Follow-up of T1D patients

We first construct the follow-up *time* for T1D patients from entry to death, *i.e.* disregarding cancer occurrence as termination of follow-up. We later discuss this omission. Subsequently, we tabulate the number of cancer cases and construct the T1D part of the analysis dataset.

3.2.1.1 Lexis object of total follow-up

Intuitively, the follow-up would be for T1D patients without cancer[1] from date of diagnosis of T1D (or 1.1.1995) until the end of the follow-up period. However, if we follow persons for *any* type of cancer regardless of other previous cancers, we should logically also follow persons for (other) cancers even if they have a cancer previous to the diagnosis of diabetes (which is what will be done).

The end of the follow-up period is in this case the end of 2012, but not for the diabetes patients:

```
> max( crg$dodm, na.rm=T )
[1] 2011.996
> max( crg$doca, na.rm=T )
[1] 2012.998
> max( crg$dodd, na.rm=T )
[1] 2012.998
> with( subset( crg, !is.na(dodm) & is.na(doca) ),
+       max( dodd, na.rm=T ) )
[1] 2011.996
```

Diabetes patients are only followed till the end of 2011. If we were to extend follow-up till the end of 2012, we would include too much risk time among diabetes patients, namely that after death (wihout cancer) in 2012, and thus underestimate the rates of cancer among T1D pateinst in 2012. The error would however likely be small because we would only erroneousle include a very small amount of follow-up.

We can assess how much this is by looking at the year 2011 and tabulate the entire follow-up during the year and relate to the extra follow-up (without events) we would see if we did not know about the deaths:

```
> dim( crg )
[1] 1065719      13
```

[1] Meaning either any cancer or a specific cancer type

```
> sb <- subset( crg,
+         # Only the T1D persons
+         ( t1d %in% c("30","35","40") ) &
+         # diagnosed with dm before end of 2011
+         ( dodm < 2012 ) &
+         # not diagnosed with cancer before 2011
+         ( pmin( doca, 2011.001, na.rm=TRUE ) > 2011 ) &
+         # and no cancer before diabetes
+         ( pmin( doca, dodm, na.rm=TRUE ) < doca | is.na(dodm) | is.na(doca) ) &
+         # and alive sometime in 2011
+         ( pmin( dodd, 2011.001, na.rm=TRUE ) > 2011 ) )
> dim( sb )
[1] 46018    13
> real.fu <- with( sb, pmin( 2012, dodd, doca, na.rm=TRUE ) -
+                      pmax( 2011, dodm, na.RM=TRUE ) )
> summary( real.fu )
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.002053 1.000000 1.000000 0.969400 1.000000 1.000000
```

The persons in `sb` those at risk of cancer at some time during 2011. If we did know the data of death of these persons we would erroneously include risk time after death until the end of 2011. Thus it only concerns those with a death recorded in 2011:

```
> dsb <- subset( sb, !is.na(dodd) & dodd>2011 & dodd<2012 )
> false.fu <- with( dsb, 2012 - dodd )
> summary( false.fu )
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.01232 0.25330 0.53520 0.51390 0.78710 0.99790
> sum( false.fu )
[1] 112.5352
> sum(  real.fu )
[1] 44611.33
> sum( false.fu ) /
+ sum(  real.fu )
[1] 0.002522571
```

So we see that missing out on the deaths among the T1D patients for the year 2011 would overestimate the person-years by some 0.3% and hence *under*estimate the rate by a similar amount. We are likely to commit an error of the same order of magnitude by including the follow-up of the diabetes patients during 2012. Thus the error by including 2012 in the follow-up of the diabetes patients would be tiny, but we would gain another year's worth of cancer events and thus strengthen the study.

We then tabulate the follow-up for these persons until death or end of 2012, knowing that a few will be counted at risk after their (unknown) death date. The point of this is that while tabulation of events is simple (see below), the person-years to be used for a given type of event (cancer diagnosis of a certain type) is from date of entry to date of exit or event date, and hence requires a complete tabulation of person-years for every type of cancer.

However, the correct risk time for a specific type of cancer can also be computed by subtracting the follow-up time after event (for those persons with an event) from the total lived follow-up, and since the tabulation of follow-up after any of the relatively few events is computationally a rather minor task, this is preferable to a tabulation of the follow-up of the entire cohort of the T1D patients to different endpoints.

However, the follow-up after cancer is most likely so small that it will not matter much anyway; we shall look into this in more detail below.

In order to tabulate the follow-up of the persons we set up a Lexis object for the T1D persons. First we need to subset the data frame to the T1D patients, and then shave it down to one record per patient. Note that we use `aggregate` with `FUN=min`, so that the date of cancer diagnosis in the resulting dataset corresponds to the date of the *first* cancer in a patient:

```
> t1dfr <- subset( crg, t1d %in% c("30","35","40") )
> table( table(t1dfr$id) )
    1     2     3
49552    79     2
> t1fu <- aggregate( x = t1dfr[,c("dobt","doe","doca","dodm","dox","dodd")],
+                   by = t1dfr[,c("id","sex","t1d","dmprev")],
+                  FUN = min )
> table( table(t1fu$id) )
    1
49633
> str( t1fu )
'data.frame':        49633 obs. of  10 variables:
 $ id    : num  15386837 15386934 15387012 15387068 15387287 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ t1d   : Factor w/ 5 levels "NoDM","30","35",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ dmprev: Factor w/ 3 levels "Inc","Prv","Pop": 1 1 1 1 1 1 1 1 1 1 ...
 $ dobt  : num  1978 1999 1986 2004 1975 ...
 $ doe   : num  1995 2009 1995 2010 2004 ...
 $ doca  : num  NA NA NA NA NA NA NA NA NA NA ...
 $ dodm  : num  1995 2009 1995 2010 2004 ...
 $ dox   : num  2013 2013 2013 2013 2013 ...
 $ dodd  : num  NA NA NA NA NA NA NA NA NA NA ...
> summary( t1fu )
       id              sex           t1d          dmprev          dobt           doe
 Min.   :15386193   M:23200   NoDM:    0   Inc:35123   Min.   :1934   Min.   :1995
 1st Qu.:15904736   F:26433   30  :23028   Prv:14510   1st Qu.:1964   1st Qu.:1995
 Median :16431359             35  :11302   Pop:    0   Median :1971   Median :2001
 Mean   :16423549             40  :15303               Mean   :1972   Mean   :2001
 3rd Qu.:16938796             T2D :    0               3rd Qu.:1979   3rd Qu.:2007
 Max.   :17452016                                      Max.   :2011   Max.   :2012

      doca            dodm            dox            dodd
 Min.   :1956   Min.   :1942   Min.   :1995   Min.   :1995
 1st Qu.:1999   1st Qu.:1994   1st Qu.:2013   1st Qu.:2002
 Median :2006   Median :2001   Median :2013   Median :2006
 Mean   :2003   Mean   :2000   Mean   :2013   Mean   :2005
 3rd Qu.:2010   3rd Qu.:2007   3rd Qu.:2013   3rd Qu.:2009
 Max.   :2013   Max.   :2012   Max.   :2013   Max.   :2013
 NA's   :48105                                NA's   :46595
```

As expected and as it should be, the only variables in the dataset with missing values is the date of first cancer and date of death.

Now we set up a Lexis-object representing follow-up of the T1D patients until death or end of study period (`dox`), along timescales age, calendar time and duration. Since we are only interested in person risk time here we do not define any entry or exit status:

```
> Lt1 <- Lexis( entry = list( per = doe,
+                             age = doe-dobt,
+                             dur = doe-dodm ),
+                exit = list( per = dox ),
+                data = t1fu )
> summary( Lt1 )
Transitions:
     To
From     0  Records:  Events: Risk time:  Persons:
   0 49627     49627        0   552879.5     49627
```

```
> head( Lt1 )
       per       age dur   lex.dur lex.Cst lex.Xst lex.id       id sex t1d dmprev     dobt
1 1995.194 17.418207   0 17.804244       0       0      1 15386837   M  30    Inc 1977.775
2 2008.615  9.941136   0  4.383299       0       0      2 15386934   M  30    Inc 1998.674
3 1995.095  8.670773   0 17.902806       0       0      3 15387012   M  30    Inc 1986.424
4 2009.639  5.390828   0  3.359343       0       0      4 15387068   M  30    Inc 2004.248
5 2003.952 28.561259   0  9.045859       0       0      5 15387287   M  30    Inc 1975.391
6 2009.014 11.509925   0  3.983573       0       0      6 15387518   M  30    Inc 1997.504
       doe doca     dodm      dox dodd
1 1995.194   NA 1995.194 2012.998   NA
2 2008.615   NA 2008.615 2012.998   NA
3 1995.095   NA 1995.095 2012.998   NA
4 2009.639   NA 2009.639 2012.998   NA
5 2003.952   NA 2003.952 2012.998   NA
6 2009.014   NA 2009.014 2012.998   NA
```

We must tabulate the risk time by current age, calendar time and duration, so we split the follow-up by these three factors. Due to memory restrictions we do this separately in 10 chunks of the data:

```
> S.all <- NULL
> n.chunks <- 10
> lm <- round( seq(0,nrow(Lt1),,n.chunks+1) )
> system.time(
+ for( i in 1:n.chunks )
+ {
+ whr  <- (lm[i]+1):(lm[i+1])
+ sa   <- splitLexis( Lt1[whr,],        0:120     , time.scale="age" )
+ sap  <- splitLexis(       sa ,     1990:2020    , time.scale="per" )
+ sapd <- splitLexis(       sap, c(0,1,2,5,10,15), time.scale="dur" )
+ S.all <- rbind( S.all, sapd )
+ } )
   user  system elapsed
 144.95    2.81  148.03
> summary( Lt1 )
Transitions:
     To
From     0  Records:  Events: Risk time:  Persons:
   0 49627     49627        0   552879.5     49627
> summary( S.all )
Transitions:
     To
From       0  Records:  Events: Risk time:  Persons:
   0 1284080   1284080        0   552879.5     49627
```

Finally we create a table with the total alive-follow-up among the T1D patients:

```
> system.time(
+ PYtab <- xtabs( lex.dur ~ timeBand(S.all,"age","left") +
+                           timeBand(S.all,"per","left") +
+                           timeBand(S.all,"dur","left") +
+                           sex + t1d + dmprev,
+                 data = S.all ) )
   user  system elapsed
   9.08    0.22    9.41
> names( dimnames( PYtab ) )[1:3] <- c("age","per","DMdur")
> str( PYtab )
 xtabs [1:77, 1:18, 1:6, 1:2, 1:5, 1:3] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 6
  ..$ age   : chr [1:77] "0" "1" "2" "3" ...
  ..$ per   : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ DMdur : chr [1:6] "0" "1" "2" "5" ...
  ..$ sex   : chr [1:2] "M" "F"
  ..$ t1d   : chr [1:5] "NoDM" "30" "35" "40" ...
  ..$ dmprev: chr [1:3] "Inc" "Prv" "Pop"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = lex.dur ~ timeBand(S.all, "age", "left") + timeBand(S.a
```

```
> dimnames( PYtab )
$age
 [1] "0"  "1"  "2"  "3"  "4"  "5"  "6"  "7"  "8"  "9"  "10" "11" "12" "13" "14" "15" "16"
[18] "17" "18" "19" "20" "21" "22" "23" "24" "25" "26" "27" "28" "29" "30" "31" "32" "33"
[35] "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44" "45" "46" "47" "48" "49" "50"
[52] "51" "52" "53" "54" "55" "56" "57" "58" "59" "60" "61" "62" "63" "64" "65" "66" "67"
[69] "68" "69" "70" "71" "72" "73" "74" "75" "76"

$per
 [1] "1995" "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005" "2006"
[13] "2007" "2008" "2009" "2010" "2011" "2012"

$DMdur
[1] "0"  "1"  "2"  "5"  "10" "15"

$sex
[1] "M" "F"

$t1d
[1] "NoDM" "30"   "35"   "40"   "T2D"

$dmprev
[1] "Inc" "Prv" "Pop"
```

(This could actually have been done somewhat more parsimoniously if we had used `aggregate` insted of `xtabs`).

```
> table( PYtab>0 )
 FALSE   TRUE
236663  12817
> summary( as.vector(PYtab[PYtab>0]) )
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
  0.0007   6.2670  21.8300  43.1400  58.2400 345.9000
```

Now the `PYtab` table contains person-years *alive* among T1D patients in the period 1.1.1995 to 31.12.2012.

However it is not for all follow-up time we can trust the subdivision by duration, so any analyses by DM duration must be restricted to those coded with `dmprev` equal to `Inc`:

```
> PYdur <- addmargins( apply( PYtab, c(3,5,6), sum ), margin=2:1 )
> print( round( ftable( PYdur, row.vars=c(3,2) ), 1), zero.print="-" )
            DMdur        0        1        2        5       10       15      Sum
dmprev t1d
Inc    NoDM              -        -        -        -        -        -        -
       30          15667.2  15062.6  38358.2  44507.7  22040.5   3437.5 139073.7
       35           8075.8   7759.8  19972.1  23274.6  11857.9   1894.9  72835.1
       40          11264.6  10800.2  27497.1  31035.3  13994.5   2111.2  96702.8
       T2D               -        -        -        -        -        -        -
       Sum         35007.5  33622.6  85827.4  98817.6  47893.0   7443.6 308611.6
Prv    NoDM              -        -        -        -        -        -        -
       30            428.8   1867.9  12265.0  34758.6  35034.1  42393.7 126748.1
       35            282.3   1076.7   5850.4  15170.9  14689.1  16007.1  53076.5
       40            352.5   1415.4   7492.3  18672.1  17836.5  18674.5  64443.4
       T2D               -        -        -        -        -        -        -
       Sum          1063.5   4360.0  25607.8  68601.5  67559.8  77075.3 244267.9
Pop    NoDM              -        -        -        -        -        -        -
       30                -        -        -        -        -        -        -
       35                -        -        -        -        -        -        -
       40                -        -        -        -        -        -        -
       T2D               -        -        -        -        -        -        -
       Sum               -        -        -        -        -        -        -
```

Thus we see slightly less than half of the follow-up time is among DM-patients prevalent in 1995. Hence when we show the distribution of follow-up time by diabetes duration, we exclude the prevalent cases, because we do not know the date of diagnosis with any certainty:

```
> ah <- apply( PYtab, 1, sum )
> ph <- apply( PYtab, 2, sum )
> dh <- apply( PYtab[,,,,,1], 3, sum )
> layout( rbind(1:3), widths=c(70,17,15)+2 )
> par( mar=c(3,1,1,1), oma=c(0,3,0,0), mgp=c(3,1,0)/1.6, las=1 )
> aa <- barplot( ah/1000, space=0, xaxt="n", ylim=c(0,48),
+                col=gray(0.4), border=FALSE, xlab="Age at follow-up" )
> whl <- 0:7*10+1
> axis( side=1, at=aa[whl]-0.5, labels=names(ah)[whl] )
> pp <- barplot( ph/1000, space=0, xaxt="n", ylim=c(0,48),
+                col=gray(0.4), border=FALSE, xlab="Date of follow-up" )
> whl <- 0:5*5+1
> axis( side=1, at=aa[whl]-0.5, labels=names(ph)[whl] )
> wd <- c(1,1,3,5,5)
> dd <- barplot( dh[-6]/1000/wd, width=wd, space=0, xaxt="n", ylim=c(0,48),
+                col=gray(0.4), border=FALSE, xlab="DM duration" )
> axis( side=1 )
> mtext( "PYers (1000s) in 1-year interval", side=2, line=1, outer=TRUE, las=0 )
```

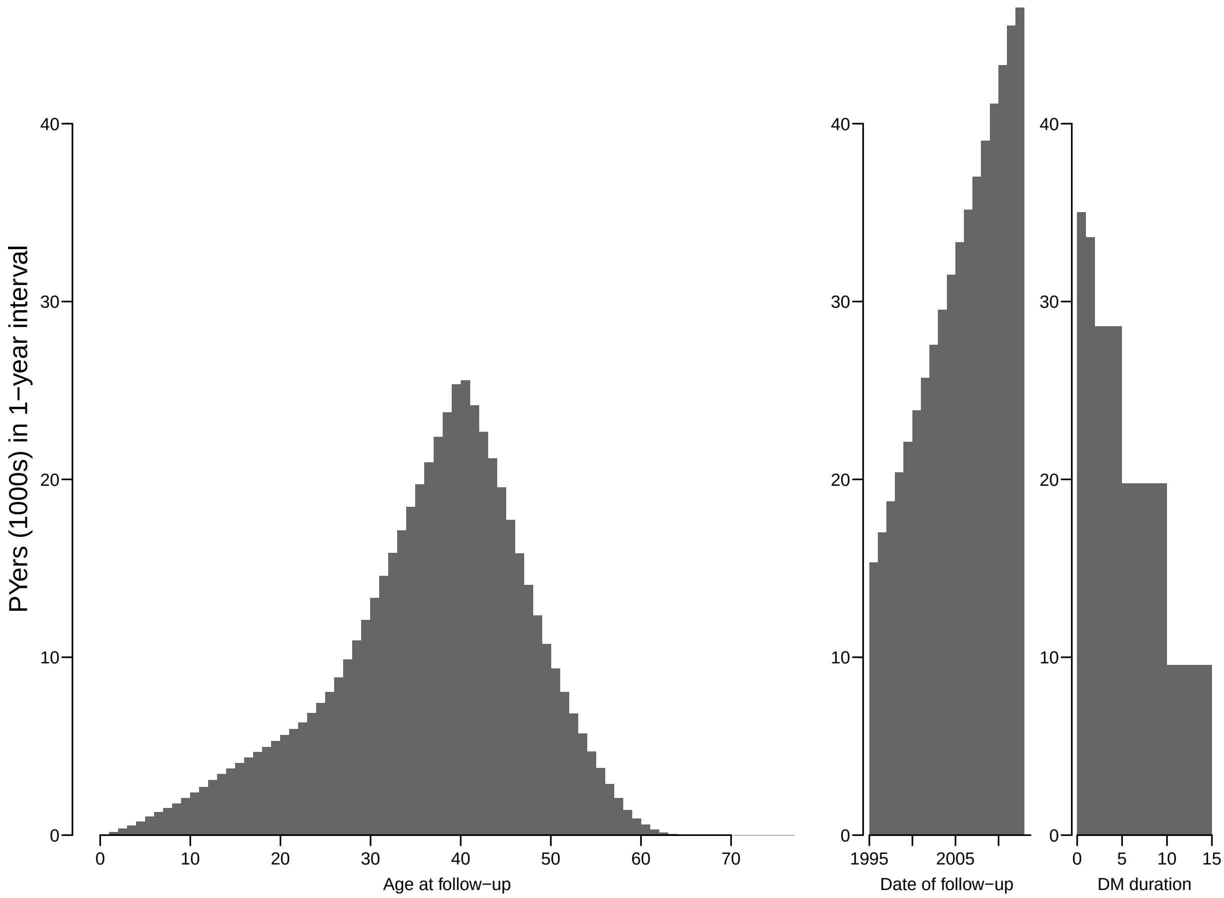

Figure 3.1: *Distribution of the follow-up time in the T1D patients (till death) by age, date and DM-duration at follow-up. The rightmost figure is only for DM-patients diagnosed after 1.1.1995, hence the smaller total area.*

From figure 4.4 it is seen that the distribution of follow-up time is largely in age-classes where cancer occurrence is quite small.

3.2.1.2 Follow-up after cancer

Formally, we will for each type of event have to subtract the follow-up after the particular type of event from this total time alive. In order to see how large this fraction is, we set up a Lexis object with the follow-up after the earliest cancer, that is the maximally possible follow-up time erroneously included.

Note that we in the `aggregate` function above used "`min`" as function, thus the `doca` in this dataset is the date of the earliest recorded cancer. Thus when constructing a `Lexis` object we only take follow-up among those with a cancer, and take the follow-up time from cancer of DM whichever is the *latest*. This will be the follow-up that should not really have been counted:

```
> Ct1 <- Lexis( entry = list( per = pmax( doca, dodm, 1995 ),
+                             age = pmax( doca, dodm, 1995 ) - dobt,
+                             dur = pmax( doca, dodm, 1995 ) - dodm ),
+                exit = list( per = dox ),
+                data = subset( t1fu, !is.na(doca) ) )
> summary( Ct1 )
Transitions:
     To
From    0  Records:  Events: Risk time:  Persons:
   0 1526       1526        0    8344.47      1526
> head( Ct1 )
         per       age       dur    lex.dur lex.Cst lex.Xst lex.id       id sex t1d dmprev
170 2010.230 39.890486 10.35729  1.160849       0       0      1 15438451   M  30    Inc
240 2011.659 25.171800  0.00000  1.338809       0       0      2 15464895   M  30    Inc
303 2001.551  9.938398  0.00000 11.446954       0       0      3 15485118   M  30    Inc
372 2011.435 28.539357  0.00000  1.563313       0       0      4 15505899   M  30    Inc
395 2006.052 26.603696  0.00000  6.945927       0       0      5 15514176   M  30    Inc
445 2002.808 21.544148  0.00000 10.190281       0       0      6 15529803   M  30    Inc
        dobt      doe     doca     dodm      dox     dodd
170 1970.339 1999.873 2010.230 1999.873 2011.391 2011.391
240 1986.487 2011.659 2008.825 2011.659 2012.998       NA
303 1991.613 2001.551 2000.749 2001.551 2012.998       NA
372 1982.895 2011.435 2011.049 2011.435 2012.998       NA
395 1979.448 2006.052 1996.248 2006.052 2012.998       NA
445 1981.264 2002.808 2000.248 2002.808 2012.998       NA
> round( c( "pct Y" = sum(Ct1$lex.du) / sum(S.all$lex.du),
+           "pct N" = with( t1fu, mean( !is.na(doca) ) ) ) * 100, 2 )
pct Y pct N
 1.51  3.08
```

From this it is seen that the total risk time is about 1.5% too large; and some 3.1% of the diabetes patients experience a cancer diagnosis. However, differences might be larger for some combinations of sex, age or calendar time, so we split and tabulate the follow-up time after cancer to make a more detailed assessment:

```
> sa   <- splitLexis( Ct1 ,     0:120 , time.scale="age" )
> sap  <- splitLexis(  sa , 1990:2020, time.scale="per" )
> sapd <- splitLexis(  sap, c(0,1,2,5,10,15), time.scale="dur" )
> summary( sapd )
Transitions:
     To
From     0  Records:  Events: Risk time:  Persons:
   0 19654     19654        0    8344.47      1526
```

```
> PCtab <- xtabs( lex.dur ~ timeBand(sapd,"age","left") +
+                           timeBand(sapd,"per","left") +
+                           timeBand(sapd,"dur","left") +
+                           sex + t1d + dmprev,
+                 data = sapd )
> names( dimnames( PCtab ) )[1:3] <- c("age","per","DMdur")
> rbind( dim( PYtab ),
+        dim( PCtab ) )
     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]   77   18    6    2    5    3
[2,]   67   18    6    2    5    3
```

We see that follow-up after cancer is represented in fewer age classes that the entire follow-up from entry to death.

Now we plot the total amount of person-time in each of the different sex, age- and period classes, against the ratio of the post-cancer follow-up:

```
> Rtab <- PCtab / PYtab[dimnames(PCtab)[[1]],,,,,]
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1 )
> plot( as.vector(Rtab)*100, as.vector(PCtab),
+       pch=16, cex=0.5, bty="n",
+       xlab="% PY missing", ylab="PY missing" )
> axis( side=1, at=1:9*10, labels=rep("",9) )
> abline( v=c(0,2,5,10,15), col="red" )
```

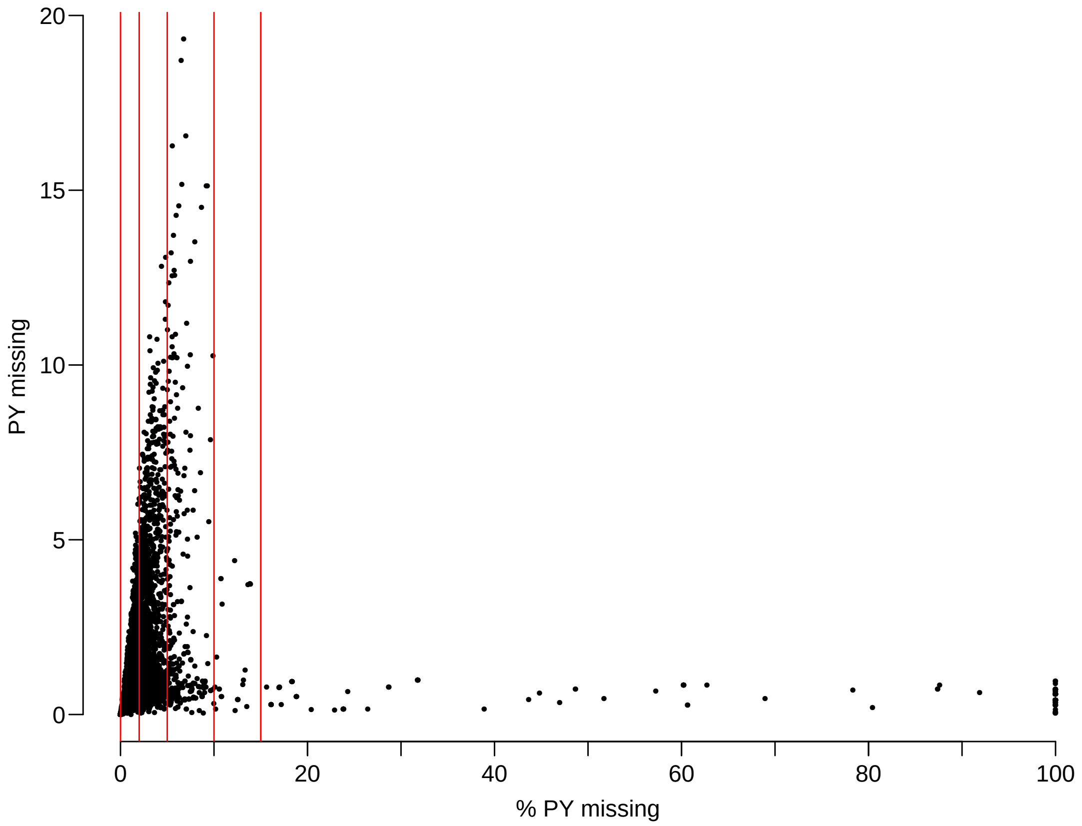

Figure 3.2: *Absolute versus relative difference between correct and approximate follow-up time.*

From figure 3.2 it is seen that except for units where only a tiny amount of risk time is missing, the excess risk time from failing to remove patients from follow-up at cancer diagnosis is generally under 5% and mostly under 2%.

Thus we shall subtract this amount of risk time from the risk time for T1D patients in the analysis of all cancers, but not in the analysis of the single cancer diagnoses as single cancer sites maximally constitute 13% (≈ 1/8) of all cases, and hence the relative error in rates will be substantially below 1%:

```
> tD <- table( t1dfr$diag )
> cbind( tD, round(tD/sum(tD)*100,1) )
     tD
21    5  0.3
22   22  1.4
26   17  1.1
28   32  2.0
33   84  5.2
51  129  8.0
70  208 12.9
82   69  4.3
83   47  2.9
84   35  2.2
91   12  0.7
92  101  6.3
101  46  2.9
103  28  1.7
113 180 11.2
121  49  3.0
131  40  2.5
132  50  3.1
133  10  0.6
139  94  5.8
241  16  1.0
242   8  0.5
243  22  1.4
249   6  0.4
251  28  1.7
999 273 16.9
```

Note that we shall not analyze the group 999 ("Other cancers"), which constitutes some 17% of all cancers, as this would not be biologically meaningful, so we construct the adjusted person-years for analysis of the category of "All cancers":

```
> PYadj <- PYtab
> PYadj[dimnames(PCtab)[[1]],,,,,] <-
+ PYtab[dimnames(PCtab)[[1]],,,,,] - PCtab
```

Thus we have the two tables of person-years classified by sex, age, calendar time, as well as by the two variables that classifies the T1D patients by the age-bracket of diagnosis and by whether they were diagnosed before or after 1.1.1995.

3.2.1.3 Cancer cases

We then set up the proper analysis table, with number of events and the unique values of `diag` in the dataset `t1dfr`; this is but a simple tabulation of the cancers occurring *after* diagnosis of diabetes from the data frame `t1dfr`:

```
> Dtab <- xtabs( !is.na(doca) ~ diag +
+                               floor(doca-dobt) +
+                               floor(doca) +
+                               I((doca>(dodm+1) ) +
+                                 (doca>(dodm+2) ) +
+                                 (doca>(dodm+5) )*3 +
+                                 (doca>(dodm+10))*5 +
+                                 (doca>(dodm+15))*5 ) +
+                               sex + t1d + dmprev,
```

```
+                data = subset( t1dfr, doca>1995 & doca>dodm ) )
> names( dimnames( Dtab) )[2:4] <-
+ names( dimnames(PYtab) )[1:3]
> dimnames( Dtab )
$diag
 [1] "21"  "22"  "26"  "28"  "33"  "51"  "70"  "82"  "83"  "84"  "91"  "92"  "101" "103"
[15] "113" "121" "131" "132" "133" "139" "241" "242" "243" "249" "251" "999"

$age
 [1] "6"  "7"  "10" "13" "15" "17" "18" "19" "20" "21" "22" "23" "24" "25" "26" "27" "28"
[18] "29" "30" "31" "32" "33" "34" "35" "36" "37" "38" "39" "40" "41" "42" "43" "44" "45"
[35] "46" "47" "48" "49" "50" "51" "52" "53" "54" "55" "56" "57" "58" "59" "60" "61" "62"
[52] "63" "64" "65" "68" "70"

$per
 [1] "1995" "1996" "1997" "1998" "1999" "2000" "2001" "2002" "2003" "2004" "2005" "2006"
[13] "2007" "2008" "2009" "2010" "2011" "2012"

$DMdur
[1] "0"  "1"  "2"  "5"  "10" "15"

$sex
[1] "M" "F"

$t1d
[1] "NoDM" "30"   "35"   "40"   "T2D"

$dmprev
[1] "Inc" "Prv" "Pop"
```

However we note that the age-dimension is a bit smaller in `Dtab`, than in the `PYtab`, so we make an amended version of `Dtab`, and make sure that all is filled with 0s where it should be:

```
> dnam <- dimnames( Dtab )
> dnam[[2]] <- dimnames( PYtab )[[1]]
> Atab <- NArray( dnam )
> str( Dtab )
 xtabs [1:26, 1:56, 1:18, 1:6, 1:2, 1:5, 1:3] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 7
  ..$ diag  : chr [1:26] "21" "22" "26" "28" ...
  ..$ age   : chr [1:56] "6" "7" "10" "13" ...
  ..$ per   : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ DMdur : chr [1:6] "0" "1" "2" "5" ...
  ..$ sex   : chr [1:2] "M" "F"
  ..$ t1d   : chr [1:5] "NoDM" "30" "35" "40" ...
  ..$ dmprev: chr [1:3] "Inc" "Prv" "Pop"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = !is.na(doca) ~ diag + floor(doca - dobt) + floor(doca)
> str( Atab )
 logi [1:26, 1:77, 1:18, 1:6, 1:2, 1:5, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 7
  ..$ diag  : chr [1:26] "21" "22" "26" "28" ...
  ..$ age   : chr [1:77] "0" "1" "2" "3" ...
  ..$ per   : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ DMdur : chr [1:6] "0" "1" "2" "5" ...
  ..$ sex   : chr [1:2] "M" "F"
  ..$ t1d   : chr [1:5] "NoDM" "30" "35" "40" ...
  ..$ dmprev: chr [1:3] "Inc" "Prv" "Pop"
> Atab[,dimnames(Dtab)[[2]],,,,,] <- Dtab
> Atab[is.na(Atab)] <- 0
```

3.2.1.4 All cancers

We also need a tabulation of "All cancers", that is the first of any of the known cancers, but this is precisely what we have in the dataset `t1fu`, where there is one record per T1D patient, and where the `doca` is the earliest of the cancer diagnosis dates:

```
> D0tab <- xtabs( !is.na(doca) ~ floor(doca-dobt) +
+                                floor(doca) +
+                                I((doca>(dodm+1) ) +
+                                  (doca>(dodm+2) ) +
+                                  (doca>(dodm+5) )*3 +
+                                  (doca>(dodm+10))*5 +
+                                  (doca>(dodm+15))*5 ) +
+                                sex + t1d + dmprev,
+                 data = subset( t1fu, doca>1995 & doca>dodm ) )
> names( dimnames(D0tab) )[1:3] <-
+ names( dimnames(PYtab) )[1:3]
```

Once this has been done, we construct the analysis data frame with two columns for person-years and a column with cases for each of the diagnosis groups.

3.2.2 The analysis data frame for T1D patients

So we start by setting up the frame with the person-years, for the analysis of all cancers (`Y0`) and for the specific cancers (`Y`), and the number of first primary cancers of any type (`D0`)

```
> Dfr <- as.data.frame( as.table(PYadj[,,,,2:4,1:2]), responseName="Y0" )
> nxt <- as.data.frame( as.table(PYtab[,,,,2:4,1:2]), responseName="Y"  )
> Dfr <- merge( Dfr, nxt, all=TRUE )
> nxt <- as.data.frame( as.table(D0tab[,,,,2:4,1:2]), responseName="d0" )
> Dfr <- merge( Dfr, nxt, all=TRUE )
> Dfr$d0[is.na(Dfr$d0)] <- 0
> head( Dfr )
  age  per DMdur sex t1d dmprev        Y0         Y d0
1   0 1995     0   M  30    Inc 0.8062971 0.8062971  0
2   0 1995     0   M  30    Prv 0.8596851 0.8596851  0
3   0 1995     0   M  35    Inc 0.0000000 0.0000000  0
4   0 1995     0   M  35    Prv 0.0000000 0.0000000  0
5   0 1995     0   M  40    Inc 0.0000000 0.0000000  0
6   0 1995     0   M  40    Prv 0.0000000 0.0000000  0
```

Then we attach the number of cases for the specific cancers column by column to get the final analysis dataset:

```
> for( i in dimnames(Atab)[[1]] )
+ {
+ nxt <- as.data.frame( as.table(Atab[i,,,,,2:4,1:2]),
+                       responseName=paste("d",i,sep="") )
+ Dfr <- merge( Dfr, nxt, all=TRUE )
+ }
> str( Dfr )
'data.frame':         99792 obs. of  35 variables:
 $ age   : Factor w/ 77 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ per   : Factor w/ 18 levels "1995","1996",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ DMdur : Factor w/ 6 levels "0","1","2","5",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ sex   : Factor w/ 2 levels "M","F": 2 2 2 2 2 2 1 1 1 1 ...
 $ t1d   : Factor w/ 3 levels "30","35","40": 1 1 2 2 3 3 1 1 2 2 ...
 $ dmprev: Factor w/ 2 levels "Inc","Prv": 1 2 1 2 1 2 1 2 1 2 ...
 $ Y0    : num  0.13 0 0 0 0 ...
 $ Y     : num  0.13 0 0 0 0 ...
 $ d0    : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
 $ d21   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d22   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d26   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d28   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d33   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d51   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d70   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d82   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d83   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d84   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d91   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d92   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d101  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d103  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d113  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d121  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d131  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d132  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d133  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d139  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d241  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d242  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d243  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d249  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d251  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d999  : num  0 0 0 0 0 0 0 0 0 0 ...
> apply( as.matrix( Dfr[,-(1:8)] ), 2, sum )
  d0  d21  d22  d26  d28  d33  d51  d70  d82  d83  d84  d91  d92 d101 d103 d113 d121 d131
1042    5   20   15   24   77   80  184   48   40   25   12   37   38   24   73   35   15
d132 d133 d139 d241 d242 d243 d249 d251 d999
  40    9   30   13    7   18    4   23  217
```

3.3 The population rates

From the orginal dataset with the entire cancer registry we can tabulate tumours in the different groups for the entire population, and use the corresponding population figures.

First we tabulate the cancer registry by diagnosis group:

```
> system.time(
+ Ctab <- xtabs( !is.na( doca ) ~ diag +
+                                   floor( doca-dobt ) +
+                                   floor( doca ) +
+                                   sex,
+                 data = subset( crg, doca>1995 & doca>dobt ) ) )
   user  system elapsed
   5.21    0.20    5.43
> str( Ctab )
 xtabs [1:26, 1:108, 1:18, 1:2] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 4
  ..$ diag             : chr [1:26] "21" "22" "26" "28" ...
  ..$ floor(doca - dobt): chr [1:108] "0" "1" "2" "3" ...
  ..$ floor(doca)       : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex              : chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = !is.na(doca) ~ diag + floor(doca - dobt) + floor(doca)
```

However, we also want to tabulate the the "all cancers" group, that is the occurrence of the first of all "real" cancers. To that end we create a dataset of the first cancers:

```
> ca.only <- subset( crg, !is.na( doca ) )
> system.time(
+ ca.first <- aggregate( x = ca.only[,c("dobt","doca")],
+                       by = ca.only[,c("id","sex")],
+                      FUN = min ) )
   user  system elapsed
  37.24    0.07   37.31
> dim( ca.first )
[1] 615153      4
> COtab <- xtabs( !is.na( doca ) ~ floor( doca-dobt ) +
+                                  floor( doca ) +
+                                  sex,
+                  data = subset( ca.first, doca>1995 & doca>dobt ) )
> str( COtab )
 xtabs [1:108, 1:18, 1:2] 6 11 7 13 9 14 4 6 1 3 ...
 - attr(*, "dimnames")=List of 3
  ..$ floor(doca - dobt): chr [1:108] "0" "1" "2" "3" ...
  ..$ floor(doca)       : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex               : chr [1:2] "M" "F"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = !is.na(doca) ~ floor(doca - dobt) + floor(doca) + sex,
```

The other part of the data needed is the poplation risk time, where we use the risk time of the entire population.

```
> data( Y.dk )
> str( Y.dk )
'data.frame':        16800 obs. of  6 variables:
 $ sex  : num  1 1 1 1 1 1 1 1 1 1 ...
 $ A    : num  0 0 1 1 2 2 3 3 4 4 ...
 $ P    : num  1971 1971 1971 1971 1971 ...
 $ C    : num  1971 1970 1970 1969 1969 ...
 $ Y    : num  19195 17944 17969 18165 18179 ...
 $ upper: num  0 1 0 1 0 1 0 1 0 1 ...
 - attr(*, "Contents")= chr "Population risk time in Denmark, in triangles of a Lexis diagram"
> Ptab <- xtabs( Y ~ A + P + sex, data= subset( Y.dk, P>1994 & P<2013 ) )
> str( Ptab )
 xtabs [1:100, 1:18, 1:2] 35899 35347 35031 34349 33308 ...
 - attr(*, "dimnames")=List of 3
  ..$ A  : chr [1:100] "0" "1" "2" "3" ...
  ..$ P  : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "1" "2"
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = Y ~ A + P + sex, data = subset(Y.dk, P > 1994 & P < 201
```

In order to align these two tables with each other and with the data from the T1D patients we change the dimnames and the names of these, and finally we restrict both tables to ages 0–84:

```
> dimnames( Ptab )[["sex"]] <- dimnames( Ctab )[["sex"]]
> names( dimnames(COtab) )[1:2] <-
+ names( dimnames(Ctab)  )[2:3] <-
+ names( dimnames(Ptab)  )[1:2] <- c("age","per")
> COtab <- COtab[ 1:85,,]
>  Ctab <-  Ctab[,1:85,,]
>  Ptab <-  Ptab[ 1:85,,]
> str( COtab )
 num [1:85, 1:18, 1:2] 6 11 7 13 9 14 4 6 1 3 ...
 - attr(*, "dimnames")=List of 3
  ..$ age: chr [1:85] "0" "1" "2" "3" ...
  ..$ per: chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
> str( Ctab )
```

```
 num [1:26, 1:85, 1:18, 1:2] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 4
  ..$ diag: chr [1:26] "21" "22" "26" "28" ...
  ..$ age : chr [1:85] "0" "1" "2" "3" ...
  ..$ per : chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex : chr [1:2] "M" "F"
> str( Ptab )
 num [1:85, 1:18, 1:2] 35899 35347 35031 34349 33308 ...
 - attr(*, "dimnames")=List of 3
  ..$ age: chr [1:85] "0" "1" "2" "3" ...
  ..$ per: chr [1:18] "1995" "1996" "1997" "1998" ...
  ..$ sex: chr [1:2] "M" "F"
```

With these tables in place we can set up the analysis dataset for the entire population with the same variables as for the T1D patients above, so that we eventually can stack these two.

Note that while we shall use different sets of person years for all cancers and for specific sites for the T1D patients (in columns `Y0` and `Y`, respectively), this is not the case for the total population rates, but we will need two separate columns of person-years (with identical numbers) in the dataset in order to stack it together with the T1D data for the final analysis dataset. Likewise we shall also need the columns `t1d`, `DMdur` and `dmprev`, however with a separate factor level, which is going to be used in the analysis of the rates:

```
> Pfr <- as.data.frame( as.table(Ptab), responseName="Y0" )
> Pfr$Y <- Pfr$Y0
> nxt <- as.data.frame( as.table(C0tab), responseName="d0" )
> Pfr <- merge( Pfr, nxt, all=TRUE )
> Pfr$t1d    <- factor("NoDM" )
> Pfr$dmprev <- factor( "Pop" )
> Pfr$DMdur  <- factor( "NoDM" )
> str( Pfr )
'data.frame':        3060 obs. of  9 variables:
 $ age   : Factor w/ 85 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ per   : Factor w/ 18 levels "1995","1996",..: 1 1 2 2 3 3 4 4 5 5 ...
 $ sex   : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
 $ Y0    : num  34126 35899 33539 35489 32936 ...
 $ Y     : num  34126 35899 33539 35489 32936 ...
 $ d0    : num  8 6 6 4 5 4 10 8 11 5 ...
 $ t1d   : Factor w/ 1 level "NoDM": 1 1 1 1 1 1 1 1 1 1 ...
 $ dmprev: Factor w/ 1 level "Pop": 1 1 1 1 1 1 1 1 1 1 ...
 $ DMdur : Factor w/ 1 level "NoDM": 1 1 1 1 1 1 1 1 1 1 ...
```

Then we attach the number of cases for the specific cancers column by column to get the final analysis dataset for the population:

```
> for( i in dimnames(Ctab)[[1]] )
+ {
+ nxt <- as.data.frame( as.table(Ctab[i,,,]),
+                       responseName=paste("d",i,sep="") )
+ Pfr <- merge( Pfr, nxt, all=TRUE )
+ }
> Afr <- rbind( Pfr, Dfr )
> str( Afr )
'data.frame':        102852 obs. of  35 variables:
 $ age   : Factor w/ 85 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ per   : Factor w/ 18 levels "1995","1996",..: 1 1 2 2 3 3 4 4 5 5 ...
 $ sex   : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
 $ Y0    : num  34126 35899 33539 35489 32936 ...
 $ Y     : num  34126 35899 33539 35489 32936 ...
 $ d0    : num  8 6 6 4 5 4 10 8 11 5 ...
 $ t1d   : Factor w/ 4 levels "NoDM","30","35",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ dmprev: Factor w/ 3 levels "Pop","Inc","Prv": 1 1 1 1 1 1 1 1 1 1 ...
 $ DMdur : Factor w/ 7 levels "NoDM","0","1",..: 1 1 1 1 1 1 1 1 1 1 ...
```

```
 $ d21   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d22   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d26   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d28   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d33   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d51   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d70   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d82   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d83   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d84   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d91   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d92   : num  0 0 0 1 0 0 0 0 0 0 ...
 $ d101  : num  0 2 0 1 0 1 3 0 1 1 ...
 $ d103  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d113  : num  1 2 0 2 1 0 1 0 3 2 ...
 $ d121  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d131  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d132  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d133  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d139  : num  2 0 3 0 0 2 1 0 1 0 ...
 $ d241  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d242  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d243  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d249  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d251  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d999  : num  5 2 3 0 4 1 5 8 6 2 ...
```

3.4 The collected analysis datset

We now have a dataset with events and person-years in the T1D population, classified by sex, age and period of follow-up, diagnosis age and prevalence at 1.1.1995, and of course T1D status (no/diag-age).

```
> print(
+ round( ftable( xtabs( cbind(D=d0,Y=Y0/1000) ~ t1d + dmprev, data=Afr ),
+                 col.vars=3:2 ), 1 ), zero.print="." )
                 D                           Y
     dmprev     Pop      Inc      Prv      Pop      Inc      Prv
t1d
NoDM        442530.0       .        .    95492.4       .        .
30                 .    113.0    163.0        .    137.9    125.4
35                 .    124.0    131.0        .     71.8     52.1
40                 .    255.0    256.0        .     94.7     62.6
> print(
+ round( ftable( xtabs( cbind(D=d0,Y=Y0/1000) ~ t1d + dmprev + DMdur, data=Afr ),
+                 col.vars=c(4,1) ), 1 ), zero.print="." )
                       D                                  Y
             t1d      NoDM       30       35       40     NoDM       30       35       40
dmprev DMdur
Pop    NoDM       442530.0        .        .        .  95492.4        .        .        .
       0                 .        .        .        .        .        .        .        .
       1                 .        .        .        .        .        .        .        .
       2                 .        .        .        .        .        .        .        .
       5                 .        .        .        .        .        .        .        .
       10                .        .        .        .        .        .        .        .
       15                .        .        .        .        .        .        .        .
Inc    NoDM              .        .        .        .        .        .        .        .
       0                 .      9.0      9.0     27.0        .     15.6      8.0     11.1
       1                 .      4.0      7.0     15.0        .     15.0      7.7     10.6
       2                 .     24.0     21.0     54.0        .     38.1     19.7     27.0
       5                 .     39.0     37.0     89.0        .     44.1     23.0     30.4
       10                .     29.0     45.0     57.0        .     21.8     11.6     13.5
```

```
        15            .         8.0      5.0     13.0      .       3.4     1.8     2.0
Prv     NoDM          .          .        .        .       .        .       .       .
        0             .          .       1.0       .       .       0.4     0.3     0.3
        1             .         1.0       .        .       .       1.9     1.1     1.4
        2             .         3.0      6.0     17.0      .      12.2     5.8     7.4
        5             .        29.0     19.0     43.0      .      34.5    15.0    18.3
        10            .        39.0     33.0     72.0      .      34.7    14.4    17.3
        15            .        91.0     72.0    124.0      .      41.7    15.6    17.8
```

However, we must render age and period continuous variables that take on the mean value of the variables in each observational unit.

```
> Afr <- transform( Afr, A = as.numeric(as.character(age))+0.5,
+                        P = as.numeric(as.character(per))+0.5 )
> with( Afr, table( A ) )
A
 0.5  1.5  2.5  3.5  4.5  5.5  6.5  7.5  8.5  9.5 10.5 11.5 12.5 13.5 14.5 15.5 16.5 17.5
1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332
18.5 19.5 20.5 21.5 22.5 23.5 24.5 25.5 26.5 27.5 28.5 29.5 30.5 31.5 32.5 33.5 34.5 35.5
1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332
36.5 37.5 38.5 39.5 40.5 41.5 42.5 43.5 44.5 45.5 46.5 47.5 48.5 49.5 50.5 51.5 52.5 53.5
1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332
54.5 55.5 56.5 57.5 58.5 59.5 60.5 61.5 62.5 63.5 64.5 65.5 66.5 67.5 68.5 69.5 70.5 71.5
1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332 1332
72.5 73.5 74.5 75.5 76.5 77.5 78.5 79.5 80.5 81.5 82.5 83.5 84.5
1332 1332 1332 1332 1332   36   36   36   36   36   36   36   36
> with( Afr, table( P ) )
P
1995.5 1996.5 1997.5 1998.5 1999.5 2000.5 2001.5 2002.5 2003.5 2004.5 2005.5 2006.5 2007.5
  5714   5714   5714   5714   5714   5714   5714   5714   5714   5714   5714   5714   5714
2008.5 2009.5 2010.5 2011.5 2012.5
  5714   5714   5714   5714   5714
```

The analysis dataset has been set up as a big table, but it is superfluous to include the units where there are no person-years anyway:

```
> with( Afr, ftable( D0=d0>0, Y0=Y0>0, Y=Y>0 ) )
            Y FALSE  TRUE
D0    Y0
FALSE FALSE   86975    61
      TRUE        0 11915
TRUE  FALSE       0     0
      TRUE        0  3901
> Afr <- subset( Afr, Y>0 )
> str( Afr )
'data.frame':        15877 obs. of  37 variables:
 $ age   : Factor w/ 85 levels "0","1","2","3",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ per   : Factor w/ 18 levels "1995","1996",..: 1 1 2 2 3 3 4 4 5 5 ...
 $ sex   : Factor w/ 2 levels "M","F": 2 1 2 1 2 1 2 1 2 1 ...
 $ Y0    : num  34126 35899 33539 35489 32936 ...
 $ Y     : num  34126 35899 33539 35489 32936 ...
 $ d0    : num  8 6 6 4 5 4 10 8 11 5 ...
 $ t1d   : Factor w/ 4 levels "NoDM","30","35",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ dmprev: Factor w/ 3 levels "Pop","Inc","Prv": 1 1 1 1 1 1 1 1 1 1 ...
 $ DMdur : Factor w/ 7 levels "NoDM","0","1",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ d21   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d22   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d26   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d28   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d33   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d51   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d70   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d82   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d83   : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
 $ d84   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d91   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d92   : num  0 0 0 1 0 0 0 0 0 0 ...
 $ d101  : num  0 2 0 1 0 1 3 0 1 1 ...
 $ d103  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d113  : num  1 2 0 2 1 0 1 0 3 2 ...
 $ d121  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d131  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d132  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d133  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d139  : num  2 0 3 0 0 2 1 0 1 0 ...
 $ d241  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d242  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d243  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d249  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d251  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d999  : num  5 2 3 0 4 1 5 8 6 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  1996 1996 1996 1996 1998 ...
```

3.5 Names and definition of sites

We do not have power to analyse the single subsites of colon, so we introduce two new sites: 24 — Colon (the sum of `d241`, `d242`, `d243` and `d249`), and also 251 — Colorectal, the sum of these *and* `d25`:

```
> Afr$d24  <- with( Afr, d241+d242+d243+d249 )
> Afr$d25  <- Afr$d251
> Afr$d251 <- with( Afr, d24+d25 )
> save( Afr, file="../data/Afrtmp.Rda" )
```

We now rename the diagnosis groups to comply with the NorCAN numbering. For annotation of the cancer sites in question, we get the names of the diagnostic groups:

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

```
> load( file="../data/Afrtmp.Rda" )
> dk.ana <- Afr[,c(3,36,37,7:9,4:6,10:34,38,39)]
> nn <- c("T1D","DMprev","DMdur","y0","y")
> cbind( names( dk.ana )[4:8],   nn )
                  nn
[1,] "t1d"    "T1D"
[2,] "dmprev" "DMprev"
[3,] "DMdur"  "DMdur"
[4,] "Y0"     "y0"
[5,] "Y"      "y"
>        names( dk.ana )[4:8] <- nn
> names( dk.ana )
 [1] "sex"    "A"      "P"      "T1D"    "DMprev" "DMdur"  "y0"     "y"      "d0"
[10] "d21"    "d22"    "d26"    "d28"    "d33"    "d51"    "d70"    "d82"    "d83"
[19] "d84"    "d91"    "d92"    "d101"   "d103"   "d113"   "d121"   "d131"   "d132"
[28] "d133"   "d139"   "d241"   "d242"   "d243"   "d249"   "d251"   "d24"    "d25"
```

Now compute which of the column names in `dk.ana` match a name in the `conv` data frame; these should be transferred to the final analysis dataset, and given names as the corresponding NorCAN names.

```
> ( wh <- match( conv$DKnam, names(dk.ana) ) )
 [1]  9 NA NA NA NA NA 10 11 NA 35 36 12 NA 13 NA NA 14 NA 16 17 18 19 NA 20 21 NA 22 23 15
[30] NA 24 25 NA NA 27 26 28 29 NA NA 34
```

```
> dk.ana <- dk.ana[,c(1:8,wh[!is.na(wh)])]
> names( dk.ana )[-(1:8)] <- conv$NCnam[!is.na(wh)]
```

Before we finally we save the data frame, we make a revision of the factor `DMdur`, since we are never going to use the duration information for the prevalent cases anyway.

```
> with( dk.ana, table( T1D, DMdur ) )
      DMdur
T1D    NoDM    0    1    2    5   10   15
  NoDM 3060    0    0    0    0    0    0
  30      0 1175 1177 1371 1565 1525 1598
  35      0  228  229  325  420  402  623
  40      0  228  228  324  420  396  583

> with( dk.ana, table( DMprev, DMdur ) )

      DMdur
DMprev NoDM    0    1    2    5   10   15
   Pop 3060    0    0    0    0    0    0
   Inc    0 1546 1463 1551 1371  821  277
   Prv    0   85  171  469 1034 1502 2527

> ndur <- nlevels(dk.ana$DMdur)
> xx <- factor( ifelse( dk.ana$DMprev=="Prv", ndur+1, as.integer(dk.ana$DMdur) ),
+               levels = 1:(ndur+1),
+               labels = c(levels(dk.ana$DMdur),"Unkn") )
> dk.ana$DMdur <- factor( xx )
> with( dk.ana, table( DMprev, DMdur ) )

      DMdur
DMprev NoDM    0    1    2    5   10   15 Unkn
   Pop 3060    0    0    0    0    0    0    0
   Inc    0 1546 1463 1551 1371  821  277    0
   Prv    0    0    0    0    0    0    0 5788

> with( dk.ana, table(    T1D, DMdur ) )

      DMdur
T1D    NoDM    0    1    2    5   10   15 Unkn
  NoDM 3060    0    0    0    0    0    0    0
  30      0 1114 1054 1050  890  540  192 3571
  35      0  216  205  251  241  141   43 1130
  40      0  216  204  250  240  140   42 1087

> tt <- xtabs( cbind(D=d0,Y=y/1000) ~ DMdur, data=dk.ana )
> tt <- tt[c(1:8,8),]
> tt[8,] <- apply( tt[2:7,], 2, sum )
> rownames(tt)[8] <- "All dur"
> cbind( Cancers=round( tt[,1], 0 ), "PY(1000)"=round( tt[,2], 1 ) )

        Cancers PY(1000)
NoDM     442530  95492.4
0            45     35.0
1            26     33.6
2            99     85.8
5           165     98.8
10          131     47.9
15           26      7.4
All dur     492    308.6
Unkn        550    244.3
```

Finally we create the country-code and save the data:

```
> dk.ana$Cnt <- "DNK"
> save( dk.ana, file="../data/DKana.Rda" )
```

g

3.6 Analysis of SMR

We load the dataset and the dataframe with the tumour labels in it:

```
> load( file="../data/DKana.Rda" )
> load( file="../data/conv.Rda" )
> dk.ana <- subset( dk.ana, y0>0 )
> dk.ana$T1 <- Relevel( dk.ana$T1D, list( NoDM=1, T1D=2:4 ) )
```

3.6.1 All cancers

We first set out the simplest possible analysis with age-period-cohort effects for the baseline rates. First we devise a couple of knots for the splines:

```
> library( splines )
> ( a.kn <- seq(5,80,,7) )
[1]  5.0 17.5 30.0 42.5 55.0 67.5 80.0
> ( p.kn <- seq(1996,2009,,4) )
[1] 1996.000 2000.333 2004.667 2009.000
> ( c.kn <- seq(1920,1990,,6) )
[1] 1920 1934 1948 1962 1976 1990
```

We then fit 3 models for the RR for T1D patients relative to the general population, using a common shape of the underlying cancer incidence rates as an age-period cohort model with $1 + (7-1) + (4-1) + (6-1) - 1 = 14$ parameters[2]. The first model (`m3`) is one with separate effects for persons diagnosed in ages < 30, 30–35 and 35–40, the second (`m1`) is a simplification where the three groups are pooled, and the third (`mc`) is an extension of `m3` where each group is subdivided by diabetes status at 1.1.1995:

```
> m3 <- glm( d0 ~ Ns(   A, knots=a.kn ) +
+                 Ns( P   , knots=p.kn ) +
+                 Ns( P-A, knots=c.kn ) +
+                 T1D,
+            offset = log(y0),
+            family = poisson,
+              data = dk.ana )
> m1 <- update( m3, . ~ . - T1D + T1 )
> mc <- update( m3, . ~ . - T1D + factor(interaction(T1D,DMprev)) )
> anova( mc, m3, m1, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    factor(interaction(T1D, DMprev))
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
Model 3: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     15796      26605
2     15799      26607 -3  -1.9505   0.5827
3     15801      26609 -2  -1.8206   0.4024
```

The tests of the models show that there is no substantial difference between them, and that does not either shown up if we list the estimated RRs from the models:

[2]There is first an intercept, then the three natural splines, where k knots gives $k-1$ parameters, and finally 1 aliased parameter from the linear relationship between $P-A$ and P and A.

```
> round( rbind( ci.exp( m3, subset="T1" ),
+               ci.exp( m1, subset="T1" ),
+               ci.exp( mc, subset="T1" ) ), 3 )
                                     exp(Est.)  2.5% 97.5%
T1D30                                    1.103 0.980 1.241
T1D35                                    0.983 0.869 1.111
T1D40                                    1.057 0.969 1.153
T1T1D                                    1.049 0.987 1.115
factor(interaction(T1D, DMprev))30.Inc   1.155 0.960 1.390
factor(interaction(T1D, DMprev))35.Inc   1.034 0.867 1.234
factor(interaction(T1D, DMprev))40.Inc   1.105 0.977 1.250
factor(interaction(T1D, DMprev))30.Prv   1.070 0.917 1.248
factor(interaction(T1D, DMprev))35.Prv   0.939 0.791 1.114
factor(interaction(T1D, DMprev))40.Prv   1.013 0.896 1.145
```

In summary, the comparison of the models show that there is no heterogeneity between the six groups of T1D patients w.r.t. the occurrence of cancer; which is also evident in figure 7.1, where the RR estimates from the three models are shown together:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind( ci.exp( m3, subset="T1"),
+                 ci.exp( mc, subset="T1" ),
+                 ci.exp( m1, subset="T1") ),
+          y = c(4,3,2,c(4,3,2)-0.2,c(4,3,2)-0.4,0.5),
+          txt = c("","","",
+                  "<30","30-35","35-40",
+                  "","","","Pooled"),
+          xlog=TRUE, xtic=c(5:10,15,20)/10, grid=TRUE,
+          xlab="RR of any cancer, T1D vs. population",
+          lwd=4, cex=2, vref=1, col=rep(gray(c(5,8,0)/10),c(3,6,1)) )
```

3.6.2 Analysis by duration

We have also devised a variable indication different time bands after diagnosis:

```
> round(
+ ftable( xtabs( cbind(d0,y0/1000) ~ DMdur + DMprev, data=dk.ana ),
+         col.vars=3:2 ), 1 )
                    d0                           V2
      DMprev       Pop      Inc      Prv        Pop      Inc      Prv
DMdur
NoDM          442530.0      0.0      0.0    95492.4      0.0      0.0
0                  0.0     45.0      0.0        0.0     34.6      0.0
1                  0.0     26.0      0.0        0.0     33.3      0.0
2                  0.0     99.0      0.0        0.0     84.9      0.0
5                  0.0    165.0      0.0        0.0     97.5      0.0
10                 0.0    131.0      0.0        0.0     46.9      0.0
15                 0.0     26.0      0.0        0.0      7.2      0.0
Unkn               0.0      0.0    550.0        0.0      0.0    240.1
```

The above analyses showed virtually no difference between the groups of patients by age at inclusion, so we pool these groups and restrict the analysis to DM patients diagnosed after 1995. Also we fit models with smaller datasets corresponding to more restrictive definitions of T1D:

```
> levels( dk.ana$T1D )
[1] "NoDM" "30"   "35"   "40"
> levels( dk.ana$DMprev )
[1] "Pop" "Inc" "Prv"
```

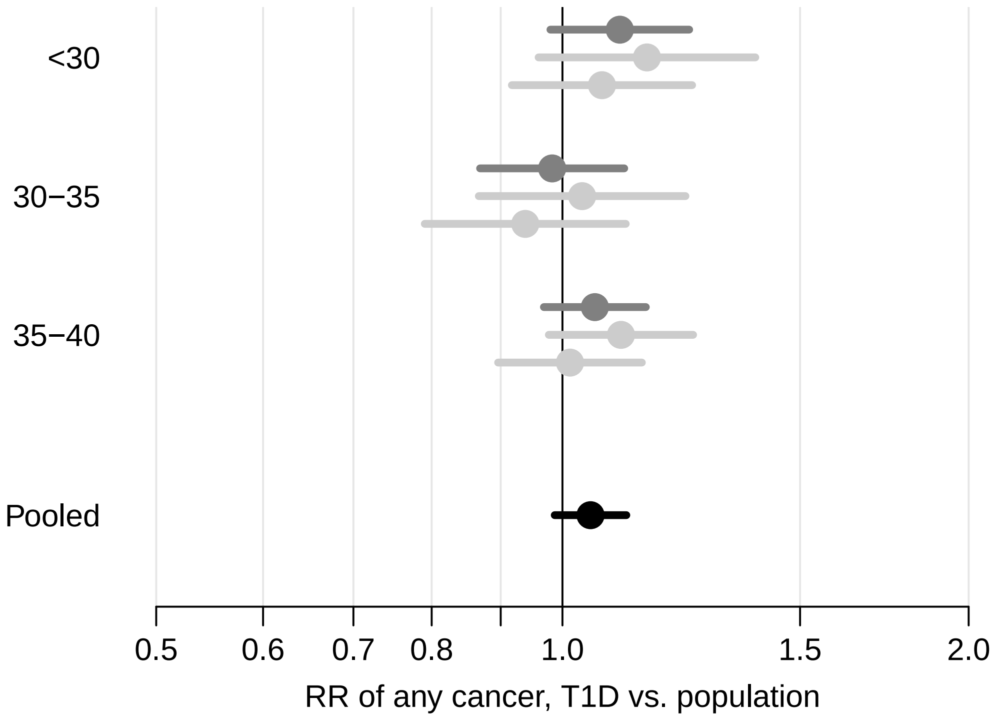

Figure 3.3: *Estimated RRs relative to the general population, the different shades of gray correspond to the different models. The upper of the light gray bars are for T1D patients diagnosed after 1995, the lower for those diagnosed before 1.1.1995. There are no significant differences anywhere.*

```
> md40 <- update( m3, . ~ . - T1D + DMdur,
+                    data=subset( dk.ana, DMprev!="Prv" & y0>0 ) )
> md35 <- update( md40, data=subset( dk.ana, DMprev!="Prv" & y0>0 & T1D %in% levels(T1D)[1:3] ) )
> md30 <- update( md40, data=subset( dk.ana, DMprev!="Prv" & y0>0 & T1D %in% levels(T1D)[1:2] ) )
> anova( md40, update( md40, . ~ . - DMdur + DMprev ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMprev
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     10069      24960
2     10074      24967 -5  -7.5658   0.1818
> anova( md35, update( md35, . ~ . - DMdur + DMprev ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMprev
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      8977      24290
2      8982      24295 -5  -5.6827   0.3383
> anova( md30, update( md30, . ~ . - DMdur + DMprev ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
```

```
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMprev
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      7880      23826
2      7885      23829 -5  -2.9219    0.712
```

We see that there are formally no effect of duration, but this might well be because of the many degrees of freedom, for the most liberal definition of T1D (and hence also that which likely includes the the largest number of T2D patients), the χ^2 is over 7, hence significant even it were on 2 d.f.

```
> round( cbind( RR40 <- ci.exp( md40, subset="DMdur" ),
+               RR35 <- ci.exp( md35, subset="DMdur" ),
+               RR30 <- ci.exp( md30, subset="DMdur" ) ), 2 )
        exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
DMdur0       1.46 1.09  1.96      1.27 0.80  2.01      1.46 0.76  2.80
DMdur1       0.81 0.55  1.18      0.73 0.41  1.33      0.61 0.23  1.63
DMdur2       1.02 0.84  1.24      0.98 0.74  1.32      1.19 0.80  1.78
DMdur5       1.07 0.92  1.25      1.02 0.81  1.28      1.15 0.84  1.58
DMdur10      1.19 1.00  1.41      1.32 1.05  1.66      1.13 0.78  1.62
DMdur15      1.10 0.75  1.62      1.07 0.62  1.84      1.46 0.73  2.92
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind(NA,RR40),
+          txt=c("Years since DM","0-1","1-2","2-5","5-10","10-15","15+"),
+          xlog=TRUE, xtic=c(3:10,15,20,25,30)/10, grid=TRUE,
+          xlab="RR of any cancer, T1D vs. population",
+          lwd=4, cex=2, vref=1, y=c(6.7,6:1)+0.2 )
> linesEst( rbind(NA,RR35),
+           lwd=3, cex=1.5, col=gray(0.5), y=7:1 )
> linesEst( rbind(NA,RR30),
+           lwd=3, cex=1.5, col=gray(0.7), y=7:1-0.15 )
```

From figure 7.2 it seems that there is an ascertainment effect for T1DM as well as what have been shown for all diabetes under one. However this is clearly attenuated if a stricter definition of T1D is applied, so we may conjecture that any ascertainment bias is if not confined to, the at least more pronounced among T2D patients.

3.6.3 Site-specific analyses

3.6.3.1 Analyses of site specific cancers

We first set up an array to hold the resulting RRs for each of the sites that we analyse, we do the analyses by sex, but also make a pooled analysis, except for the sex-specific cancers (including breast):

```
> ( vnam <- names(dk.ana)[9:32] )
 [1] "d0"  "d6"  "d7"  "d9"  "d10" "d11" "d13" "d16" "d18" "d19" "d20" "d22"
[13] "d24" "d25" "d27" "d28" "d29" "d32" "d33" "d36" "d37" "d38" "d40" "d52"
> site <- conv[match(vnam,conv$NCnam),"Clab"]
> RRtab <- NArray( list( site = site,
+                        sex = c(levels(dk.ana$sex),"Both"),
+                       what = c("N.pop","N.T1","RR","lo","hi") ) )
> dimnames( RRtab )
$site
 [1] "All sites"            "Oesophagus"            "Stomach"
 [4] "Colon"                "Rectum"                "Liver"
 [7] "Pancreas"             "Lung"                  "Breast"
[10] "Cervix uteri"         "Corpus uteri"          "Ovary"
[13] "Prostate"             "Testis"                "Kidney"
[16] "Bladder"              "Melanoma of skin"      "Brain, CNS"
```

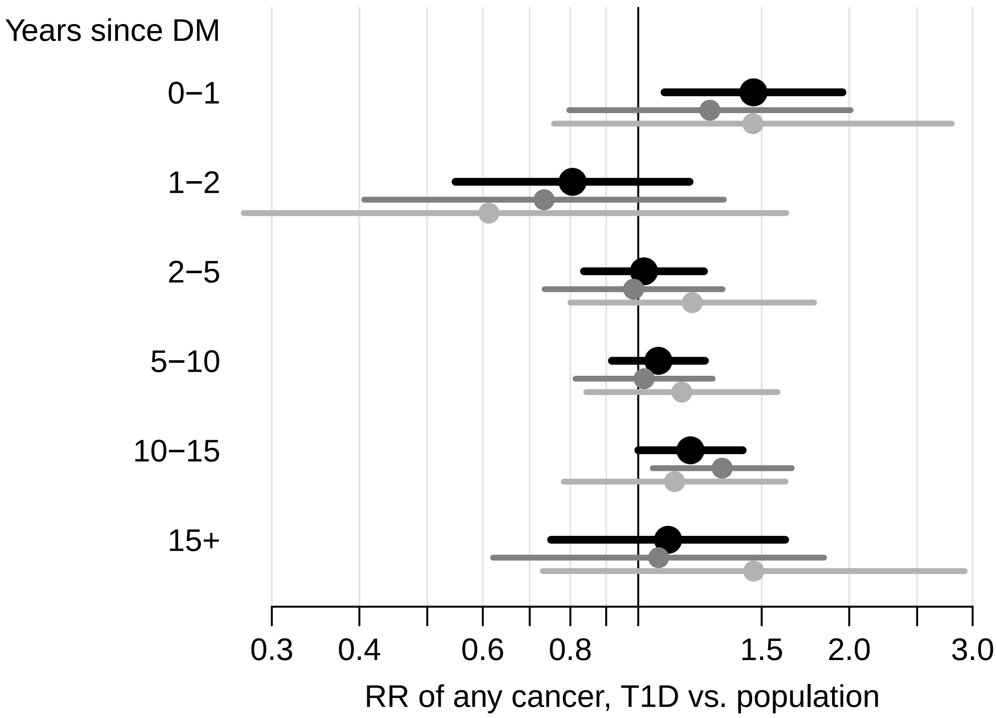

Figure 3.4: *The effect of duration on the RR of cancer. The two gray sets of effecs are from the models where data are further restricted to patients diagnosed under 35 and 30, respectively.*

```
[19] "Thyroid"              "Non-Hodgkin lymphoma" "Hodgkin lymphoma"
[22] "Multiple myeloma"     "Leukaemia"            "Colorectal"

$sex
[1] "M"    "F"    "Both"

$what
[1] "N.pop" "N.T1"  "RR"    "lo"    "hi"
```

With this fixed we can the make a loop doing the analysis for all sites:

```
> system.time(
+ for( i in 1:length(vnam) )
+    {
+    aset <- dk.ana[,c(vnam[i],
+                      paste("y",if(i==1) "0", sep=""),
+                      "A","P","T1","sex")]
+    names( aset )[1:2] <- c("D","Y")
+    mB <- glm( D ~ Ns   ( A, knots=a.kn ) +
+                   Ns( P   , knots=p.kn ) +
+                   Ns( P-A, knots=c.kn ) +
+                   T1,
+            offset = log(Y),
+            family = poisson,
+              data = aset )
+    mM <- update( mB, data=subset(aset,sex=="M") )
+    mF <- update( mB, data=subset(aset,sex=="F") )
+    RRtab[i,"M"    ,3:5] <- ci.exp( mM, subset="T1" )
+    RRtab[i,"F"    ,3:5] <- ci.exp( mF, subset="T1" )
+    RRtab[i,"Both",3:5] <- ci.exp( mB, subset="T1" )
+    RRtab[i,,1:2] <- addmargins( with( aset, tapply(D,list(sex,T1),sum) ), 1 )
+    } )
```

```
   user  system elapsed
  32.37    0.37   32.76
> RRorg <- RRtab
> RRtab <- RRorg
> for(i in 1:dim(RRtab)[1])
+ for(j in 1:dim(RRtab)[2]) if(          RRtab[i,j,5]==Inf   ) RRtab[i, j    ,] <- NA
> for(i in 1:dim(RRtab)[1]) if(any(is.na(RRtab[i, ,5]==Inf))) RRtab[i,"Both",] <- NA
> round( ftable( RRtab ), 2 )
                          what     N.pop      N.T1         RR         lo         hi
site                     sex
All sites                M     222404.00    401.00       1.10       1.00       1.21
                         F     220126.00    641.00       1.00       0.92       1.08
                         Both  442530.00   1042.00       1.05       0.99       1.12
Oesophagus               M       5039.00      4.00       0.58       0.22       1.55
                         F       1826.00      1.00       0.42       0.06       3.00
                         Both    6865.00      5.00       0.53       0.22       1.28
Stomach                  M       5833.00      6.00       0.72       0.32       1.61
                         F       2986.00     14.00       2.65       1.55       4.52
                         Both    8819.00     20.00       1.46       0.94       2.27
Colon                    M      20463.00     26.00       1.29       0.88       1.91
                         F      20769.00     16.00       0.73       0.44       1.19
                         Both   41232.00     42.00       1.00       0.74       1.35
Rectum                   M      12585.00     13.00       0.90       0.52       1.55
                         F       8224.00     10.00       0.82       0.44       1.52
                         Both   20809.00     23.00       0.86       0.57       1.29
Liver                    M       3371.00      9.00       2.07       1.07       4.00
                         F       1596.00      6.00       2.48       1.10       5.60
                         Both    4967.00     15.00       2.20       1.32       3.67
Pancreas                 M       6856.00     15.00       1.87       1.12       3.12
                         F       6576.00      9.00       1.37       0.71       2.66
                         Both   13432.00     24.00       1.64       1.09       2.45
Lung                     M      37120.00     39.00       1.24       0.91       1.71
                         F      30664.00     38.00       0.98       0.71       1.35
                         Both   67784.00     77.00       1.10       0.88       1.38
Breast                   M            NA        NA         NA         NA         NA
                         F      70345.00    184.00       0.74       0.64       0.86
                         Both         NA        NA         NA         NA         NA
Cervix uteri             M            NA        NA         NA         NA         NA
                         F       7098.00     48.00       0.85       0.64       1.13
                         Both         NA        NA         NA         NA         NA
Corpus uteri             M            NA        NA         NA         NA         NA
                         F      11658.00     40.00       2.24       1.64       3.07
                         Both         NA        NA         NA         NA         NA
Ovary                    M            NA        NA         NA         NA         NA
                         F      10067.00     25.00       1.01       0.68       1.50
                         Both         NA        NA         NA         NA         NA
Prostate                 M      50154.00     12.00       0.53       0.30       0.93
                         F            NA        NA         NA         NA         NA
                         Both         NA        NA         NA         NA         NA
Testis                   M       5210.00     37.00       0.78       0.56       1.07
                         F            NA        NA         NA         NA         NA
                         Both         NA        NA         NA         NA         NA
Kidney                   M       6371.00     23.00       1.66       1.10       2.51
                         F       3674.00     15.00       2.27       1.36       3.80
                         Both   10045.00     38.00       1.83       1.33       2.52
Bladder                  M      20887.00     17.00       0.93       0.57       1.50
                         F       7051.00      7.00       0.98       0.46       2.06
                         Both   27938.00     24.00       0.93       0.62       1.38
Melanoma of skin         M      10532.00     21.00       0.50       0.32       0.77
                         F      12539.00     59.00       0.69       0.54       0.90
                         Both   23071.00     80.00       0.65       0.52       0.80
Brain, CNS               M       7005.00     31.00       1.15       0.81       1.64
                         F       6017.00     42.00       1.49       1.09       2.02
                         Both   13022.00     73.00       1.32       1.05       1.67
Thyroid                  M        879.00      6.00       1.16       0.52       2.59
                         F       2239.00     29.00       1.63       1.13       2.35
```

```
                   Both        3118.00      35.00      1.59      1.13      2.22
Non-Hodgkin lymphoma M         7976.00      26.00      1.32      0.90      1.95
                   F           6485.00      14.00      0.93      0.55      1.57
                   Both       14461.00      40.00      1.14      0.83      1.56
Hodgkin lymphoma   M           1331.00      11.00      1.42      0.78      2.58
                   F            968.00       4.00      0.58      0.22      1.56
                   Both        2299.00      15.00      1.02      0.61      1.70
Multiple myeloma   M           3571.00       4.00      1.00      0.37      2.68
                   F           2608.00       5.00      1.51      0.62      3.66
                   Both        6179.00       9.00      1.23      0.64      2.37
Leukaemia          M           7612.00      14.00      1.03      0.61      1.74
                   F           5399.00      16.00      1.43      0.87      2.34
                   Both       13011.00      30.00      1.20      0.83      1.72
Colorectal         M          33048.00      39.00      1.13      0.82      1.55
                   F          28993.00      26.00      0.76      0.51      1.12
                   Both       62041.00      65.00      0.94      0.74      1.20
```

Of course we would also like to see the results as a forest plot, so we extract the relevant quantities for doing this:

```
> eM <- RRtab[,"M",3:5]
> eF <- RRtab[,"F",3:5]
> eB <- RRtab[,"Both",3:5]
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( eB, y=nrow(eM):1, txtpos=nrow(eM):1,
+          col="lightgray", xlog=TRUE,
+          xtic=c(1:10/10,1.5,2:7), xlim=c(0.095,7),
+          grid=TRUE, vref=1, xlab="Cancer incidence RR, T1D vs. population" )
> linesEst( eF, y=nrow(eM):1-0.2, col="red" )
> linesEst( eM, y=nrow(eM):1+0.2, col="blue" )
> text( rep(0.095,dim(RRtab)[1]), dim(RRtab)[1]:1+0.2, RRtab[,"M",2], col="blue", adj=1, cex=0.7 )
> text( rep(0.095,dim(RRtab)[1]), dim(RRtab)[1]:1-0.2, RRtab[,"F",2], col="red" , adj=1, cex=0.7 )
```

We see that the only sites with appreaciable increased RR and sufficiently narrow confidence intervals are liver, pancreas, corpus uteri and kidney; whereas there seems to a lower risk for melanoma, breast and testis cancer among T1D patients.

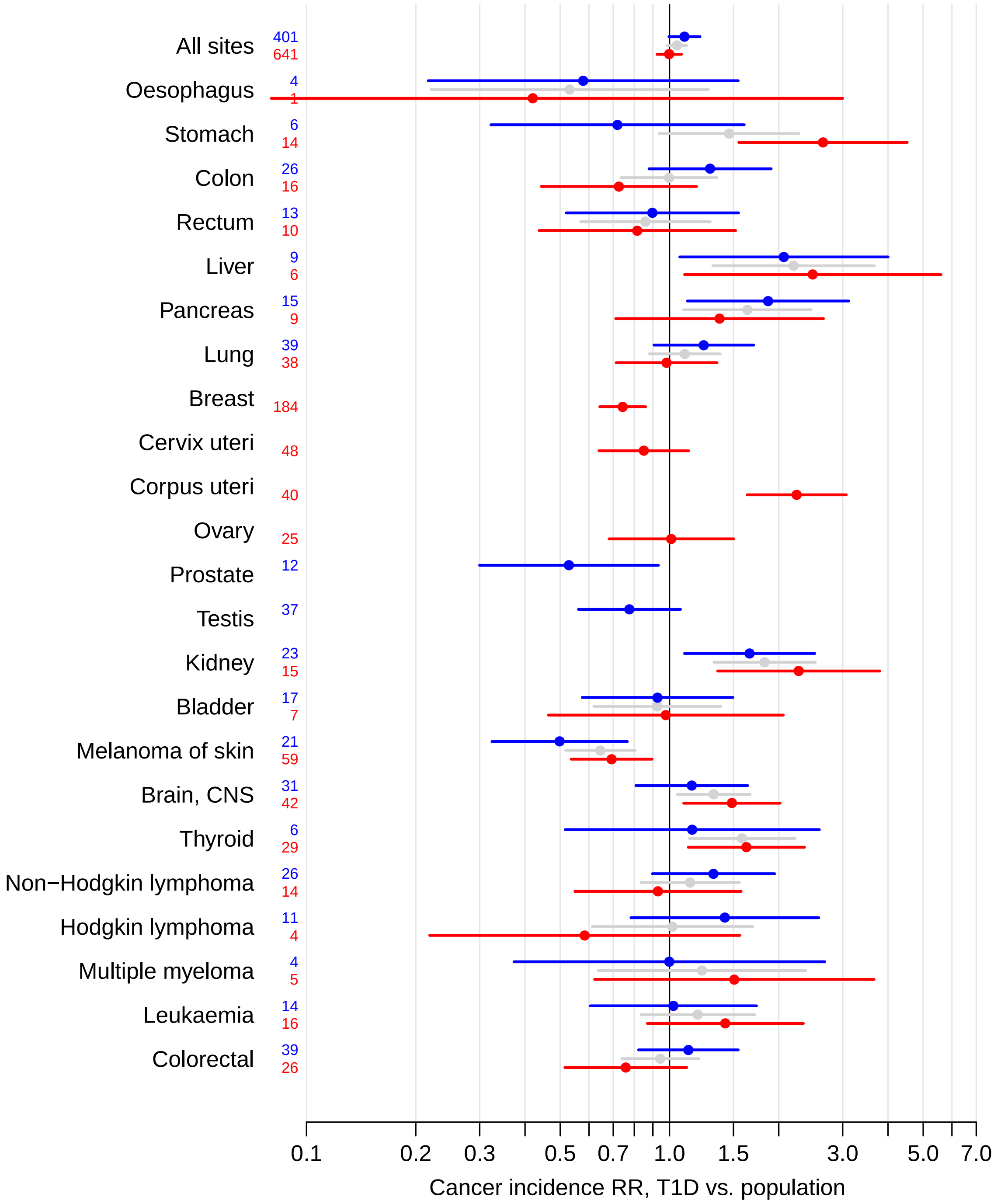

Figure 3.5: *RRs of cancer incidence among T1D patients (i.e. diagnosed < 40 years of age) in Denmark relative to the general population. The numbers to the left are the number of cancers observed among the T1D patients. Men: Blue, Women: Red, Both sexes: Light gray.*

Chapter 4

Swedish data

4.1 Reading T1D data

We have the follow-up of some 93,000 diabetes patients in a file:

```
> library( Epi )
> clear()
> SE <- read.csv("../data/diabetesalla.csv",header=T,as.is=TRUE,na.strings="")
> ( names( SE ) <- tolower( names(SE) ) )
 [1] "lpnr"            "sant_diadat"     "sant_foddat"     "diaalder"        "diaalder39"
 [6] "diaalder34"      "diaalder29"      "dodsdat_d"       "canc_alder"      "icd7"
[11] "candat"          "lip"             "oralcavity_all"  "tongue"          "mouth"
[16] "salivary"        "pharynx"         "esophagus"       "stomach"         "smallintest"
[21] "colon"           "colorect"        "rectal"          "liver"           "gall"
[26] "pancreas"        "nose"            "larynx"          "lung"            "pleura"
[31] "bone"            "melanoma"        "softtissue"      "breast"          "othfemgen"
[36] "cervix"          "corpus"          "othuterus"       "ovary"           "penis"
[41] "prostate"        "testis"          "kidney"          "bladder"         "eye"
[46] "brain"           "thyroid"         "hodgkin"         "nonhodgkin"      "multmyelom"
[51] "leukemia"        "othleukemia"     "acuteleuk"       "allcancer"       "slutuppf_2"
[56] "f_traff"         "persondagar"     "kon_ny"
> str(SE[,1:14])
'data.frame':       93402 obs. of  14 variables:
 $ lpnr          : int  24 31 34 50 64 66 90 103 114 132 ...
 $ sant_diadat   : chr  "1995-10-29" "1999-05-10" "1978-05-22" "1971-10-12" ...
 $ sant_foddat   : chr  "1959-10-29" "1960-05-10" "1949-05-22" "1964-10-12" ...
 $ diaalder      : int  36 39 29 7 34 29 36 11 37 36 ...
 $ diaalder39    : int  1 1 1 1 1 1 1 1 1 1 ...
 $ diaalder34    : int  0 0 1 1 1 1 0 1 0 0 ...
 $ diaalder29    : int  0 0 1 1 0 1 0 1 0 0 ...
 $ dodsdat_d     : chr  NA NA "1984-10-11" NA ...
 $ canc_alder    : int  NA NA NA NA NA NA NA NA NA NA ...
 $ icd7          : int  NA NA NA NA NA NA NA NA NA NA ...
 $ candat        : chr  NA NA NA NA ...
 $ lip           : int  NA NA NA NA NA NA NA NA NA NA ...
 $ oralcavity_all: int  NA NA NA NA NA NA NA NA NA NA ...
 $ tongue        : int  NA NA NA NA NA NA NA NA NA NA ...
> ( wh <- union( grep( "dat", names(SE) ),
+                grep( "f_" , names(SE) ) ) )
[1]  2  3  8 11 55 56
> for( i in wh ) SE[,i] <- cal.yr( as.Date(SE[,i]) )
> head(SE[,wh])
  sant_diadat sant_foddat dodsdat_d candat slutuppf_2  f_traff
1    1995.823    1959.823        NA     NA   2011.996 2011.996
2    1999.352    1960.355        NA     NA   2011.996 2011.996
3    1978.386    1949.387  1984.776     NA   2011.996 1984.776
4    1971.777    1964.779        NA     NA   2011.996 2011.996
5    2010.537    1976.538        NA     NA   2011.996 2011.996
6    2004.305    1975.303        NA     NA   2011.996 2011.996
```

```
> cbind( names(SE)[wh], nnam<-c("dodm","dob","dodd","doca","dend","dox") )
     [,1]          [,2]
[1,] "sant_diadat" "dodm"
[2,] "sant_foddat" "dob"
[3,] "dodsdat_d"   "dodd"
[4,] "candat"      "doca"
[5,] "slutuppf_2"  "dend"
[6,] "f_traff"     "dox"
>        names(SE)[wh]<-nnam
> head(SE[,wh])
      dodm      dob     dodd doca     dend      dox
1 1995.823 1959.823       NA   NA 2011.996 2011.996
2 1999.352 1960.355       NA   NA 2011.996 2011.996
3 1978.386 1949.387 1984.776   NA 2011.996 1984.776
4 1971.777 1964.779       NA   NA 2011.996 2011.996
5 2010.537 1976.538       NA   NA 2011.996 2011.996
6 2004.305 1975.303       NA   NA 2011.996 2011.996
```

Now we have one record per person with T1 diabetes with all the relevant dates.

However, the date of birth is artifically created from date of DM diagnosis and age at diagnosis; hence to avoid strange “bumps” in data we create an artficial date of birth. Since the date of birth is made by subtracting the persons integer age at date of diagnosis, the “real” date of birth is somewhere between 0 and 1 year earlier (because the exact age of a, say, 57 year old is somewhere between 57 and $58-\epsilon$):

```
> SE$dob <- SE$dob - runif(nrow(SE))
```

This revised dataset is now going to be used for:

1. Tabulation of cancer cases by type of cancer, sex, diabetes duration, age at diabetes diagnosis, age and date of diagnosis.

2. Tabulation of follow-up time by sex, diabetes duration, age at diabetes diagnosis, age and date of diagnosis.

4.1.1 Checking of data

```
> pairs( SE[!is.na(SE$doca),wh[-5]], gap=0,
+        panel=function(x,y) {points(x,y,pch=16,cex=0.5) ; abline(0,1,col="red")} )
```

```
> pairs( SE[sample(1:nrow(SE),5000),wh[-5]], gap=0,
+        panel=function(x,y) {points(x,y,pch=16,cex=0.5) ; abline(0,1,col="red")} )
```

Also we would like to see the distribution of dates of diagnosis of diabetes and cancer as well as the ages of these:

```
> par( mfrow=c(2,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> with( SE, hist(pmax(dodm,1940),breaks=1940:2012,col="black",
+              main="Date of DM diag") )
> with( SE, hist(dodm-dob,breaks=-1:41,col="black",
+              main="Age at DM diag") )
> with( SE, hist(doca,breaks=1960:2012,col="black",
+              main="Date of Cancer diag") )
> with( SE, hist(doca-dob,breaks=0:100,col="black",
+              main="Age at Cancer diag") )
```

From figure 8.1 we see there is a strange excess of patients diagnosed in 2010.

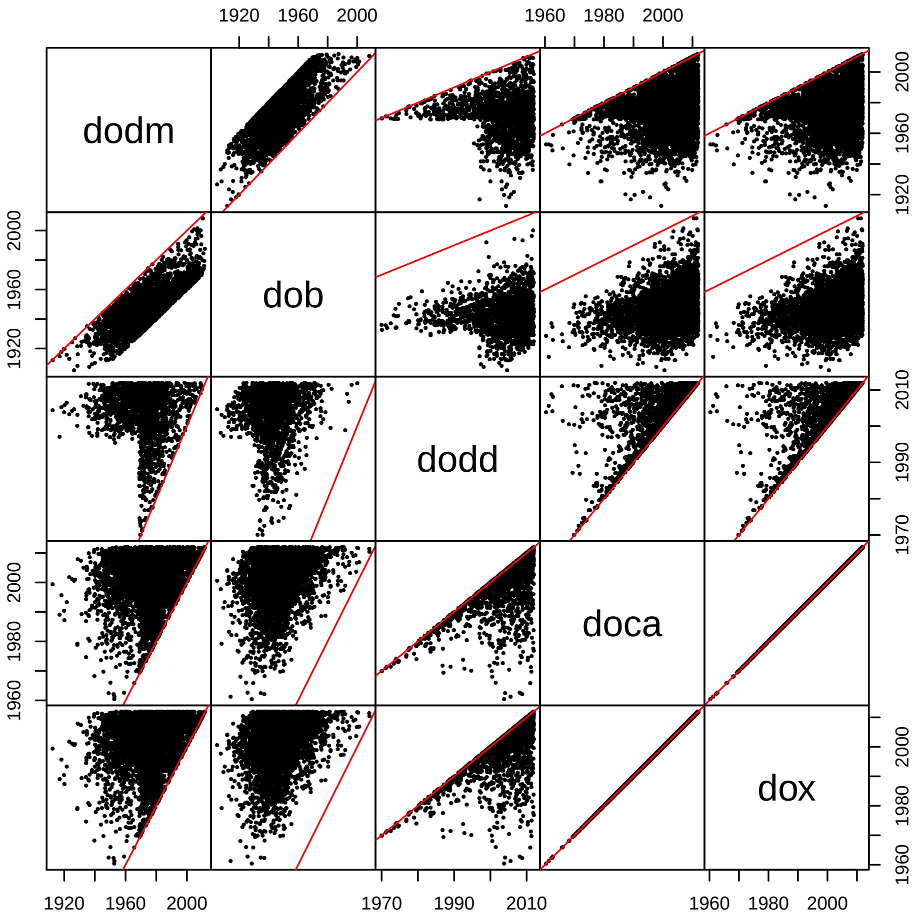

Figure 4.1: *Pairwise plots for all diabetes patients with a cancer diagnosis. The red lines are identity lines, so all points should be on one side of this line.*

4.1.2 Tumour coding

We shall rename the variables in the dataset so that they correspond to the numbering used in NORDCAN, so we align the names of the Swedish data with those in the conversion data frame, and the rename the columns in the Swedish data correspondingly (after checking that the alignment is correct):

```
> load("../data/conv.Rda")
> names(SE)
 [1] "lpnr"          "dodm"          "dob"            "diaalder"      "diaalder39"
 [6] "diaalder34"    "diaalder29"    "dodd"           "canc_alder"    "icd7"
[11] "doca"          "lip"           "oralcavity_all" "tongue"        "mouth"
[16] "salivary"      "pharynx"       "esophagus"      "stomach"       "smallintest"
[21] "colon"         "colorect"      "rectal"         "liver"         "gall"
[26] "pancreas"      "nose"          "larynx"         "lung"          "pleura"
[31] "bone"          "melanoma"      "softtissue"     "breast"        "othfemgen"
[36] "cervix"        "corpus"        "othuterus"      "ovary"         "penis"
[41] "prostate"      "testis"        "kidney"         "bladder"       "eye"
[46] "brain"         "thyroid"       "hodgkin"        "nonhodgkin"    "multmyelom"
[51] "leukemia"      "othleukemia"   "acuteleuk"      "allcancer"     "dend"
[56] "dox"           "persondagar"   "kon_ny"
> wh <- c(54,12,14,16,15,17,18,19,20,21,23:30,
+          34,36,37,39,38,41,42,40,43,44,32,
```

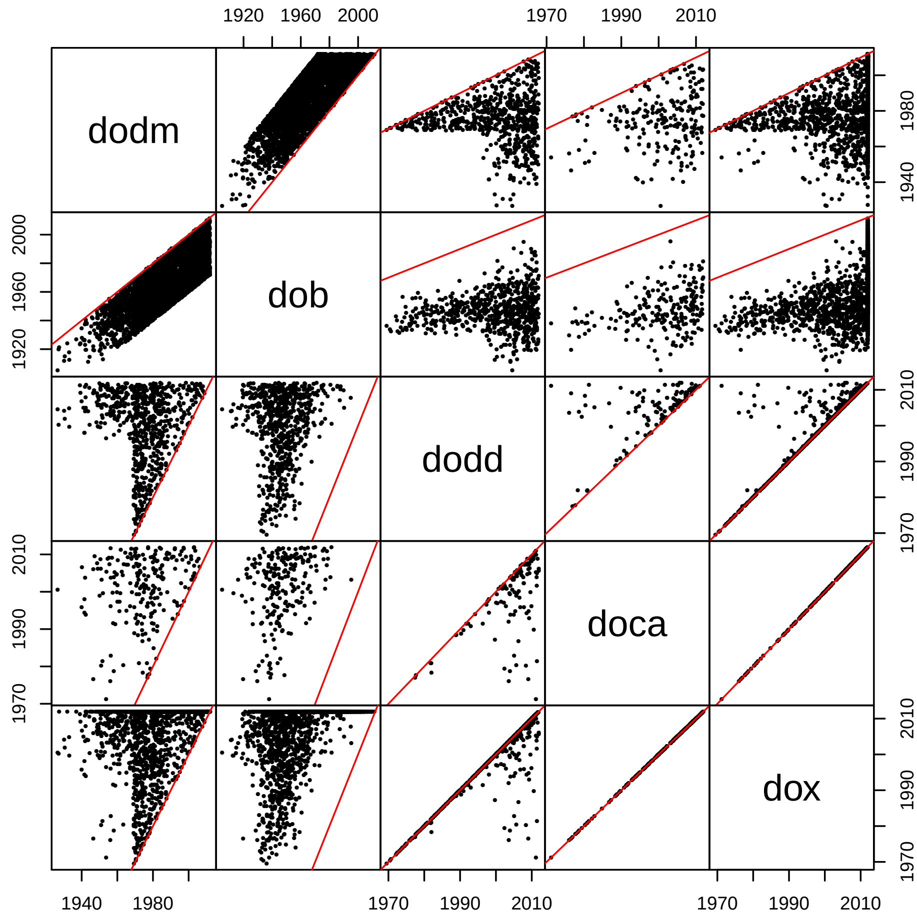

Figure 4.2: *Pairwise plots for a random sample of 5000 diabetes patients. The red lines are identity lines, so all points should be on one side of this line.*

```
+             45:47,31,NA,49,48,50,51,NA,13,22)
> cbind( conv[,c("NCnam","Clab")],
+        nam.SE = names(SE)[wh] )
   NCnam                         Clab       nam.SE
1     d0                    All sites    allcancer
2     d1                          Lip          lip
3     d2                       Tongue       tongue
4     d3              Salivary glands     salivary
5     d4                        Mouth        mouth
6     d5                      Pharynx      pharynx
7     d6                  Oesophagus     esophagus
8     d7                      Stomach      stomach
9     d8              Small intestine  smallintest
10    d9                        Colon        colon
11   d10                       Rectum       rectal
12   d11                        Liver        liver
13   d12                  Gallbladder         gall
14   d13                     Pancreas     pancreas
15   d14               Nose, sinuses         nose
16   d15                       Larynx       larynx
17   d16                         Lung         lung
18   d17                       Pleura       pleura
19   d18                       Breast       breast
20   d19                 Cervix uteri       cervix
```

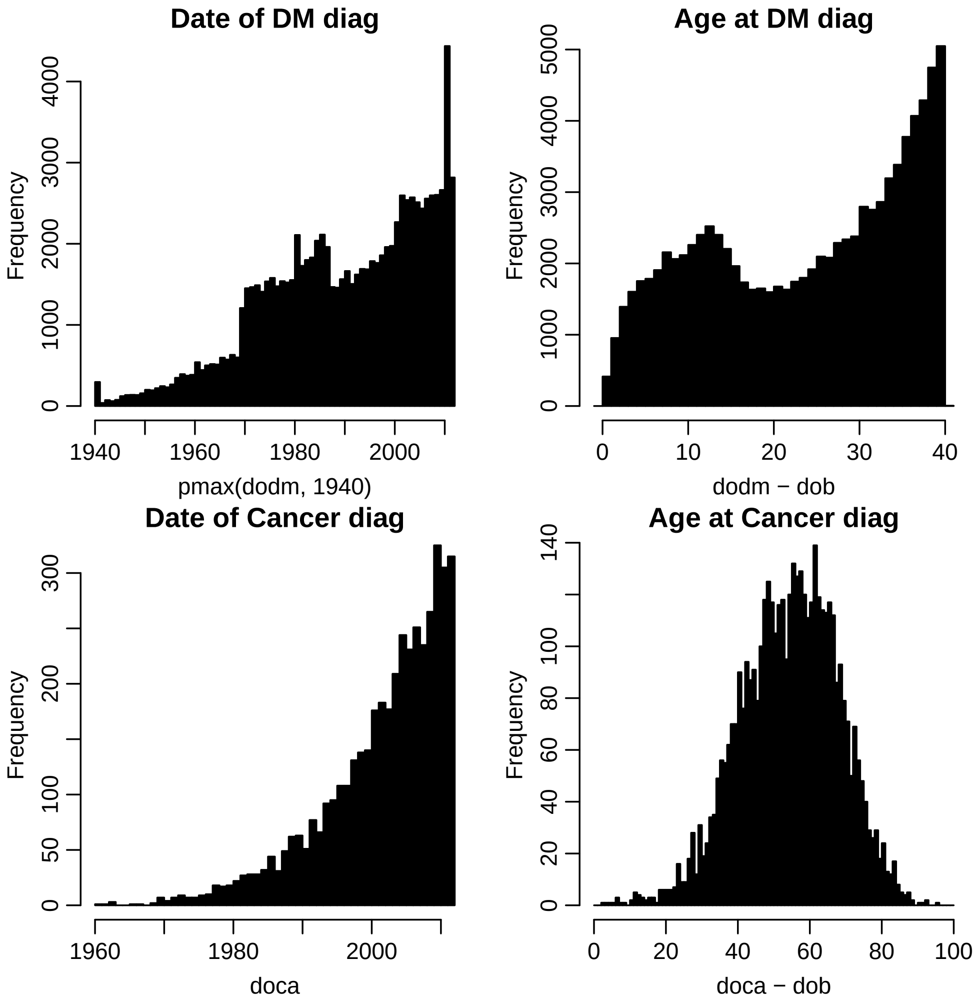

Figure 4.3: *Histograms of dates and ages of diagnosis of diabetes and cancer in the Swedish diabetes data. It is seen that the period 1970–87 used a more sensitive definition of diabetes.*

```
21   d20                  Corpus uteri         corpus
22   d22                         Ovary          ovary
23   d23   Other female genital organs      othuterus
24   d24                      Prostate       prostate
25   d25                        Testis         testis
26   d26                    Penis etc.          penis
27   d27                        Kidney         kidney
28   d28                       Bladder        bladder
29   d29             Melanoma of skin        melanoma
30   d31                           Eye            eye
31   d32                   Brain, CNS           brain
32   d33                       Thyroid        thyroid
33   d34                          Bone           bone
34   d35                  Soft tissues           <NA>
35   d36          Non-Hodgkin lymphoma     nonhodgkin
36   d37              Hodgkin lymphoma        hodgkin
37   d38              Multiple myeloma     multmyelom
38   d40                     Leukaemia       leukemia
39   d48 Other and unspecified cancers           <NA>
40   d51                    Oral etc. oralcavity_all
41   d52                    Colorectal       colorect
> names(SE)[wh[!is.na(wh)]] <- d.SE <- conv$NCnam[!is.na(wh)]
> names( SE )
```

```
 [1] "lpnr"         "dodm"         "dob"          "diaalder"     "diaalder39"   "diaalder34"
 [7] "diaalder29"   "dodd"         "canc_alder"   "icd7"         "doca"         "d1"
[13] "d51"          "d2"           "d4"           "d3"           "d5"           "d6"
[19] "d7"           "d8"           "d9"           "d52"          "d10"          "d11"
[25] "d12"          "d13"          "d14"          "d15"          "d16"          "d17"
[31] "d34"          "d29"          "softtissue"   "d18"          "othfemgen"    "d19"
[37] "d20"          "d23"          "d22"          "d26"          "d24"          "d25"
[43] "d27"          "d28"          "d31"          "d32"          "d33"          "d37"
[49] "d36"          "d38"          "d40"          "othleukemia"  "acuteleuk"    "d0"
[55] "dend"         "dox"          "persondagar"  "kon_ny"
> d.SE
 [1] "d0"  "d1"  "d2"  "d3"  "d4"  "d5"  "d6"  "d7"  "d8"  "d9"  "d10" "d11" "d12" "d13" "d14" "d15"
[17] "d16" "d17" "d18" "d19" "d20" "d22" "d23" "d24" "d25" "d26" "d27" "d28" "d29" "d31" "d32" "d33"
[33] "d34" "d36" "d37" "d38" "d40" "d51" "d52"
```

The variable `d.SE` now contains the (nicely sorted) names of the variables in the `SE` data frame that holds the tumour counts for each type of tumour, that will be available from NORDCAN data.

4.2 T1D analysis dataset

On the basis of this total dataset with all cancers and diabetes diagnoses, we now set up an analysis dataset, which for the tabulated follow-up of the T1D patients will have the following variables:

- Response variables

  `dxx` no. of events in diagnosis group `xx` (cancer type)

  `Y` person-years (corresponding to each diagnosis group)

- Classification (explanatory) variables:

  `sex` sex

  `age` age at follow-up, 1-year classes

  `per` date of follow-up, 1 year classes

  `DMdur` time since diagnosis, left end point of intervals $0, 1, 2, 5, 10, 15, 30$

  `T1D` T1 diabetes status: 30 — DM diagnosis < 30, 35 — DM diagnosis aged 30–34, 40 — DM diagnosis aged 35–39

  `DMprev` T1D present at 1.1.1987: Inc/Prv — essentially an indicator of whether a particular DM patient is followed from diagnosis.

We furthermore only include follow-up from 1987 (which is the period from which we have population rates — see below), and we define sex as a factor, and the grouping of patients by interval of age at diagnosis:

```
> table(SE$doca>1969,exclude=NULL)
FALSE  TRUE  <NA>
    9  4421 88972
> SE <- transform( SE, sex = factor( kon_ny, labels=c("M","F") ),
+                       doe = pmax( dodm, 1987, na.rm=TRUE ),
+                       T1D = pmin( pmax( ceiling((dodm-dob)/5)*5, 30 ), 40 ),
+                    DMprev = factor( dodm<1987, labels=c("Inc","Prv") ) )
> table( SE$T1D )
```

```
   30    35    40
56473 14993 21936
```

The analysis dataset will be constructed from multi-way tables of cases and person-years derived from the dataset of cases.

For the comparison population (which will be the entire population) we will construct a similar dataset. In the population dataset there will of course be a separate value for each of the variables `DMdur`, `T1D` and `DMprev` — these variables will enter the analysis models, and must therefore also be present in the population dataset, but with a special value that allows for a special (reference) level of their effect for the population part of data. Stacking the two datasets appropriately will then enable a joint analysis, similar to an SMR-analysis.

4.2.1 Follow-up of T1D patients

We first construct the follow-up *time* for T1D patients from entry to death, *i.e.* disregarding cancer occurrence as termination of follow-up. We later discuss this omission. Subsequently, we tabulate the number of cancer cases and construct the T1D part of the analysis dataset.

4.2.1.1 Lexis object of total follow-up time

Intuitively, the follow-up would be for T1D patients without cancer[1] from date of diagnosis of T1D (or 1.1.1995) until the end of the follow-up period. However, if we follow persons for *any* type of cancer regardless of other previous cancers, we should logically also follow persons for (other) cancers even if they have a cancer previous to the diagnosis of diabetes (which is what will be done).

We then tabulate the follow-up for the diabetes persons until death or end of 2011 (`dend`). The point of this is that while tabulation of events is simple (see below), the person-years to be used for a given type of event (cancer diagnosis of a certain type) is from date of entry to date of exit or event date, and hence in principle requires a complete tabulation of person-years for every type of cancer.

But as discussed with the Danish data, we will just use follow-up till death for each of the specific cancer sites, but until cancer for the category of all cancers. In the `SE` dataset, the variable `dox` is the exitdate for follwo-up for all cancers, so when doing follow-up for the specific cancers we need to follow persons till `pmin(dend,dodd)`, but for all cancers only till `dox`.

Now we set up a Lexis-object representing follow-up of the T1D patients until death or end of study period (`dox`), along timescales age, calendar time and duration. We define the state of observation to be "texttttnoCA", because we wll introduce follow-up after 1st cancer later as a separate state of follow-up, and thus tabulate the follow-up time by this too:

```
> whv <- c("dob","dodm","doe","doca","dend","dodd","sex","T1D","DMprev","lpnr")
> Lt1 <- Lexis( entry = list( per = doe,
+                             age = doe-dob,
+                             dur = doe-dodm ),
+                exit = list( per = pmin(dend,dodd,na.rm=TRUE) ),
+        entry.status = factor("noCa"),
+         exit.status = factor("noCa"),
+                data = subset( SE, doe<doca | is.na(doca),
+                                   select = whv ) )
```

[1] Meaning either any cancer or a specific cancer type

Those excluded are those with all follow-up before 1.1.1987 plus a few persons with date of DM diagnosis equal to date of death or end of study.

```
> summary( Lt1 )
Transitions:
     To
From    noCa  Records:  Events: Risk time:  Persons:
  noCa 91056     91056        0    1392555     91056
> system.time(
+ Lc1 <- cutLexis( Lt1, cut = Lt1$doca,
+                  new.state = "anyCa",
+                  precursor = "noCa" ) )
   user  system elapsed
  9.631   0.021   9.649
> ( ss <- summary( Lc1 )$Transitions )
Transitions:
     To
From     noCa anyCa  Records:  Events: Risk time:  Persons:
  noCa  87052  4004     91056     4004  1369807.0     91056
  anyCa     0  4004      4004        0    22748.1      4004
  Sum   87052  8008     95060     4004  1392555.1     91056
> round( ss[2,5] / ss[3,5] * 100, 1 )
[1] 1.6
```

This Lexis object now represents follow-up through two states; before and after any cancer; we see that the fraction of the follow-up after (any) cancer diagnosis is 1.6%.

We must tabulate the risk time by current age, calendar time and duration, so we split the follow-up by these three factors. Due to memory restrictions we do this separately in 10 chunks of the data:

```
> S.all <- NULL
> n.chunks <- 10
> lm <- round( seq(0,nrow(Lc1),,n.chunks+1) )
> cat( format(Sys.time()) )
2015-05-06 13:40:18
> system.time(
+ for( i in 1:n.chunks )
+ {
+ whr  <- (lm[i]+1):(lm[i+1])
+ sa   <- splitLexis( Lc1[whr,],        0:120        , time.scale="age" )
+ sap  <- splitLexis(        sa ,      1987:2012      , time.scale="per" )
+ sapd <- splitLexis(        sap, c(0,1,2,5,10,15,30), time.scale="dur" )
+ S.all <- rbind( S.all, sapd )
+ cat( "chunk", i, ", now", nrow(S.all), "rows, at", format(Sys.time()), "\n" )
+ } )
chunk 1 , now 307130 rows, at 2015-05-06 13:40:38
chunk 2 , now 614245 rows, at 2015-05-06 13:40:58
chunk 3 , now 923362 rows, at 2015-05-06 13:41:18
chunk 4 , now 1230174 rows, at 2015-05-06 13:41:38
chunk 5 , now 1536599 rows, at 2015-05-06 13:41:58
chunk 6 , now 1842666 rows, at 2015-05-06 13:42:18
chunk 7 , now 2150075 rows, at 2015-05-06 13:42:38
chunk 8 , now 2458792 rows, at 2015-05-06 13:42:59
chunk 9 , now 2764681 rows, at 2015-05-06 13:43:20
chunk 10 , now 3073565 rows, at 2015-05-06 13:43:41
   user  system elapsed
202.128   0.328 202.387
> summary( Lc1 )
Transitions:
     To
From     noCa anyCa  Records:  Events: Risk time:  Persons:
  noCa  87052  4004     91056     4004  1369807.0     91056
  anyCa     0  4004      4004        0    22748.1      4004
  Sum   87052  8008     95060     4004  1392555.1     91056
```

```
> summary( S.all )
Transitions:
     To
From        noCa anyCa  Records:  Events: Risk time:  Persons:
  noCa   3019566  4004   3023570     4004  1369807.0     91056
  anyCa        0 49995     49995        0    22748.1      4004
  Sum    3019566 53999   3073565     4004  1392555.1     91056
> save( S.all, file="../data/S.all.Rda" )
```

Finally we use this to create a data frame with the total alive-follow-up among the T1D patients:

```
> # load( file="../data/S.all.Rda" )
> S.all <- transform( S.all, A = floor(age),
+                            P = floor(per),
+                        DMdur = (dur>=1) +
+                                (dur>=2) +
+                                (dur>=5)*3 +
+                                (dur>=10)*5 +
+                                (dur>=15)*5 +
+                                (dur>=30)*15 )
> system.time(
+ PYtab <- aggregate( S.all$lex.dur,
+                    by = S.all[,c("A","P","sex","T1D","DMdur","DMprev","lex.Cst")],
+                   FUN = sum ) )
   user  system elapsed
 23.822   0.000  23.806
> # Only follow-up time prior to first cancer
> PY.noc <- subset( PYtab, lex.Cst=="noCa" )[,c("A","P","sex","T1D","DMdur","DMprev","x")]
> names( PY.noc )
[1] "A"      "P"      "sex"    "T1D"    "DMdur"  "DMprev" "x"
> dim( PY.noc )
[1] 24376     7
> # Aggregate the follow-up time before AND after cancer diagnosis
> PY.all <- aggregate( PYtab$x,
+                     by = PYtab[,c("A","P","sex","T1D","DMdur","DMprev")],
+                    FUN = sum )
> dim( PY.all )
[1] 24411     7
> names( PY.all )[match("x",names(PY.all))] <- "y"
> names( PY.noc )[match("x",names(PY.noc))] <- "y0"
> PYtab <- merge( PY.all, PY.noc )
> names( PYtab )[c(1:2)] <- c("age","per")
> str( PYtab )
'data.frame':        24376 obs. of  8 variables:
 $ age   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ per   : num  1987 1987 1987 1988 1988 ...
 $ sex   : Factor w/ 2 levels "M","F": 2 1 1 2 1 2 1 2 1 2 ...
 $ T1D   : num  30 30 30 30 30 30 30 30 30 30 ...
 $ DMdur : num  0 0 0 0 0 0 0 0 0 0 ...
 $ DMprev: Factor w/ 2 levels "Inc","Prv": 2 1 2 1 1 1 1 1 1 1 ...
 $ y     : num  1.35 1.84 2.17 3.89 3.29 ...
 $ y0    : num  1.35 1.84 2.17 3.89 3.29 ...
> head(PYtab)
  age  per sex T1D DMdur DMprev        y       y0
1   0 1987   F  30     0    Prv 1.349857 1.349857
2   0 1987   M  30     0    Inc 1.841272 1.841272
3   0 1987   M  30     0    Prv 2.166590 2.166590
4   0 1988   F  30     0    Inc 3.886703 3.886703
5   0 1988   M  30     0    Inc 3.292552 3.292552
6   0 1989   F  30     0    Inc 1.217798 1.217798
```

```
> ah <- with( PYtab, tapply( y, age, sum ) )
> ph <- with( PYtab, tapply( y, per, sum ) )
> dh <- with( PYtab, tapply( y, DMdur, sum ) )
> layout( rbind(1:3), widths=c(length(names(ah)),
+                              length(names(ph)),
+                              30)+2 )
> par( mar=c(3,1,1,1), oma=c(0,3,0,0), mgp=c(3,1,0)/1.6, las=1 )
> aa <- barplot( ah/1000, space=0, xaxt="n", ylim=c(0,85),
+               col=gray(0.4), border=FALSE, xlab="Age at follow-up" )
> whl <- 0:9*10+1
> axis( side=1, at=aa[whl]-0.5, labels=names(ah)[whl] )
> pp <- barplot( ph/1000, space=0, xaxt="n", ylim=c(0,85),
+               col=gray(0.4), border=FALSE, xlab="Date of follow-up" )
> whl <- 0:8*5+2
> axis( side=1, at=aa[whl]-0.5, labels=names(ph)[whl] )
> wd <- c(1,1,3,5,5,15)
> dd <- barplot( dh[-7]/1000/wd, width=wd, space=0, xaxt="n", ylim=c(0,85),
+               col=gray(0.4), border=FALSE, xlab="DM duration" )
> axis( side=1 )
> mtext( "PYears (1000s) in 1-year intervals", side=2, line=1, outer=TRUE, las=0 )
```

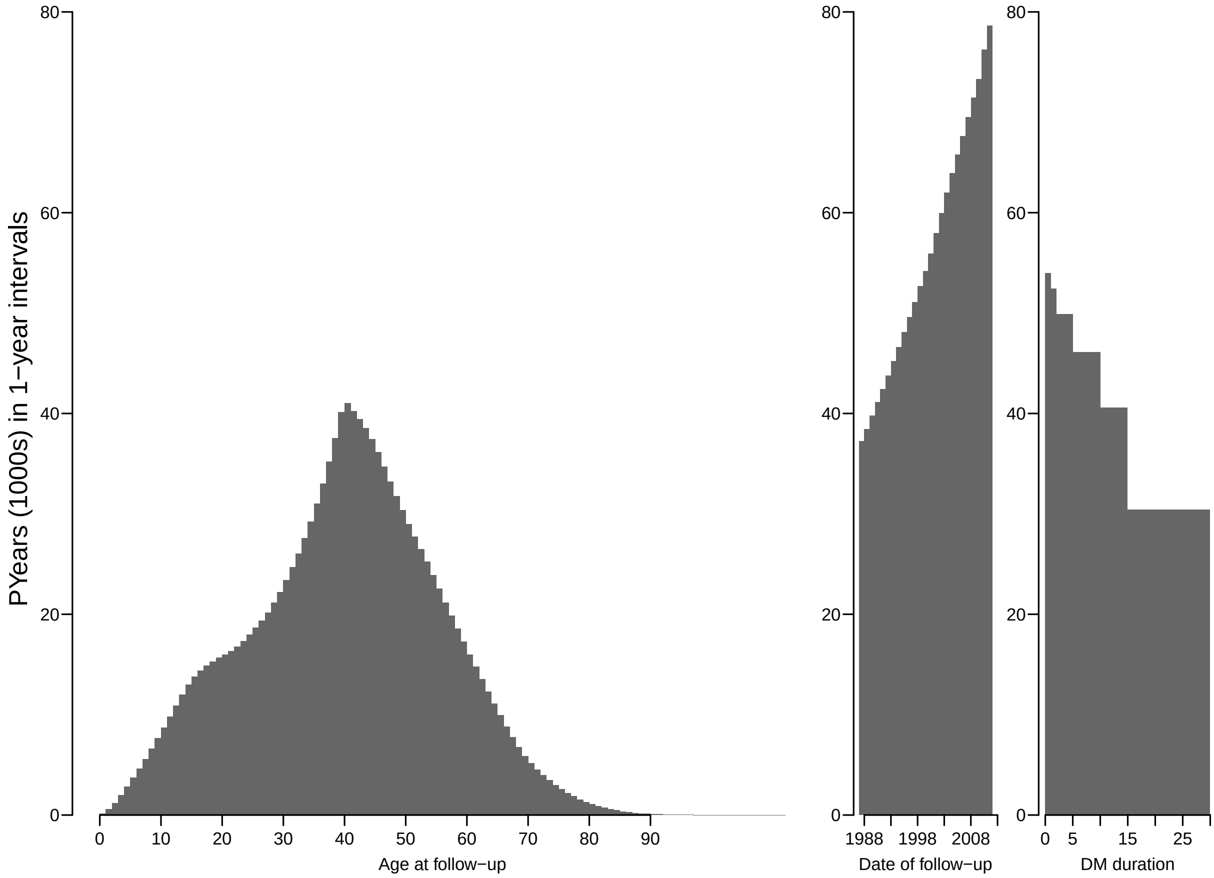

Figure 4.4: *Distribution of the follow-up time in the T1D patients (till death) by age, date and DM-duration at follow-up.*

From figure 4.4 it is seen that the distribution of follow-up time is largely in age-classes where cancer occurrence is quite small.

4.2.1.2 Number of cancer cases

Since we now are going to tabulate the cancer cases we must classify them by age, date and diabetes duration *at date of cancer*, so we define these for all persons in the data frame:

```
> SEca <- transform( subset( SE, !is.na(doca) & doca>doe ),
+                    A = floor(doca-dob),
+                    P = floor(doca),
+                DMdur = ((doca-dodm)>=1) +
+                        ((doca-dodm)>=2) +
+                        ((doca-dodm)>=5)*3 +
+                        ((doca-dodm)>=10)*5 +
+                        ((doca-dodm)>=15)*5 +
+                        ((doca-dodm)>=30)*15 )
> dim( SEca )
[1] 4083   65
> system.time(
+ Catab <- aggregate( SEca[,d.SE],
+                     by = SEca[,c("A","P","sex","T1D","DMdur","DMprev")],
+                    FUN = sum ) )
   user  system elapsed
  0.923   0.000   0.923
> dim( Catab )
[1] 3040   45
> names( Catab )[1:2] <- c("age","per")
> names( Catab )
 [1] "age"    "per"    "sex"    "T1D"    "DMdur"  "DMprev" "d0"     "d1"     "d2"     "d3"
[11] "d4"     "d5"     "d6"     "d7"     "d8"     "d9"     "d10"    "d11"    "d12"    "d13"
[21] "d14"    "d15"    "d16"    "d17"    "d18"    "d19"    "d20"    "d22"    "d23"    "d24"
[31] "d25"    "d26"    "d27"    "d28"    "d29"    "d31"    "d32"    "d33"    "d34"    "d36"
[41] "d37"    "d38"    "d40"    "d51"    "d52"
> names( PYtab )
[1] "age"    "per"    "sex"    "T1D"    "DMdur"  "DMprev" "y"      "y0"
> dim( PYtab )
[1] 24376     8
> dim( Catab )
[1] 3040   45
> CaFU <- merge( PYtab, Catab, all=TRUE )
> dim( CaFU )
[1] 24376    47
> names( CaFU )
 [1] "age"    "per"    "sex"    "T1D"    "DMdur"  "DMprev" "y"      "y0"     "d0"     "d1"
[11] "d2"     "d3"     "d4"     "d5"     "d6"     "d7"     "d8"     "d9"     "d10"    "d11"
[21] "d12"    "d13"    "d14"    "d15"    "d16"    "d17"    "d18"    "d19"    "d20"    "d22"
[31] "d23"    "d24"    "d25"    "d26"    "d27"    "d28"    "d29"    "d31"    "d32"    "d33"
[41] "d34"    "d36"    "d37"    "d38"    "d40"    "d51"    "d52"
> save( CaFU, file="../data/SE-CaFU.Rda" )
> load( file="../data/SE-CaFU.Rda" )
```

Thus we now have the dataset of events and follow-up among the T1D patients.

4.3 Reading population data

Now we load the population rates from NORDCAN; they are in a dataframe called `Pfr`:

```
> load( file="../data/Sweden-NORCAN.RData" )
> str( Pfr )
```

```
'data.frame':        850 obs. of  62 variables:
 $ ageg  : int  1 2 3 4 5 6 7 8 9 10 ...
 $ A     : num  2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5 42.5 47.5 ...
 $ per   : num  1987 1987 1987 1987 1987 ...
 $ P     : num  1988 1988 1988 1988 1988 ...
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ t1d   : Factor w/ 1 level "NoDM": 1 1 1 1 1 1 1 1 1 1 ...
 $ dmprev: Factor w/ 1 level "Pop": 1 1 1 1 1 1 1 1 1 1 ...
 $ Y0    : num  251121 245929 272783 289835 314322 ...
 $ Y     : num  251121 245929 272783 289835 314322 ...
 $ d0    : num  74 30 28 54 65 100 137 206 317 417 ...
 $ d1    : int  0 0 0 0 0 0 0 0 1 4 ...
 $ d2    : int  0 0 0 0 0 0 0 3 1 7 ...
 $ d3    : int  0 0 0 2 0 0 2 3 2 0 ...
 $ d4    : int  0 0 0 0 1 1 1 4 4 5 ...
 $ d5    : int  0 0 0 0 0 0 0 1 5 3 ...
 $ d6    : int  0 0 0 0 0 0 2 0 2 6 ...
 $ d7    : int  0 0 0 0 0 0 1 5 16 17 ...
 $ d8    : int  0 0 0 0 1 1 0 0 2 6 ...
 $ d9    : int  0 0 0 0 0 1 5 10 14 17 ...
 $ d10   : int  0 0 0 0 0 1 3 4 14 19 ...
 $ d11   : int  1 0 0 0 1 0 3 2 1 4 ...
 $ d12   : int  0 0 0 0 0 0 0 1 3 3 ...
 $ d13   : int  0 0 0 1 0 1 1 2 7 15 ...
 $ d14   : int  0 0 0 0 1 0 0 0 0 3 ...
 $ d15   : int  0 0 0 0 0 0 0 0 2 2 ...
 $ d16   : int  0 0 0 0 1 0 4 7 20 37 ...
 $ d17   : int  0 0 0 0 0 0 0 0 4 8 ...
 $ d18   : int  0 0 0 0 0 0 0 0 0 2 ...
 $ d19   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d20   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d21   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d22   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d23   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d24   : int  0 0 0 0 0 0 0 1 3 12 ...
 $ d25   : int  2 1 0 7 17 31 32 40 24 12 ...
 $ d26   : int  1 1 0 0 0 0 0 4 2 1 ...
 $ d27   : int  4 1 0 0 1 0 2 5 17 28 ...
 $ d28   : int  1 0 0 0 1 2 6 10 24 35 ...
 $ d29   : int  0 0 0 4 8 15 17 26 44 51 ...
 $ d30   : int  0 0 0 2 1 3 8 12 6 16 ...
 $ d31   : int  4 0 0 0 0 2 2 0 2 4 ...
 $ d32   : int  20 12 11 14 11 18 27 26 30 33 ...
 $ d33   : int  0 0 0 2 0 2 3 6 10 2 ...
 $ d34   : int  0 2 2 5 2 0 0 2 0 0 ...
 $ d35   : int  4 0 0 1 3 2 0 1 5 8 ...
 $ d36   : int  5 1 3 4 6 5 2 12 23 21 ...
 $ d37   : int  0 0 1 8 5 12 11 9 6 2 ...
 $ d38   : int  0 0 0 0 0 0 0 1 4 10 ...
 $ d40   : int  25 12 8 4 5 3 6 8 7 12 ...
 $ d41   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d42   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d43   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d44   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d45   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d46   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d47   : int  NA NA NA NA NA NA NA NA NA NA ...
 $ d48   : int  7 0 3 2 1 3 7 13 18 28 ...
 $ d49   : int  74 30 28 56 66 103 145 218 323 433 ...
 $ d50   : int  74 30 28 54 65 100 137 206 317 417 ...
 $ d51   : int  0 0 0 2 1 1 3 11 13 19 ...
 $ d52   : int  0 0 0 0 0 2 8 14 28 36 ...
 $ DMdur : Factor w/ 1 level "NoDM": 1 1 1 1 1 1 1 1 1 1 ...
 - attr(*, "status.info")= chr "Survo data file FICA: record=171 bytes, M1=74 L=64  M=61 N=850"
 - attr(*, "status.description")= chr
 - attr(*, "status.varname")= chr  "AGEG     Age group                 (####)                     "
```

```
 - attr(*, "status.vartype")= chr  "1A      " "4A-     " "4A-     " "4A-     " ...
 - attr(*, "status.varlen")= int  1 4 4 4 1 1 1 4 8 4 ...
> Pfr[1:10,1:13]
   ageg    A  per      P sex  t1d dmprev     Y0      Y  d0 d1 d2 d3
1     1  2.5 1987 1987.5   M NoDM    Pop 251121 251121  74  0  0  0
2     2  7.5 1987 1987.5   M NoDM    Pop 245929 245929  30  0  0  0
3     3 12.5 1987 1987.5   M NoDM    Pop 272783 272783  28  0  0  0
4     4 17.5 1987 1987.5   M NoDM    Pop 289835 289835  54  0  0  2
5     5 22.5 1987 1987.5   M NoDM    Pop 314322 314322  65  0  0  0
6     6 27.5 1987 1987.5   M NoDM    Pop 285257 285257 100  0  0  0
7     7 32.5 1987 1987.5   M NoDM    Pop 294904 294904 137  0  0  2
8     8 37.5 1987 1987.5   M NoDM    Pop 318215 318215 206  0  3  3
9     9 42.5 1987 1987.5   M NoDM    Pop 337540 337540 317  1  1  2
10   10 47.5 1987 1987.5   M NoDM    Pop 253929 253929 417  4  7  0
> Pfr <- transform( Pfr, y = Y, y0 = Y0, DMprev=dmprev, T1D=t1d )
```

The final analysis dataset is now contructed by stacking the follow-up for the T1D patients and the population as a whole:

```
> CaFU <- transform( CaFU, A = age+0.5,
+                          P = per+0.5,
+                        T1D = factor(T1D),
+                      DMdur = factor(DMdur) )
> ( avars <- c("sex","A","P","T1D","DMprev","DMdur","y0","y",d.SE) )
 [1] "sex"    "A"      "P"      "T1D"    "DMprev" "DMdur"  "y0"     "y"      "d0"     "d1"
[11] "d2"     "d3"     "d4"     "d5"     "d6"     "d7"     "d8"     "d9"     "d10"    "d11"
[21] "d12"    "d13"    "d14"    "d15"    "d16"    "d17"    "d18"    "d19"    "d20"    "d22"
[31] "d23"    "d24"    "d25"    "d26"    "d27"    "d28"    "d29"    "d31"    "d32"    "d33"
[41] "d34"    "d36"    "d37"    "d38"    "d40"    "d51"    "d52"
> se.ana <- rbind( Pfr[,avars], CaFU[,avars] )
> se.ana[is.na(se.ana)] <- 0
> str( se.ana )
'data.frame':        25226 obs. of  47 variables:
 $ sex   : Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A     : num  2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5 42.5 47.5 ...
 $ P     : num  1988 1988 1988 1988 1988 ...
 $ T1D   : Factor w/ 4 levels "NoDM","30","35",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ DMprev: Factor w/ 3 levels "Pop","Inc","Prv": 1 1 1 1 1 1 1 1 1 1 ...
 $ DMdur : Factor w/ 8 levels "NoDM","0","1",..: 1 1 1 1 1 1 1 1 1 1 ...
 $ y0    : num  251121 245929 272783 289835 314322 ...
 $ y     : num  251121 245929 272783 289835 314322 ...
 $ d0    : num  74 30 28 54 65 100 137 206 317 417 ...
 $ d1    : num  0 0 0 0 0 0 0 0 1 4 ...
 $ d2    : num  0 0 0 0 0 0 0 3 1 7 ...
 $ d3    : num  0 0 0 2 0 0 2 3 2 0 ...
 $ d4    : num  0 0 0 0 1 1 1 4 4 5 ...
 $ d5    : num  0 0 0 0 0 0 0 1 5 3 ...
 $ d6    : num  0 0 0 0 0 0 2 0 2 6 ...
 $ d7    : num  0 0 0 0 0 0 1 5 16 17 ...
 $ d8    : num  0 0 0 0 1 1 0 0 2 6 ...
 $ d9    : num  0 0 0 0 0 1 5 10 14 17 ...
 $ d10   : num  0 0 0 0 0 1 3 4 14 19 ...
 $ d11   : num  1 0 0 0 1 0 3 2 1 4 ...
 $ d12   : num  0 0 0 0 0 0 0 1 3 3 ...
 $ d13   : num  0 0 0 1 0 1 1 2 7 15 ...
 $ d14   : num  0 0 0 0 1 0 0 0 0 3 ...
 $ d15   : num  0 0 0 0 0 0 0 0 2 2 ...
 $ d16   : num  0 0 0 0 1 0 4 7 20 37 ...
 $ d17   : num  0 0 0 0 0 0 0 0 4 8 ...
 $ d18   : num  0 0 0 0 0 0 0 0 0 2 ...
 $ d19   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d20   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d22   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d23   : num  0 0 0 0 0 0 0 0 0 0 ...
 $ d24   : num  0 0 0 0 0 0 0 1 3 12 ...
```

```
 $ d25    : num  2 1 0 7 17 31 32 40 24 12 ...
 $ d26    : num  1 1 0 0 0 0 0 4 2 1 ...
 $ d27    : num  4 1 0 0 1 0 2 5 17 28 ...
 $ d28    : num  1 0 0 0 1 2 6 10 24 35 ...
 $ d29    : num  0 0 0 4 8 15 17 26 44 51 ...
 $ d31    : num  4 0 0 0 0 2 2 0 2 4 ...
 $ d32    : num  20 12 11 14 11 18 27 26 30 33 ...
 $ d33    : num  0 0 0 2 0 2 3 6 10 2 ...
 $ d34    : num  0 2 2 5 2 0 0 2 0 0 ...
 $ d36    : num  5 1 3 4 6 5 2 12 23 21 ...
 $ d37    : num  0 0 1 8 5 12 11 9 6 2 ...
 $ d38    : num  0 0 0 0 0 0 0 1 4 10 ...
 $ d40    : num  25 12 8 4 5 3 6 8 7 12 ...
 $ d51    : num  0 0 0 2 1 1 3 11 13 19 ...
 $ d52    : num  0 0 0 0 0 2 8 14 28 36 ...
> table(se.ana$T1D)
 NoDM    30    35    40
  850 15114  4764  4498
> t( xtabs( as.matrix(se.ana[,avars[-(1:8)]]) ~ se.ana$T1D ) )
     se.ana$T1D
        NoDM     30     35     40
  d0  959456   1886    848   1349
  d1    3296      7      0      4
  d2    3425     11      5      4
  d3    2110      8      0      3
  d4    4140     15      5      9
  d5    5534     18     10     13
  d6    8554     14      9     20
  d7   24970     57     18     28
  d8    5095     11      2      8
  d9   72201    134     54    105
  d10  43256     43     38     63
  d11  12499     19     12     30
  d12  10468      3      1      4
  d13  23452     30     17     35
  d14   1512      3      2      1
  d15   4473      5      4      7
  d16  72653    113     62     88
  d17   2675      1      4      6
  d18 134488    359    168    190
  d19  11115     43     19     23
  d20  28485     68     27     56
  d22  20622     48     25     41
  d23   4107      3      4      7
  d24 169292    139     79    126
  d25   6301     47      6      4
  d26   1714      7      4      9
  d27  23112     20     24     55
  d28  50477     63     28     69
  d29  41974    109     41     79
  d31   2638      9      1      8
  d32  30402     93     37     47
  d33   7734     40     15     11
  d34   1772      5      1      2
  d36  30805     85     33     49
  d37   4385     17      6      6
  d38  12843     20      4     15
  d40  23299     44     12     10
  d51  18505     59     20     33
  d52 115457    177     92    168
> round(
+ t( xtabs( as.matrix(se.ana[,avars[  7:8 ]]) ~ se.ana$T1D ) ), 1 )
    se.ana$T1D
            NoDM          30          35          40
  y0 217267131.0    890244.5    206037.3    273525.1
  y  217267131.0    901445.9    210648.5    280445.1
```

```
> se.ana <- transform( se.ana,
+                      DMprev = Relevel( T1D, list( Pop=1, Inc=2:4 ) ),
+                         Cnt = "SWE" )
> save( se.ana, file="../data/SEana.Rda" )
```

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

4.4 Analysis of Swedish data

We load the dataset and the data frame with the tumour labels in it:

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

We define a factor pooling the ages at diagnosis, `t1`:

```
> se.ana <- transform( subset( se.ana, y0>0 & P>1987 ),
+                      T1  = Relevel( T1D,
+                                     list(NoDM=1,T1D=2:nlevels(T1D)) ) )
> with( se.ana,
+ round( cbind( tt <- addmargins( tapply(d0, list( P, T1D ), sum ), 1 ),
+               T1D = apply( tt[,-1], 1, sum ) ) ) )
         NoDM   30  35   40  T1D
1987.5  34084   22  14   13   49
1988.5  34000   22  13   27   62
1989.5  34023   28  18   17   63
1990.5  34493   30   8   13   51
1991.5  34756   30  21   26   77
1992.5  34844   32  14   20   66
1993.5  35392   38  17   36   91
1994.5  35835   45  23   26   94
1995.5  35354   50  16   42  108
1996.5  35577   42  32   34  108
1997.5  35743   72  21   38  131
1998.5  36503   65  26   46  137
1999.5  37827   62  33   45  140
2000.5  37960   85  40   50  175
2001.5  38615   95  33   55  183
2002.5  39395   79  27   71  177
2003.5  40918   90  48   70  208
2004.5  42262  117  47   80  244
2005.5  42019  109  51   70  230
2006.5  41662  117  54   78  249
2007.5  41710  119  39   76  234
2008.5  42289  117  53   93  263
2009.5  44700  143  62  120  325
2010.5  44640  137  57  110  304
2011.5  44855  140  81   93  314
Sum    959456 1886 848 1349 4083
```

Thus we have 4,083 cases of cancer among the T1D patients, and close to 1 mil. in the reference population over the study period.

4.4.1 All cancers

We first set out the simplest possible analysis with age-period-cohort effects for the baseline rates. First we devise a couple of knots for the splines:

```
> with( se.ana, rbind(
+  A = quantile( rep(  A,d0), 0:10/10 ),
+  P = quantile( rep(P  ,d0), 0:10/10 ),
+  C = quantile( rep(P-A,d0), 0:10/10 ) ) )
      0%    10%    20%    30%    40%    50%    60%    70%    80%    90%   100%
A    2.5   47.5   57.5   62.5   67.5   67.5   72.5   72.5   77.5   82.5   95.5
P 1987.5 1989.5 1992.5 1995.5 1998.5 2000.5 2003.5 2005.5 2007.5 2009.5 2011.5
C 1905.0 1916.0 1921.0 1925.0 1928.0 1932.0 1937.0 1941.0 1946.0 1954.0 2009.0
> ( a.kn <- seq(5,80,,7) )
[1]  5.0 17.5 30.0 42.5 55.0 67.5 80.0
> ( p.kn <- seq(1989,2009,,5) )
[1] 1989 1994 1999 2004 2009
> ( c.kn <- seq(1915,1960,,6) )
[1] 1915 1924 1933 1942 1951 1960
```

We then fit models for the RR for T1D patients relative to the general population, using a common shape of the underlying cancer incidence rates as an age-period cohort model with $1+(7-1)+(5-1)+(6-1)-1=15$ parameters[2]. The first model (`m3`) is one with separate effects for persons diagnosed in ages <30, 30–35 and 35–40, the second (`m1`) is a simplification where the three groups are pooled:

```
> m3 <- glm( d0 ~ Ns(   A, knots=a.kn ) +
+                 Ns( P  , knots=p.kn ) +
+                 Ns( P-A, knots=c.kn ) +
+                 T1D + sex,
+            offset = log(y0),
+            family = poisson,
+              data = se.ana )
> m1 <- update( m3, . ~. - T1D + T1 )
> anova( m3, m1, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D + sex
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    sex + T1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     25207      59599
2     25209      59602 -2  -2.7938   0.2474
> round( rbind( ci.exp( m3, subset="T1" ),
+               ci.exp( m1, subset="T1" ) ), 3 )
      exp(Est.)  2.5% 97.5%
T1D30     0.971 0.928 1.016
T1D35     1.033 0.966 1.105
T1D40     1.016 0.963 1.071
T1T1D     0.998 0.968 1.029
```

It appears that the risk of cancer is the same in all groups of age at diagnosis; which is also evident in figure 7.1, where the RR estimates from the three models are shown together:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind( ci.exp( m3, subset="T1D"),
+                 ci.exp( m1, subset="T1") ),
+            y = 4:1,
+            txt = c("<30","30-35","35-40","Pooled"),
+            xlog=TRUE, xtic=7:15/10, grid=TRUE,
+            xlab="RR of any cancer, T1D vs. population",
+            lwd=4, cex=2, vref=1 )
```

[2]There is first an intercept, then the three natural splines, where k knots gives $k-1$ parameters, and finally 1 aliased parameter from the linear relationship between $P-A$ and P and A.

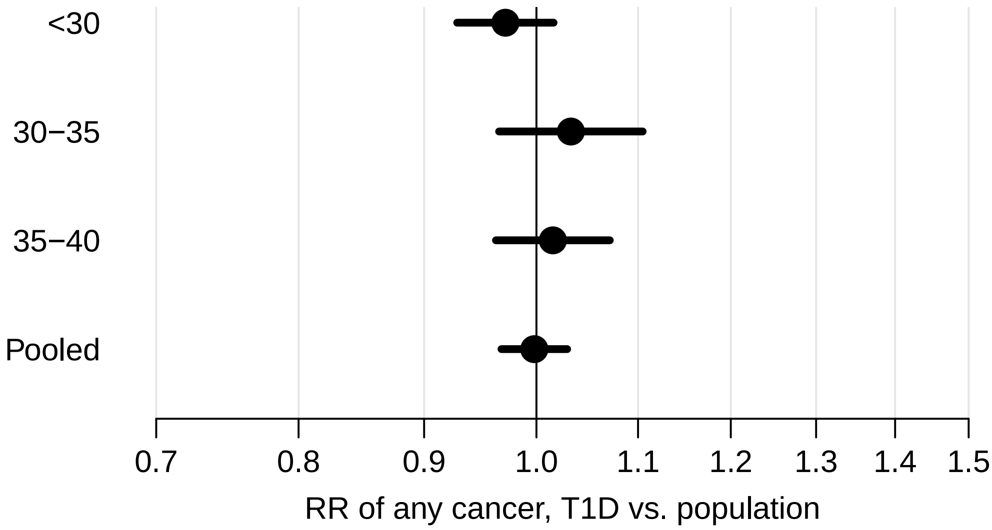

Figure 4.5: *Estimated RRs relative to the general population in two different models. There is no significant differences betwen the ages at DM diagnosis.*

4.4.2 Analysis by duration

We have also have a variable `DMdur` indicating different time bands after diagnosis:

```
> tt <- xtabs( cbind(d0,y0=y0/1000,r=y0) ~ T1D + DMdur, data=se.ana )
> tt[,,3] <- tt[,,"d0"]/tt[,,"y0"]*100
> round( ftable(tt), 1 )
                 d0       y0        r
T1D  DMdur
NoDM NoDM  959456.0 217267.1    441.6
     0          0.0      0.0      NaN
     1          0.0      0.0      NaN
     2          0.0      0.0      NaN
     5          0.0      0.0      NaN
     10         0.0      0.0      NaN
     15         0.0      0.0      NaN
     30         0.0      0.0      NaN
30   NoDM       0.0      0.0      NaN
     0         27.0     30.5     88.5
     1          9.0     29.5     30.6
     2         32.0     83.9     38.1
     5         73.0    130.8     55.8
     10        93.0    119.6     77.7
     15       490.0    295.5    165.8
     30      1162.0    200.4    579.7
35   NoDM       0.0      0.0      NaN
     0         19.0      8.9    214.5
     1          9.0      8.7    103.6
     2         27.0     25.0    107.9
     5         67.0     39.4    170.2
     10        95.0     34.6    274.2
     15       386.0     70.3    548.7
     30       245.0     19.1   1282.3
40   NoDM       0.0      0.0      NaN
     0         40.0     14.6    274.8
     1         16.0     14.2    112.8
     2         59.0     40.3    146.5
     5        147.0     59.3    247.9
     10       169.0     46.9    360.3
     15       622.0     82.1    757.9
     30       296.0     16.2   1823.2
```

```
> md40 <- update( m1  , . ~ . - T1 + DMdur )
> md35 <- update( md40, data=subset(se.ana,T1D %in% levels(T1D)[-   4 ]) )
> md30 <- update( md40, data=subset(se.ana,T1D %in% levels(T1D)[-(3:4)]) )
```

We see that there a very strong effect of duration, and that it is the same regardless of the age cut-off:

```
> round( cbind( RR40 <- ci.exp( md40, subset=c("t1","DMdur") ),
+               RR35 <- ci.exp( md35, subset=c("t1","DMdur") ),
+               RR30 <- ci.exp( md30, subset=c("t1","DMdur") ) ), 3 )
        exp(Est.)  2.5% 97.5% exp(Est.)  2.5% 97.5% exp(Est.)  2.5% 97.5%
DMdur0      2.736 2.214 3.380     3.251 2.435 4.341     3.613 2.477 5.270
DMdur1      1.026 0.733 1.436     1.216 0.766 1.930     1.175 0.611 2.258
DMdur2      1.060 0.885 1.269     1.191 0.923 1.537     1.275 0.901 1.803
DMdur5      1.212 1.080 1.361     1.293 1.096 1.527     1.350 1.073 1.699
DMdur10     1.156 1.042 1.282     1.254 1.087 1.447     1.215 0.992 1.489
DMdur15     1.008 0.958 1.060     1.052 0.985 1.124     1.053 0.964 1.150
DMdur30     0.905 0.863 0.949     0.883 0.838 0.931     0.891 0.841 0.943
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind(NA,RR40),
+          txt=c("Years since DM","0-1","1-2","2-5","5-10","10-15","15-30","30+"),
+              xlog=TRUE, xtic=c(c(5:10,15,20,25)/10,3:6), grid=TRUE,
+              xlab="RR of any cancer, T1D vs. population",
+              lwd=4, cex=2, vref=1, y=c(7.7,7:1)+0.2 )
> linesEst( rbind(NA,RR35),
+           lwd=3, cex=1.5, col=gray(0.5), y=8:1 )
> linesEst( rbind(NA,RR30),
+           lwd=3, cex=1.5, col=gray(0.7), y=8:1-0.15 )
```

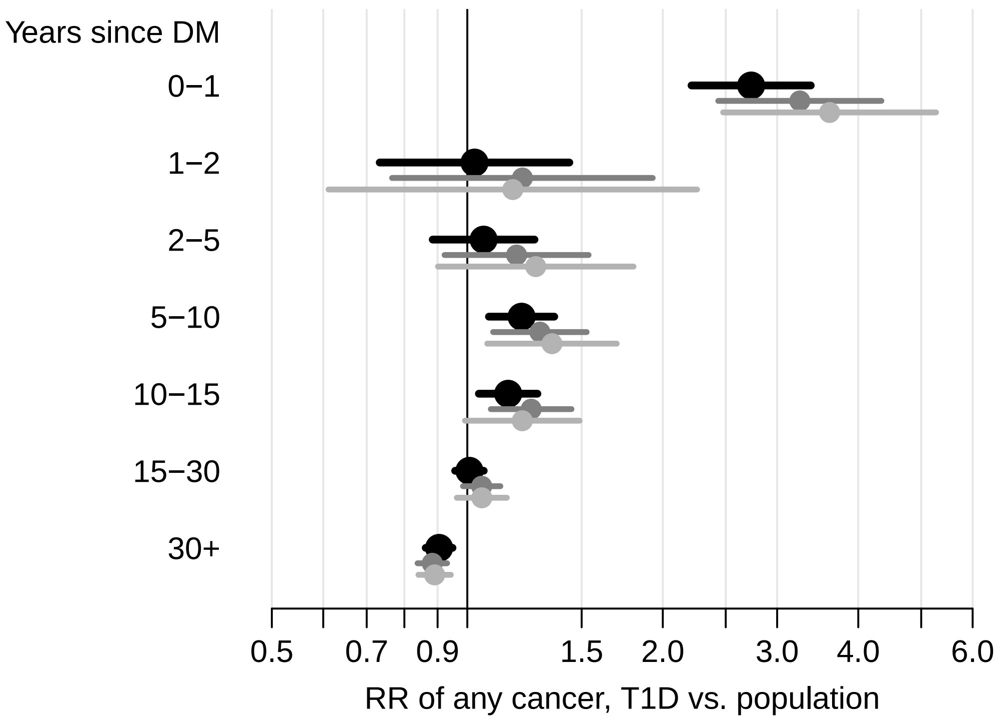

Figure 4.6: *The effect of duration on the RR of cancer. The two gray sets of effects are from the models where data are further restricted to patients diagnosed under 35 and 30, respectively.*

From figure 7.2 it seems that there is a strong ascertainment effect for T1DM as well as what have been shown for all diabetes under one, and it is not attenuated if a stricter definition of T1D is applied.

4.4.3 Site-specific analyses

4.4.3.1 Analyses of site specific cancers

With the above results in mind, we first set up an array to hold the resulting RRs for each of the sites that we analyse. We also do the analyses by sex, but also make a pooled analysis, except for the sex-specific cancers (including breast):

```
> wh <- c(9,46,15,16,18,19,47,20,22,25,37,27:30,32,33,35,36,39,40,42:45)
> ( vnam <- names(se.ana)[wh] )
 [1] "d0"  "d51" "d6"  "d7"  "d9"  "d10" "d52" "d11" "d13" "d16" "d29" "d18" "d19" "d20"
[15] "d22" "d24" "d25" "d27" "d28" "d32" "d33" "d36" "d37" "d38" "d40"
```

```
> cbind( wh, site <- conv[match(vnam,conv$NCnam),"Clab"] )
```

```
      wh
 [1,] "9"  "All sites"
 [2,] "46" "Oral etc."
 [3,] "15" "Oesophagus"
 [4,] "16" "Stomach"
 [5,] "18" "Colon"
 [6,] "19" "Rectum"
 [7,] "47" "Colorectal"
 [8,] "20" "Liver"
 [9,] "22" "Pancreas"
[10,] "25" "Lung"
[11,] "37" "Melanoma of skin"
[12,] "27" "Breast"
[13,] "28" "Cervix uteri"
[14,] "29" "Corpus uteri"
[15,] "30" "Ovary"
[16,] "32" "Prostate"
[17,] "33" "Testis"
[18,] "35" "Kidney"
[19,] "36" "Bladder"
[20,] "39" "Brain, CNS"
[21,] "40" "Thyroid"
[22,] "42" "Non-Hodgkin lymphoma"
[23,] "43" "Hodgkin lymphoma"
[24,] "44" "Multiple myeloma"
[25,] "45" "Leukaemia"
```

```
> RRtab <- NArray( list( site = site,
+                        sex = c(levels(se.ana$sex),"Both"),
+                       what = c("N.Pop","N.T1D","RR","lo","hi") ) )
> dimnames( RRtab )
```

```
$site
 [1] "All sites"            "Oral etc."            "Oesophagus"
 [4] "Stomach"              "Colon"                "Rectum"
 [7] "Colorectal"           "Liver"                "Pancreas"
[10] "Lung"                 "Melanoma of skin"     "Breast"
[13] "Cervix uteri"         "Corpus uteri"         "Ovary"
[16] "Prostate"             "Testis"               "Kidney"
[19] "Bladder"              "Brain, CNS"           "Thyroid"
[22] "Non-Hodgkin lymphoma" "Hodgkin lymphoma"     "Multiple myeloma"
[25] "Leukaemia"

$sex
[1] "M"    "F"    "Both"

$what
[1] "N.Pop" "N.T1D" "RR"    "lo"    "hi"
```

With this fixed we can the make a loop doing the analysis for all sites:

```
> system.time(
+ for( i in 1:length(vnam) )
+    {
+    cat( i, wh[i], site[i], "\n" )
+    aset <- se.ana[,c(vnam[i],
+                      paste("y",if(i==1) "0", sep=""),
+                      "A","P","T1","sex")]
+    names( aset )[1:2] <- c("D","Y")
+    mB <- glm( D ~ Ns  ( A, knots=a.kn ) +
+                   Ns( P  , knots=p.kn ) +
+                   Ns( P-A, knots=c.kn ) +
+                   T1,
+            offset = log(Y),
+            family = poisson,
+              data = aset )
+    mM <- update( mB, data=subset(aset,sex=="M") )
+    mF <- update( mB, data=subset(aset,sex=="F") )
+    RRtab[i,"M"   ,3:5] <- ci.exp( mM, subset="T1" )
+    RRtab[i,"F"   ,3:5] <- ci.exp( mF, subset="T1" )
+    RRtab[i,"Both",3:5] <- ci.exp( mB, subset="T1" )
+    RRtab[i,,1:2] <- addmargins( with( aset, tapply(D,list(sex,T1),sum) ), 1 )
+    } )
1 9 All sites
2 46 Oral etc.
3 15 Oesophagus
4 16 Stomach
5 18 Colon
6 19 Rectum
7 47 Colorectal
8 20 Liver
9 22 Pancreas
10 25 Lung
11 37 Melanoma of skin
12 27 Breast
13 28 Cervix uteri
14 29 Corpus uteri
15 30 Ovary
16 32 Prostate
17 33 Testis
18 35 Kidney
19 36 Bladder
20 39 Brain, CNS
21 40 Thyroid
22 42 Non-Hodgkin lymphoma
23 43 Hodgkin lymphoma
24 44 Multiple myeloma
25 45 Leukaemia
   user  system elapsed
 28.900   0.085  28.978
> RRorg <- RRtab
> RRtab <- RRorg
> for(i in 1:dim(RRtab)[1])
+ for(j in 1:dim(RRtab)[2]) if(          RRtab[i,j,5]==Inf  ) RRtab[i, j    ,] <- NA
> for(i in 1:dim(RRtab)[1]) if(any(is.na(RRtab[i, ,5]==Inf))) RRtab[i,"Both",] <- NA
> round( ftable( RRtab ), 2 )
                     what      N.Pop    N.T1D       RR       lo       hi
site                 sex
All sites            M     503211.00  1892.00     0.90     0.86     0.94
                     F     456245.00  2191.00     1.13     1.08     1.18
                     Both  959456.00  4083.00     1.01     0.98     1.04
Oral etc.            M      11894.00    76.00     1.17     0.93     1.46
                     F       6611.00    36.00     1.21     0.87     1.68
                     Both   18505.00   112.00     1.21     1.01     1.46
Oesophagus           M       6253.00    29.00     1.04     0.72     1.49
                     F       2301.00    14.00     2.01     1.19     3.41
                     Both    8554.00    43.00     1.30     0.96     1.75
```

Stomach	M	15546.00	64.00	1.21	0.95	1.55
	F	9424.00	39.00	1.47	1.07	2.02
	Both	24970.00	103.00	1.34	1.10	1.62
Colon	M	35888.00	174.00	1.26	1.08	1.46
	F	36313.00	119.00	1.10	0.92	1.32
	Both	72201.00	293.00	1.20	1.07	1.34
Rectum	M	24608.00	85.00	0.82	0.67	1.02
	F	18648.00	59.00	0.91	0.71	1.18
	Both	43256.00	144.00	0.88	0.74	1.03
Colorectal	M	60496.00	259.00	1.07	0.95	1.21
	F	54961.00	178.00	1.03	0.89	1.19
	Both	115457.00	437.00	1.07	0.97	1.17
Liver	M	7689.00	44.00	1.44	1.07	1.94
	F	4810.00	17.00	1.23	0.76	1.98
	Both	12499.00	61.00	1.42	1.10	1.82
Pancreas	M	11655.00	52.00	1.08	0.82	1.42
	F	11797.00	30.00	0.84	0.59	1.20
	Both	23452.00	82.00	0.98	0.79	1.22
Lung	M	42561.00	135.00	0.79	0.67	0.94
	F	30092.00	128.00	1.01	0.85	1.20
	Both	72653.00	263.00	0.90	0.79	1.01
Melanoma of skin	M	20955.00	124.00	0.92	0.77	1.10
	F	21019.00	105.00	0.80	0.66	0.97
	Both	41974.00	229.00	0.86	0.76	0.98
Breast	M	791.00	4.00	1.11	0.41	2.96
	F	133697.00	713.00	0.95	0.89	1.03
	Both	134488.00	717.00	0.88	0.82	0.94
Cervix uteri	M	NA	NA	NA	NA	NA
	F	11115.00	85.00	1.13	0.91	1.40
	Both	NA	NA	NA	NA	NA
Corpus uteri	M	NA	NA	NA	NA	NA
	F	28485.00	151.00	1.37	1.16	1.60
	Both	NA	NA	NA	NA	NA
Ovary	M	NA	NA	NA	NA	NA
	F	20622.00	114.00	1.19	0.99	1.44
	Both	NA	NA	NA	NA	NA
Prostate	M	169292.00	344.00	0.51	0.46	0.56
	F	NA	NA	NA	NA	NA
	Both	NA	NA	NA	NA	NA
Testis	M	6301.00	57.00	0.98	0.76	1.28
	F	NA	NA	NA	NA	NA
	Both	NA	NA	NA	NA	NA
Kidney	M	13733.00	62.00	0.96	0.75	1.23
	F	9379.00	37.00	1.17	0.84	1.61
	Both	23112.00	99.00	1.05	0.86	1.28
Bladder	M	37452.00	125.00	0.87	0.73	1.04
	F	13025.00	35.00	0.87	0.62	1.21
	Both	50477.00	160.00	0.92	0.79	1.08
Brain, CNS	M	14209.00	79.00	0.86	0.69	1.08
	F	16193.00	98.00	1.09	0.90	1.33
	Both	30402.00	177.00	0.97	0.84	1.13
Thyroid	M	2203.00	15.00	1.01	0.61	1.68
	F	5531.00	51.00	1.45	1.10	1.91
	Both	7734.00	66.00	1.28	1.00	1.63
Non-Hodgkin lymphoma	M	17115.00	111.00	1.29	1.07	1.55
	F	13690.00	56.00	1.07	0.82	1.39
	Both	30805.00	167.00	1.22	1.05	1.42
Hodgkin lymphoma	M	2474.00	20.00	1.17	0.75	1.82
	F	1911.00	9.00	0.82	0.42	1.58
	Both	4385.00	29.00	1.04	0.72	1.50
Multiple myeloma	M	7137.00	26.00	0.86	0.58	1.26
	F	5706.00	13.00	0.72	0.42	1.25
	Both	12843.00	39.00	0.82	0.60	1.13
Leukaemia	M	13480.00	38.00	0.61	0.44	0.84
	F	9819.00	28.00	0.76	0.52	1.10
	Both	23299.00	66.00	0.68	0.53	0.86

Of course we would also like to see the results as a forest plot, so we extract the relevant quantities for doing this:

```
> eM <- RRtab[,"M",3:5]
> eF <- RRtab[,"F",3:5]
> eB <- RRtab[,"Both",3:5]
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( eB, y=nrow(eM):1, txtpos=nrow(eM):1,
+          col="lightgray", xlog=TRUE,
+          xtic=c(1:10/10,1.5,2:7), xlim=c(0.05,7),
+          grid=TRUE, vref=1, xlab="Cancer incidence RR, T1D vs. population" )
> linesEst( eF, y=nrow(eM):1-0.2, col="red" )
> linesEst( eM, y=nrow(eM):1+0.2, col="blue" )
> text( rep(0.2,dim(RRtab)[1]), dim(RRtab)[1]:1+0.2, RRtab[,"M",2], col="blue", adj=1, cex=0.7 )
> text( rep(0.2,dim(RRtab)[1]), dim(RRtab)[1]:1-0.2, RRtab[,"F",2], col="red" , adj=1, cex=0.7 )
```

We see that the only sites with appreciable increased RR and sufficiently narrow confidence intervals are colon, liver (men only), corpus uteri and thyroid (women only); whereas there seems to a lower risk of melanoma, breast and prostate cancer and leukaemia among T1D patients.

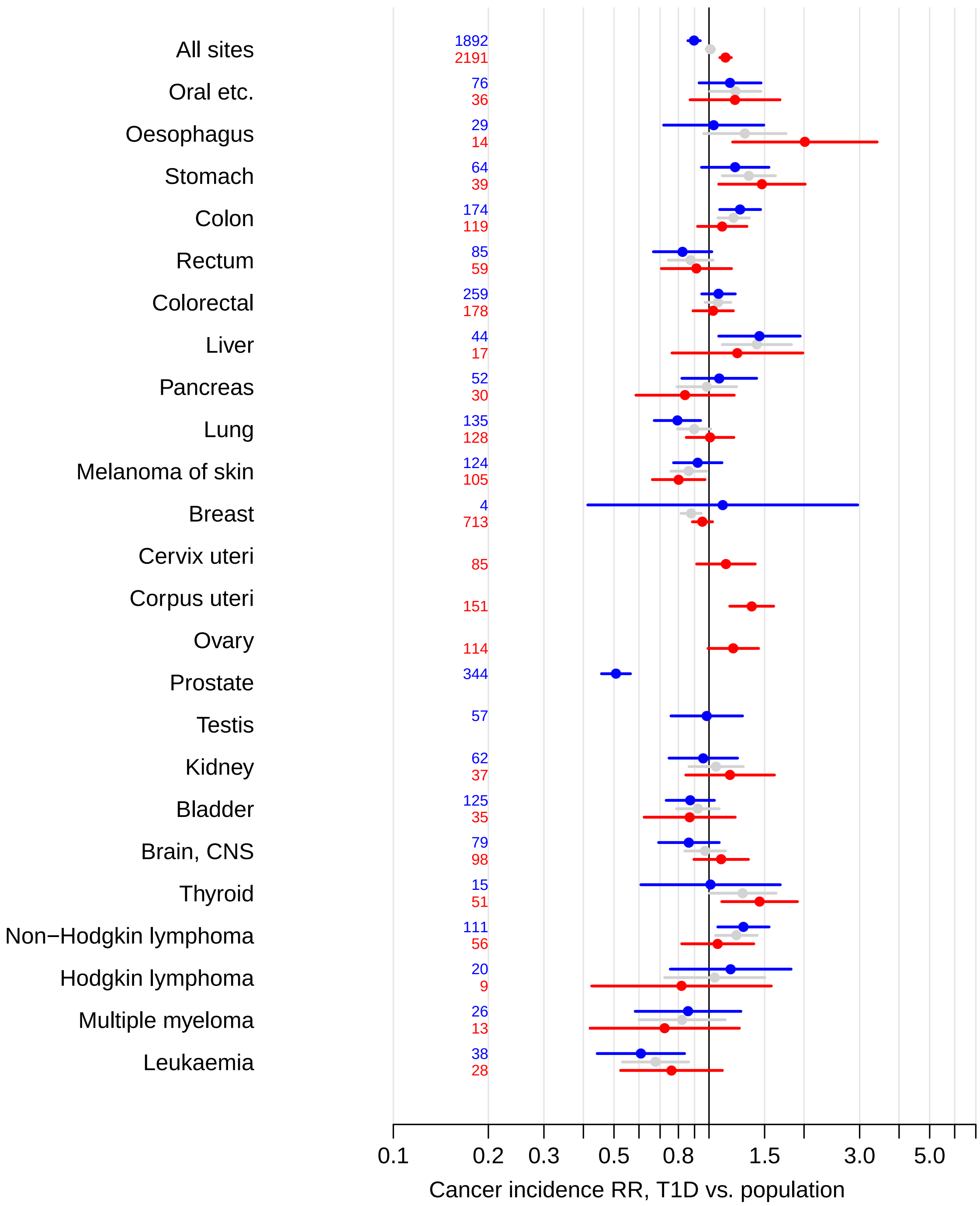

Figure 4.7: *RRs of cancer incidence among T1D patients (i.e. diagnosed* $<$ *40 years of age) in Sweden relative to the general population. The numbers to the left are the number of cancers observed among the T1D patients. Men: Blue, Women: Red, Both sexes: Light gray.*

Chapter 5

Finnish T1D data

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

5.1 The updated dataset

The Finnish dataset has been updated at the end of May 2013; now compatible with the agreed standard, but with the twist that we need colorectal cancer as group in `d52`

```
> clear()
> load( file="../data/FInew.Rda" )
> lls()
  name   mode class      size
1 fi.ana list data.frame 30538 50
> names( fi.ana )
 [1] "sex"     "A"       "P"       "DMdur"   "T1D"     "DMprev"  "y0"      "y"
 [9] "d0"      "d1"      "d2"      "d3"      "d4"      "d5"      "d6"      "d7"
[17] "d8"      "d9"      "d10"     "d11"     "d12"     "d13"     "d14"     "d15"
[25] "d16"     "d17"     "d18"     "d19"     "d20"     "d21"     "d22"     "d23"
[33] "d24"     "d25"     "d26"     "d27"     "d28"     "d29"     "d31"     "d32"
[41] "d33"     "d34"     "d35"     "d36"     "d37"     "d38"     "d40"     "d48"
[49] "DMdur.1" "T1"
> fi.ana <- fi.ana[,-49]
> fi.ana$d52 <- fi.ana$d9 + fi.ana$d10
```

Also we note that we have the right coding of the varibales:

```
> levels( fi.ana$DMprev )[3] <- "Prv"
> with( fi.ana, ftable( DMdur, DMprev, T1D, row.vars=1 ) )
      DMprev  Pop                     Inc                     Prv
      T1D     NoDM   30   35   40 NoDM   30   35   40 NoDM   30   35   40
DMdur
NoDM          1360    0    0    0    0    0    0    0    0    0    0    0
0                0    0    0    0    0 2471  535  540    0   59   12   12
1                0    0    0    0    0 2413  520  524    0  119   26   25
2                0    0    0    0    0 2500  654  657    0  312   77   75
5                0    0    0    0    0 2431  727  731    0  636  163  161
10               0    0    0    0    0 2076  619  620    0  636  163  161
15               0    0    0    0    0 2038  817  827    0 1326  383  379
30               0    0    0    0    0  706  222  222    0 1015  298  290
```

Finally note that the numbering of the cancer sites already is as in NordCAN:

```
> load( file="../data/conv.Rda" )
> wh.fi <- which( substr( names(fi.ana), 1, 1 )=="d" )
> tum <- cbind( apply( fi.ana[fi.ana$T1D!="NoDM",wh.fi], 2, sum, na.rm=TRUE ),
+               apply( fi.ana[fi.ana$T1D=="NoDM",wh.fi], 2, sum, na.rm=TRUE ) )
> dd <- data.frame( tum,
+         round(tum/tum[rep(1,nrow(tum)),]*100,2),
+         site = conv[match(rownames(tum),conv$NCnam),"Clab"],
+         stringsAsFactors=FALSE)
> names( dd ) <- c("N DM","N pop","% DM","% pop")
> dd
    N DM  N pop   % DM  % pop                          NA
d0  2408 707026 100.00 100.00                   All sites
d1    10   4959   0.42   0.70                         Lip
d2    14   2857   0.58   0.40                      Tongue
d3    11   1755   0.46   0.25             Salivary glands
d4    20   2791   0.83   0.39                       Mouth
d5    12   2916   0.50   0.41                     Pharynx
d6    21   8114   0.87   1.15                  Oesophagus
d7    97  35881   4.03   5.07                     Stomach
d8    13   2488   0.54   0.35             Small intestine
d9   122  39333   5.07   5.56                       Colon
d10   87  27486   3.61   3.89                      Rectum
d11   42   9607   1.74   1.36                       Liver
d12   18   8738   0.75   1.24                 Gallbladder
d13   95  26541   3.95   3.75                    Pancreas
d14    5   1288   0.21   0.18              Nose, sinuses
d15   15   4866   0.62   0.69                      Larynx
d16  177  85554   7.35  12.10                        Lung
d17    5   2068   0.21   0.29                      Pleura
d18  548 102164  22.76  14.45                      Breast
d19   36   6803   1.50   0.96                Cervix uteri
d20   97  22465   4.03   3.18                Corpus uteri
d21    1    505   0.04   0.07                        <NA>
d22   80  16605   3.32   2.35                       Ovary
d23    7   3197   0.29   0.45  Other female genital organs
d24  148  88045   6.15  12.45                    Prostate
d25   23   2947   0.96   0.42                      Testis
d26    4    764   0.17   0.11                  Penis etc.
d27  114  22598   4.73   3.20                      Kidney
d28   65  25679   2.70   3.63                     Bladder
d29  119  21261   4.94   3.01            Melanoma of skin
d31    9   1989   0.37   0.28                         Eye
d32   67  25224   2.78   3.57                  Brain, CNS
d33  122  11315   5.07   1.60                     Thyroid
d34   10   1820   0.42   0.26                        Bone
d35   29   5031   1.20   0.71                Soft tissues
d36  102  26149   4.24   3.70         Non-Hodgkin lymphoma
d37   23   5081   0.96   0.72             Hodgkin lymphoma
d38   22   9667   0.91   1.37             Multiple myeloma
d40   72  18558   2.99   2.62                   Leukaemia
d48   87  21917   3.61   3.10 Other and unspecified cancers
d52  209  66819   8.68   9.45                  Colorectal
```

We see that the only major discrepancy in the site-distribution of tumours between is for Thyroid cancers.

There are also cases with unknown duration (although they have e acoded duration), we tag them as "`Unkn`" in the `DMdur` variable:

```
> round( ftable( xtabs( cbind( d0, PY=y/1000 ) ~ DMdur + T1D + DMprev, data = fi.ana ),
+                 col.vars=3:2, row.vars=c(4,1) ) )
         DMprev     Pop                           Inc                           Prv
         T1D       NoDM     30      35      40   NoDM     30      35      40   NoDM     30      35      40
   DMdur
d0 NoDM          707026      0       0       0      0      0       0       0      0      0       0       0
   0                  0      0       0       0      0     25      14      38      0      0       0       0
```

```
   1               0      0      0      0      0     10     14     21      0      0      2      1
   2               0      0      0      0      0     30     35     56      0      0      0      2
   5               0      0      0      0      0     60     57    128      0      2      2      3
   10              0      0      0      0      0     58     86    151      0      6      6     10
   15              0      0      0      0      0    349    270    358      0     53     23     47
   30              0      0      0      0      0    147     77     81      0    125     26     35
PY NoDM       198320      0      0      0      0      0      0      0      0      0      0      0
   0               0      0      0      0      0     41     15     18      0      0      0      0
   1               0      0      0      0      0     39     14     17      0      1      0      0
   2               0      0      0      0      0    104     36     44      0      6      1      1
   5               0      0      0      0      0    138     45     55      0     14      2      3
   10              0      0      0      0      0    105     32     37      0     14      2      3
   15              0      0      0      0      0    185     50     50      0     40      5      6
   30              0      0      0      0      0     32      7      6      0     28      3      3
```

```
> fi.ana$DMdur <- with( fi.ana, factor( ifelse( DMprev=="Prv1972",
+                                               "Unkn",
+                                               as.character(DMdur) ),
+                                       levels = c("NoDM","Unkn",levels(DMdur)[-1]) ) )
> round( ftable( xtabs( cbind( d0, PY=y/1000 ) ~ DMdur + T1D + DMprev, data = fi.ana ),
+                col.vars=3:2, row.vars=c(4,1) ) )
        DMprev    Pop                        Inc                        Prv
        T1D      NoDM     30     35     40  NoDM     30     35     40  NoDM     30     35     40
   DMdur
d0 NoDM        707026      0      0      0      0      0      0      0      0      0      0      0
   Unkn            0      0      0      0      0      0      0      0      0      0      0      0
   0               0      0      0      0      0     25     14     38      0      0      0      0
   1               0      0      0      0      0     10     14     21      0      0      2      1
   2               0      0      0      0      0     30     35     56      0      0      0      2
   5               0      0      0      0      0     60     57    128      0      2      2      3
   10              0      0      0      0      0     58     86    151      0      6      6     10
   15              0      0      0      0      0    349    270    358      0     53     23     47
   30              0      0      0      0      0    147     77     81      0    125     26     35
PY NoDM       198320      0      0      0      0      0      0      0      0      0      0      0
   Unkn            0      0      0      0      0      0      0      0      0      0      0      0
   0               0      0      0      0      0     41     15     18      0      0      0      0
   1               0      0      0      0      0     39     14     17      0      1      0      0
   2               0      0      0      0      0    104     36     44      0      6      1      1
   5               0      0      0      0      0    138     45     55      0     14      2      3
   10              0      0      0      0      0    105     32     37      0     14      2      3
   15              0      0      0      0      0    185     50     50      0     40      5      6
   30              0      0      0      0      0     32      7      6      0     28      3      3
```

Also we make a brife overview of the data by calendar time:

```
> round( ftable( xtabs( cbind( d0, PY=y/1000 ) ~ floor(P) + T1D, data = fi.ana ),
+                col.vars=3:2 ) )
              d0                      PY
        T1D  NoDM    30    35    40  NoDM    30    35    40
floor(P)
1972        11487     0     2     2  4627     3     1     1
1973        11559     0     1     1  4651     4     1     1
1974        11773     1     1     1  4674     5     1     1
1975        12060     1     1     2  4693     6     1     2
1976        12587     0     0     0  4705     7     1     2
1977        12903     0     3     1  4717     8     2     2
1978        13158     6     1     4  4729     8     2     2
1979        13500     3     1     5  4740     9     2     2
1980        13892     3     1     5  4753    10     2     3
1981        13961     8     8     4  4771    11     3     3
1982        14320     6     4    13  4796    11     3     3
1983        14546     7     5     7  4822    12     3     3
1984        14648     5     2     8  4845    13     3     4
1985        14978     5     3     7  4863    13     4     4
1986        15404     8     6    10  4877    14     4     4
1987        15707    11     6     5  4888    15     4     4
1988        15718     9     6    15  4901    15     4     5
```

```
1989           15658    19    10    18  4916    16     4     5
1990           16034    11     8    14  4936    17     4     5
1991           16489     4     5    11  4960    17     5     5
1992           17184    14    16    27  4985    18     5     5
1993           17196    15    11    18  5007    18     5     5
1994           17886    15    11    24  5026    19     5     6
1995           18267    17    20    25  5042    20     5     6
1996           19296    14    16    26  5056    21     6     7
1997           19312    27    21    37  5069    21     6     7
1998           19747    27    13    31  5080    22     6     7
1999           19959    21    21    31  5089    23     7     8
2000           20389    34    20    32  5098    24     7     8
2001           20715    32    21    34  5109    25     7     9
2002           21495    30    21    37  5120    26     8     9
2003           22003    38    22    37  5131    28     8     9
2004           23410    38    25    47  5145    29     9    10
2005           23905    48    45    49  5160    30     9    11
2006           23697    47    36    57  5175    32    10    11
2007           23140    52    36    54  5193    33    10    12
2008           23689    63    46    51  5212    35    11    13
2009           24873    66    46    72  5233    36    12    13
2010           24879    77    47    53  5252    37    12    14
2011           25602    83    44    56  5272    33    11    13
2012               0     0     0     0     0     0     0     0
```

We actually have population rates and follow-up all the way back to 1972, so we include the entire peiod in the analysis dataset.

Finally we include a country variable, `Cnt` and save the dataset.

```
> fi.ana$Cnt <- "FIN"
> save( fi.ana, file="../data/FIana.Rda" )
```

5.2 Analysis of Finnish data

5.2.1 All cancers

For the sake of choice of parameters we list the quantiles of the events in the population:

```
> load( file="../data/FIana.Rda" )
> with( subset(fi.ana,T1D=="NoDM"),
+       rbind( A = quantile( rep(  A,d0), probs=0:10/10 ),
+              P = quantile( rep(P  ,d0), probs=0:10/10 ),
+              C = quantile( rep(P-A,d0), probs=0:10/10 ) ) )
      0%    10%    20%    30%    40%    50%    60%    70%    80%    90%   100%
A    2.5   47.5   52.5   57.5   62.5   67.5   72.5   72.5   77.5   82.5   82.5
P 1972.5 1977.5 1983.5 1987.5 1992.5 1996.5 1999.5 2003.5 2006.5 2009.5 2011.5
C 1890.0 1908.0 1915.0 1920.0 1925.0 1929.0 1933.0 1939.0 1945.0 1952.0 2009.0
```

We first set out the simplest possible analysis with age-period-cohort effects for the baseline rates. First we devise a couple of knots for the splines:

```
> library( splines )
> ( a.kn <- seq(5,80,,7) )
[1]  5.0 17.5 30.0 42.5 55.0 67.5 80.0
> ( p.kn <- seq(1975,2009,,5) )
[1] 1975.0 1983.5 1992.0 2000.5 2009.0
> ( c.kn <- seq(1895,1990,,6) )
[1] 1895 1914 1933 1952 1971 1990
```

We then fit 3 models for the RR for T1D patients relative to the general population, using a common shape of the underlying cancer incidence rates as an age-period cohort model with $1 + (7 - 1) + (5 - 1) + (6 - 1) - 1 = 15$ parameters[1]. The first model (`m3`) is one with

[1]There is first an intercept, then the three natural splines, where k knots gives $k - 1$ parameters, and finally 1 aliased parameter from the linear relationship between $P - A$ and P and A.

separate effects for persons diagnosed in ages < 30, 30–35 and 35–40, the second (`m1`) is a simplification where the three groups are pooled, and the third (`mc`) is an extension of `m3` where each group is subdivided by diabetes status at 1.1.1995:

```
> m3 <- glm( d0 ~ Ns(   A, knots=a.kn ) +
+                 Ns( P   , knots=p.kn ) +
+                 Ns( P-A, knots=c.kn ) +
+                 T1D,
+            offset = log(y0),
+            family = poisson,
+              data = fi.ana )
> m1 <- update( m3, . ~ . - T1D + Relevel(T1D,list(1,2:4)) )
> mc <- update( m3, . ~ . - T1D + factor(interaction(T1D,DMprev)) )
> anova( mc, m3, m1, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    factor(interaction(T1D, DMprev))
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
Model 3: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    Relevel(T1D, list(1, 2:4))
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     30517      70539
2     30520      70543 -3  -4.2970   0.2311
3     30522      70543 -2  -0.1201   0.9417
```

The tests of the models show that there is no substantial difference between them, and that does not either shown up if we list the estimated RRs from the models:

```
> round( RR <- rbind( ci.exp( m3, subset="T1" ),
+                     ci.exp( mc, subset="T1" ),
+                     ci.exp( m1, subset="T1" ) ), 3 )
                                       exp(Est.)  2.5% 97.5%
T1D30                                      1.085 1.015 1.160
T1D35                                      1.102 1.018 1.193
T1D40                                      1.084 1.017 1.157
factor(interaction(T1D, DMprev))30.Inc     1.106 1.026 1.193
factor(interaction(T1D, DMprev))35.Inc     1.130 1.040 1.228
factor(interaction(T1D, DMprev))40.Inc     1.090 1.018 1.166
factor(interaction(T1D, DMprev))30.Prv     1.015 0.879 1.172
factor(interaction(T1D, DMprev))35.Prv     0.897 0.695 1.158
factor(interaction(T1D, DMprev))40.Prv     1.042 0.855 1.270
Relevel(T1D, list(1, 2:4))30+35+40         1.089 1.046 1.134
```

In summary, the comparison of the models show that there is no heterogeneity between the six groups of T1D patients w.r.t. the occurrence of cancer; which is also evident in figure 7.1, where the RR estimates from the three models are shown together:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( RR, y = c(4,3,2,c(4,3,2)-0.25,c(4,3,2)-0.4,0.5),
+          txt = c("","","",
+                  "<30","30-35","35-40",
+                  "","","","Pooled"),
+          xlog=TRUE, xtic=c(5:20)/10, grid=TRUE,
+          xlab="RR of any cancer, T1D vs. population",
+          lwd=3, cex=1.5, vref=1, col=rep(gray(c(4,7,0)/10),c(3,6,1)) )
```

5.2.2 Analysis by duration

We have also devised a variable indicating different time bands after diagnosis:

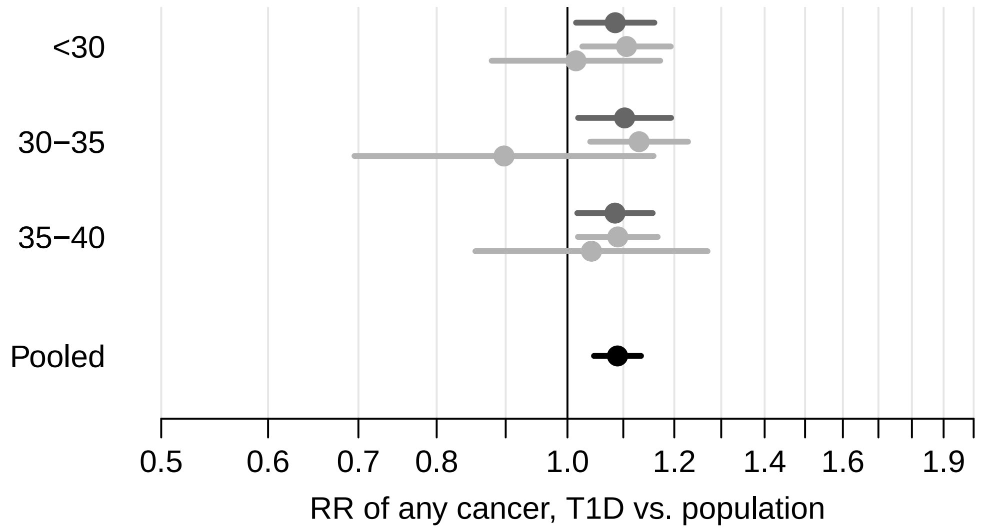

Figure 5.1: *Finland: Estimated RRs relative to the general population, the different shades of gray correpsond to the different models. The upper of the light gray bars are for T1D patients diagnosed after 1972, the lower for those diagnosed before 1.1.1972. There are no significant differences anywhere.*

```
> round(
+ ftable( xtabs( cbind(d0,y0=y0/1000) ~ DMdur + DMprev, data=fi.ana ),
+          col.vars=3:2 ), 1 )
                   d0                             y0
      DMprev      Pop      Inc      Prv      Pop      Inc      Prv
DMdur
NoDM         707026.0      0.0      0.0 198320.5      0.0      0.0
Unkn              0.0      0.0      0.0      0.0      0.0      0.0
0                 0.0     77.0      0.0      0.0     73.0      0.5
1                 0.0     45.0      3.0      0.0     69.0      1.3
2                 0.0    121.0      2.0      0.0    182.6      8.6
5                 0.0    245.0      7.0      0.0    236.0     19.1
10                0.0    295.0     22.0      0.0    172.3     18.6
15                0.0    977.0    123.0      0.0    279.1     50.7
30                0.0    305.0    186.0      0.0     42.6     31.4
```

The above analyses showed virtually no difference between the groups of patients by age at inclusion, so we pool these groups and restrict the analysis to DM patients diagnosed after 1972. Also we fit models with smaller datasets corresponding to more restrictive definitions of T1D:

```
> levels( fi.ana$T1D )
[1] "NoDM" "30"   "35"   "40"
> levels( fi.ana$DMprev )
[1] "Pop" "Inc" "Prv"
> md40 <- update( m3, . ~ . - T1D + DMdur,
+                        data=subset( fi.ana, DMprev!="Prv1972" & y0>0 ) )
> md35 <- update( md40, data=subset( fi.ana, DMprev!="Prv1972" & y0>0 & T1D %in% levels(T1D)[1:3] )
> md30 <- update( md40, data=subset( fi.ana, DMprev!="Prv1972" & y0>0 & T1D %in% levels(T1D)[1:2] )
> anova( md40, update( md40, . ~ . - DMdur + DMprev ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
```

```
    DMprev
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     30516      70512
2     30521      70540 -5  -28.187 3.346e-05
> anova( md35, update( md35, . ~ . - DMdur + DMprev ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMprev
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     25292      67915
2     25297      67927 -5  -12.151  0.03278
> anova( md30, update( md30, . ~ . - DMdur + DMprev ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMprev
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     20076      65997
2     20081      66014 -5  -16.336 0.005946
```

We see that there is significant effects of duration, for all three definitions of diabetes:

```
> round( cbind( RR40 <- ci.exp( md40, subset="DMdur" ),
+               RR35 <- ci.exp( md35, subset="DMdur" ),
+               RR30 <- ci.exp( md30, subset="DMdur" ) ), 2 )
        exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
DMdur0       2.00 1.60  2.50      1.98 1.44  2.71      2.45 1.65  3.63
DMdur1       1.20 0.91  1.60      1.28 0.87  1.89      0.96 0.52  1.79
DMdur2       0.98 0.82  1.16      1.02 0.80  1.30      0.93 0.65  1.33
DMdur5       1.08 0.95  1.22      1.03 0.86  1.23      1.06 0.83  1.36
DMdur10      1.18 1.06  1.32      1.13 0.96  1.32      0.91 0.71  1.16
DMdur15      1.08 1.02  1.15      1.12 1.04  1.20      1.14 1.03  1.26
DMdur30      1.00 0.92  1.09      1.01 0.91  1.12      1.04 0.92  1.17
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind(NA,RR40),
+          txt=c("Years since DM","0-1","1-2","2-5","5-10","10-15","15-30","30+"),
+              xlog=TRUE, xtic=c(3:10,15,20,25,30)/10, grid=TRUE,
+              xlab="RR of any cancer, T1D vs. population",
+              lwd=4, cex=2, vref=1, y=c(6.7,6:0)+0.2 )
> linesEst( rbind(RR35),
+            lwd=3, cex=1.5, col=gray(0.5), y=6:0 )
> linesEst( rbind(RR30),
+            lwd=3, cex=1.5, col=gray(0.7), y=6:0-0.15 )
```

From figure 7.2 we see an ascertainment effect for T1DM as well as what have been shown for all diabetes under one, and it seems to be the same even if a stricter definition of T1D is applied.

5.2.3 Site-specific analyses

5.2.3.1 Analyses of site specific cancers

We first set up an array to hold the resulting RRs for each of the sites that we analyse, we do the analyses by sex, but also make a pooled analysis, except for the sex-specific cancers (including breast):

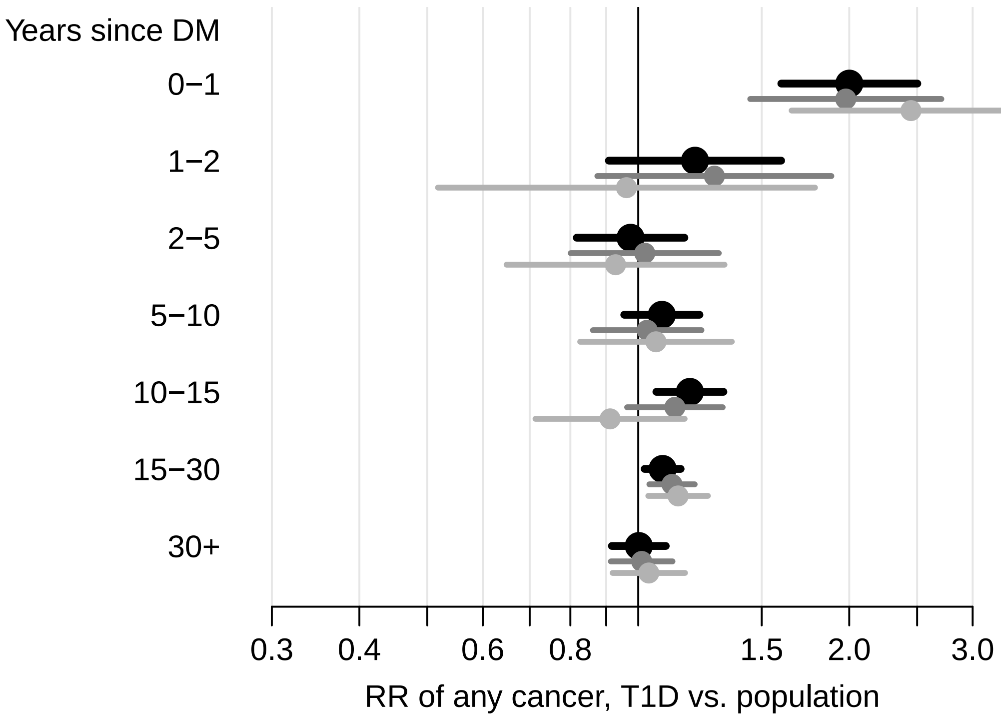

Figure 5.2: *Finland: The effect of duration on the RR of cancer. The two gray sets of effects are from the models where data are further restricted to patients diagnosed under 35 and 30, respectively.*

```
> wh <- c(9,15,16,18,19,20,22,25,38,27:29,31,33,34,36,37,40,41,44:47)
> ( vnam <- names(fi.ana)[wh] )
 [1] "d0"  "d6"  "d7"  "d9"  "d10" "d11" "d13" "d16" "d29" "d18" "d19" "d20" "d22" "d24"
[15] "d25" "d27" "d28" "d32" "d33" "d36" "d37" "d38" "d40"
> site <- conv[match(vnam,conv$NCnam),"Clab"]
> data.frame( wh, vnam, site )
   wh vnam                site
1   9   d0           All sites
2  15   d6          Oesophagus
3  16   d7             Stomach
4  18   d9               Colon
5  19  d10              Rectum
6  20  d11               Liver
7  22  d13            Pancreas
8  25  d16                Lung
9  38  d29     Melanoma of skin
10 27  d18              Breast
11 28  d19        Cervix uteri
12 29  d20        Corpus uteri
13 31  d22               Ovary
14 33  d24            Prostate
15 34  d25              Testis
16 36  d27              Kidney
17 37  d28             Bladder
18 40  d32          Brain, CNS
19 41  d33             Thyroid
20 44  d36 Non-Hodgkin lymphoma
21 45  d37     Hodgkin lymphoma
22 46  d38     Multiple myeloma
23 47  d40           Leukaemia
> RRtab <- NArray( list( site = site,
+                         sex = c(levels(fi.ana$sex),"Both"),
```

```
+                           what = c("N.pop","N.T1","RR","lo","hi") ) )
> dimnames( RRtab )
$site
 [1] "All sites"             "Oesophagus"            "Stomach"
 [4] "Colon"                 "Rectum"                "Liver"
 [7] "Pancreas"              "Lung"                  "Melanoma of skin"
[10] "Breast"                "Cervix uteri"          "Corpus uteri"
[13] "Ovary"                 "Prostate"              "Testis"
[16] "Kidney"                "Bladder"               "Brain, CNS"
[19] "Thyroid"               "Non-Hodgkin lymphoma" "Hodgkin lymphoma"
[22] "Multiple myeloma"      "Leukaemia"

$sex
[1] "M"    "F"    "Both"

$what
[1] "N.pop" "N.T1"  "RR"    "lo"    "hi"
```

With this fixed we can the make a loop doing the analysis for all sites:

```
> system.time(
+ for( i in 1:length(vnam) )
+    {
+    aset <- fi.ana[,c(vnam[i],
+                      paste("y",if(i==1) "0", sep=""),
+                      "A","P","T1","sex")]
+    names( aset )[1:2] <- c("D","Y")
+    mB <- glm( D ~ Ns  ( A, knots=a.kn ) +
+                   Ns( P  , knots=p.kn ) +
+                   Ns( P-A, knots=c.kn ) +
+                   T1,
+            offset = log(Y),
+            family = poisson,
+              data = aset )
+    mM <- update( mB, data=subset(aset,sex=="M") )
+    mF <- update( mB, data=subset(aset,sex=="F") )
+    RRtab[i,"M"    ,3:5] <- ci.exp( mM, subset="T1" )
+    RRtab[i,"F"    ,3:5] <- ci.exp( mF, subset="T1" )
+    RRtab[i,"Both",3:5] <- ci.exp( mB, subset="T1" )
+    RRtab[i,,1:2] <- addmargins( with( aset, tapply(D,list(sex,T1),sum) ), 1 )
+    } )
   user  system elapsed
  59.78    0.98   60.87
> RRorg <- RRtab
> RRtab <- RRorg
> for(i in 1:dim(RRtab)[1])
+ for(j in 1:dim(RRtab)[2]) if(          RRtab[i,j,5]==Inf  ) RRtab[i, j    ,] <- NA
> for(i in 1:dim(RRtab)[1]) if(any(is.na(RRtab[i, ,5]==Inf))) RRtab[i,"Both",] <- NA
> round( ftable( RRtab ), 2 )
                     what     N.pop      N.T1        RR        lo        hi
site                 sex
All sites            M    359417.00   1000.00      1.18      1.11      1.26
                     F    347609.00   1408.00      1.00      0.95      1.06
                     Both 707026.00   2408.00      1.09      1.05      1.13
Oesophagus           M      4723.00     16.00      1.33      0.81      2.18
                     F      3391.00      5.00      1.20      0.50      2.90
                     Both   8114.00     21.00      1.18      0.77      1.81
Stomach              M     20048.00     47.00      1.48      1.11      1.97
                     F     15833.00     50.00      1.83      1.39      2.43
                     Both  35881.00     97.00      1.61      1.32      1.97
Colon                M     17793.00     56.00      1.23      0.95      1.60
                     F     21540.00     66.00      1.15      0.90      1.46
                     Both  39333.00    122.00      1.18      0.99      1.41
Rectum               M     14678.00     46.00      1.31      0.98      1.75
                     F     12808.00     41.00      1.17      0.86      1.59
                     Both  27486.00     87.00      1.22      0.98      1.50
```

```
Liver                M       5809.00    34.00   2.65   1.88   3.71
                     F       3798.00     8.00   1.05   0.52   2.10
                     Both    9607.00    42.00   1.94   1.43   2.63
Pancreas             M      13050.00    54.00   1.89   1.45   2.47
                     F      13491.00    41.00   1.58   1.16   2.15
                     Both   26541.00    95.00   1.70   1.39   2.08
Lung                 M      68937.00   119.00   1.20   1.00   1.44
                     F      16617.00    58.00   1.15   0.89   1.49
                     Both   85554.00   177.00   1.14   0.99   1.33
Melanoma of skin     M      10682.00    60.00   1.17   0.90   1.50
                     F      10579.00    59.00   0.86   0.66   1.11
                     Both   21261.00   119.00   0.99   0.83   1.19
Breast               M        428.00     2.00   1.30   0.32   5.32
                     F     101736.00   546.00   0.88   0.81   0.96
                     Both  102164.00   548.00   1.00   0.92   1.09
Cervix uteri         M            NA       NA     NA     NA     NA
                     F       6803.00    36.00   0.99   0.71   1.38
                     Both         NA       NA     NA     NA     NA
Corpus uteri         M            NA       NA     NA     NA     NA
                     F      22465.00    97.00   1.29   1.05   1.57
                     Both         NA       NA     NA     NA     NA
Ovary                M            NA       NA     NA     NA     NA
                     F      16605.00    80.00   1.16   0.93   1.45
                     Both         NA       NA     NA     NA     NA
Prostate             M      88045.00   148.00   0.79   0.67   0.93
                     F            NA       NA     NA     NA     NA
                     Both         NA       NA     NA     NA     NA
Testis               M       2947.00    23.00   0.89   0.59   1.34
                     F            NA       NA     NA     NA     NA
                     Both         NA       NA     NA     NA     NA
Kidney               M      12724.00    67.00   1.75   1.37   2.22
                     F       9874.00    47.00   1.67   1.25   2.23
                     Both   22598.00   114.00   1.65   1.38   1.99
Bladder              M      19488.00    49.00   1.33   1.00   1.76
                     F       6191.00    16.00   1.34   0.82   2.19
                     Both   25679.00    65.00   1.24   0.97   1.58
Brain, CNS           M      10524.00    35.00   0.69   0.50   0.96
                     F      14700.00    32.00   0.38   0.27   0.54
                     Both   25224.00    67.00   0.51   0.40   0.64
Thyroid              M       2563.00    18.00   1.28   0.80   2.04
                     F       8752.00   104.00   1.54   1.27   1.87
                     Both   11315.00   122.00   1.60   1.34   1.92
Non-Hodgkin lymphoma M      13454.00    56.00   1.08   0.83   1.41
                     F      12695.00    46.00   0.92   0.69   1.23
                     Both   26149.00   102.00   0.99   0.81   1.20
Hodgkin lymphoma     M       2869.00    14.00   0.83   0.49   1.40
                     F       2212.00     9.00   0.64   0.33   1.23
                     Both    5081.00    23.00   0.72   0.48   1.09
Multiple myeloma     M       4625.00    14.00   1.35   0.80   2.28
                     F       5042.00     8.00   0.79   0.39   1.58
                     Both    9667.00    22.00   1.05   0.69   1.60
Leukaemia            M      10013.00    39.00   1.39   1.01   1.90
                     F       8545.00    33.00   1.30   0.92   1.83
                     Both   18558.00    72.00   1.32   1.05   1.67
```

Of course we would also like to see the results as a forest plot, so we extract the relevant quantities for doing this:

```
> eM <- RRtab[,"M",3:5]
> eF <- RRtab[,"F",3:5]
> eB <- RRtab[,"Both",3:5]
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( eB, y=nrow(eM):1, txtpos=nrow(eM):1,
+          col="lightgray", xlog=TRUE,
+          xtic=c(1:10/10,1.5,2:7), xlim=c(0.095,7),
+          grid=TRUE, vref=1, xlab="Cancer incidence RR, T1D vs. population" )
```

```
> linesEst( eF, y=nrow(eM):1-0.2, col="red" )
> linesEst( eM, y=nrow(eM):1+0.2, col="blue" )
> text( rep(0.095,dim(RRtab)[1]), dim(RRtab)[1]:1+0.2, RRtab[,"M",2], col="blue", adj=1, cex=0.7 )
> text( rep(0.095,dim(RRtab)[1]), dim(RRtab)[1]:1-0.2, RRtab[,"F",2], col="red" , adj=1, cex=0.7 )
```

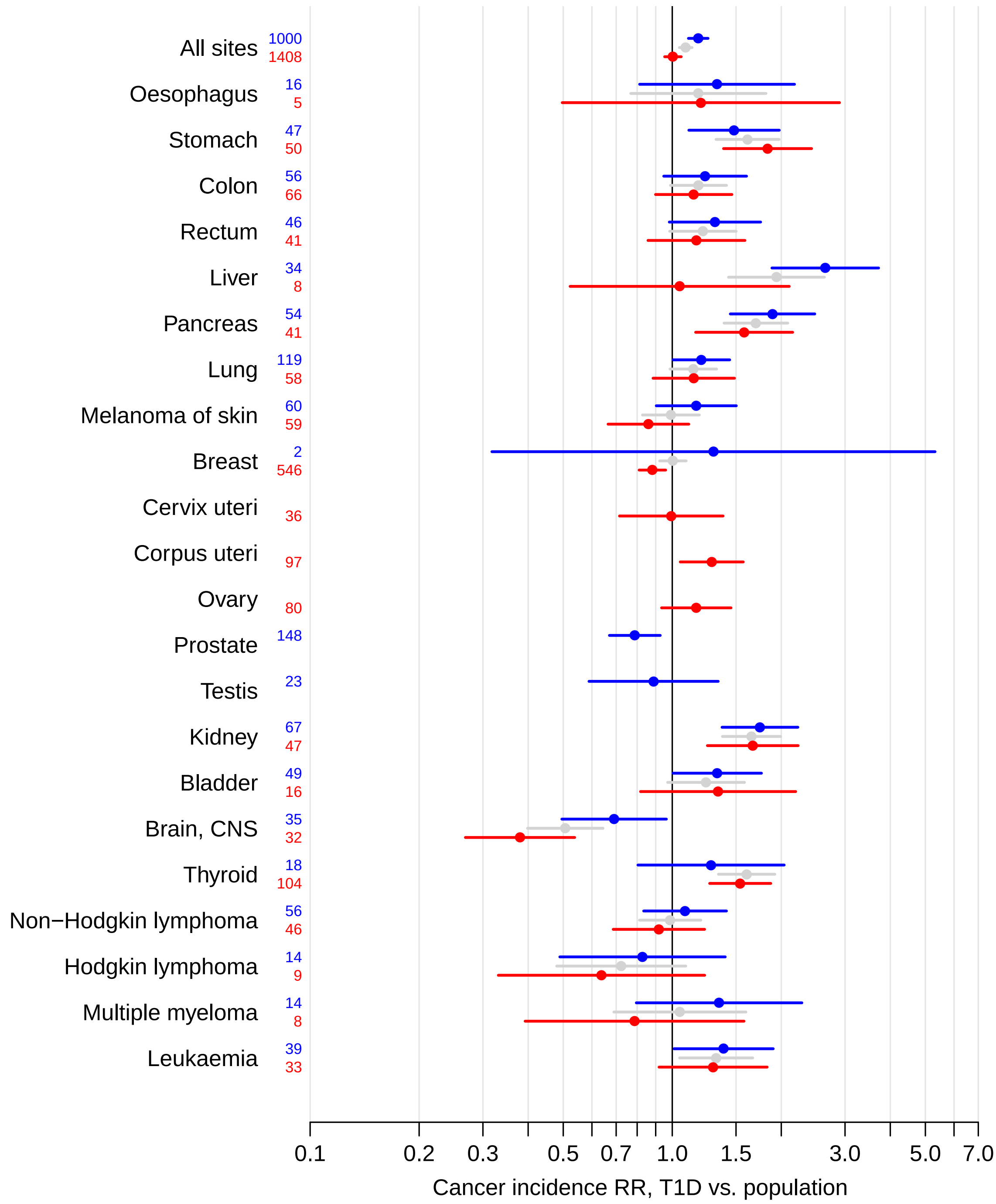

Figure 5.3: *RRs of cancer incidence among T1D patients (i.e. diagnosed* $<$ *40 years of age) in Sweden relative to the general population. The numbers to the left are the number of cancers observed among the T1D patients. Men: Blue, Women: Red, Both sexes: Light gray.*

We see that the only sites with appreaciable increased RR and sufficiently narrow confidence intervals are rectum, liver, pancreas, lung, corpus uteri, kidney and thyroid; whereas there seems to a lower risk for breast and prostate and brain cancer among T1D patients in Finland.

Chapter 6

Scottish T1D data

The following is a technical account of the approach to construction and analysis of data from Scotland.

```
> library( Epi )
> library( foreign )
> sessionInfo()
R version 3.2.3 (2015-12-10)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 14.04.3 LTS

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C               LC_TIME=en_US.UTF-8
 [4] LC_COLLATE=en_US.UTF-8     LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C                  LC_ADDRESS=C
[10] LC_TELEPHONE=C             LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

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

other attached packages:
[1] foreign_0.8-66 Epi_1.1.71

loaded via a namespace (and not attached):
[1] cmprsk_2.2-7     MASS_7.3-44      parallel_3.2.3   survival_2.38-3 etm_0.6-2       splines_3.2.3
[7] grid_3.2.3       lattice_0.20-31
```

6.1 Naming of sites and files

First we read the file with the SCottish coding of the sites in order to make sure that the naming is consistent with the standard for the study:

```
> scn <- read.table( "../data/scotcodes.txt" )
> names( scn ) <- c("site","code","N")
> scn$code <- tolower( gsub( "D0", "D", scn$code ) )
> head( scn )
      site code   N
1 Allsites   d0 533
2      Lip   d1   0
3   Tongue   d2   5
4 Salivary   d3   1
5    Mouth   d4   5
6  Pharynx   d5   1
```

To check the consistency with the reference we load this too:

```
> load( file="../data/conv.Rda" )
> head( conv )
  DKnam NCnam            Clab
1    d0    d0       All sites
2   d11    d1             Lip
3   d12    d2          Tongue
4   d14    d3 Salivary glands
5   d13    d4           Mouth
6   d15    d5         Pharynx
> merge( conv, scn, by.x="NCnam", by.y="code" )
   NCnam DKnam                           Clab        site   N
1     d0    d0                      All sites    Allsites 533
2     d1   d11                            Lip         Lip   0
3    d10   d25                         Rectum      Rectum  21
4    d11   d26                          Liver       Liver  15
5    d12   d27                    Gallbladder GallBladder   0
6    d13   d28                       Pancreas    Pancreas  12
7    d14   d31                  Nose, sinuses        Nose   0
8    d15   d32                         Larynx      Larynx   4
9    d16   d33                           Lung        Lung  56
10   d17   d36                         Pleura      Pleura   1
11   d18   d70                         Breast      Breast 100
12   d19   d82                   Cervix uteri      Cervix  14
13    d2   d12                         Tongue      Tongue   5
14   d20   d83                   Corpus uteri Endometrium  12
15   d22   d84                          Ovary       Ovary  11
16   d23   d85   Other female genital organs   OthFemGen   2
17   d24   d91                       Prostate    Prostate  12
18   d25   d92                         Testis      Testis  20
19   d26   d93                     Penis etc.       Penis  34
20   d27  d101                         Kidney      Kidney  18
21   d28  d103                        Bladder     Bladder  17
22   d29   d51               Melanoma of skin    Melanoma  31
23    d3   d14                Salivary glands    Salivary   1
24   d31  d111                            Eye         Eye   2
25   d32  d113                     Brain, CNS       Brain  46
26   d33  d121                        Thyroid     Thyroid  16
27   d34   d40                           Bone        Bone   0
28   d35   d63                   Soft tissues  SoftTissue   4
29   d36  d132            Non-Hodgkin lymphoma   OtherNHL  29
30   d37  d131               Hodgkin lymphoma     Hodgkin   6
31   d38  d133               Multiple myeloma     Myeloma   3
32    d4   d13                          Mouth       Mouth   5
33   d40  d139                      Leukaemia   Leukaemia  22
34   d48  d140 Other and unspecified cancers Unspecified  19
35    d5   d15                        Pharynx     Pharynx   1
36   d51  d151                     Oral etc.  OralCavity  12
37   d52  d251                     Colorectal  Colorectal  48
38    d6   d21                     Oesophagus  Oesophagus  16
39    d7   d22                        Stomach     Stomach   9
40    d8   d23                Small intestine Smallintest   2
41    d9   d24                          Colon       Colon  27
```

6.2 Reading the data

Then we can read the data proper from Scotland and generate the correct names:

```
> scana <- read.csv( file="../data/ScotlandT1D.csv", header=TRUE )
> names( scana )
 [1] "per"     "Age"     "SEX"     "dur"     "t1d"     "DMprev"  "Pyr"     "D0"      "PyrD0"   "D01"
[11] "PyrD01"  "D02"     "PyrD02"  "D03"     "PyrD03"  "D04"     "PyrD04"  "D05"     "PyrD05"  "D06"
[21] "PyrD06"  "D07"     "PyrD07"  "D08"     "PyrD08"  "D09"     "PyrD09"  "D10"     "PyrD10"  "D11"
[31] "PyrD11"  "D12"     "PyrD12"  "D13"     "PyrD13"  "D14"     "PyrD14"  "D15"     "PyrD15"  "D16"
```

```
[41] "PyrD16" "D17"    "PyrD17" "D18"    "PyrD18" "D19"    "PyrD19" "D20"    "PyrD20" "D22"
[51] "PyrD22" "D23"    "PyrD23" "D24"    "PyrD24" "D25"    "PyrD25" "D26"    "PyrD26" "D27"
[61] "PyrD27" "D28"    "PyrD28" "D29"    "PyrD29" "D31"    "PyrD31" "D32"    "PyrD32" "D33"
[71] "PyrD33" "D34"    "PyrD34" "D35"    "PyrD35" "D36"    "PyrD36" "D38"    "PyrD38" "D40"
[81] "PyrD40" "D48"    "PyrD48" "D51"    "PyrD51" "D52"    "PyrD52"
> names( scana )[1:5] <- c("P","A","sex","DMdur","T1D")
```

The “PyrDxx” variables contain the person-years lived *after* diagnosis of cancer. So we tabulate this and see how large a percentage of the total follow-up time we erroneouly include:

```
> names( scana )[wh<-7+(0:40)*2]
 [1] "Pyr"    "PyrD0"  "PyrD01" "PyrD02" "PyrD03" "PyrD04" "PyrD05" "PyrD06" "PyrD07" "PyrD08"
[11] "PyrD09" "PyrD10" "PyrD11" "PyrD12" "PyrD13" "PyrD14" "PyrD15" "PyrD16" "PyrD17" "PyrD18"
[21] "PyrD19" "PyrD20" "PyrD22" "PyrD23" "PyrD24" "PyrD25" "PyrD26" "PyrD27" "PyrD28" "PyrD29"
[31] "PyrD31" "PyrD32" "PyrD33" "PyrD34" "PyrD35" "PyrD36" "PyrD38" "PyrD40" "PyrD48" "PyrD51"
[41] "PyrD52"
> tt <- apply( scana[,wh], 2, sum )
> cbind( N=round(tt), pct=round(tt/tt[1]*100,2) )
              N    pct
Pyr    87267584 100.00
PyrD0    225853   0.26
PyrD01      477   0.00
PyrD02     1404   0.00
PyrD03      393   0.00
PyrD04     1748   0.00
PyrD05     1902   0.00
PyrD06     7043   0.01
PyrD07     7356   0.01
PyrD08      703   0.00
PyrD09    20032   0.02
PyrD10    10902   0.01
PyrD11     2727   0.00
PyrD12     1326   0.00
PyrD13     5558   0.01
PyrD14      309   0.00
PyrD15     2596   0.00
PyrD16    40874   0.05
PyrD17     1431   0.00
PyrD18    32136   0.04
PyrD19     2735   0.00
PyrD20     4440   0.01
PyrD22     5371   0.01
PyrD23     1093   0.00
PyrD24    21644   0.02
PyrD25     1792   0.00
PyrD26    23852   0.03
PyrD27     5413   0.01
PyrD28    13663   0.02
PyrD29     7268   0.01
PyrD31      526   0.00
PyrD32     7255   0.01
PyrD33     1434   0.00
PyrD34      373   0.00
PyrD35     1018   0.00
PyrD36     7881   0.01
PyrD38     3087   0.00
PyrD40     6073   0.01
PyrD48    12934   0.01
PyrD51     5299   0.01
PyrD52    30542   0.03
```

So we see that it is only for all cancers that the person-years lived after cancer constitutes any appreaciable part of the total amount of FU in the Scottish population (and then only 0.26%).

We can now finalize the data so the are in the same form as the other datasets before we save it; first we tabulate how the three DM-classification variables relate in order to get the recoding correct:

```
> with( scana, ftable(DMdur,DMprev,T1D,row.vars=1) )
      DMprev   -1                      0                      1
      T1D      -1   30   35   40   -1   30   35   40   -1   30   35   40
DMdur
-1           3094    0    0    0    0    0    0    0    0    0    0    0
0               0    0    0    0    0 1053  238  238    0    2    0    3
1               0    0    0    0    0  997  224  224    0   66   14   14
2               0    0    0    0    0  989  264  264    0  262   66   67
5               0    0    0    0    0  825  244  244    0  619  178  180
10              0    0    0    0    0  472  134  134    0  969  289  289
15              0    0    0    0    0  126   30   30    0 1237  457  457
> sc.ana <- transform( scana, y = Pyr,
+                            y0 = Pyr-PyrD0,
+                             A = A+0.5,
+                             P = P+0.5,
+                           sex = factor( sex, labels=c("M","F") ),
+                           T1D = factor( T1D,
+                                         levels=c(-1,30,35,40,Inf),
+                                         labels=c("NoDM","30","35","40","Inf") ),
+                         DMdur = factor( DMdur ),
+                        DMprev = factor( DMprev, labels=c("Pop","Inc","Prv") ) )
> levels( sc.ana$DMdur )[1] <- "NoDM"
```

So for the rest of the analyses we do not need the `Pyr`-variables, so we leave them out

```
> names( sc.ana )[wh]
 [1] "Pyr"    "PyrD0"  "PyrD01" "PyrD02" "PyrD03" "PyrD04" "PyrD05" "PyrD06" "PyrD07" "PyrD08"
[11] "PyrD09" "PyrD10" "PyrD11" "PyrD12" "PyrD13" "PyrD14" "PyrD15" "PyrD16" "PyrD17" "PyrD18"
[21] "PyrD19" "PyrD20" "PyrD22" "PyrD23" "PyrD24" "PyrD25" "PyrD26" "PyrD27" "PyrD28" "PyrD29"
[31] "PyrD31" "PyrD32" "PyrD33" "PyrD34" "PyrD35" "PyrD36" "PyrD38" "PyrD40" "PyrD48" "PyrD51"
[41] "PyrD52"
> sc.ana <- sc.ana[,-wh]
> names( sc.ana )[match("D0",names(sc.ana))] <- "D00" # Otherwise d0 is lost
> names( sc.ana ) <- gsub( "D0", "D", names(sc.ana) )
> names( sc.ana )[wh <- 7:46]
 [1] "D0"  "D1"  "D2"  "D3"  "D4"  "D5"  "D6"  "D7"  "D8"  "D9"  "D10" "D11" "D12" "D13" "D14" "D15"
[17] "D16" "D17" "D18" "D19" "D20" "D22" "D23" "D24" "D25" "D26" "D27" "D28" "D29" "D31" "D32" "D33"
[33] "D34" "D35" "D36" "D38" "D40" "D48" "D51" "D52"
> names( sc.ana )[wh] <- tolower( names( sc.ana )[wh] )
> sc.ana <- subset( sc.ana, y0>0 )
> names( sc.ana )
 [1] "P"      "A"      "sex"    "DMdur"  "T1D"    "DMprev" "d0"     "d1"     "d2"     "d3"
[11] "d4"     "d5"     "d6"     "d7"     "d8"     "d9"     "d10"    "d11"    "d12"    "d13"
[21] "d14"    "d15"    "d16"    "d17"    "d18"    "d19"    "d20"    "d22"    "d23"    "d24"
[31] "d25"    "d26"    "d27"    "d28"    "d29"    "d31"    "d32"    "d33"    "d34"    "d35"
[41] "d36"    "d38"    "d40"    "d48"    "d51"    "d52"    "y"      "y0"
> with( sc.ana, ftable( DMdur, DMprev, T1D, row.vars=1 ) )
      DMprev  Pop                          Inc                          Prv
      T1D     NoDM   30   35   40  Inf NoDM   30   35   40  Inf NoDM   30   35   40  Inf
DMdur
NoDM          3094    0    0    0    0    0    0    0    0    0    0    0    0    0    0
0                0    0    0    0    0    0 1041  204  204    0    0    2    0    3    0
1                0    0    0    0    0    0  997  224  224    0    0   65   14   14    0
2                0    0    0    0    0    0  989  264  264    0    0  262   66   67    0
5                0    0    0    0    0    0  825  244  244    0    0  619  178  178    0
10               0    0    0    0    0    0  472  134  134    0    0  969  289  289    0
15               0    0    0    0    0    0  126   30   30    0    0 1237  457  457    0
```

Finally we put in the country-variable and save data

```
> sc.ana$Cnt <- "SCO"
> save( sc.ana, file="../data/SCana.Rda" )
```

6.3 Analysis of SMR

We load the dataset and the dataframe with the tumour labels in it:

```
> load( file="../data/SCana.Rda" )
> levels( sc.ana$T1D )
[1] "NoDM" "30"   "35"   "40"   "Inf"
> levels( sc.ana$DMdur )
[1] "NoDM" "0"    "1"    "2"    "5"    "10"   "15"
> sc.ana <- subset( sc.ana, y>0 )
> round( addmargins( xtabs( d0 ~ P + T1D, data=sc.ana ), 1 ) )
        T1D
P          NoDM     30     35     40    Inf
  1995.5  24129      7      3      4      0
  1996.5  25614      6      3      6      0
  1997.5  24913      5      2      9      0
  1998.5  24531      7      2      6      0
  1999.5  24679      9      6     11      0
  2000.5  25023      9      4      8      0
  2001.5  25118      4      2     12      0
  2002.5  25700     14      4     13      0
  2003.5  25762     16     11     10      0
  2004.5  26310      8     12     15      0
  2005.5  25927     14      7      7      0
  2006.5  26556      5     13     19      0
  2007.5  27136     18     10     16      0
  2008.5  27768     12     13     15      0
  2009.5  28146     17     10     19      0
  2010.5  27967     24      8     26      0
  2011.5  28273     12     14     26      0
  Sum    443552    187    124    222      0
> round( addmargins( xtabs( y0 ~ P + T1D, data=sc.ana ), 1 ), 1 )
        T1D
P                NoDM         30         35         40        Inf
  1995.5  5091461.3     9356.8     1703.0     1855.7        0.0
  1996.5  5079611.8     9844.0     1830.6     2002.3        0.0
  1997.5  5070610.7    10308.4     1938.5     2161.7        0.0
  1998.5  5064583.9    10788.4     2042.7     2300.3        0.0
  1999.5  5059422.0    11257.6     2162.0     2458.5        0.0
  2000.5  5050174.1    11765.7     2287.6     2622.2        0.0
  2001.5  5051504.7    12273.4     2394.1     2772.2        0.0
  2002.5  5041800.9    12767.4     2492.1     2943.7        0.0
  2003.5  5044338.8    13256.9     2602.1     3123.2        0.0
  2004.5  5065145.0    13750.1     2712.4     3292.8        0.0
  2005.5  5081762.6    14279.2     2823.4     3472.4        0.0
  2006.5  5103607.0    14836.4     2928.7     3652.5        0.0
  2007.5  5130564.2    15364.5     3043.3     3814.0        0.0
  2008.5  5154420.2    15891.1     3144.6     3983.6        0.0
  2009.5  5179873.5    16404.8     3241.0     4168.9        0.0
  2010.5  5207980.5    16933.4     3352.7     4336.2        0.0
  2011.5  5240694.2    17496.0     3480.1     4462.4        0.0
  Sum    86717555.4   226574.1    44178.9    53422.7        0.0
> load( file="../data/conv.Rda" )
```

6.3.1 All cancers

We first set out the simplest possible analysis with age-period-cohort effects for the baseline rates. First we devise a couple of knots for the splines:

```
> library( splines )
> with( sc.ana, rbind(
+  A = quantile( rep(  A,d0), 0:10/10 ),
+  P = quantile( rep(P  ,d0), 0:10/10 ),
+  C = quantile( rep(P-A,d0), 0:10/10 ) ) )
      0%    10%    20%    30%    40%    50%    60%    70%    80%    90%   100%
A    0.5   48.5   56.5   61.5   65.5   69.5   72.5   76.5   79.5   84.5   90.5
P 1995.5 1996.5 1998.5 2000.5 2002.5 2003.5 2005.5 2007.5 2008.5 2010.5 2011.5
C 1905.0 1919.0 1923.0 1927.0 1931.0 1934.0 1938.0 1943.0 1948.0 1957.0 2011.0
> ( a.kn <- seq(5,85,,7) )
[1]  5.00000 18.33333 31.66667 45.00000 58.33333 71.66667 85.00000
> ( p.kn <- seq(1996,2010,,4) )
[1] 1996.000 2000.667 2005.333 2010.000
> ( c.kn <- seq(1910,1990,,6) )
[1] 1910 1926 1942 1958 1974 1990
```

We then fit 3 models for the RR for T1D patients relative to the general population, using a common shape of the underlying cancer incidence rates as an age-period cohort model with $1 + (7-1) + (4-1) + (6-1) - 1 = 14$ parameters[1]. The first model (`m3`) is one with separate effects for persons diagnosed in ages < 30, 30–35, 35–40, 40–45 and 45+, the second (`m1`) is a simplification where these groups are pooled:

```
> m3 <- glm( d0 ~ Ns(   A, knots=a.kn ) +
+                 Ns( P  , knots=p.kn ) +
+                 Ns( P-A, knots=c.kn ) +
+                 T1D,
+            offset = log(y),
+            family = poisson,
+              data = sc.ana )
> m1  <- update( m3, . ~ . - T1D + Relevel(T1D,list(1,2:4)) )
> mp  <- update( m3, . ~ . - T1D + DMprev )
> anova( m3, m1, mp, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    Relevel(T1D, list(1, 2:4))
Model 3: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMprev
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     14893      26892
2     14895      26898 -2  -6.7189 0.034754
3     14894      26888  1   9.7262 0.001817
```

The tests of the models show that there is a detectable difference between the age at diagnosis groups, and it appears that the risk is higher the older the patients are at diagnosis. However it also seems that the risk is higher among more recenty diagnosed T1D patients compared to those prevalent at study strat in 1995:

```
> round( rbind( ci.exp( m3, subset="T1" ),
+               ci.exp( m1, subset="T1" ),
+               ci.exp( mp, subset="DM" ) ), 3 )
                                   exp(Est.)  2.5% 97.5%
T1D30                                  0.880 0.763 1.016
T1D35                                  1.051 0.881 1.253
T1D40                                  1.134 0.994 1.293
Relevel(T1D, list(1, 2:4))30+35+40     1.013 0.930 1.103
DMprevInc                              1.270 1.084 1.488
DMprevPrv                              0.936 0.846 1.035
```

[1] There is first an intercept, then the three natural splines, where k knots gives $k-1$ parameters, and finally 1 aliased parameter from the linear relationship between $P - A$ and P and A.

In summary, the comparison of the models show that there is some heterogeneity between groups of T1D patients w.r.t. the occurrence of cancer; which is also evident in figure 7.1, where the RR estimates from the three models are shown together:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind( ci.exp( m3, subset="T1D"),
+                 ci.exp( m1, subset="T1D") ),
+          y = 4:1,
+          txt = c("<30","30-35","35-40","Pooled"),
+          xlog=TRUE, xtic=c(5:10,15,20)/10, grid=TRUE,
+          xlab="RR of any cancer, T1D vs. population",
+          lwd=4, cex=2, vref=1 )
```

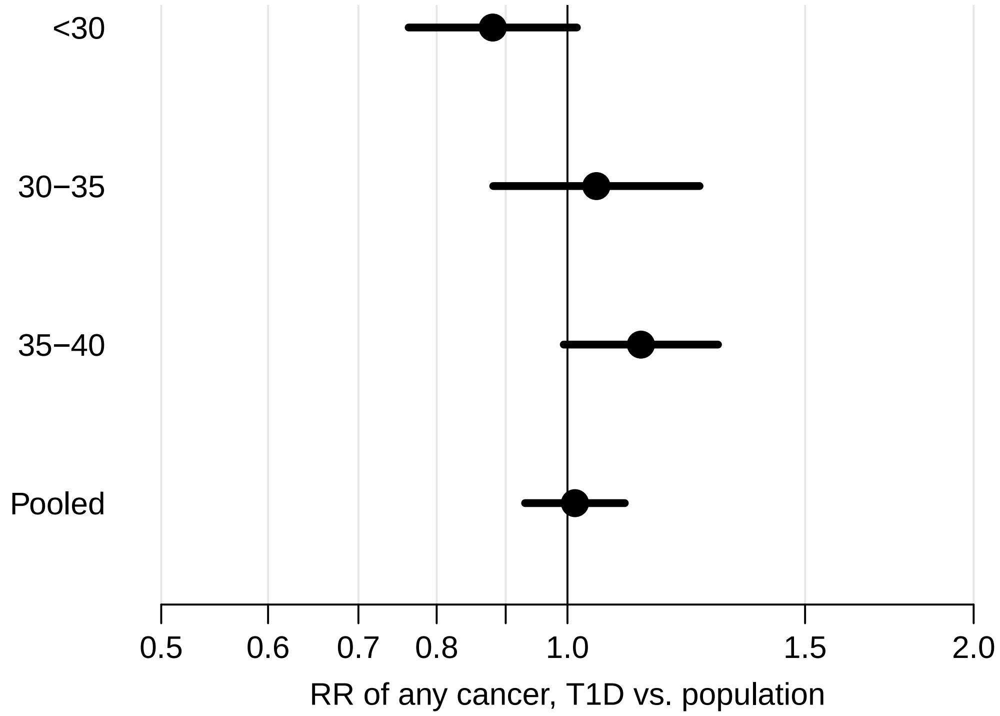

Figure 6.1: *Estimated RRs relative to the general Scottish population in two different models.*

6.3.2 Analysis by duration

We have also devised a variable indicating different time bands after diagnosis:

```
> round( ftable( xtabs( cbind(Ca=d0,PY=y/1000) ~ DMprev + DMdur,
+                       data=sc.ana ), col.vars=c(1,3) ), 1 )
      DMprev      Pop                Inc                Prv
                   Ca        PY       Ca        PY       Ca        PY
DMdur
NoDM         443552.0   86941.3      0.0       0.0      0.0       0.0
0                 0.0       0.0     25.0       8.2      0.0       0.0
1                 0.0       0.0     16.0      15.0      1.0       0.7
2                 0.0       0.0     24.0      38.4      8.0       6.9
5                 0.0       0.0     41.0      43.2     22.0      26.3
10                0.0       0.0     38.0      20.0     68.0      48.3
15                0.0       0.0     10.0       2.3    280.0     117.0
```

We now fit models with the smaller datasets corresponding to more restrictive definitions of T1D:

```
> md40 <- update( m3, . ~ . - T1D + DMdur,
+                        data=subset( sc.ana, T1D %in% levels(T1D)[1:4] & DMdur != "Unkn" ) )
> md35 <- update( md40, data=subset( sc.ana, T1D %in% levels(T1D)[1:3] & DMdur != "Unkn" ) )
> md30 <- update( md40, data=subset( sc.ana, T1D %in% levels(T1D)[1:2] & DMdur != "Unkn" ) )
> anova( m3, md40, update( md40, . ~ . + T1D ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 3: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur + T1D
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     14893      26892
2     14890      26848  3   43.895 1.589e-09
3     14888      26843  2    4.384    0.1117
> anova(     md35, update( md35, . ~ . + T1D ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur + T1D
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     12782      25959
2     12781      25957  1   1.6584   0.1978
```

We see that there a very strong effect of duration, but that it is the same regardless of the age cut-off:

```
> round( cbind( RR40 <- ci.exp( md40, subset="DMdur" ),
+               RR35 <- ci.exp( md35, subset="DMdur" ),
+               RR30 <- ci.exp( md30, subset="DMdur" ) ), 2 )
        exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
DMdur0       5.24 3.54  7.75      6.10 3.61 10.29      5.57 2.66 11.69
DMdur1       1.75 1.09  2.81      2.14 1.15  3.99      2.36 1.06  5.25
DMdur2       0.98 0.70  1.39      0.75 0.43  1.32      0.92 0.46  1.85
DMdur5       0.94 0.73  1.20      0.82 0.57  1.18      0.81 0.49  1.32
DMdur10      1.12 0.92  1.35      1.10 0.85  1.41      1.09 0.79  1.51
DMdur15      0.91 0.81  1.03      0.85 0.74  0.99      0.77 0.64  0.93
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind(NA,RR40),
+          txt=c("Years since DM","0-1","1-2","2-5","5-10","10-15","15+"),
+              xlog=TRUE, xtic=c(c(5:10,15,20,25)/10,3:6), grid=TRUE,
+              xlab="RR of any cancer, T1D vs. population",
+              lwd=4, cex=2, vref=1, y=c(6.7,6:1)+0.15 )
> linesEst( rbind(NA,RR35),
+           lwd=3, cex=1.5, col=gray(0.5), y=7:1-0.00 )
> linesEst( rbind(NA,RR30),
+           lwd=3, cex=1.5, col=gray(0.7), y=7:1-0.15 )
```

From figure 7.2 it seems that there is a strong ascertainment effect for T1DM as well as what have been shown for all diabetes under one, and that it is not changes if a stricter definition of T1D is applied.

6.3.3 Site-specific analyses

6.3.3.1 Analyses of site specific cancers

We first set up an array to hold the resulting simple RRs for each of the sites; we do the analyses by sex, but also make a pooled analysis, except for the sex-specific cancers (including breast):

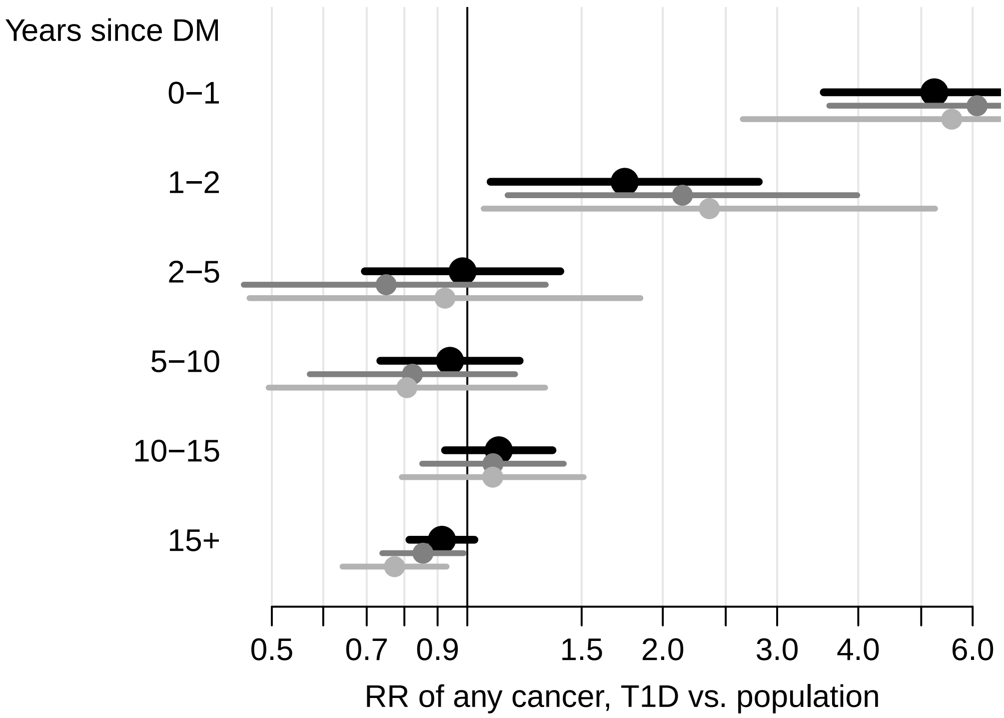

Figure 6.2: *The effect of duration on the RR of cancer in teh Scottish population. The gray sets of effecs are from the models where data are further restricted to patients diagnosed under 35 and 30, respectively.*

```
> wh <- c(46:7)
> vnam <- names(sc.ana)[wh]
> site <- conv[match(vnam,conv$NCnam),"Clab"]
> data.frame( vnam, site )
   vnam                          site
1   d52                    Colorectal
2   d51                     Oral etc.
3   d48 Other and unspecified cancers
4   d40                     Leukaemia
5   d38              Multiple myeloma
6   d36          Non-Hodgkin lymphoma
7   d35                  Soft tissues
8   d34                          Bone
9   d33                       Thyroid
10  d32                    Brain, CNS
11  d31                           Eye
12  d29              Melanoma of skin
13  d28                       Bladder
14  d27                        Kidney
15  d26                    Penis etc.
16  d25                        Testis
17  d24                      Prostate
18  d23    Other female genital organs
19  d22                         Ovary
20  d20                  Corpus uteri
21  d19                  Cervix uteri
22  d18                        Breast
23  d17                        Pleura
24  d16                          Lung
25  d15                        Larynx
26  d14                 Nose, sinuses
27  d13                      Pancreas
```

```
28  d12                  Gallbladder
29  d11                        Liver
30  d10                       Rectum
31   d9                        Colon
32   d8              Small intestine
33   d7                      Stomach
34   d6                   Oesophagus
35   d5                      Pharynx
36   d4                        Mouth
37   d3              Salivary glands
38   d2                       Tongue
39   d1                          Lip
40   d0                    All sites
```

```
> RRtab <- NArray( list( site = site,
+                          sex = c(levels(sc.ana$sex),"Both"),
+                         what = c("N.Pop","N.T1D","RR","lo","hi") ) )
> str( RRtab )
```

```
 logi [1:40, 1:3, 1:5] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ site: chr [1:40] "Colorectal" "Oral etc." "Other and unspecified cancers" "Leukaemia" ...
  ..$ sex : chr [1:3] "M" "F" "Both"
  ..$ what: chr [1:5] "N.Pop" "N.T1D" "RR" "lo" ...
```

With this fixed we can the make a loop doing the analysis for all sites:

```
> system.time(
+ for( i in 1:length(vnam) )
+    { # i <- 1
+    aset <- sc.ana[,c(vnam[i],
+                      paste("y",if(i==1) "0", sep=""),
+                      "A","P","DMprev","sex")]
+    names( aset )[1:2] <- c("D","Y")
+    mB <- glm( D ~ Ns   ( A, knots=a.kn ) +
+                   Ns( P   , knots=p.kn ) +
+                   Ns( P-A, knots=c.kn ) +
+                   Relevel(DMprev,list(1,2:3)),
+            offset = log(Y),
+            family = poisson,
+              data = aset )
+    mM <- update( mB, data=subset(aset,sex=="M") )
+    mF <- update( mB, data=subset(aset,sex=="F") )
+    RRtab[i,"M"   ,3:5] <- ci.exp( mM, subset="DMprev" )
+    RRtab[i,"F"   ,3:5] <- ci.exp( mF, subset="DMprev" )
+    RRtab[i,"Both",3:5] <- ci.exp( mB, subset="DMprev" )
+    RRtab[i,,1:2] <- addmargins( with( aset,
+                                 tapply(D,list(sex,Relevel(DMprev,list(1,2:3))),sum) ), 1 )
+    } )
   user  system elapsed
 30.555   0.037  30.585
```

```
> RRorg <- RRtab
> RRtab <- RRorg
> for(i in 1:dim(RRtab)[1])
+ for(j in 1:dim(RRtab)[2]) if(         RRtab[i,j,5]==Inf  ) RRtab[i, j    ,] <- NA
> for(i in 1:dim(RRtab)[1]) if(any(is.na(RRtab[i, ,5]==Inf))) RRtab[i,"Both",] <- NA
> round( ftable( RRtab ), 2 )
```

```
                                   what        N.Pop        N.T1D           RR           lo
site                               sex
Colorectal                         M     3.203200e+04 3.300000e+01 1.210000e+00 8.600000e-01 1.7
                                   F     2.828700e+04 1.500000e+01 8.900000e-01 5.300000e-01 1.4
                                   Both  6.031900e+04 4.800000e+01 1.110000e+00 8.300000e-01 1.4
Oral etc.                          M     6.921000e+03 8.000000e+00 5.700000e-01 2.800000e-01 1.1
                                   F     3.609000e+03 4.000000e+00 9.200000e-01 3.400000e-01 2.4
                                   Both  1.053000e+04 1.200000e+01 7.000000e-01 3.900000e-01 1.2
Other and unspecified cancers M     1.214100e+04 9.000000e+00 1.030000e+00 5.400000e-01 1.9
                                   F     1.334700e+04 1.000000e+01 1.530000e+00 8.200000e-01 2.8
```

	Both	2.548800e+04	1.900000e+01	1.250000e+00	8.000000e-01	1.9
Leukaemia	M	6.744000e+03	1.300000e+01	1.460000e+00	8.500000e-01	2.5
	F	5.076000e+03	9.000000e+00	1.800000e+00	9.300000e-01	3.4
	Both	1.182000e+04	2.200000e+01	1.620000e+00	1.070000e+00	2.4
Multiple myeloma	M	3.143000e+03	3.000000e+00	1.010000e+00	3.300000e-01	3.1
	F	NA	NA	NA	NA	
	Both	NA	NA	NA	NA	
Non-Hodgkin lymphoma	M	7.707000e+03	2.100000e+01	1.470000e+00	9.600000e-01	2.2
	F	7.594000e+03	8.000000e+00	1.020000e+00	5.100000e-01	2.0
	Both	1.530100e+04	2.900000e+01	1.340000e+00	9.300000e-01	1.9
Soft tissues	M	1.129000e+03	1.000000e+00	3.800000e-01	5.000000e-02	2.7
	F	8.470000e+02	3.000000e+00	2.060000e+00	6.600000e-01	6.4
	Both	1.976000e+03	4.000000e+00	1.010000e+00	3.800000e-01	2.7
Bone	M	NA	NA	NA	NA	
	F	NA	NA	NA	NA	
	Both	NA	NA	NA	NA	
Thyroid	M	7.250000e+02	4.000000e+00	1.520000e+00	5.700000e-01	4.0
	F	1.980000e+03	1.200000e+01	1.770000e+00	1.000000e+00	3.1
	Both	2.705000e+03	1.600000e+01	1.590000e+00	9.700000e-01	2.6
Brain, CNS	M	6.172000e+03	1.800000e+01	1.050000e+00	6.600000e-01	1.6
	F	7.875000e+03	2.800000e+01	1.590000e+00	1.090000e+00	2.3
	Both	1.404700e+04	4.600000e+01	1.300000e+00	9.700000e-01	1.7
Eye	M	5.240000e+02	1.000000e+00	9.100000e-01	1.300000e-01	6.5
	F	5.300000e+02	1.000000e+00	1.400000e+00	2.000000e-01	1.0
	Both	1.054000e+03	2.000000e+00	1.120000e+00	2.800000e-01	4.5
Melanoma of skin	M	6.247000e+03	1.600000e+01	9.400000e-01	5.700000e-01	1.5
	F	8.114000e+03	1.500000e+01	6.900000e-01	4.200000e-01	1.1
	Both	1.436100e+04	3.100000e+01	7.800000e-01	5.500000e-01	1.1
Bladder	M	1.849800e+04	1.100000e+01	8.300000e-01	4.600000e-01	1.5
	F	8.557000e+03	6.000000e+00	1.530000e+00	6.900000e-01	3.4
	Both	2.705500e+04	1.700000e+01	1.060000e+00	6.600000e-01	1.7
Kidney	M	6.328000e+03	1.400000e+01	1.480000e+00	8.700000e-01	2.5
	F	4.368000e+03	4.000000e+00	1.010000e+00	3.800000e-01	2.7
	Both	1.069600e+04	1.800000e+01	1.410000e+00	8.800000e-01	2.2
Penis etc.	M	4.662500e+04	3.300000e+01	7.300000e-01	5.200000e-01	1.0
	F	NA	NA	NA	NA	
	Both	NA	NA	NA	NA	
Testis	M	3.345000e+03	2.000000e+01	8.800000e-01	5.600000e-01	1.3
	F	NA	NA	NA	NA	
	Both	NA	NA	NA	NA	
Prostate	M	4.248400e+04	1.100000e+01	5.400000e-01	3.000000e-01	9.7
	F	NA	NA	NA	NA	
	Both	NA	NA	NA	NA	
Other female genital organs	M	NA	NA	NA	NA	
	F	2.179000e+03	2.000000e+00	6.900000e-01	1.700000e-01	2.7
	Both	NA	NA	NA	NA	
Ovary	M	NA	NA	NA	NA	
	F	1.063500e+04	1.100000e+01	7.400000e-01	4.100000e-01	1.3
	Both	NA	NA	NA	NA	
Corpus uteri	M	NA	NA	NA	NA	
	F	8.752000e+03	1.200000e+01	1.430000e+00	8.100000e-01	2.5
	Both	NA	NA	NA	NA	
Cervix uteri	M	NA	NA	NA	NA	
	F	5.383000e+03	1.400000e+01	7.100000e-01	4.200000e-01	1.2
	Both	NA	NA	NA	NA	
Breast	M	3.360000e+02	1.000000e+00	2.690000e+00	3.700000e-01	1.9
	F	6.257900e+04	9.900000e+01	9.100000e-01	7.500000e-01	1.1
	Both	6.291500e+04	1.000000e+02	7.800000e-01	6.400000e-01	9.5
Pleura	M	2.387000e+03	1.000000e+00	8.800000e-01	1.200000e-01	6.2
	F	NA	NA	NA	NA	
	Both	NA	NA	NA	NA	
Lung	M	4.430700e+04	3.700000e+01	1.410000e+00	1.020000e+00	1.9
	F	3.614600e+04	1.900000e+01	1.120000e+00	7.200000e-01	1.7
	Both	8.045300e+04	5.600000e+01	1.310000e+00	1.010000e+00	1.7
Larynx	M	4.026000e+03	4.000000e+00	8.200000e-01	3.100000e-01	2.2
	F	1.060000e+03	0.000000e+00	0.000000e+00	0.000000e+00	5.08

```
                              Both        5.086000e+03  4.000000e+00  7.600000e-01  2.900000e-01  2.0
Nose, sinuses                 M                     NA            NA            NA            NA
                              F                     NA            NA            NA            NA
                              Both                  NA            NA            NA            NA
Pancreas                      M           5.265000e+03  7.000000e+00  1.490000e+00  7.100000e-01  3.1
                              F           5.758000e+03  5.000000e+00  1.950000e+00  8.100000e-01  4.7
                              Both        1.102300e+04  1.200000e+01  1.700000e+00  9.600000e-01  3.0
Gallbladder                   M                     NA            NA            NA            NA
                              F           1.533000e+03  0.000000e+00  0.000000e+00  0.000000e+00 2.43
                              Both                  NA            NA            NA            NA
Liver                         M           3.482000e+03  1.200000e+01  3.460000e+00  1.960000e+00  6.1
                              F           1.899000e+03  3.000000e+00  2.880000e+00  9.200000e-01  8.9
                              Both        5.381000e+03  1.500000e+01  3.580000e+00  2.150000e+00  5.9
Rectum                        M           1.255200e+04  1.500000e+01  1.190000e+00  7.200000e-01  1.9
                              F           8.917000e+03  6.000000e+00  8.500000e-01  3.800000e-01  1.9
                              Both        2.146900e+04  2.100000e+01  1.100000e+00  7.200000e-01  1.6
Colon                         M           1.994800e+04  1.800000e+01  1.180000e+00  7.400000e-01  1.8
                              F           1.966200e+04  9.000000e+00  8.700000e-01  4.500000e-01  1.6
                              Both        3.961000e+04  2.700000e+01  1.070000e+00  7.300000e-01  1.5
Small intestine               M                     NA            NA            NA            NA
                              F           6.430000e+02  2.000000e+00  3.000000e+00  7.400000e-01  1.2
                              Both                  NA            NA            NA            NA
Stomach                       M           8.787000e+03  5.000000e+00  8.500000e-01  3.500000e-01  2.0
                              F           5.770000e+03  4.000000e+00  1.500000e+00  5.600000e-01  4.0
                              Both        1.455700e+04  9.000000e+00  1.100000e+00  5.700000e-01  2.1
Oesophagus                    M           8.557000e+03  9.000000e+00  1.070000e+00  5.500000e-01  2.0
                              F           5.272000e+03  7.000000e+00  3.840000e+00  1.820000e+00  8.0
                              Both        1.382900e+04  1.600000e+01  1.700000e+00  1.040000e+00  2.7
Pharynx                       M           2.733000e+03  1.000000e+00  1.700000e-01  2.000000e-02  1.2
                              F           1.048000e+03  0.000000e+00  0.000000e+00  0.000000e+00 4.54
                              Both        3.781000e+03  1.000000e+00  1.500000e-01  2.000000e-02  1.0
Mouth                         M           2.167000e+03  4.000000e+00  1.040000e+00  3.900000e-01  2.7
                              F           1.327000e+03  1.000000e+00  8.100000e-01  1.100000e-01  5.7
                              Both        3.494000e+03  5.000000e+00  1.050000e+00  4.400000e-01  2.5
Salivary glands               M                     NA            NA            NA            NA
                              F           3.490000e+02  1.000000e+00  1.830000e+00  2.600000e-01  1.3
                              Both                  NA            NA            NA            NA
Tongue                        M           1.796000e+03  3.000000e+00  7.700000e-01  2.500000e-01  2.4
                              F           9.880000e+02  2.000000e+00  1.580000e+00  3.900000e-01  6.3
                              Both        2.784000e+03  5.000000e+00  1.030000e+00  4.300000e-01  2.4
Lip                           M                     NA            NA            NA            NA
                              F                     NA            NA            NA            NA
                              Both                  NA            NA            NA            NA
All sites                     M           2.170780e+05  2.530000e+02  1.050000e+00  9.300000e-01  1.1
                              F           2.264740e+05  2.800000e+02  1.030000e+00  9.200000e-01  1.1
                              Both        4.435520e+05  5.330000e+02  1.010000e+00  9.300000e-01  1.1
```

Of course we would also like to see the results as a forest plot, so we extract the relevant quantities for doing this:

```
> eM <- RRtab[,"M",3:5]
> eF <- RRtab[,"F",3:5]
> eB <- RRtab[,"Both",3:5]
> nr <- nrow( eM )
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( eB, y=1:nr, txtpos=1:nr,
+          col="lightgray", xlog=TRUE,
+          xtic=c(1:10/10,1.5,2:7), xlim=c(0.095,7),
+          grid=TRUE, vref=1, xlab="Cancer incidence RR, T1D vs. population" )
> linesEst( eF, y=1:nr-0.2, col="red" )
> linesEst( eM, y=1:nr+0.2, col="blue" )
> text( rep(0.095, nr), nr:1+0.2, RRtab[,"M",2],
+       col="blue", adj=1, cex=0.7 )
> text( rep(0.095, nr), nr:1-0.2, RRtab[,"F",2],
+       col="red" , adj=1, cex=0.7 )
```

We see that the only sites with appreaciable increased RR and sufficiently narrow confidence intervals are liver, corpus uteri, thyroid (women) and leukaemia; whereas there seems to be a lower risk for prostate cancer among T1D patients.

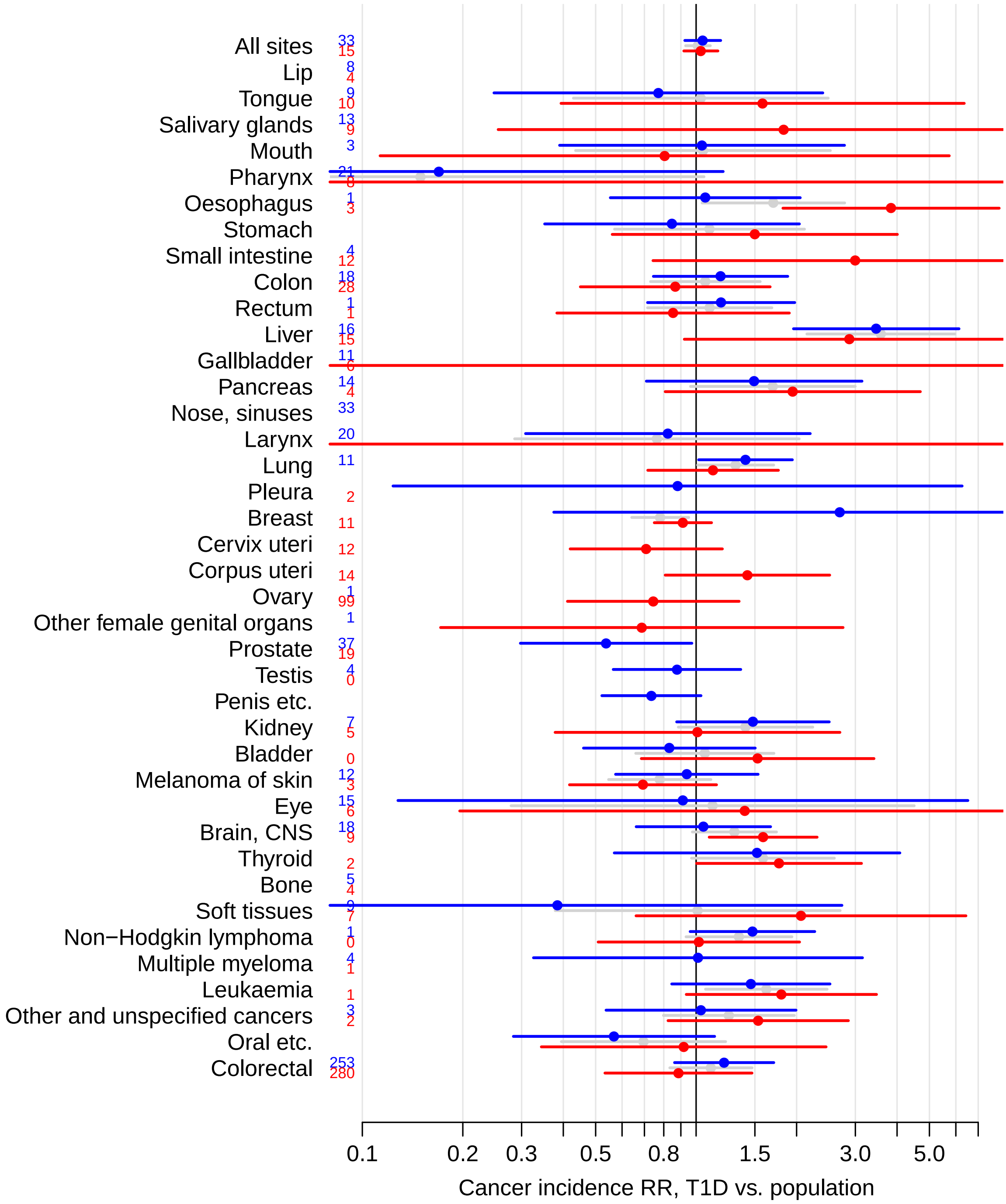

Figure 6.3: *RRs of cancer incidence among T1D patients (i.e. diagnosed* $<$ *40 years of age) in Scotland relative to the general population. The numbers to the left are the number of cancers observed among the T1D patients. Men: Blue, Women: Red, Both sexes: Light gray.*

Chapter 7

Australian T1D data

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

7.1 Naming of sites and files

We start by laying out the correct naming of the files with incidence and population data from each site, and link these to the common naming of response variables in the project:

```
> clear()
> load( "../data/conv.Rda" )
> conv
   DKnam NCnam                         Clab
1     d0    d0                    All sites
2    d11    d1                          Lip
3    d12    d2                       Tongue
4    d14    d3              Salivary glands
5    d13    d4                        Mouth
6    d15    d5                      Pharynx
7    d21    d6                   Oesophagus
8    d22    d7                      Stomach
9    d23    d8              Small intestine
10   d24    d9                        Colon
11   d25   d10                       Rectum
12   d26   d11                        Liver
13   d27   d12                  Gallbladder
14   d28   d13                     Pancreas
15   d31   d14               Nose, sinuses
16   d32   d15                       Larynx
17   d33   d16                         Lung
18   d36   d17                       Pleura
19   d70   d18                       Breast
20   d82   d19                 Cervix uteri
21   d83   d20                 Corpus uteri
22   d84   d22                        Ovary
23   d85   d23  Other female genital organs
24   d91   d24                     Prostate
25   d92   d25                       Testis
26   d93   d26                   Penis etc.
27  d101   d27                       Kidney
28  d103   d28                      Bladder
29   d51   d29            Melanoma of skin
30  d111   d31                          Eye
31  d113   d32                  Brain, CNS
32  d121   d33                      Thyroid
```

```
33   d40   d34                         Bone
34   d63   d35                 Soft tissues
35  d132   d36           Non-Hodgkin lymphoma
36  d131   d37             Hodgkin lymphoma
37  d133   d38             Multiple myeloma
38  d139   d40                    Leukaemia
39  d140   d48 Other and unspecified cancers
40  d151   d51                    Oral etc.
41  d251   d52                   Colorectal
```

Here are the names of variable names, the names of the site-specific files with incidence data (`fnam`), and the names of the site-specific population rate files (`pnam`):

```
> dnam <- c(0,6,7,52,11,13,16,18,19,20,22,24,
+           25,27,28,29,32,33,36,37,38)
> dnam <- paste( "d", dnam, sep="" )
> fnam <- c(
+ "allcancer",
+ "oesophagus",
+ "stomach",
+ "colorectal",
+ "liver",
+ "pancreas",
+ "lung",
+ "breast",
+ "cervical",
+ "uterine",
+ "ovarian",
+ "prostate",
+ "testicular",
+ "kidney",
+ "bladder",
+ "melanoma",
+ "brain",
+ "thyroid",
+ "nhlymphoma",
+ "hlymphoma",
+ "myeloma")
> pnam <- c(
+ "all",
+ "oesophageal",
+ "stomach",
+ "bowel",
+ "liver",
+ "pancreatic",
+ "lung",
+ "breast",
+ "cervical",
+ "uterine",
+ "ovarian",
+ "prostate",
+ "testicular",
+ "kidney",
+ "bladder",
+ "melanoma",
+ "brain",
+ "thyroid",
+ "nhlymphoma",
+ "hlymphoma",
+ "myeloma")
```

We can now list the filenames used for the population rates (`fnam`), from the T1D rates (`pnam`) together with the eventual variable names that we shall use (which also will be used in the joint analysis) and finally the labelling of the estimates as we shall use them in the reporting of teh analysis. The latter two are from the object `conv`:

```
> cbind( fnam, pnam, dnam, conv$Clab[match(dnam,conv$NCnam)] )
      fnam          pnam          dnam
 [1,] "allcancer"   "all"         "d0"  "All sites"
 [2,] "oesophagus"  "oesophageal" "d6"  "Oesophagus"
 [3,] "stomach"     "stomach"     "d7"  "Stomach"
 [4,] "colorectal"  "bowel"       "d52" "Colorectal"
 [5,] "liver"       "liver"       "d11" "Liver"
 [6,] "pancreas"    "pancreatic"  "d13" "Pancreas"
 [7,] "lung"        "lung"        "d16" "Lung"
 [8,] "breast"      "breast"      "d18" "Breast"
 [9,] "cervical"    "cervical"    "d19" "Cervix uteri"
[10,] "uterine"     "uterine"     "d20" "Corpus uteri"
[11,] "ovarian"     "ovarian"     "d22" "Ovary"
[12,] "prostate"    "prostate"    "d24" "Prostate"
[13,] "testicular"  "testicular"  "d25" "Testis"
[14,] "kidney"      "kidney"      "d27" "Kidney"
[15,] "bladder"     "bladder"     "d28" "Bladder"
[16,] "melanoma"    "melanoma"    "d29" "Melanoma of skin"
[17,] "brain"       "brain"       "d32" "Brain, CNS"
[18,] "thyroid"     "thyroid"     "d33" "Thyroid"
[19,] "nhlymphoma"  "nhlymphoma"  "d36" "Non-Hodgkin lymphoma"
[20,] "hlymphoma"   "hlymphoma"   "d37" "Hodgkin lymphoma"
[21,] "myeloma"     "myeloma"     "d38" "Multiple myeloma"
```

7.2 Cancer incidences in T1 patients

7.2.1 Person-years

The Australian incidence data is provided in form of a tabulation of the follow up of *all* known T1D patients and a tabulation of the patients with known date of diagnosis subdivided by duration during follow-up.

However, we want the cases subdivided by *whether* DM duration is available or not; this is done by adding an extra level to the duration classification. This mean that we must take the non-duration classified cases and classify them *only* by sex, age, period and age band at diagnosis [1]

First we read the person-years (and deaths in the variable `_d`) and make the relevant tabulation, and subsequent subtract the person-years among those with duration from that among those without duration:

```
> pya <- read.dta( file="../data/aus/all/person-years.dta" )
> pyd <- read.dta( file="../data/aus/dmdur/person-years.dta" )
> names( pya )
[1] "asex"       "t1d_status" "_age"       "_year"      "_d"         "_y"
> names( pyd )
[1] "asex"       "t1d_status" "_age"       "_year"      "dmdur"      "_d"         "_y"
> names( pyd )[-5] <- names( pya ) <- c("sex","T1D","A","P","dd","y")
> names( pyd )[ 5] <- "DMdur"
> pyD <- aggregate( pyd[,6:7], pyd[,1:4], sum )
> c( sum( pya$y ), sum( pyd$y ), sum( pyD$y ) )
[1] 752864 214558 214558
> names( pyD )
[1] "sex" "T1D" "A"   "P"   "dd"  "y"
```

[1]Superficially this sounds a bit odd, since duration of diabetes should be known if both age at diagnosis and current age were known. However the classification by age at inclusion is quite broad.

```
> names( pyD )[5:6] <- c("DD","Y")
> pyp <- merge( pya, pyD, all=TRUE )
> pyp$y[is.na(pyp$y)] <- 0
> pyp$Y[is.na(pyp$Y)] <- 0
> c( sum(pyp$y), sum(pyp$Y) )
[1] 752864 214558
> pyp <- transform( pyp,
+                      dd = dd-DD,
+                       y =  y- Y,
+                   DMdur = factor("Unkn") )
```

So pyp now contains the person-years (y) and the no. of deaths (dd) in the group of patients *without* duration information, whereas pyd contains the person-years (y) and the no. of deaths (dd) among those *with* duration information, further classified by diabetes duration.

Then we can stack the two datsets, and check that the total risk time is the same in the new dataset as in the original dataset:

```
> py <- rbind( subset( pyp, y>0, select=-c(DD,Y) ),
+              transform( subset( pyd, y>0 ), DMdur=factor(DMdur) ) )
> round( addmargins( xx <- xtabs( y ~ DMdur, data=py )  ), 1 )
DMdur
    Unkn        0        1        2        5       10       15       30      Sum
538306.0  23366.2  23371.5  59518.0  61403.1  16758.3  22292.8   7848.2 752864.0
> sum( xx[-1] )
[1] 214558
```

7.2.2 Cancer cases

The next exercise is to read the datasets with cases among *all* diabetes patients and those with known duration and merge them to one where the duration is a factor with a separate level for those with unknown duration.

We have the correct names of the variables that we need in the variables pnam, dnam and fnam, so we make a loop over thes and read the two dataset for each site, and do the correct subtractions:

```
> au.ana <- py[,c(1,3,4,2,7,6,5)]
> names( au.ana )
[1] "sex"   "A"     "P"     "T1D"   "DMdur" "y"     "dd"
> durlab <- levels( au.ana$DMdur )
> for( dn in dnam ) # ( dn <- dnam[4] )
+    {
+    # Cancers in ALL DM patients
+    da <- read.dta( file=paste("../data/aus/all/",
+                              wh<-fnam[match(dn,dnam)],
+                              ".dta", sep="" ) )
+    print( paste(dn,wh) )
+    nnam <- c("sex","T1D","A","P",dn)
+    names(da) <- nnam
+    # Cancers in DM patients with known duration
+    dd <- read.dta( file=paste("../data/aus/dmdur/",
+                              fnam[match(dn,dnam)],
+                              ".dta", sep="" ) )
+    nnam <- c("sex","T1D","A","P","dur",dn)
+    names(dd) <- nnam
+    dd <- transform( dd, DMdur = factor( 1 + (dur>0.99)
+                                           + (dur>1.99)
+                                           + (dur>4.99)
+                                           + (dur>9.99)
```

```
+                                        + (dur>14.99)
+                                        + (dur>29.99),
+                                     levels = 0:7,
+                                     labels = durlab) )
+    # Aggregate the cancers with known duration in duration classes
+    dd <- aggregate( dd[,dn], dd[,c(nnam[1:4],"DMdur")], FUN=sum )
+    names( dd )[6] <- dn
+    dD <- aggregate( dd[,dn], dd[,  nnam[1:4]         ], FUN=sum )
+    # Merge and subtract from alll cancers in DM ptt
+    dx <- merge( da, dD, all=TRUE )
+    dx[is.na(dx)] <- 0
+    dx[,dn] <- dx[,dn] - dx[,"x"]
+    dx$DMdur <- factor( 0, levels = 0:7, labels = durlab )
+    dx <- subset( dx, select=-x )
+    au.ana <- merge( rbind( dx, dd ), au.ana, all=TRUE )
+    au.ana[is.na(au.ana)] <- 0
+ }
[1] "d0 allcancer"
[1] "d6 oesophagus"
[1] "d7 stomach"
[1] "d52 colorectal"
[1] "d11 liver"
[1] "d13 pancreas"
[1] "d16 lung"
[1] "d18 breast"
[1] "d19 cervical"
[1] "d20 uterine"
[1] "d22 ovarian"
[1] "d24 prostate"
[1] "d25 testicular"
[1] "d27 kidney"
[1] "d28 bladder"
[1] "d29 melanoma"
[1] "d32 brain"
[1] "d33 thyroid"
[1] "d36 nhlymphoma"
[1] "d37 hlymphoma"
[1] "d38 myeloma"
> names(au.ana)
 [1] "sex"   "T1D"   "A"     "P"     "DMdur" "d38"   "d37"   "d36"   "d33"   "d32"   "d29"   "d28"
[13] "d27"   "d25"   "d24"   "d22"   "d20"   "d19"   "d18"   "d16"   "d13"   "d11"   "d52"   "d7"
[25] "d6"    "d0"    "y"     "dd"
> nlev <- c(30,35,40,45,Inf)
> cbind( levels( au.ana$T1D ),   nlev )
              nlev
[1,] "<30"    "30"
[2,] "30-<35" "35"
[3,] "35'<40" "40"
[4,] "40-<45" "45"
[5,] ">45"    "Inf"
>         levels( au.ana$T1D ) <- nlev
> levels( au.ana$sex ) <- c("M","F")
> au.ana <- transform( au.ana, A = A+0.5,
+                              P = P+0.5 )
> summary( au.ana )
 sex        T1D            A                 P               DMdur            d38
 M:9074   30 :7882   Min.   :  0.50   Min.    :1998   Unkn    :4734   Min.    :0.0000000
 F:9492   35 :2913   1st Qu.: 30.50   1st Qu.:2000   15      :2494   1st Qu.:0.0000000
          40 :2805   Median : 45.50   Median :2004   30      :2402   Median :0.0000000
          45 :2602   Mean   : 44.66   Mean    :2003   5       :2082   Mean    :0.0009157
          Inf:2364   3rd Qu.: 59.50   3rd Qu.:2006   2       :1926   3rd Qu.:0.0000000
                     Max.   :103.50   Max.    :2008   10      :1781   Max.    :1.0000000
                                                     (Other):3147
      d37                 d36                 d33                 d32                 d29
 Min.   :0.0000000   Min.   :-1.000000   Min.   :-1.000000   Min.   :-1.000000   Min.   :-1.00000
```

```
 1st Qu.:0.0000000   1st Qu.: 0.000000   1st Qu.: 0.000000   1st Qu.: 0.000000   1st Qu.: 0.00000
 Median :0.0000000   Median : 0.000000   Median : 0.000000   Median : 0.000000   Median : 0.00000
 Mean   :0.0007541   Mean   : 0.005386   Mean   : 0.004578   Mean   : 0.002693   Mean   : 0.01454
 3rd Qu.:0.0000000   3rd Qu.: 0.000000   3rd Qu.: 0.000000   3rd Qu.: 0.000000   3rd Qu.: 0.00000
 Max.   :1.0000000   Max.   : 3.000000   Max.   : 2.000000   Max.   : 2.000000   Max.   : 4.00000

      d28                 d27                 d25                 d24
 Min.   :-1.000000   Min.   :-1.000000   Min.   :-1.000000   Min.   :-1.000000
 1st Qu.: 0.000000   1st Qu.: 0.000000   1st Qu.: 0.000000   1st Qu.: 0.000000
 Median : 0.000000   Median : 0.000000   Median : 0.000000   Median : 0.000000
 Mean   : 0.001777   Mean   : 0.003717   Mean   : 0.001293   Mean   : 0.007056
 3rd Qu.: 0.000000   3rd Qu.: 0.000000   3rd Qu.: 0.000000   3rd Qu.: 0.000000
 Max.   : 1.000000   Max.   : 2.000000   Max.   : 2.000000   Max.   : 6.000000

      d22                 d20                 d19                d18                d16
 Min.   :-1.000000   Min.   :-1.000000   Min.   :0.000000   Min.   :-1.0000   Min.   :-1.000000
 1st Qu.: 0.000000   1st Qu.: 0.000000   1st Qu.:0.000000   1st Qu.: 0.0000   1st Qu.: 0.000000
 Median : 0.000000   Median : 0.000000   Median :0.000000   Median : 0.0000   Median : 0.000000
 Mean   : 0.001993   Mean   : 0.002909   Mean   :0.001077   Mean   : 0.0175   Mean   : 0.007487
 3rd Qu.: 0.000000   3rd Qu.: 0.000000   3rd Qu.:0.000000   3rd Qu.: 0.0000   3rd Qu.: 0.000000
 Max.   : 2.000000   Max.   : 2.000000   Max.   :1.000000   Max.   : 4.0000   Max.   : 3.000000

      d13                 d11                 d52                d7                  d6
 Min.   :-1.000000   Min.   :-1.000000   Min.   :-1.00000   Min.   :-1.000000   Min.   :0.000000
 1st Qu.: 0.000000   1st Qu.: 0.000000   1st Qu.: 0.00000   1st Qu.: 0.000000   1st Qu.:0.000000
 Median : 0.000000   Median : 0.000000   Median : 0.00000   Median : 0.000000   Median :0.000000
 Mean   : 0.003501   Mean   : 0.002801   Mean   : 0.01287   Mean   : 0.002101   Mean   :0.001347
 3rd Qu.: 0.000000   3rd Qu.: 0.000000   3rd Qu.: 0.00000   3rd Qu.: 0.000000   3rd Qu.:0.000000
 Max.   : 2.000000   Max.   : 2.000000   Max.   : 3.00000   Max.   : 2.000000   Max.   :2.000000

       d0                  y                  dd
 Min.   :-1.0000   Min.   :  0.000   Min.   : 0.0000
 1st Qu.: 0.0000   1st Qu.:  1.383   1st Qu.: 0.0000
 Median : 0.0000   Median :  6.034   Median : 0.0000
 Mean   : 0.1129   Mean   : 40.551   Mean   : 0.3822
 3rd Qu.: 0.0000   3rd Qu.: 22.872   3rd Qu.: 0.0000
 Max.   :11.0000   Max.   :639.337   Max.   :58.0000
```

The dataset `au.ana` (in principle) now contains the number of cases and amount of person-years for the different types of cancer in Australia, among T1D patients.

However, due to the uncertainty of the NDSS scheme before 2000, we exclude cases and follow-up for the period before 1 January 2000:

```
> au.ana <- subset( au.ana, P>2000 )
> ftable( xtabs( d0 ~ sex + P + T1D, data = au.ana ), col.vars=c(1,3) )
       sex  M                 F
       T1D 30 35 40 45 Inf 30 35 40 45 Inf
P
2000.5      12 14 14 23   3 10 19 13 25   1
2001.5      14 10 22 29   0 14 12 25 29   1
2002.5      19  7 17 16   1 30 16 20 30   5
2003.5      16 15 12 46   5 24 15 26 25   1
2004.5      13 18 25 36   4 26 22 21 23   2
2005.5      27 13 23 42   1 28 17 29 28   1
2006.5      19 15 34 50   3 28 16 21 37   0
2007.5      16 11 29 54   5 28 20 29 27   3
2008.5      24 18 39 46   3 34 22 29 28   3
```

7.2.2.1 Fishy data

However there is still a small problem:

```
> table(au.ana$d0)
```

```
   -1     0     1     2     3     4     5     6     7     9    10    11 
   27 13265   799   201    79    36    17     4     3     1     1     2 
> names( au.ana )[wh<-6:26]
 [1] "d38" "d37" "d36" "d33" "d32" "d29" "d28" "d27" "d25" "d24" "d22" "d20" "d19" "d18" "d16" "d13"
[17] "d11" "d52" "d7"  "d6"  "d0" 
> subset( au.ana, apply( au.ana[,wh], 1, min )<0 )
        sex T1D    A      P DMdur d38 d37 d36 d33 d32 d29 d28 d27 d25 d24 d22 d20 d19 d18 d16 d13 d11
775       M  30 13.5 2007.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
910       M  30 15.5 2005.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
1468      M  30 22.5 2002.5  Unkn   0   0  -1   0   0   0   0   0   0   0   0   0   0   0   0   0   0
1874      M  30 27.5 2000.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0
1881      M  30 27.5 2001.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
3063      M  30 46.5 2001.5  Unkn   0   0   0   0   0   0  -1   0   0   0   0   0   0   0   0   0   0
3213      M  30 51.5 2001.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0   0
4082      M  35 33.5 2008.5  Unkn   0   0   0   0  -1   0   0   0   0   0   0   0   0   0   0   0   0
5283      M  35 89.5 2008.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
5808      M  40 45.5 2003.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0
5874      M  40 47.5 2001.5  Unkn   0   0   0   0   0   0   0   0  -1   0   0   0   0   0   0   0   0
6223      M  40 59.5 2000.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
6792      M  45 43.5 2002.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0
7137      M  45 51.5 2000.5  Unkn   0   0   0   1   0   0   0   0   0   0   0   0   0   0   0  -1   0
7248      M  45 54.5 2001.5  Unkn   0   0   1   0   0   0   0   0   0   0   0   0   0   0  -1   0   0
7787      M  45 77.5 2001.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0
7932      M Inf 46.5 2000.5  Unkn   0   0   0   0   0   0   0   0  -1   0   0   0   0   0   0   0   0
9298      F  30  5.5 2004.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
9318      F  30  5.5 2008.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
9335      F  30  6.5 2000.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
9785      F  30 13.5 2000.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
9796      F  30 13.5 2002.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
9814      F  30 13.5 2005.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
10125     F  30 17.5 2005.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
10759     F  30 25.5 2000.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
11011     F  30 28.5 2000.5  Unkn   0   0   0  -1   0   0   0   0   0   0   0   0   0   1   0   0   0
11046     F  30 28.5 2005.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
14111     F  35 53.5 2001.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
14200     F  35 59.5 2000.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0   0   0
15406     F  40 50.5 2005.5  Unkn   0   0   0   0   0  -1   0   0   0   0   0   0   0   1   0   0   0
16155     F  45 40.5 2007.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0
16330     F  45 44.5 2000.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0  -1   0
17283     F  45 78.5 2002.5  Unkn   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0
        d52 d7 d6 d0            y dd
775       0  0  0 -1 5.144285e+01  0
910       0  0  0 -1 1.331403e+02  0
1468      0  0  0 -1 3.095708e+02  2
1874      0  0  0 -1 4.298508e+02  1
1881      0  0  0 -1 3.958138e+02  2
3063      0  0  0 -1 5.587935e-09  0
3213      0  0  0 -1 2.600958e-02  0
4082      0  0  0 -1 1.009856e+01  0
5283      0 -1  0  0 0.000000e+00  0
5808      0  0  0 -1 1.885024e+02  0
5874      0  0  0 -1 2.671533e+02  8
6223     -1  0  0 -1 0.000000e+00  0
6792      0  0  0 -1 3.078029e+00  0
7137      0  0  0  1 3.319083e+02  6
7248      0  0  0  2 5.967201e+02 14
7787      0  0  0 -1 0.000000e+00  0
7932      0  0  0 -1 0.000000e+00  0
9298      0  0  0 -1 7.029433e-01  0
9318      0  0  0 -1 4.733744e+00  0
9335      0  0  0 -1 4.266667e+01  0
9785      0  0  0 -1 1.945661e+02  0
9796      0  0  0 -1 1.453101e+02  0
9814      0  0  0 -1 8.191513e+01  0
10125     0  0  0 -1 1.900999e+02  0
10759     0  0  0 -1 3.610062e+02  0
```

```
11011   0  0  0  0 4.483457e+02  4
11046   0  0  0 -1 3.361485e+02  1
14111   0  0  0 -1 0.000000e+00  0
14200   0  0  0  0 0.000000e+00  0
15406   0  0  0  0 2.281520e+02  2
16155   0  0  0 -1 2.138946e+00  0
16330   0  0  0 -1 5.734839e+01  0
17283  -1  0  0 -1 0.000000e+00  0
```

We will have to look into this at a later time, but for now we bypass it bluntly, by replacing negative or missing counts by 0:

```
> for( i in wh ) au.ana[,i] <- pmax( au.ana[,i], 0, na.rm=TRUE )
> summary( au.ana )
 sex         T1D            A                P              DMdur            d38
 M:7043   30 :6033   Min.   :  0.50   Min.   :2000   Unkn    :3759   Min.   :0.000000
 F:7392   35 :2270   1st Qu.: 31.50   1st Qu.:2002   15      :1906   1st Qu.:0.000000
          40 :2170   Median : 46.50   Median :2004   30      :1892   Median :0.000000
          45 :2000   Mean   : 45.42   Mean   :2005   5       :1670   Mean   :0.001039
          Inf:1962   3rd Qu.: 60.50   3rd Qu.:2006   2       :1487   3rd Qu.:0.000000
                     Max.   :103.50   Max.   :2008   10      :1393   Max.   :1.000000
                                                     (Other):2328
      d37                 d36                d33                d32                d29
 Min.   :0.0000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.00000
 1st Qu.:0.0000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.00000
 Median :0.0000000   Median :0.000000   Median :0.000000   Median :0.000000   Median :0.00000
 Mean   :0.0009699   Mean   :0.005542   Mean   :0.004849   Mean   :0.002355   Mean   :0.01469
 3rd Qu.:0.0000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.00000
 Max.   :1.0000000   Max.   :2.000000   Max.   :2.000000   Max.   :1.000000   Max.   :4.00000

      d28                d27                d25                d24                d22
 Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000   Min.   :0.000000
 1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.000000
 Median :0.000000   Median :0.000000   Median :0.000000   Median :0.000000   Median :0.000000
 Mean   :0.002286   Mean   :0.003949   Mean   :0.001593   Mean   :0.008521   Mean   :0.002355
 3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.000000
 Max.   :1.000000   Max.   :2.000000   Max.   :2.000000   Max.   :6.000000   Max.   :2.000000

      d20                d19                d18               d16                d13
 Min.   :0.000000   Min.   :0.000000   Min.   :0.00000   Min.   :0.000000   Min.   :0.000000
 1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000
 Median :0.000000   Median :0.000000   Median :0.00000   Median :0.000000   Median :0.000000
 Mean   :0.003325   Mean   :0.001178   Mean   :0.01884   Mean   :0.008452   Mean   :0.004087
 3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000
 Max.   :2.000000   Max.   :1.000000   Max.   :4.00000   Max.   :3.000000   Max.   :2.000000

      d11               d52               d7                 d6                 d0
 Min.   :0.00000   Min.   :0.00000   Min.   :0.000000   Min.   :0.000000   Min.   : 0.0000
 1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000000   1st Qu.:0.000000   1st Qu.: 0.0000
 Median :0.00000   Median :0.00000   Median :0.000000   Median :0.000000   Median : 0.0000
 Mean   :0.00291   Mean   :0.01386   Mean   :0.002771   Mean   :0.001663   Mean   : 0.1214
 3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000000   3rd Qu.:0.000000   3rd Qu.: 0.0000
 Max.   :2.00000   Max.   :3.00000   Max.   :2.000000   Max.   :2.000000   Max.   :11.0000

       y                 dd
 Min.   :  0.000   Min.   : 0.0000
 1st Qu.:  1.540   1st Qu.: 0.0000
 Median :  7.455   Median : 0.0000
 Mean   : 41.319   Mean   : 0.4242
 3rd Qu.: 26.107   3rd Qu.: 0.0000
 Max.   :611.554   Max.   :58.0000
```

The dataset `au.ana`, thus now contains the number of cancers and person-years lived by the Australian T1D population 2000–2008 incl., classified by sex, age (`A`), calendar time (`P`) (both in 1-year classes, coded as the mispoint of the intervals), as well as by diabetes duration, a factor taking the values of the left endpoint of the intervals or the value "Unkn" for those with unknown duration of diabetes.

7.3 Australian population rates

Here we access the Excel sheets downloaded from Australian Institute of Health and Welfare, at https://www.aihw.gov.au/acim-books/. The filenames have been changed after download to simplify the code that runs through them.

7.3.1 Population risk time

We start by getting the population figures (we use the mid-year population numbers as given in any of the sheets).

```
> library( XLConnect )
> cpop <- readWorksheetFromFile( paste("../data/aus/pop/all.xls",sep=""),
+                                  sheet = "Populations",
+                               startRow = 13 )
> cpop <- subset( cpop, Year %in% 2000:2008 )
> str( cpop )
'data.frame':       9 obs. of  65 variables:
 $ Year     : num  2000 2001 2002 2003 2004 ...
 $ X0.4     : num  653221 653053 650563 650616 651502 ...
 $ X5.9     : num  688195 689098 686788 682601 679483 ...
 $ X10.14   : num  680131 688396 695811 703262 708387 ...
 $ X15.19   : num  671954 684154 689986 693648 697859 ...
 $ X20.24   : num  649535 654544 668791 686750 703491 ...
 $ X25.29   : num  716339 694298 682089 676288 675089 ...
 $ X30.34   : num  704211 722451 738914 747724 748782 ...
 $ X35.39   : num  744059 736877 728346 720877 720531 ...
 $ X40.44   : num  715742 729922 745106 755254 759473 ...
 $ X45.49   : num  663238 670907 681079 692759 706985 ...
 $ X50.54   : num  630498 648130 644584 647251 652232 ...
 $ X55.59   : num  487075 509420 545884 578102 597807 ...
 $ X60.64   : num  398235 411183 423058 433865 450490 ...
 $ X65.69   : num  329907 333321 341402 350695 361124 ...
 $ X70.74   : num  297685 301501 301422 299204 297740 ...
 $ X75.79   : num  218191 225821 231304 237596 243017 ...
 $ X80.84   : num  118211 127383 135732 143958 152166 ...
 $ X85.     : num  77038 81367 84624 87147 89793 ...
 $ Col20    : logi  NA NA NA NA NA NA ...
 $ Total    : num  9443465 9561826 9675483 9787597 9895951 ...
 $ Col22    : logi  NA NA NA NA NA NA ...
 $ Year.1   : num  2000 2001 2002 2003 2004 ...
 $ X0.4.1   : num  620507 620632 618479 618520 618674 ...
 $ X5.9.1   : num  653189 653425 650643 647081 645030 ...
 $ X10.14.1: num  648099 655629 662322 667589 671148 ...
 $ X15.19.1: num  643833 655832 661925 666720 670030 ...
 $ X20.24.1: num  630318 635585 646892 663262 677235 ...
 $ X25.29.1: num  721080 699510 681715 673022 668629 ...
 $ X30.34.1: num  714004 735150 751776 761226 760318 ...
 $ X35.39.1: num  752101 746155 737748 730935 730858 ...
 $ X40.44.1: num  724739 740243 755459 765722 770759 ...
 $ X45.49.1: num  670207 679338 689625 702917 717332 ...
 $ X50.54.1: num  619246 643855 643712 650127 657784 ...
 $ X55.59.1: num  470468 492559 532020 566080 589127 ...
 $ X60.64.1: num  394318 405285 416226 427212 444830 ...
 $ X65.69.1: num  342887 344577 352056 360951 371550 ...
 $ X70.74.1: num  331527 332562 329725 325975 322990 ...
 $ X75.79.1: num  285927 290027 292051 294773 296501 ...
 $ X80.84.1: num  188803 200436 209425 218712 227491 ...
```

```
 $ X85..1  : num  174084 182075 187928 192316 196485 ...
 $ Col42   : logi  NA NA NA NA NA NA ...
 $ Total.1 : num  9585337 9712875 9819727 9933140 10036771 ...
 $ Col44   : logi  NA NA NA NA NA NA ...
 $ Year.2  : num  2000 2001 2002 2003 2004 ...
 $ X0.4.2  : num  1273728 1273685 1269042 1269136 1270176 ...
 $ X5.9.2  : num  1341384 1342523 1337431 1329682 1324513 ...
 $ X10.14.2: num  1328230 1344025 1358133 1370851 1379535 ...
 $ X15.19.2: num  1315787 1339986 1351911 1360368 1367889 ...
 $ X20.24.2: num  1279853 1290129 1315683 1350012 1380726 ...
 $ X25.29.2: num  1437419 1393808 1363804 1349310 1343718 ...
 $ X30.34.2: num  1418215 1457601 1490690 1508950 1509100 ...
 $ X35.39.2: num  1496160 1483032 1466094 1451812 1451389 ...
 $ X40.44.2: num  1440481 1470165 1500565 1520976 1530232 ...
 $ X45.49.2: num  1333445 1350245 1370704 1395676 1424317 ...
 $ X50.54.2: num  1249744 1291985 1288296 1297378 1310016 ...
 $ X55.59.2: num  957543 1001979 1077904 1144182 1186934 ...
 $ X60.64.2: num  792553 816468 839284 861077 895320 ...
 $ X65.69.2: num  672794 677898 693458 711646 732674 ...
 $ X70.74.2: num  629212 634063 631147 625179 620730 ...
 $ X75.79.2: num  504118 515848 523355 532369 539518 ...
 $ X80.84.2: num  307014 327819 345157 362670 379657 ...
 $ X85..2  : num  251122 263442 272552 279463 286278 ...
 $ Col64   : logi  NA NA NA NA NA NA ...
 $ Total.2 : num  19028802 19274701 19495210 19720737 19932722 ...
> cpop[1:3,1:19]
   Year   X0.4   X5.9 X10.14 X15.19 X20.24 X25.29 X30.34 X35.39 X40.44 X45.49 X50.54 X55.59 X60.64
33 2000 653221 688195 680131 671954 649535 716339 704211 744059 715742 663238 630498 487075 398235
34 2001 653053 689098 688396 684154 654544 694298 722451 736877 729922 670907 648130 509420 411183
35 2002 650563 686788 695811 689986 668791 682089 738914 728346 745106 681079 644584 545884 423058
   X65.69 X70.74 X75.79 X80.84  X85.
33 329907 297685 218191 118211 77038
34 333321 301501 225821 127383 81367
35 341402 301422 231304 135732 84624
> cpop[1:3,1:19+22]
   Year.1 X0.4.1 X5.9.1 X10.14.1 X15.19.1 X20.24.1 X25.29.1 X30.34.1 X35.39.1 X40.44.1 X45.49.1
33   2000 620507 653189   648099   643833   630318   721080   714004   752101   724739   670207
34   2001 620632 653425   655629   655832   635585   699510   735150   746155   740243   679338
35   2002 618479 650643   662322   661925   646892   681715   751776   737748   755459   689625
   X50.54.1 X55.59.1 X60.64.1 X65.69.1 X70.74.1 X75.79.1 X80.84.1 X85..1
33   619246   470468   394318   342887   331527   285927   188803 174084
34   643855   492559   405285   344577   332562   290027   200436 182075
35   643712   532020   416226   352056   329725   292051   209425 187928
> cpop[,1:5]
   Year   X0.4   X5.9 X10.14 X15.19
33 2000 653221 688195 680131 671954
34 2001 653053 689098 688396 684154
35 2002 650563 686788 695811 689986
36 2003 650616 682601 703262 693648
37 2004 651502 679483 708387 697859
38 2005 656043 677441 710978 705932
39 2006 664456 678901 710385 714616
40 2007 686251 680272 709912 729591
41 2008 710252 683671 710306 743757
> md <- reshape( data = cpop[,1:19],
+             varying = -1,
+             v.names = "y",
+           direction = "long" )
> fd <- reshape( data = cpop[,c(1,2:19+22)],
+             varying = -1,
+             v.names = "y",
+           direction = "long" )
> au.pop <- transform( rbind( cbind( sex=1, md ),
+                             cbind( sex=2, fd ) ),
+                       sex = factor( sex, labels=c("M","F") ),
```

```
+                        A = time*5-2.5,
+                        P = Year+0.5 )[,c("sex","A","P","y")]
> str( au.pop )
'data.frame':        324 obs. of  4 variables:
 $ sex: Factor w/ 2 levels "M","F": 1 1 1 1 1 1 1 1 1 1 ...
 $ A  : num  2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5 7.5 ...
 $ P  : num  2000 2002 2002 2004 2004 ...
 $ y  : num  653221 653053 650563 650616 651502 ...
> head( au.pop )
    sex   A      P      y
1.1   M 2.5 2000.5 653221
2.1   M 2.5 2001.5 653053
3.1   M 2.5 2002.5 650563
4.1   M 2.5 2003.5 650616
5.1   M 2.5 2004.5 651502
6.1   M 2.5 2005.5 656043
> round( addmargins( xtabs( y/1000 ~ A + P, data=au.pop ) ), 1 )
      P
A        2000.5   2001.5   2002.5   2003.5   2004.5   2005.5   2006.5   2007.5   2008.5      Sum
  2.5    1273.7   1273.7   1269.0   1269.1   1270.2   1277.5   1294.5   1336.5   1383.1  11647.4
  7.5    1341.4   1342.5   1337.4   1329.7   1324.5   1321.0   1324.3   1327.6   1334.7  11983.1
  12.5   1328.2   1344.0   1358.1   1370.9   1379.5   1384.8   1383.5   1382.8   1383.1  12315.1
  17.5   1315.8   1340.0   1351.9   1360.4   1367.9   1379.5   1392.7   1420.7   1447.6  12376.5
  22.5   1279.9   1290.1   1315.7   1350.0   1380.7   1414.6   1448.4   1483.1   1526.4  12489.0
  27.5   1437.4   1393.8   1363.8   1349.3   1343.7   1352.6   1381.6   1431.0   1500.0  12553.2
  32.5   1418.2   1457.6   1490.7   1509.0   1509.1   1500.5   1474.1   1457.1   1458.3  13274.6
  37.5   1496.2   1483.0   1466.1   1451.8   1451.4   1468.4   1508.8   1555.7   1589.6  13471.0
  42.5   1440.5   1470.2   1500.6   1521.0   1530.2   1527.6   1516.4   1504.2   1499.4  13510.1
  47.5   1333.4   1350.2   1370.7   1395.7   1424.3   1451.1   1477.7   1509.8   1537.8  12850.9
  52.5   1249.7   1292.0   1288.3   1297.4   1310.0   1325.5   1347.8   1373.1   1397.8  11881.6
  57.5    957.5   1002.0   1077.9   1144.2   1186.9   1226.4   1258.0   1254.2   1268.6  10375.8
  62.5    792.6    816.5    839.3    861.1    895.3    935.3    978.8   1055.2   1117.6   8291.6
  67.5    672.8    677.9    693.5    711.6    732.7    754.9    773.1    800.6    827.2   6644.3
  72.5    629.2    634.1    631.1    625.2    620.7    619.1    624.1    640.2    657.9   5681.7
  77.5    504.1    515.8    523.4    532.4    539.5    543.8    546.7    547.7    546.9   4800.4
  82.5    307.0    327.8    345.2    362.7    379.7    391.9    401.4    410.9    420.6   3347.2
  87.5    251.1    263.4    272.6    279.5    286.3    302.2    318.7    337.1    352.6   2663.4
  Sum   19028.8  19274.7  19495.2  19720.7  19932.7  20176.8  20451.0  20827.6  21249.2 180156.8
```

The dataset `au.pop` now has the Australian population risk time by sex, age (5-year classes, `A`) and calendar time (one-year classes, `P`), the latter two coded as the midpoint if the intervals.

7.3.2 Population cancer cases

Then we get the number of cancers and attach these to the dataset with the person-years:

```
> for( pn in pnam ) #  (pn <- pnam[4])
+    {
+ dn <- dnam[match(pn,pnam)]
+ cat( dn, "\n" ) ; flush.console()
+ cpop <- readWorksheetFromFile( paste("../data/aus/pop/",pn,".xls",sep=""),
+                                sheet = "Raw data",
+                             startRow = 2 )
+ cpop <- subset( cpop, diagnosis_year %in% 2000:2008 )[1:9,]
+ for( i in 1:ncol(cpop) ) cpop[,i] <- round( as.numeric(cpop[,i]))
+ names( cpop )
+ mc <-  2:19
+ fc <- 21:38
+ md <- reshape( data = cpop[,c(1,mc)],
+             varying = -1,
+             v.names = dn,
+           direction = "long" )
```

```
+ fd <- reshape( data = cpop[,c(1,fc)],
+              varying = -1,
+              v.names = dn,
+            direction = "long" )
+ res <- transform( rbind( cbind( sex=1, md ),
+                          cbind( sex=2, fd ) ),
+                   sex = factor( sex, labels=c("M","F") ),
+                     A = time*5-2.5,
+                     P = diagnosis_year+0.5 )[,c("sex","A","P",dn)]
+ names(au.pop)
+ names(res)
+ dim(au.pop)
+ dim(res)
+ au.pop <- merge( au.pop, res, all.x=TRUE )
+ cat( names(au.pop), "\n" )
+     }
d0
sex A P y d0
d6
sex A P y d0 d6
d7
sex A P y d0 d6 d7
d52
sex A P y d0 d6 d7 d52
d11
sex A P y d0 d6 d7 d52 d11
d13
sex A P y d0 d6 d7 d52 d11 d13
d16
sex A P y d0 d6 d7 d52 d11 d13 d16
d18
sex A P y d0 d6 d7 d52 d11 d13 d16 d18
d19
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19
d20
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20
d22
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22
d24
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24
d25
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25
d27
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27
d28
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27 d28
d29
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27 d28 d29
d32
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27 d28 d29 d32
d33
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27 d28 d29 d32 d33
d36
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27 d28 d29 d32 d33 d36
d37
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27 d28 d29 d32 d33 d36 d37
d38
sex A P y d0 d6 d7 d52 d11 d13 d16 d18 d19 d20 d22 d24 d25 d27 d28 d29 d32 d33 d36 d37 d38
```

7.4 The analysis dataset

We can finally put this together, remembering to put in the primary exposure variable, T1D status, and the country variable, "`Cnt`":

```
> au.ana <- rbind( au.ana,
+                  cbind( au.pop, T1D = factor("NoDM"),
+                                 DMdur = factor("NoDM"),
+                                    dd = NA ) )
> au.ana <- transform( au.ana,
+                       T1D = relevel( T1D  , "NoDM" ),
+                     DMdur = Relevel( DMdur, c("NoDM","Unkn") ) )
> au.ana <- transform( au.ana,
+                    DMprev = Relevel( T1D, list("Pop" = "NoDM",
+                                                "Inc" = 2:nlevels(T1D)) ),
+                        y0 = y,
+                       Cnt = "AUS" )
> summ <- function(x) sum(x,na.rm=TRUE)
> with( au.ana, addmargins( tapply(d0, list( P, T1D ), summ ), 1, FUN=summ ) )
         NoDM  30  35  40  45 Inf
2000.5  87635  26  33  28  49   5
2001.5  90130  31  23  48  59   1
2002.5  93886  51  23  37  48   6
2003.5  95473  40  30  39  71   6
2004.5 100314  40  40  46  59   6
2005.5 103274  59  30  52  70   2
2006.5 105915  47  31  55  87   3
2007.5 109557  45  31  58  82   8
2008.5 113187  59  41  68  74   6
summ   899371 398 282 431 599  43
> with( au.ana, addmargins( tapply(d0, list( A, T1D ), summ ), 1, FUN=summ ) )
        NoDM  30  35  40  45 Inf
0.5       NA   0  NA  NA  NA  NA
1.5       NA   0  NA  NA  NA  NA
2.5     2452   1  NA  NA  NA  NA
3.5       NA   0  NA  NA  NA  NA
4.5       NA   0  NA  NA  NA  NA
5.5       NA   2  NA  NA  NA  NA
6.5       NA   1  NA  NA  NA  NA
7.5     1246   0  NA  NA  NA  NA
8.5       NA   0  NA  NA  NA  NA
9.5       NA   1  NA  NA  NA  NA
10.5      NA   0  NA  NA  NA  NA
11.5      NA   0  NA  NA  NA  NA
12.5    1498   1  NA  NA  NA  NA
13.5      NA   8  NA  NA  NA  NA
14.5      NA   1  NA  NA  NA  NA
15.5      NA   2  NA  NA  NA  NA
16.5      NA   3  NA  NA  NA  NA
17.5    2992   3  NA  NA  NA  NA
18.5      NA   0  NA  NA  NA  NA
19.5      NA   2  NA  NA  NA  NA
20.5      NA   2  NA  NA  NA  NA
21.5      NA   3  NA  NA  NA  NA
22.5    4885   6  NA  NA  NA  NA
23.5      NA   3   0  NA  NA  NA
24.5      NA   8   0   0  NA  NA
25.5      NA   7   0   0  NA  NA
26.5      NA   6   1   0  NA  NA
27.5    8020   9   0   0  NA  NA
28.5      NA  10   0   0  NA  NA
29.5      NA   9   0   0  NA  NA
30.5      NA   5   0   0  NA  NA
31.5      NA  10   0   0  NA  NA
32.5   12771   8   0   0   0  NA
33.5      NA  11   4   0   0  NA
34.5      NA  19   3   0   0  NA
35.5      NA  16   4   1   0   0
36.5      NA  11   3   1   0   0
37.5   19340  22   4   1   0   0
38.5      NA  10   5   2   0   0
```

```
39.5       NA  12   9   8   0   0
40.5       NA  22  15  10   1   0
41.5       NA  20  10   6   1   0
42.5    30371  25  19   5   2   0
43.5       NA  15  18   5   6   0
44.5       NA  15  12  13   7   0
45.5       NA  10  19  16   8   0
46.5       NA  11  12   7   8   1
47.5    45468   7  19  30   7   0
48.5       NA  10  20  25  14   0
49.5       NA   4  16  22  18   1
50.5       NA   5  16  21  23   1
51.5       NA   2  13  31  17   1
52.5    65178   2   7  30  27   1
53.5       NA   2  16  34  26   0
54.5       NA   1   7  19  29   1
55.5       NA   4   4  23  42   1
56.5       NA   1   2  27  51   0
57.5    88777   1   0  23  40   4
58.5       NA   0   2  17  38   1
59.5       NA   0   1  19  39   0
60.5       NA   1   1   3  47   3
61.5       NA   2   3   3  38   1
62.5   102983   3   2   4  25   3
63.5       NA   3   0   5  29   3
64.5       NA   3   1   3  12   2
65.5       NA   1   2   0   8   4
66.5       NA   0   2   1   4   2
67.5   114444   4   3   1   4   2
68.5       NA   0   0   1   2   1
69.5       NA   0   0   3   3   2
70.5       NA   2   1   1   2   1
71.5       NA   0   1   2   3   1
72.5   118130   0   1   2   3   0
73.5       NA   0   0   2   4   2
74.5       NA   1   0   0   1   0
75.5       NA   2   0   1   1   0
76.5       NA   1   1   0   2   0
77.5   117974   1   0   1   2   1
78.5       NA   1   0   0   2   1
79.5       NA   0   1   0   0   1
80.5       NA   1   0   1   0   0
81.5       NA   1   1   0   0   1
82.5    90393   0   0   0   2   0
83.5       NA   1   0   1   1   0
84.5       NA   0   0   0   0   0
85.5       NA   1   0   0   0   0
86.5       NA   0   0   0   0   0
87.5    72449   0   1   0   0  NA
88.5       NA   0   0   0   0   0
89.5       NA   0   0   0   0   0
90.5       NA   0   0   0   0  NA
91.5       NA   0  NA   0   0  NA
92.5       NA   0   0  NA   0  NA
93.5       NA   0   0  NA   0  NA
94.5       NA   0   0   0   0  NA
95.5       NA  NA  NA   0  NA  NA
96.5       NA   0   0  NA   0  NA
102.5      NA   0  NA  NA   0  NA
103.5      NA  NA  NA   0   0  NA
summ   899371 398 282 431 599  43
> summary( au.ana )
 sex          T1D             A                  P              DMdur              d38
 M:7205    NoDM: 324   Min.   :  0.50   Min.   :2000   Unkn    :3759   Min.   :  0.0000
 F:7554    30   :6033   1st Qu.: 31.50   1st Qu.:2002   15      :1906   1st Qu.:  0.0000
           35   :2270   Median : 46.50   Median :2004   30      :1892   Median :  0.0000
```

```
         40  :2170   Mean   : 45.41   Mean   :2005   5       :1670   Mean   :  0.7576
         45  :2000   3rd Qu.: 60.50   3rd Qu.:2006   2       :1487   3rd Qu.:  0.0000
         Inf :1962   Max.   :103.50   Max.   :2008   10      :1393   Max.   :133.0000
                                                     (Other):2652
      d37              d36              d33               d32              d29
 Min.   : 0.0000   Min.   :  0.000   Min.   :  0.0000   Min.   :  0.0000   Min.   :  0.000
 1st Qu.: 0.0000   1st Qu.:  0.000   1st Qu.:  0.0000   1st Qu.:  0.0000   1st Qu.:  0.000
 Median : 0.0000   Median :  0.000   Median :  0.0000   Median :  0.0000   Median :  0.000
 Mean   : 0.2933   Mean   :  2.349   Mean   :  0.9172   Mean   :  0.8619   Mean   :  6.082
 3rd Qu.: 0.0000   3rd Qu.:  0.000   3rd Qu.:  0.0000   3rd Qu.:  0.0000   3rd Qu.:  0.000
 Max.   :38.0000   Max.   :305.000   Max.   :195.0000   Max.   :107.0000   Max.   :802.000

      d28              d27              d25               d24               d22
 Min.   :  0.000   Min.   :  0.000   Min.   :  0.0000   Min.   :   0.000   Min.   :  0.0000
 1st Qu.:  0.000   1st Qu.:  0.000   1st Qu.:  0.0000   1st Qu.:   0.000   1st Qu.:  0.0000
 Median :  0.000   Median :  0.000   Median :  0.0000   Median :   0.000   Median :  0.0000
 Mean   :  1.368   Mean   :  1.405   Mean   :  0.3985   Mean   :   9.465   Mean   :  0.7519
 3rd Qu.:  0.000   3rd Qu.:  0.000   3rd Qu.:  0.0000   3rd Qu.:   0.000   3rd Qu.:  0.0000
 Max.   :349.000   Max.   :243.000   Max.   :136.0000   Max.   :4125.000   Max.   :172.0000

      d20              d19               d18                d16                d13
 Min.   :  0.000   Min.   :  0.0000   Min.   :   0.000   Min.   :   0.000   Min.   :  0.000
 1st Qu.:  0.000   1st Qu.:  0.0000   1st Qu.:   0.000   1st Qu.:   0.000   1st Qu.:  0.000
 Median :  0.000   Median :  0.0000   Median :   0.000   Median :   0.000   Median :  0.000
 Mean   :  1.088   Mean   :  0.4508   Mean   :   7.567   Mean   :   5.575   Mean   :  1.328
 3rd Qu.:  0.000   3rd Qu.:  0.0000   3rd Qu.:   0.000   3rd Qu.:   0.000   3rd Qu.:  0.000
 Max.   :334.000   Max.   :107.0000   Max.   :1885.000   Max.   :1154.000   Max.   :260.000

      d11                d52               d7                d6                 d0
 Min.   :  0.0000   Min.   :   0.000   Min.   :  0.000   Min.   :  0.0000   Min.   :   0.00
 1st Qu.:  0.0000   1st Qu.:   0.000   1st Qu.:  0.000   1st Qu.:  0.0000   1st Qu.:   0.00
 Median :  0.0000   Median :   0.000   Median :  0.000   Median :  0.0000   Median :   0.00
 Mean   :  0.6414   Mean   :   8.046   Mean   :  1.178   Mean   :  0.7217   Mean   :  61.06
 3rd Qu.:  0.0000   3rd Qu.:   0.000   3rd Qu.:  0.000   3rd Qu.:  0.0000   3rd Qu.:   0.00
 Max.   :136.0000   Max.   :1322.000   Max.   :233.000   Max.   :136.0000   Max.   :9802.00

       y                   dd           DMprev           y0                Cnt
 Min.   :     0.0   Min.   : 0.0000   Pop:  324   Min.   :     0.0   AUS:14759
 1st Qu.:     1.6   1st Qu.: 0.0000   Inc:14435   1st Qu.:     1.6
 Median :     7.9   Median : 0.0000               Median :     7.9
 Mean   : 12247.0   Mean   : 0.4242               Mean   : 12247.0
 3rd Qu.:    29.0   3rd Qu.: 0.0000               3rd Qu.:    29.0
 Max.   :800808.0   Max.   :58.0000               Max.   :800808.0
                    NA's   :324
> save( au.ana, file="../data/AUana.Rda" )
```

7.5 Analysis of SMR

We load the dataset and the dataframe with the tumour labels in it:

```
> load( file="../data/AUana.Rda" )
> levels( au.ana$T1D )
[1] "NoDM" "30"   "35"   "40"   "45"   "Inf"
> levels( au.ana$DMdur )
[1] "NoDM" "Unkn" "0"    "1"    "2"    "5"    "10"   "15"   "30"
> au.ana <- subset( au.ana, y>0 & P>1997 & A<85 )
> round( addmargins( xtabs( d0 ~ P + T1D, data=au.ana ), 1 ) )
        T1D
P         NoDM    30    35    40    45   Inf
  2000.5  80927   26    33    28    49     5
  2001.5  83093   31    23    48    59     1
  2002.5  86480   49    22    37    48     6
```

```
  2003.5  88023     39     29     39     71      6
  2004.5  92575     39     40     46     59      6
  2005.5  94913     59     30     52     70      2
  2006.5  97165     47     31     55     87      3
  2007.5 100209     45     31     58     82      8
  2008.5 103537     58     41     68     74      6
  Sum    826922    393    280    431    599     43
> round( addmargins( xtabs( y0 ~ P + T1D, data=au.ana ), 1 ), 1 )
        T1D
P                NoDM          30          35          40          45         Inf
  2000.5  18777680.0     30817.6      8651.0      8763.9      8988.9       357.2
  2001.5  19011259.0     32664.4      8875.9      8917.4      9073.1       378.5
  2002.5  19222658.0     34507.4      9075.3      9023.2      9108.9       398.3
  2003.5  19441274.0     36379.7      9239.7      9090.1      9116.1       424.3
  2004.5  19646444.0     38314.5      9396.8      9181.3      9095.3       443.1
  2005.5  19874653.0     40166.6      9523.6      9245.2      9068.9       461.8
  2006.5  20132274.0     41997.0      9662.6      9280.4      9034.6       483.0
  2007.5  20490486.0     43889.4      9811.4      9339.6      9033.4       497.1
  2008.5  20896643.0     45743.8      9936.5      9368.5      8999.2       509.0
  Sum    177493371.0    344480.4     84172.6     82209.6     81518.4      3952.1
> load( file="../data/conv.Rda" )
```

7.5.1 All cancers

We first set out the simplest possible analysis with age-period-cohort effects for the baseline rates. First we devise a couple of knots for the splines:

```
> library( splines )
> with( au.ana, rbind(
+  A = quantile( rep(  A,d0), 0:10/10 ),
+  P = quantile( rep(P   ,d0), 0:10/10 ),
+  C = quantile( rep(P-A,d0), 0:10/10 ) ) )
      0%    10%    20%    30%    40%    50%    60%    70%    80%    90%   100%
A    2.5   42.5   52.5   57.5   62.5   67.5   67.5   72.5   77.5   82.5   83.5
P 2000.5 2001.5 2002.5 2002.5 2003.5 2004.5 2005.5 2006.5 2007.5 2008.5 2008.5
C 1918.0 1924.0 1928.0 1931.0 1935.0 1939.0 1943.0 1947.0 1952.0 1960.0 2006.0
> ( a.kn <- seq(5,80,,7) )
[1]  5.0 17.5 30.0 42.5 55.0 67.5 80.0
> ( p.kn <- seq(1998,2008,,4) )
[1] 1998.000 2001.333 2004.667 2008.000
> ( c.kn <- seq(1925,1960,,6) )
[1] 1925 1932 1939 1946 1953 1960
```

We then fit 3 models for the RR for T1D patients relative to the general population, using a common shape of the underlying cancer incidence rates as an age-period cohort model with $1 + (7 - 1) + (4 - 1) + (6 - 1) - 1 = 14$ parameters[2]. The first model (`m3`) is one with separate effects for persons diagnosed in ages < 30, 30–35, 35–40, 40–45 and 45+, the second (`m1`) is a simplification where these groups are pooled:

```
> m3 <- glm( d0 ~ Ns(   A, knots=a.kn ) +
+                 Ns( P   , knots=p.kn ) +
+                 Ns( P-A, knots=c.kn ) +
+                 T1D,
+            offset = log(y),
+            family = poisson,
+              data = au.ana )
> m1  <- update( m3, . ~ . - T1D + DMprev )
> m1r <- update( m1, data=subset( au.ana, T1D %in% levels(T1D)[1:4] ) )
> anova( m1, m3, test="Chisq" )
```

[2]There is first an intercept, then the three natural splines, where k knots gives $k - 1$ parameters, and finally 1 aliased parameter from the linear relationship between $P - A$ and P and A.

```
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMprev
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     14267      61046
2     14263      61041  4   4.3721    0.358
```

The tests of the models show that there is no detectable difference between the groups, but it does appear that the risk is lower the older the patients are at diagnosis:

```
> round( rbind( ci.exp( m3 , subset="T1" ),
+               ci.exp( m1 , subset="DMprev" ),
+               ci.exp( m1r, subset="DMprev" ) ), 3 )
          exp(Est.)  2.5% 97.5%
T1D30         1.073 0.972 1.185
T1D35         1.071 0.952 1.204
T1D40         1.067 0.971 1.172
T1D45         0.967 0.893 1.048
T1DInf        0.941 0.698 1.268
DMprevInc     1.029 0.982 1.079
DMprevInc     1.070 1.009 1.135
```

In summary, the comparison of the models show that there is no heterogeneity between the five groups of T1D patients w.r.t. the occurrence of cancer; which is also evident in figure 7.1, where the RR estimates from the three models are shown together:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind( ci.exp( m3, subset="T1D"),
+                 ci.exp( m1, subset="DMprev") ),
+          y = 6:1,
+          txt = c("<30","30-35","35-40","40-45","45+","Pooled"),
+          xlog=TRUE, xtic=c(5:10,15,20)/10, grid=TRUE,
+          xlab="RR of any cancer, T1D vs. population",
+          lwd=4, cex=2, vref=1 )
```

7.5.2 Analysis by duration

We have also devised a variable indicating different time bands after diagnosis:

```
> round( xtabs( cbind(Ca=d0,PY=y/1000) ~ DMdur, data=au.ana ), 1 )
DMdur         Ca       PY
  NoDM 826922.0 177493.4
  Unkn   1296.0    401.2
  0        33.0     17.5
  1        14.0     19.1
  2        58.0     56.9
  5        93.0     59.5
  10       48.0     15.2
  15      123.0     19.9
  30       81.0      7.1
```

The above analyses showed no statistically significant difference between the groups of patients by age at inclusion, but in line with what was found in previous analyses of cancer occurrende in `all` Danish diabetes patients, we fit models with the smaller datasets corresponding to more restrictive definitions of T1D:

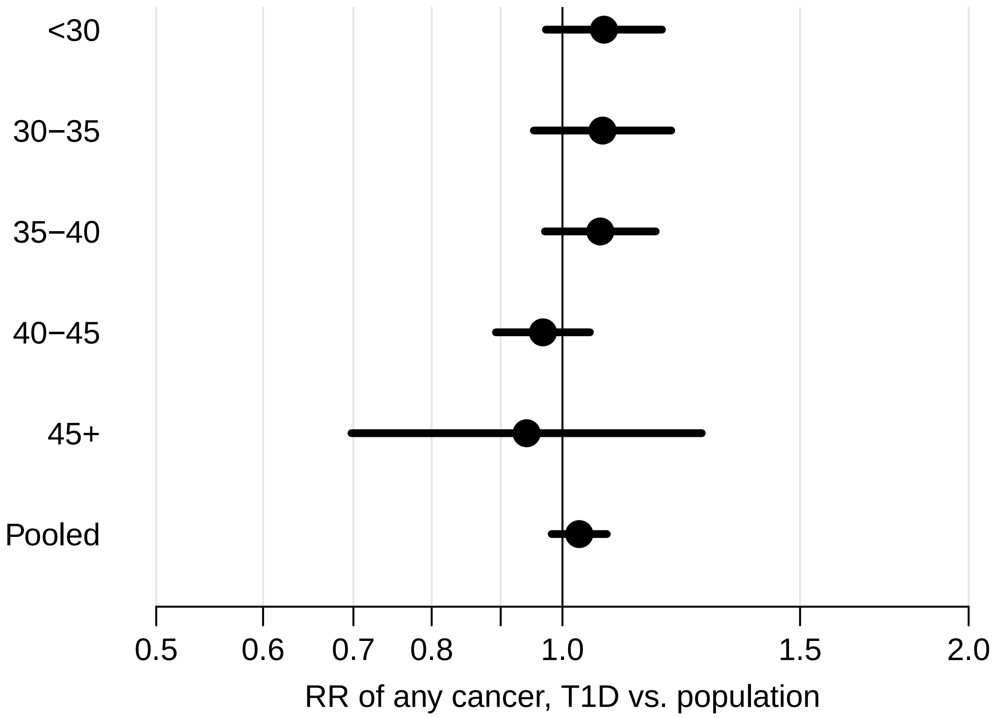

Figure 7.1: *Estimated RRs relative to the general population in two different models.*

```
> md45 <- update( m3, . ~ . - T1D + DMdur,
+                       data=subset( au.ana, T1D %in% levels(T1D)[1:5] & DMdur != "Unkn" ) )
> md40 <- update( md45, data=subset( au.ana, T1D %in% levels(T1D)[1:4] & DMdur != "Unkn" ) )
> md35 <- update( md45, data=subset( au.ana, T1D %in% levels(T1D)[1:3] & DMdur != "Unkn" ) )
> md30 <- update( md45, data=subset( au.ana, T1D %in% levels(T1D)[1:2] & DMdur != "Unkn" ) )
> anova( md45, update( md45, . ~ . - DMdur + T1D ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1      9496      59225
2      9499      59249 -3  -23.987 2.514e-05
> anova( md40, update( md40, . ~ . - DMdur + T1D ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      8076      58632
2      8080      58650 -4  -17.845 0.001323
> anova( md35, update( md35, . ~ . - DMdur + T1D ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      6507      58103
2      6512      58131 -5  -28.647 2.72e-05
```

```
> anova( md30, update( md30, . ~ . - DMdur + T1D ), test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    DMdur
Model 2: d0 ~ Ns(A, knots = a.kn) + Ns(P, knots = p.kn) + Ns(P - A, knots = c.kn) +
    T1D
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1      4872      57696
2      4878      57724 -6  -27.323 0.000126
```

We see that there a very strong effect of duration, and it is the same regardless of the age cut-off:

```
> round( cbind( RR45 <- ci.exp( md45, subset="DMdur" ),
+               RR40 <- ci.exp( md40, subset="DMdur" ),
+               RR35 <- ci.exp( md35, subset="DMdur" ),
+               RR30 <- ci.exp( md30, subset="DMdur" ) ), 2 )
        exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5% exp(Est.) 2.5%
DMdur0       3.22 2.23  4.63      3.22 2.14  4.85      3.99 2.63  6.06      4.21 2.61
DMdur1       1.24 0.74  2.10      1.03 0.54  1.98      0.61 0.23  1.62      0.64 0.21
DMdur2       1.15 0.87  1.51      1.11 0.81  1.53      1.05 0.71  1.56      1.20 0.77
DMdur5       1.17 0.95  1.44      1.09 0.85  1.41      1.04 0.75  1.43      0.87 0.55
DMdur10      1.16 0.87  1.55      1.19 0.82  1.73      0.70 0.36  1.34      0.42 0.14
DMdur15      1.04 0.87  1.24      1.15 0.92  1.44      1.11 0.82  1.49      0.93 0.60
DMdur30      0.99 0.80  1.23      1.01 0.80  1.29      1.04 0.79  1.38      1.06 0.76
        97.5%
DMdur0   6.77
DMdur1   1.98
DMdur2   1.88
DMdur5   1.36
DMdur10  1.31
DMdur15  1.42
DMdur30  1.48
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( rbind(NA,RR45),
+          txt=c("Years since DM","0-1","1-2","2-5","5-10","10-15","15-30","30+"),
+             xlog=TRUE, xtic=c(c(5:10,15,20,25)/10,3:6), grid=TRUE,
+             xlab="RR of any cancer, T1D vs. population",
+             lwd=4, cex=2, vref=1, y=c(7.7,7:1)+0.21 )
> linesEst( rbind(NA,RR40),
+           lwd=3, cex=1.5, col=gray(0.5), y=8:1+0.07 )
> linesEst( rbind(NA,RR35),
+           lwd=3, cex=1.5, col=gray(0.5), y=8:1-0.07 )
> linesEst( rbind(NA,RR30),
+           lwd=3, cex=1.5, col=gray(0.7), y=8:1-0.21 )
```

From figure 7.2 it seems that there is a strong ascertainment effect for T1DM as well as what have been shown for all diabetes under one, and it is not attenuated if a stricter definition of T1D is applied.

7.5.3 Site-specific analyses

7.5.3.1 Analyses of site specific cancers

We first set up an array to hold the resulting simple RRs for each of the sites; we do the analyses by sex, but also make a pooled analysis, except for the sex-specific cancers (including breast):

```
> wh <- c(26:6)
> vnam <- names(au.ana)[wh]
> site <- conv[match(vnam,conv$NCnam),"Clab"]
> data.frame( vnam, site )
```

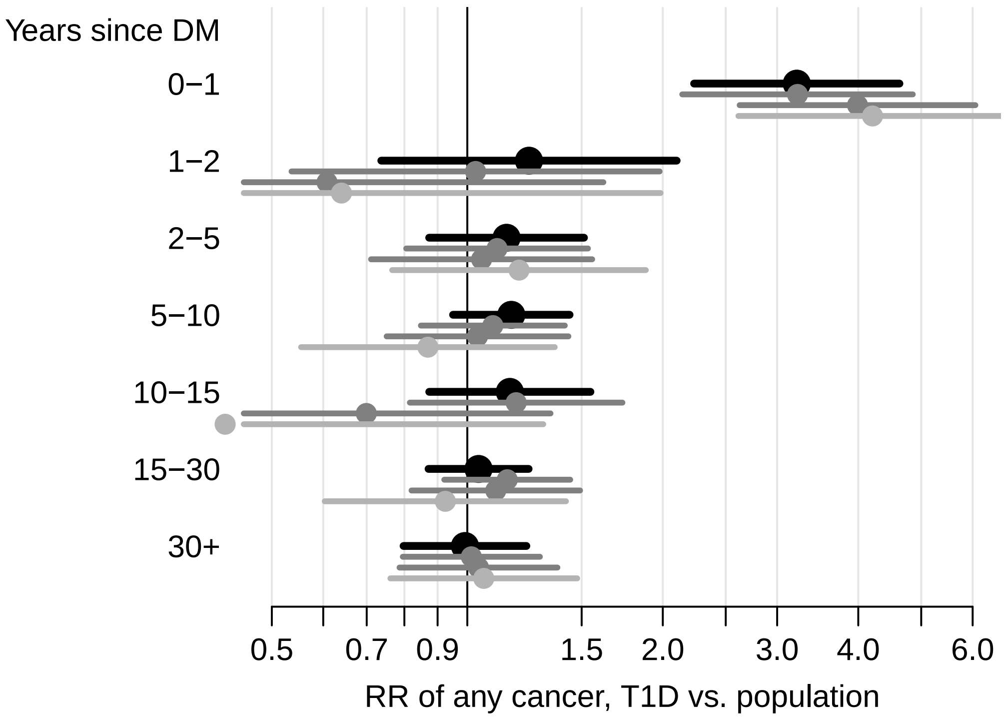

Figure 7.2: *The effect of duration on the RR of cancer. The three gray sets of effecs are from the models where data are further restricted to patients diagnosed under 40, 35 and 30, respectively.*

```
   vnam                 site
1    d0            All sites
2    d6           Oesophagus
3    d7              Stomach
4   d52           Colorectal
5   d11                Liver
6   d13             Pancreas
7   d16                 Lung
8   d18               Breast
9   d19         Cervix uteri
10  d20         Corpus uteri
11  d22                Ovary
12  d24             Prostate
13  d25               Testis
14  d27               Kidney
15  d28              Bladder
16  d29     Melanoma of skin
17  d32           Brain, CNS
18  d33              Thyroid
19  d36 Non-Hodgkin lymphoma
20  d37     Hodgkin lymphoma
21  d38     Multiple myeloma
> RRtab <- NArray( list( site = site,
+                         sex = c(levels(au.ana$sex),"Both"),
+                        what = c("N.Pop","N.T1D","RR","lo","hi") ) )
> dimnames( RRtab )
$site
 [1] "All sites"            "Oesophagus"           "Stomach"
 [4] "Colorectal"           "Liver"                "Pancreas"
 [7] "Lung"                 "Breast"               "Cervix uteri"
[10] "Corpus uteri"         "Ovary"                "Prostate"
```

```
[13] "Testis"              "Kidney"              "Bladder"
[16] "Melanoma of skin"    "Brain, CNS"          "Thyroid"
[19] "Non-Hodgkin lymphoma" "Hodgkin lymphoma"   "Multiple myeloma"

$sex
[1] "M"    "F"    "Both"

$what
[1] "N.Pop" "N.T1D" "RR"    "lo"    "hi"
```

With this fixed we can the make a loop doing the analysis for all sites:

```
> system.time(
+ for( i in 1:length(vnam) )
+    { # i <- 1
+    aset <- au.ana[,c(vnam[i],
+                      paste("y",if(i==1) "0", sep=""),
+                      "A","P","DMprev","sex")]
+    names( aset )[1:2] <- c("D","Y")
+    mB <- glm( D ~ Ns  ( A, knots=a.kn ) +
+                   Ns( P  , knots=p.kn ) +
+                   Ns( P-A, knots=c.kn ) +
+                   DMprev,
+            offset = log(Y),
+            family = poisson,
+              data = aset )
+    mM <- update( mB, data=subset(aset,sex=="M") )
+    mF <- update( mB, data=subset(aset,sex=="F") )
+    RRtab[i,"M"   ,3:5] <- ci.exp( mM, subset="DMprev" )
+    RRtab[i,"F"   ,3:5] <- ci.exp( mF, subset="DMprev" )
+    RRtab[i,"Both",3:5] <- ci.exp( mB, subset="DMprev" )
+    RRtab[i,,1:2] <- addmargins( with( aset, tapply(D,list(sex,DMprev),sum) ), 1 )
+    } )
   user  system elapsed
 13.694   0.024  13.715
> RRorg <- RRtab
> RRtab <- RRorg
> for(i in 1:dim(RRtab)[1])
+ for(j in 1:dim(RRtab)[2]) if(          RRtab[i,j,5]==Inf  ) RRtab[i, j    ,] <- NA
> for(i in 1:dim(RRtab)[1]) if(any(is.na(RRtab[i, ,5]==Inf))) RRtab[i,"Both",] <- NA
> round( ftable( RRtab ), 2 )
                     what      N.Pop     N.T1D        RR        lo        hi
site                 sex
All sites            M      466839.00    874.00      0.97      0.91      1.04
                     F      360083.00    872.00      1.08      1.01      1.15
                     Both   826922.00   1746.00      1.03      0.98      1.08
Oesophagus           M        6672.00     17.00      1.35      0.84      2.18
                     F        2692.00      6.00      2.26      1.01      5.06
                     Both     9364.00     23.00      1.60      1.06      2.41
Stomach              M       10195.00     21.00      1.24      0.81      1.90
                     F        5108.00     18.00      2.35      1.48      3.73
                     Both    15303.00     39.00      1.64      1.19      2.24
Colorectal           M       60661.00    109.00      1.05      0.87      1.27
                     F       46444.00     90.00      1.36      1.10      1.67
                     Both   107105.00    199.00      1.19      1.03      1.37
Liver                M        6348.00     31.00      2.22      1.56      3.16
                     F        2359.00     11.00      3.27      1.81      5.93
                     Both     8707.00     42.00      2.54      1.87      3.44
Pancreas             M        9111.00     41.00      2.65      1.95      3.61
                     F        7686.00     18.00      2.14      1.35      3.41
                     Both    16797.00     59.00      2.53      1.96      3.27
Lung                 M       48270.00     72.00      1.09      0.87      1.38
                     F       27576.00     49.00      1.28      0.97      1.70
                     Both    75846.00    121.00      1.19      0.99      1.42
Breast               M             NA        NA        NA        NA        NA
                     F      104938.00    271.00      0.89      0.79      1.00
```

```
                       Both           NA       NA      NA      NA      NA
Cervix uteri           M              NA       NA      NA      NA      NA
                       F         6373.00    17.00    0.72    0.45    1.15
                       Both           NA       NA      NA      NA      NA
Corpus uteri           M              NA       NA      NA      NA      NA
                       F        15048.00    48.00    1.46    1.10    1.94
                       Both           NA       NA      NA      NA      NA
Ovary                  M              NA       NA      NA      NA      NA
                       F        10124.00    34.00    1.49    1.06    2.08
                       Both           NA       NA      NA      NA      NA
Prostate               M       130567.00   121.00    0.58    0.48    0.69
                       F              NA       NA      NA      NA      NA
                       Both           NA       NA      NA      NA      NA
Testis                 M         5847.00    23.00    0.87    0.58    1.31
                       F              NA       NA      NA      NA      NA
                       Both           NA       NA      NA      NA      NA
Kidney                 M        12665.00    36.00    1.16    0.83    1.60
                       F         6739.00    21.00    1.58    1.03    2.42
                       Both     19404.00    57.00    1.32    1.02    1.71
Bladder                M        13058.00    26.00    1.63    1.11    2.40
                       F         4039.00     7.00    1.89    0.90    3.98
                       Both     17097.00    33.00    1.79    1.27    2.51
Melanoma of skin       M        48755.00   122.00    0.91    0.76    1.09
                       F        35878.00    90.00    0.83    0.68    1.02
                       Both     84633.00   212.00    0.89    0.77    1.01
Brain, CNS             M         7157.00    20.00    1.00    0.65    1.55
                       F         5034.00    14.00    1.18    0.70    1.99
                       Both     12191.00    34.00    1.08    0.77    1.52
Thyroid                M         3308.00    19.00    1.56    1.00    2.46
                       F         9936.00    51.00    1.29    0.98    1.69
                       Both     13244.00    70.00    1.34    1.06    1.69
Non-Hodgkin lymphoma   M        17950.00    49.00    1.15    0.86    1.52
                       F        14023.00    31.00    1.16    0.81    1.65
                       Both     31973.00    80.00    1.16    0.93    1.44
Hodgkin lymphoma       M         2326.00    12.00    1.44    0.81    2.53
                       F         1918.00     2.00    0.30    0.08    1.22
                       Both      4244.00    14.00    0.94    0.56    1.59
Multiple myeloma       M         5740.00    13.00    1.29    0.75    2.23
                       F         4177.00     2.00    0.36    0.09    1.42
                       Both      9917.00    15.00    0.98    0.59    1.62
```

Of course we would also like to see the results as a forest plot, so we extract the relevant quantities for doing this:

```
> eM <- RRtab[,"M",3:5]
> eF <- RRtab[,"F",3:5]
> eB <- RRtab[,"Both",3:5]
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( eB, y=nrow(eM):1, txtpos=nrow(eM):1,
+          col="lightgray", xlog=TRUE,
+          xtic=c(1:10/10,1.5,2:7), xlim=c(0.095,7),
+          grid=TRUE, vref=1, xlab="Cancer incidence RR, T1D vs. population" )
> linesEst( eF, y=nrow(eM):1-0.2, col="red" )
> linesEst( eM, y=nrow(eM):1+0.2, col="blue" )
> text( rep(0.095,dim(RRtab)[1]), dim(RRtab)[1]:1+0.2, RRtab[,"M",2],
+       col="blue", adj=1, cex=0.7 )
> text( rep(0.095,dim(RRtab)[1]), dim(RRtab)[1]:1-0.2, RRtab[,"F",2],
+       col="red" , adj=1, cex=0.7 )
```

We see that the only sites with appreaciable increased RR and sufficiently narrow confidence intervals are colorectal, liver, corpus uteri, ovary and thyroid; whereas there seems to be a lower risk for cancer of the breast, cervix uteri and prostate cancer among T1D patients.

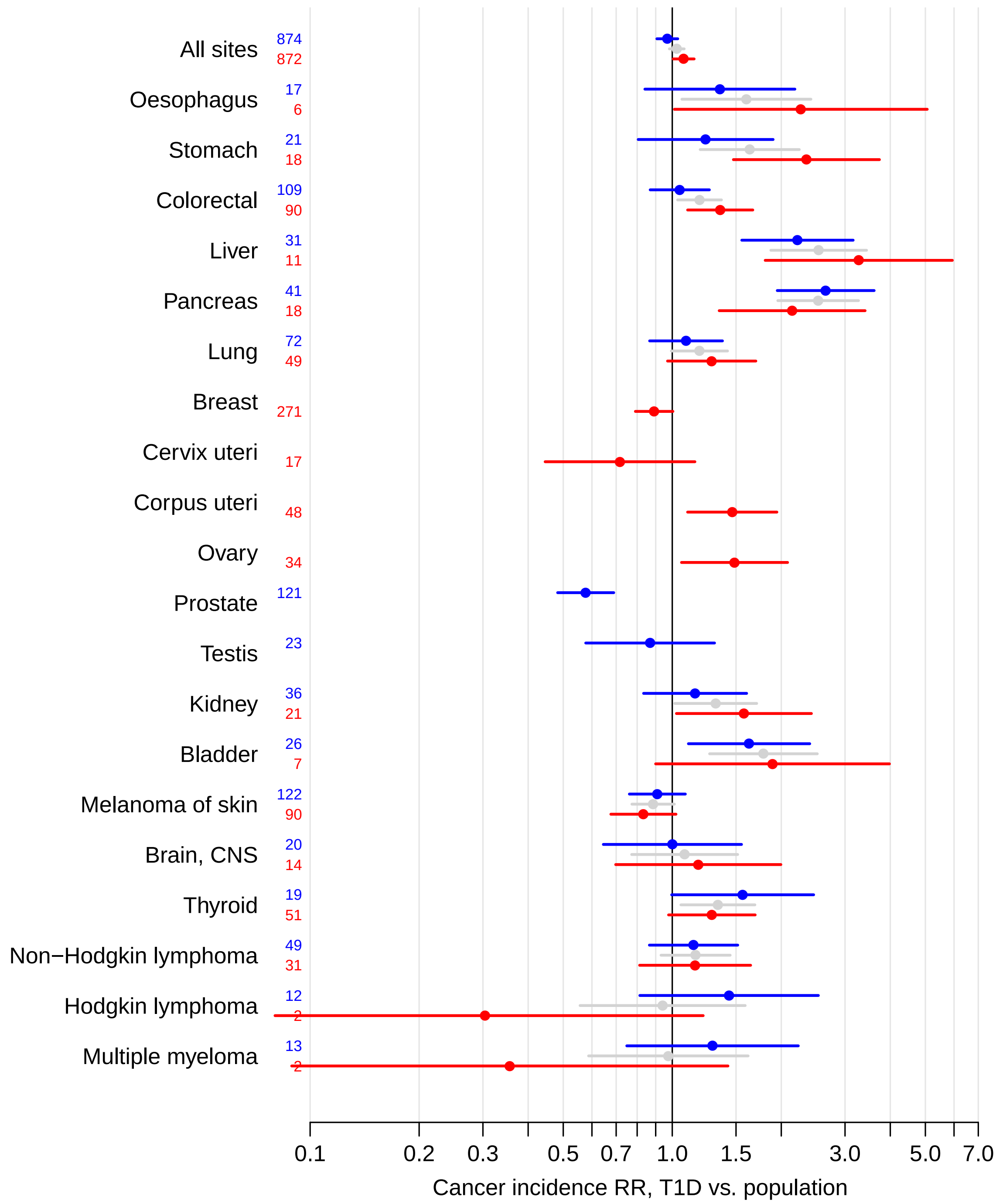

Figure 7.3: *RRs of cancer incidence among T1D patients (i.e. diagnosed* $<$ *40 years of age) in Australia relative to the general population. The numbers to the left are the number of cancers observed among the T1D patients. Men: Blue, Women: Red, Both sexes: Light gray.*

Chapter 8

Joint analyses of 5 countries' data

Here we do all analyses to determine overall estimates of the HR of different cancers associated with the presence of T1D and the duration of it.

8.1 Merging of data frames

8.1.1 Total data sets

We first load the data sets from each country and create an overview of the tumours available in each of the countries:

```
> load( "../data/DKana.Rda" )
> load( "../data/SEana.Rda" )
> load( "../data/FIana.Rda" )
> load( "../data/SCana.Rda" )
> load( "../data/AUana.Rda" )
```

8.1.2 Merging the data sets

Then we check which types of cancer that are common for all countries:

```
> load( "../data/conv.Rda" )
> str( conv )
'data.frame':        41 obs. of  3 variables:
 $ DKnam: chr  "d0" "d11" "d12" "d14" ...
 $ NCnam: chr  "d0" "d1" "d2" "d3" ...
 $ Clab : chr  "All sites" "Lip" "Tongue" "Salivary glands" ...
> dk.wh <- match( names(dk.ana), conv$NCnam ) ; (dk.wh <- dk.wh[!is.na(dk.wh)])
 [1]  1  7  8 10 11 12 14 17 19 20 21 22 24 25 27 28 29 31 32 35 36 37 38 41
> se.wh <- match( names(se.ana), conv$NCnam ) ; (se.wh <- se.wh[!is.na(se.wh)])
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
[33] 33 35 36 37 38 40 41
> fi.wh <- match( names(fi.ana), conv$NCnam ) ; (fi.wh <- fi.wh[!is.na(fi.wh)])
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
[33] 33 34 35 36 37 38 39 41
> sc.wh <- match( names(sc.ana), conv$NCnam ) ; (sc.wh <- sc.wh[!is.na(sc.wh)])
 [1]  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
[33] 33 34 35 37 38 39 40 41
> au.wh <- match( names(au.ana), conv$NCnam ) ; (au.wh <- au.wh[!is.na(au.wh)])
 [1] 37 36 35 32 31 29 28 27 25 24 22 21 20 19 17 14 12 41  8  7  1
```

```
> dk.ca <-
+ se.ca <-
+ fi.ca <-
+ sc.ca <-
+ au.ca <- rep( "  ", nrow(conv) )
> dk.ca[dk.wh] <- "DK"
> se.ca[se.wh] <- "SE"
> fi.ca[fi.wh] <- "FI"
> sc.ca[sc.wh] <- "SC"
> au.ca[au.wh] <- "AU"
> which.ctrib <- paste( dk.ca, se.ca, fi.ca, sc.ca, au.ca )
> cbind( conv[,3:2], which.ctrib )
```

```
                          Clab NCnam    which.ctrib
1                    All sites    d0 DK SE FI SC AU
2                          Lip    d1    SE FI SC
3                       Tongue    d2    SE FI SC
4              Salivary glands    d3    SE FI SC
5                        Mouth    d4    SE FI SC
6                      Pharynx    d5    SE FI SC
7                   Oesophagus    d6 DK SE FI SC AU
8                      Stomach    d7 DK SE FI SC AU
9              Small intestine    d8    SE FI SC
10                       Colon    d9 DK SE FI SC
11                      Rectum   d10 DK SE FI SC
12                       Liver   d11 DK SE FI SC AU
13                 Gallbladder   d12    SE FI SC
14                    Pancreas   d13 DK SE FI SC AU
15               Nose, sinuses   d14    SE FI SC
16                      Larynx   d15    SE FI SC
17                        Lung   d16 DK SE FI SC AU
18                      Pleura   d17    SE FI SC
19                      Breast   d18 DK SE FI SC AU
20                Cervix uteri   d19 DK SE FI SC AU
21                Corpus uteri   d20 DK SE FI SC AU
22                       Ovary   d22 DK SE FI SC AU
23   Other female genital organs d23    SE FI SC
24                    Prostate   d24 DK SE FI SC AU
25                      Testis   d25 DK SE FI SC AU
26                  Penis etc.   d26    SE FI SC
27                      Kidney   d27 DK SE FI SC AU
28                     Bladder   d28 DK SE FI SC AU
29            Melanoma of skin   d29 DK SE FI SC AU
30                         Eye   d31    SE FI SC
31                 Brain, CNS    d32 DK SE FI SC AU
32                     Thyroid   d33 DK SE FI SC AU
33                        Bone   d34    SE FI SC
34                Soft tissues   d35       FI SC
35          Non-Hodgkin lymphoma d36 DK SE FI SC AU
36            Hodgkin lymphoma   d37 DK SE FI    AU
37            Multiple myeloma   d38 DK SE FI SC AU
38                   Leukaemia   d40 DK SE FI SC
39 Other and unspecified cancers d48       FI SC
40                   Oral etc.   d51    SE    SC
41                  Colorectal   d52 DK SE FI SC AU
```

We see that the Danish and Australian data contain fewer sites than the Swedish, Finnish and Scottish. The sites available in the Australian data are all the major sites of primary relevance, except for leukemia and the subdivision of colorectal in colon and rectum.

We shall therefore restrict the analysis to the sites defined in the Danish data, meaning that colon, rectum and leukemia will be based on 4 countries (not Australia) as will Hodgkin's lymphoma (not Scotland).

```
> ( vars <- names(dk.ana) )
```

```
 [1] "sex"    "A"      "P"      "T1D"    "DMprev" "DMdur"  "y0"     "y"      "d0"     "d6"
[11] "d7"     "d9"     "d10"    "d11"    "d13"    "d16"    "d18"    "d19"    "d20"    "d22"
[21] "d24"    "d25"    "d27"    "d28"    "d29"    "d32"    "d33"    "d36"    "d37"    "d38"
[31] "d40"    "d52"    "Cnt"
> wh.ca <- vars[substr(vars,1,1)=="d"]
> ( wh.ca <- wh.ca[c(1:5,24,6:8,17,9:16,18:23)] )
 [1] "d0"  "d6"  "d7"  "d9"  "d10" "d52" "d11" "d13" "d16" "d29" "d18" "d19" "d20" "d22" "d24" "d25"
[17] "d27" "d28" "d32" "d33" "d36" "d37" "d38" "d40"
> cbind( wh.ca, conv[match(wh.ca,conv$NCnam),c("NCnam","Clab")] )
   wh.ca NCnam                 Clab
1     d0    d0            All sites
7     d6    d6           Oesophagus
8     d7    d7              Stomach
10    d9    d9                Colon
11   d10   d10               Rectum
41   d52   d52           Colorectal
12   d11   d11                Liver
14   d13   d13             Pancreas
17   d16   d16                 Lung
29   d29   d29     Melanoma of skin
19   d18   d18               Breast
20   d19   d19        Cervix uteri
21   d20   d20        Corpus uteri
22   d22   d22                Ovary
24   d24   d24             Prostate
25   d25   d25               Testis
27   d27   d27               Kidney
28   d28   d28              Bladder
31   d32   d32           Brain, CNS
32   d33   d33              Thyroid
35   d36   d36 Non-Hodgkin lymphoma
36   d37   d37     Hodgkin lymphoma
37   d38   d38     Multiple myeloma
38   d40   d40            Leukaemia
```

8.2 Analysis data frame

The analysis data frame is simply constructed by stacking the analysis data sets from different countries. We must however insert the missing columns from Australia and Scotland first. We restrict the analysis dataset to ages under 85 (which is the age-range present in all countries) and units with positive followup time (y> 0) and also order the levels of `DMdur` in a more sensible way, than that rendered by the `rbind`ing of the analysis data frames:

```
> au.ana$d9  <-
+ au.ana$d10 <-
+ au.ana$d40 <- NA
> sc.ana$d37 <- NA
> all.ana <- rbind( dk.ana[,vars],
+                   fi.ana[,vars],
+                   se.ana[,vars],
+                   sc.ana[,vars],
+                   au.ana[,vars] )
> all.ana <- transform( subset( all.ana, A<85 & y0>0 ),
+                       Cnt = factor( Cnt,
+                                     levels=c("DNK","FIN","SWE","SCO","AUS"),
+                                     labels=c("DK" ,"FI" ,"SE" ,"SC" ,"AU" ) ),
+                        T1 = Relevel( T1D,
+                                      list("NoDM","T1DM"=2:nlevels(T1D) ) ),
+                       T1D = Relevel( T1D, c(1:4,6,5) ),
```

```
+                     DMdur = Relevel( DMdur, c(1:7,9,8) ) )
> with( all.ana, cbind( tapply(  y, y0>0, sum ),
+                       tapply( d0, y0>0, sum ) ) )
          [,1]    [,2]
TRUE 778028445 3350432
> dim( all.ana )
[1] 99687    34
> apply( all.ana, 2, function(x) sum(is.na(x)) )
   sex      A      P    T1D DMprev  DMdur     y0      y     d0     d6     d7     d9    d10    d11
     0      0      0      0      0      0      0      0      0      0      0  14282  14282      0
   d13    d16    d18    d19    d20    d22    d24    d25    d27    d28    d29    d32    d33    d36
     0      0      0      0      0      0      0      0      0      0      0      0      0      0
   d37    d38    d40    d52    Cnt     T1
 14706      0  14282      0      0      0
```

We can now make an overview of the number of cancers by T1D status:

```
> with( all.ana, ftable( T1, T1D, DMdur ) )
          DMdur NoDM    0    1    2    5   10   15   30 Unkn
T1   T1D
NoDM NoDM       8466    0    0    0    0    0    0    0    0
     30            0    0    0    0    0    0    0    0    0
     35            0    0    0    0    0    0    0    0    0
     40            0    0    0    0    0    0    0    0    0
     45            0    0    0    0    0    0    0    0    0
     Inf           0    0    0    0    0    0    0    0    0
T1DM NoDM          0    0    0    0    0    0    0    0    0
     30            0 6856 6815 7540 8087 7375 8570 5264 4923
     35            0 1407 1432 1939 2351 2112 3208 2116 1605
     40            0 1408 1425 1941 2369 2115 3207 1772 1556
     45            0  109  109  166  229  203  357  247  470
     Inf           0  240  255  346  320  146   49    0  582
> with( all.ana, ftable( Cnt, DMprev, T1D ) )
           T1D  NoDM    30    35    40    45   Inf
Cnt DMprev
DK  Pop         3060     0     0     0     0     0
    Inc            0  4840  1097  1092     0     0
    Prv            0  3557  1089  1081     0     0
FI  Pop         1360     0     0     0     0     0
    Inc            0 14635  4094  4121     0     0
    Prv            0  4103  1122  1103     0     0
SE  Pop          850     0     0     0     0     0
    Inc            0 14738  4513  4244     0     0
    Prv            0     0     0     0     0     0
SC  Pop         2890     0     0     0     0     0
    Inc            0  4450  1100  1100     0     0
    Prv            0  3154  1004  1008     0     0
AU  Pop          306     0     0     0     0     0
    Inc            0  5953  2151  2044  1890  1938
    Prv            0     0     0     0     0     0
> S.all <- function(x) sum(x[!is.na(x)])
> S.m1  <- function(x) sum(x[-1][!is.na(x[-1])])
> ftable( addmargins( with( all.ana,
+                           tapply( d0,
+                                   list(Cnt,T1D),
+                                   sum ) ),
+                     FUN=list(S.all,S.m1), q=T ) )
          NoDM      30      35      40      45     Inf    S.m1

DK      442530     276     255     511      NA      NA    1042
FI      707026     865     612     931      NA      NA    2408
SE      959456    1878     842    1342      NA      NA    4062
SC      404707     187     124     222      NA      NA     533
AU      826922     393     280     431     599      43    1746
S.all  3340641    3599    2113    3437     599      43    9791
```

```
> ftable( addmargins( with( all.ana,
+                           tapply( d0,
+                                   list(Cnt,DMprev,T1D),
+                                   sum ) ),
+                     FUN=list(S.all,S.m1,S.m1), q=T ) )
               NoDM      30      35      40      45     Inf    S.m1

DK    Pop    442530      NA      NA      NA      NA      NA       0
      Inc        NA     113     124     255      NA      NA     492
      Prv        NA     163     131     256      NA      NA     550
      S.m1        0     276     255     511       0       0    1042
FI    Pop    707026      NA      NA      NA      NA      NA       0
      Inc        NA     679     553     833      NA      NA    2065
      Prv        NA     186      59      98      NA      NA     343
      S.m1        0     865     612     931       0       0    2408
SE    Pop    959456      NA      NA      NA      NA      NA       0
      Inc        NA    1878     842    1342      NA      NA    4062
      Prv        NA      NA      NA      NA      NA      NA       0
      S.m1        0    1878     842    1342       0       0    4062
SC    Pop    404707      NA      NA      NA      NA      NA       0
      Inc        NA      44      28      82      NA      NA     154
      Prv        NA     143      96     140      NA      NA     379
      S.m1        0     187     124     222       0       0     533
AU    Pop    826922      NA      NA      NA      NA      NA       0
      Inc        NA     393     280     431     599      43    1746
      Prv        NA      NA      NA      NA      NA      NA       0
      S.m1        0     393     280     431     599      43    1746
S.all Pop   3340641       0       0       0       0       0       0
      Inc         0    3107    1827    2943     599      43    8519
      Prv         0     492     286     494       0       0    1272
      S.m1        0    3599    2113    3437     599      43    9791
```

With these tables we restrict the analysis to be for persons diagnosed with DM before the age of 40:

```
> all.ana <- transform( subset( all.ana, T1D %in% levels(T1D)[1:4] ),
+                       T1D = factor( T1D ) )
> ftable( addmargins( with( all.ana,
+                           tapply( d0,
+                                   list(Cnt,T1D),
+                                   sum ) ),
+                     FUN=list(S.all,S.m1), q=T ) )
          NoDM      30      35      40    S.m1

DK      442530     276     255     511    1042
FI      707026     865     612     931    2408
SE      959456    1878     842    1342    4062
SC      404707     187     124     222     533
AU      826922     393     280     431    1104
S.all  3340641    3599    2113    3437    9149
> round(
+ ftable( addmargins( with( all.ana,
+                           tapply( y/1000,
+                                   list(Cnt,T1D),
+                                   sum ) ),
+                     FUN=list(S.all,S.m1), q=T ) ), 1 )
           NoDM        30        35        40      S.m1

DK      95492.4     265.8     125.9     161.1     552.8
FI     198320.5     747.6     211.4     242.9    1201.9
SE     217267.1     900.8     210.3     280.0    1391.1
SC      85386.6     227.6      44.6      54.1     326.3
AU     177493.4     344.5      84.2      82.2     510.9
S.all  773960.0    2486.3     676.4     820.4    3983.0
```

Thus the final analysis dataset has 9,149 cancers in 4.0 mil. person-years among T1D patients defined as persons diagnosed under the age of 40.

8.2.0.1 Non-sex-specific cancers

Besides the category of "all cancers" in the variable `d0`, we also want a category of all those cancers that are common for both sexes; that is all cancers *except* cancer of breast (`d18`), cervix (`d19`), endometrium (`d20`), ovary (`d22`), prostate (`d24`) and testis (`d25`). We construct this variable as `d00` and add a name to it in the `conv` data frame and the vector `wh.ca` with the names of variables to be analyzed:

```
> conv <- rbind( conv[1,], c("","d00","Non-sex-specific"), conv[-1,] )
> wh.ca <- c(wh.ca[1],"d00",wh.ca[-1])
> all.ana$d00 <- with( all.ana, pmax(d0-d18-d19-d20-d22-d24-d25,0,na.rm=TRUE) )
> save( all.ana, wh.ca, conv, file = "../data/ALLana.Rda" )
> # load( file = "../data/ALLana.Rda" )
```

8.2.1 Overview of the analysis data frame

We provide an overview of the dataset; the number of units in the data frame, the number of cancers and the amount of follow-up time (person-years):

```
> for( cn in levels(all.ana$Cnt) )
+    {
+    cat( "\n", rep("-",70), "\n", cn, "\n", sep="" )
+    print( round(
+    ftable( addmargins(
+            xtabs( cbind(D=d0,Y=y/1000) ~ floor(P) + T1D,
+                   data=subset( all.ana, Cnt==cn ) ),
+                       margin=1 ),
+            col.vars=3:2 ) ) )
+    }
----------------------------------------------------------------------
DK
                   D                           Y
         T1D    NoDM    30    35    40  NoDM    30    35    40
floor(P)
1995           20980     5     4     6  5144     8     3     4
1996           19233     6     4     9  5172     8     4     5
1997           21484     0     4    12  5192     9     4     5
1998           23752     7     4    14  5209    10     5     6
1999           22490    10     8    16  5225    11     5     6
2000           20531     7     7     7  5241    12     6     7
2001           22708    10    11    22  5260    12     6     7
2002           25185    12     7    27  5277    13     6     8
2003           21605    16     4    21  5291    14     7     9
2004           24628    11    21    24  5304    15     7     9
2005           25198    13     8    26  5317    16     8    10
2006           26202    19    17    35  5332    17     8    10
2007           26515    18    17    34  5355    18     8    11
2008           27705    17    23    45  5386    19     9    12
2009           29491    30    26    57  5414    20     9    12
2010           28147    25    25    47  5437    21    10    13
2011           28622    29    27    48  5458    22    10    14
2012           28054    41    38    61  5478    22    11    14
Sum           442530   276   255   511 95492   266   126   161

----------------------------------------------------------------------
FI
                   D                           Y
         T1D    NoDM    30    35    40  NoDM    30    35    40
floor(P)
1972           11487     0     2     2  4627     3     1     1
1973           11559     0     1     1  4651     4     1     1
1974           11773     1     1     1  4674     5     1     1
```

```
1975          12060      1      1      2   4693      6      1      2
1976          12587      0      0      0   4705      7      1      2
1977          12903      0      3      1   4717      8      2      2
1978          13158      6      1      4   4729      8      2      2
1979          13500      3      1      5   4740      9      2      2
1980          13892      3      1      5   4753     10      2      3
1981          13961      8      8      4   4771     11      3      3
1982          14320      6      4     13   4796     11      3      3
1983          14546      7      5      7   4822     12      3      3
1984          14648      5      2      8   4845     13      3      4
1985          14978      5      3      7   4863     13      4      4
1986          15404      8      6     10   4877     14      4      4
1987          15707     11      6      5   4888     15      4      4
1988          15718      9      6     15   4901     15      4      5
1989          15658     19     10     18   4916     16      4      5
1990          16034     11      8     14   4936     17      4      5
1991          16489      4      5     11   4960     17      5      5
1992          17184     14     16     27   4985     18      5      5
1993          17196     15     11     18   5007     18      5      5
1994          17886     15     11     24   5026     19      5      6
1995          18267     17     20     25   5042     20      5      6
1996          19296     14     16     26   5056     21      6      7
1997          19312     27     21     37   5069     21      6      7
1998          19747     27     13     31   5080     22      6      7
1999          19959     21     21     31   5089     23      7      8
2000          20389     34     20     32   5098     24      7      8
2001          20715     32     21     34   5109     25      7      9
2002          21495     30     21     37   5120     26      8      9
2003          22003     38     22     37   5131     28      8      9
2004          23410     38     25     47   5145     29      9     10
2005          23905     48     45     49   5160     30      9     11
2006          23697     47     36     57   5175     32     10     11
2007          23140     52     36     54   5193     33     10     12
2008          23689     63     46     51   5212     35     11     13
2009          24873     66     46     72   5233     36     12     13
2010          24879     77     47     53   5252     37     12     14
2011          25602     83     44     56   5272     33     11     13
2012              0      0      0      0      0      0      0      0
Sum          707026    865    612    931 198320    748    211    243

-----------------------------------------------------------------------
SE
                  D                            Y
         T1D   NoDM     30     35     40   NoDM     30     35     40
floor(P)
1987          34084     22     14     13   8267     25      6      7
1988          34000     22     13     27   8301     26      6      7
1989          34023     28     18     17   8352     27      6      7
1990          34493     30      8     13   8412     28      6      7
1991          34756     30     21     26   8465     28      6      8
1992          34844     32     14     20   8510     29      6      8
1993          35392     38     17     36   8554     30      7      8
1994          35835     45     23     26   8610     31      7      9
1995          35354     50     16     42   8649     32      7      9
1996          35577     42     32     34   8657     33      7      9
1997          35743     71     21     38   8656     34      8     10
1998          36503     65     26     46   8656     34      8     10
1999          37827     61     33     44   8658     35      8     11
2000          37960     84     38     50   8669     36      8     11
2001          38615     94     32     55   8690     37      9     12
2002          39395     79     27     71   8716     38      9     12
2003          40918     89     47     69   8747     40      9     13
2004          42262    117     47     79   8779     41     10     13
2005          42019    109     51     70   8807     42     10     14
2006          41662    117     53     78   8848     43     10     14
2007          41710    117     39     75   8910     44     11     15
```

```
2008           42289    117     52     92   8977     45     11     16
2009           44700    142     62    119   9052     46     11     16
2010           44640    137     57    109   9128     48     12     17
2011           44855    140     81     93   9197     49     12     17
Sum           959456   1878    842   1342 217267    901    210    280

----------------------------------------------------------------------
SC
                  D                           Y
         T1D   NoDM     30     35     40   NoDM     30     35     40
floor(P)
1995           22048      7      3      4   5024      9      2      2
1996           23468      6      3      6   5011     10      2      2
1997           22772      5      2      9   5001     10      2      2
1998           22509      7      2      6   4992     11      2      2
1999           22444      9      6     11   4986     11      2      2
2000           22829      9      4      8   4975     12      2      3
2001           22863      4      2     12   4975     12      2      3
2002           23572     14      4     13   4967     13      3      3
2003           23578     16     11     10   4971     13      3      3
2004           24104      8     12     15   4993     14      3      3
2005           23743     14      7      7   5004     14      3      4
2006           24169      5     13     19   5022     15      3      4
2007           24693     18     10     16   5046     15      3      4
2008           25248     12     13     15   5068     16      3      4
2009           25523     17     10     19   5090     17      3      4
2010           25422     24      8     26   5115     17      3      4
2011           25722     12     14     26   5145     18      4      5
Sum           404707    187    124    222  85387    228     45     54

----------------------------------------------------------------------
AU
                  D                           Y
         T1D   NoDM     30     35     40   NoDM     30     35     40
floor(P)
2000           80927     26     33     28  18778     31      9      9
2001           83093     31     23     48  19011     33      9      9
2002           86480     49     22     37  19223     35      9      9
2003           88023     39     29     39  19441     36      9      9
2004           92575     39     40     46  19646     38      9      9
2005           94913     59     30     52  19875     40     10      9
2006           97165     47     31     55  20132     42     10      9
2007          100209     45     31     58  20490     44     10      9
2008          103537     58     41     68  20897     46     10      9
Sum           826922    393    280    431 177493    344     84     82
> round( ftable( xtabs( cbind(D=d0,Y=y/1000) ~ Cnt + DMdur + T1D,
+                       data=all.ana ),
+                col.vars=4:3 ) )
                   D                           Y
          T1D   NoDM     30     35     40   NoDM     30     35     40
Cnt DMdur
DK  NoDM      442530      0      0      0  95492      0      0      0
    0              0      9      9     27      0     16      8     11
    1              0      4      7     15      0     15      8     11
    2              0     24     21     54      0     38     20     27
    5              0     39     37     89      0     45     23     31
    10             0     29     45     57      0     22     12     14
    15             0      8      5     13      0      3      2      2
    30             0      0      0      0      0      0      0      0
    Unkn           0    163    131    256      0    127     53     64
FI  NoDM      707026      0      0      0 198320      0      0      0
    0              0     25     14     38      0     41     15     18
    1              0     10     16     22      0     40     14     17
    2              0     30     35     58      0    110     37     45
    5              0     62     59    131      0    153     47     57
    10             0     64     92    161      0    119     34     40
```

```
    15             0    402    293    405      0    225     55     57
    30             0    272    103    116      0     60     10      8
    Unkn           0      0      0      0      0      0      0      0
SE  NoDM      959456      0      0      0 217267      0      0      0
    0              0     27     19     40      0     31      9     15
    1              0      9      9     16      0     29      9     14
    2              0     32     27     59      0     84     25     40
    5              0     73     67    147      0    131     40     60
    10             0     93     95    169      0    120     35     48
    15             0    490    386    622      0    298     72     86
    30             0   1154    239    289      0    207     20     18
    Unkn           0      0      0      0      0      0      0      0
SC  NoDM      404707      0      0      0  85387      0      0      0
    0              0      7      7     11      0      5      1      2
    1              0      6      4      7      0     10      2      4
    2              0      8      4     20      0     28      7     10
    5              0     16     13     34      0     45     10     14
    10             0     37     24     45      0     48     10     11
    15             0    113     72    105      0     91     14     14
    30             0      0      0      0      0      0      0      0
    Unkn           0      0      0      0      0      0      0      0
AU  NoDM      826922      0      0      0 177493      0      0      0
    0              0     17      5      1      0     14      1      1
    1              0      3      1      5      0     15      2      1
    2              0     19      6     12      0     42      6      5
    5              0     19     18     22      0     39      8      7
    10             0      3      6     18      0      8      2      2
    15             0     21     22     34      0     11      2      3
    30             0     35     14     16      0      5      1      1
    Unkn           0    276    208    323      0    210     62     62
> tt <- addmargins( xtabs( cbind( N=(y>0), Ca=d0, PY=y/1000 ) ~ Cnt + T1D,
+                          data = all.ana ),
+                   margin=2 )
> dd <- addmargins( xtabs( cbind( N=(y>0), Ca=d0, PY=y/1000 ) ~ Cnt + T1D,
+                          data = subset(all.ana,T1D %in% c("30","35","40")) ),
+                   margin=2 )
> tt[,"Sum",] <- dd[,"Sum",]
> round( ftable( addmargins( tt, margin=1 ), row.vars=c(3,1) ) )
       T1D     NoDM      30      35      40     Sum
   Cnt
N  DK          3060    8397    2186    2173   12756
   FI          1360   18738    5216    5224   29178
   SE           850   14738    4513    4244   23495
   SC          2890    7604    2104    2108   11816
   AU           306    5953    2151    2044   10148
   Sum         8466   55430   16170   15793   87393
Ca DK        442530     276     255     511    1042
   FI        707026     865     612     931    2408
   SE        959456    1878     842    1342    4062
   SC        404707     187     124     222     533
   AU        826922     393     280     431    1104
   Sum      3340641    3599    2113    3437    9149
PY DK         95492     266     126     161     553
   FI        198320     748     211     243    1202
   SE        217267     901     210     280    1391
   SC         85387     228      45      54     326
   AU        177493     344      84      82     511
   Sum       773960    2486     676     820    3983
> # Percent cancers and PYs resp:
> round( 100*tt[,"Sum",-1]/tt[,"NoDM",-1], 2 )
Cnt     Ca   PY
  DK 0.24 0.58
  FI 0.34 0.61
  SE 0.42 0.64
  SC 0.13 0.38
  AU 0.13 0.29
```

```
> round( ftable( xtabs( cbind( N=(y>0), Ca=d0, PY=y/1000 ) ~ Cnt + T1D + DMdur,
+                       data = all.ana ),
+                row.vars=c(4,1,3) ) )

             T1D   NoDM     30     35     40
   Cnt DMdur
N  DK  NoDM        3060      0      0      0
       0              0   1114    216    216
       1              0   1054    205    204
       2              0   1050    251    250
       5              0    890    241    240
       10             0    540    141    140
       15             0    192     43     42
       30             0      0      0      0
       Unkn           0   3557   1089   1081
   FI  NoDM        1360      0      0      0
       0              0   2530    547    552
       1              0   2532    546    549
       2              0   2812    731    732
       5              0   3067    890    892
       10             0   2712    782    781
       15             0   3364   1200   1206
       30             0   1721    520    512
       Unkn           0      0      0      0
   SE  NoDM         850      0      0      0
       0              0   1609    312    314
       1              0   1609    312    315
       2              0   1833    436    438
       5              0   2057    560    561
       10             0   2056    560    561
       15             0   2866   1120   1119
       30             0   2708   1213    936
       Unkn           0      0      0      0
   SC  NoDM        2890      0      0      0
       0              0   1043    204    207
       1              0   1062    238    238
       2              0   1251    330    331
       5              0   1444    422    422
       10             0   1441    423    423
       15             0   1363    487    487
       30             0      0      0      0
       Unkn           0      0      0      0
   AU  NoDM         306      0      0      0
       0              0    560    128    119
       1              0    558    131    119
       2              0    594    191    190
       5              0    629    238    254
       10             0    626    206    210
       15             0    785    358    353
       30             0    835    383    324
       Unkn           0   1366    516    475
Ca DK  NoDM      442530      0      0      0
       0              0      9      9     27
       1              0      4      7     15
       2              0     24     21     54
       5              0     39     37     89
       10             0     29     45     57
       15             0      8      5     13
       30             0      0      0      0
       Unkn           0    163    131    256
   FI  NoDM      707026      0      0      0
       0              0     25     14     38
       1              0     10     16     22
       2              0     30     35     58
       5              0     62     59    131
       10             0     64     92    161
       15             0    402    293    405
```

		30	0	272	103	116
		Unkn	0	0	0	0
	SE	NoDM	959456	0	0	0
		0	0	27	19	40
		1	0	9	9	16
		2	0	32	27	59
		5	0	73	67	147
		10	0	93	95	169
		15	0	490	386	622
		30	0	1154	239	289
		Unkn	0	0	0	0
	SC	NoDM	404707	0	0	0
		0	0	7	7	11
		1	0	6	4	7
		2	0	8	4	20
		5	0	16	13	34
		10	0	37	24	45
		15	0	113	72	105
		30	0	0	0	0
		Unkn	0	0	0	0
	AU	NoDM	826922	0	0	0
		0	0	17	5	1
		1	0	3	1	5
		2	0	19	6	12
		5	0	19	18	22
		10	0	3	6	18
		15	0	21	22	34
		30	0	35	14	16
		Unkn	0	276	208	323
PY	DK	NoDM	95492	0	0	0
		0	0	16	8	11
		1	0	15	8	11
		2	0	38	20	27
		5	0	45	23	31
		10	0	22	12	14
		15	0	3	2	2
		30	0	0	0	0
		Unkn	0	127	53	64
	FI	NoDM	198320	0	0	0
		0	0	41	15	18
		1	0	40	14	17
		2	0	110	37	45
		5	0	153	47	57
		10	0	119	34	40
		15	0	225	55	57
		30	0	60	10	8
		Unkn	0	0	0	0
	SE	NoDM	217267	0	0	0
		0	0	31	9	15
		1	0	29	9	14
		2	0	84	25	40
		5	0	131	40	60
		10	0	120	35	48
		15	0	298	72	86
		30	0	207	20	18
		Unkn	0	0	0	0
	SC	NoDM	85387	0	0	0
		0	0	5	1	2
		1	0	10	2	4
		2	0	28	7	10
		5	0	45	10	14
		10	0	48	10	11
		15	0	91	14	14
		30	0	0	0	0
		Unkn	0	0	0	0
	AU	NoDM	177493	0	0	0

```
 0             0    14    1    1
 1             0    15    2    1
 2             0    42    6    5
 5             0    39    8    7
 10            0     8    2    2
 15            0    11    2    3
 30            0     5    1    1
 Unkn          0   210   62   62
```

8.2.2 Histograms of follow-up (events & PY)

In order to provide a graphical overview of how follow-up (cancer events and person-years) is distributed by age, date, DM duration and age at DM we provide histograms of these separately for each country and for the total FU:

```
> # Age at FU
> Af <- addmargins( xtabs( cbind( d0, y/1000 ) ~ floor(A) + Cnt,
+                          data=subset(all.ana,T1=="T1DM") ), 2 )
> str( Af )
 table [1:85, 1:6, 1:2] 0 0 0 0 0 0 1 1 0 0 ...
 - attr(*, "dimnames")=List of 3
  ..$ floor(A): chr [1:85] "0" "1" "2" "3" ...
  ..$ Cnt     : chr [1:6] "DK" "FI" "SE" "SC" ...
  ..$         : chr [1:2] "d0" "V2"
 - attr(*, "class")= chr [1:2] "table" "array"
> # Date of FU
> Pf <- addmargins( xtabs( cbind( d0, y/1000 ) ~ floor(P) + Cnt,
+                          data=subset(all.ana,T1=="T1DM") ), 2 )
> str( Pf )
 table [1:41, 1:6, 1:2] 0 0 0 0 0 0 0 0 0 0 ...
 - attr(*, "dimnames")=List of 3
  ..$ floor(P): chr [1:41] "1972" "1973" "1974" "1975" ...
  ..$ Cnt     : chr [1:6] "DK" "FI" "SE" "SC" ...
  ..$         : chr [1:2] "d0" "V2"
 - attr(*, "class")= chr [1:2] "table" "array"
> # Age at FU for duration-known
> af <- addmargins( xtabs( cbind( d0, y/1000 ) ~ floor(A) + Cnt,
+                          data=subset(all.ana,T1=="T1DM" & DMdur!="Unkn") ), 2 )
> # Date of FU for duration-known
> pf <- addmargins( xtabs( cbind( d0, y/1000 ) ~ floor(P) + Cnt,
+                          data=subset(all.ana,T1=="T1DM" & DMdur!="Unkn") ), 2 )
> # Duration at follow-up
> df <- addmargins( xtabs( cbind( d0, y/1000 ) ~ DMdur + Cnt,
+                          data=subset(all.ana,T1=="T1DM" & DMdur!="Unkn") ), 2 )
> # Color conventions:
> cclr <- c("red","mediumblue","darkorange","forestgreen","black")
> xclr <- c( cclr, gray(0.5) )
> # Names
> lcnt <- c("Denmark", "Finland", "Sweden", "Scotland", "Australia")
> xcnt <- c(lcnt,"All")
> what <- c("cases","PY")
> # Layout-details
> wd <- c(85,40,85,40,45)
> MM <- matrix( 1:60, 12, 5, byrow=TRUE )
> MM <- rbind( MM[1:6,],61:65,MM[7:12,])
> par( mar=c(1,3,0,0), oma=c(2,0,0,0), las=1, bty="n", mgp=c(3,1,0)/1.6 )
> layout( MM, widths=wd+3, heights=rep(c(5,2,5),c(6,1,6)) )
> for( it in 1:2 ) # it <- 1
+ for( ic in 1:6 ) # ic <- 1
+    {
+ aa <- barplot( Af[,ic,it], space=0, xaxt="n", xaxs="i",
+    #            ylim=c(0,130+170*(ic==6)),
+                col=xclr[ic], border="transparent" )
```

```
+ if( ic==6 ){
+ axis( side=1, at=seq(0,80,20) )
+ axis( side=1, at=seq(0,80,10), labels=NA )
+ axis( side=1, at=seq(0,85, 5), labels=NA, tcl=-0.3 )
+            }
+ text( par("usr")[1], par("usr")[4], paste(xcnt[ic],", ",what[it], sep=""),
+       adj=c(-0.1,1.1), col=xclr[ic] )
+
+ pp <- barplot( Pf[,ic,it], space=0, xaxt="n", xaxs="i",
+    #            ylim=c(0,700+170*(ic==6)),
+                col=xclr[ic], border="transparent" )
+ if( ic==6 ){
+ axis( side=1, at=seq(1980,2010,10)-1972, labels=seq(1980,2010,10))
+ axis( side=1, at=seq(1980,2015, 5)-1972, labels=NA, tcl=-0.3 )
+            }
+ aa <- barplot( af[,ic,it], space=0, xaxt="n", xaxs="i",
+    #            ylim=c(0,130+170*(ic==6)),
+                col=xclr[ic], border="transparent" )
+ if( ic==6 ){
+ axis( side=1, at=seq(0,80,20) )
+ axis( side=1, at=seq(0,80,10), labels=NA )
+ axis( side=1, at=seq(0,85, 5), labels=NA, tcl=-0.3 )
+            }
+ pp <- barplot( pf[,ic,it], space=0, xaxt="n", xaxs="i",
+    #            ylim=c(0,700+170*(ic==6)),
+                col=xclr[ic], border="transparent" )
+ if( ic==6 ){
+ axis( side=1, at=seq(1980,2010,10)-1972, labels=seq(1980,2010,10))
+ axis( side=1, at=seq(1980,2015, 5)-1972, labels=NA, tcl=-0.3 )
+            }
+ wi = c(1,1,3,5,5,15,15)
+ dd <- barplot( df[2:8,ic,it]/wi, width=wi, space=0, xaxt="n", xaxs="i",
+    #            ylim=c(0,300+170*(ic==6)),
+                col=xclr[ic], border="transparent" )
+ if( ic==6 ){
+ axis( side=1, at=seq(0,40,10) )
+ axis( side=1, at=seq(0,45,5), labels=NA, tcl=-0.3 )
+            }
+ }
> mtext( c("Age (all)",
+          "Date (all)",
+          "Age (w/duration)",
+          "Date (w/duration)",
+          "DM duration"),
+        side=1, line=1,
+        at=(cumsum(c(0,wd))[-6]+wd/2)/sum(wd), outer=TRUE )
```

8.2.3 Site distribution

We provide an overview of the number of cancers by country, site and sex, as well as the relative distribution of these (but only among the sites present in all countries) :

```
> mm <- as.matrix( all.ana[,wh.ca] )
> mm[is.na(mm)] <- 0
> tt <- xtabs( mm ~ sex + Cnt + T1, data=all.ana )[,,"T1DM",]
> vn <- dimnames( tt )[[3]]
> dn <- paste( vn, ": ", conv$Clab[match(vn,conv$NCnam)], sep="")
> dimnames( tt )[[3]] <- dn
> tt["M",,"d18: Breast"] <- 0
> tt["F",,"d24: Prostate"] <- 0
> print( ftable( addmargins(tt,2), col.vars=1:2 ), zero.print="." )
                         sex    M                             F
                         Cnt   DK   FI   SE   SC   AU  Sum   DK   FI   SE   SC   AU  Sum
```

```
d0: All sites                   401 1000 1882  253  504 4040  641 1408 2180  280  600 5109
d00: Non-sex-specific           352  832 1480  222  443 3329  351  680 1122  146  364 2663
d6: Oesophagus                    4   16   29    9    9   67    1    5   14    7    3   30
d7: Stomach                       6   47   64    5   12  134   14   50   39    4   13  120
d9: Colon                        26   56  173   18    .  273   16   66  119    9    .  210
d10: Rectum                      13   46   84   15    .  158   10   41   57    6    .  114
d52: Colorectal                  39  102  257   33   61  492   26  107  176   15   60  384
d11: Liver                        9   34   44   12   14  113    6    8   17    3    7   41
d13: Pancreas                    15   54   52    7   19  147    9   41   30    5    8   93
d16: Lung                        39  119  134   37   41  370   38   58  128   19   29  272
d29: Melanoma of skin            21   60  122   16   86  305   59   59  103   15   72  308
d18: Breast                       .    .    .    .    .    .  184  546  710   99  184 1723
d19: Cervix uteri                 .    .    .    .    .    .   48   36   85   14   11  194
d20: Corpus uteri                 .    .    .    .    .    .   40   97  149   12   25  323
d22: Ovary                        .    .    .    .    .    .   25   80  114   11   22  252
d24: Prostate                    12  148  341   11   41  553    .    .    .    .    .    .
d25: Testis                      37   23   57   20   22  159    .    .    .    .    .    .
d27: Kidney                      23   67   62   14   21  187   15   47   37    4   15  118
d28: Bladder                     17   49  124   11   12  213    7   16   35    6    2   66
d32: Brain, CNS                  31   35   79   18   17  180   42   32   98   28   13  213
d33: Thyroid                      6   18   15    4   15   58   29  104   51   12   45  241
d36: Non-Hodgkin lymphoma        26   56  111   21   30  244   14   46   56    8   19  143
d37: Hodgkin lymphoma            11   14   20    .   11   56    4    9    9    .    2   24
d38: Multiple myeloma             4   14   26    3    6   53    5    8   13    .    2   28
d40: Leukaemia                   14   39   38   13    .  104   16   33   28    9    .   86
> xcl <- c(1,2,5,6,23,24,25)
> conv[match(wh.ca[-xcl],conv$NCnam),]
   DKnam NCnam                 Clab
7    d21    d6           Oesophagus
8    d22    d7              Stomach
41  d251   d52           Colorectal
12   d26   d11                Liver
14   d28   d13             Pancreas
17   d33   d16                 Lung
29   d51   d29     Melanoma of skin
19   d70   d18               Breast
20   d82   d19         Cervix uteri
21   d83   d20         Corpus uteri
22   d84   d22                Ovary
24   d91   d24             Prostate
25   d92   d25               Testis
27  d101   d27               Kidney
28  d103   d28              Bladder
31  d113   d32           Brain, CNS
32  d121   d33              Thyroid
35  d132   d36 Non-Hodgkin lymphoma
> swpct <- function(mm,dm) sweep(mm,dm,apply(mm,dm,sum),"/")*100
> pp <- swpct( addmargins(tt[,,-xcl],2), 1:2 )
> print( round( ftable( addmargins(pp,3), col.vars=1:2 ), 1 ), zero.print=" " )
                          sex      M                                        F
                          Cnt     DK    FI    SE    SC    AU   Sum    DK    FI    SE    SC    AU   Su

d6: Oesophagus                   1.4   1.9   1.9   4.1   2.2   2.1   0.2   0.4   0.8   2.7   0.6   0.
d7: Stomach                      2.1   5.7   4.3   2.3   3.0   4.2   2.5   3.8   2.1   1.5   2.5   2.
d52: Colorectal                 13.7  12.3  17.2  15.1  15.2  15.3   4.7   8.0   9.6   5.7  11.4   8.
d11: Liver                       3.2   4.1   3.0   5.5   3.5   3.5   1.1   0.6   0.9   1.1   1.3   0.
d13: Pancreas                    5.3   6.5   3.5   3.2   4.8   4.6   1.6   3.1   1.6   1.9   1.5   2.
d16: Lung                       13.7  14.4   9.0  17.0  10.2  11.5   6.8   4.4   6.9   7.3   5.5   6.
d29: Melanoma of skin            7.4   7.2   8.2   7.3  21.5   9.5  10.6   4.4   5.6   5.7  13.6   6.
d18: Breast                                                        33.0  41.0  38.5  37.8  34.8  38.
d19: Cervix uteri                                                   8.6   2.7   4.6   5.3   2.1   4.
d20: Corpus uteri                                                   7.2   7.3   8.1   4.6   4.7   7.
d22: Ovary                                                          4.5   6.0   6.2   4.2   4.2   5.
d24: Prostate                    4.2  17.9  22.9   5.0  10.2  17.2
d25: Testis                     13.0   2.8   3.8   9.2   5.5   4.9
d27: Kidney                      8.1   8.1   4.2   6.4   5.2   5.8   2.7   3.5   2.0   1.5   2.8   2.
```

```
d28: Bladder                    6.0   5.9   8.3   5.0   3.0   6.6   1.3   1.2   1.9   2.3   0.4   1.
d32: Brain, CNS                10.9   4.2   5.3   8.3   4.2   5.6   7.5   2.4   5.3  10.7   2.5   4.
d33: Thyroid                    2.1   2.2   1.0   1.8   3.8   1.8   5.2   7.8   2.8   4.6   8.5   5.
d36: Non-Hodgkin lymphoma       9.1   6.8   7.4   9.6   7.5   7.6   2.5   3.5   3.0   3.1   3.6   3.
Sum                           100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.0 100.
```

We also provide a print with the sum across both sex and country for inclusion in the paper:

```
> TT <- addmargins( tt, 2 )
> TT <- t( rbind( TT["M",,], TT["F",,], TT["M","Sum",]+TT["F","Sum",] ) )
> colnames( TT )[c(1,7,13)] <- c("M: DK","   F: DK","   Total")
> rownames( TT ) <- substr( gsub("[0-9]","",rownames(TT)), 4, 50 )
> print.table( TT, zero.print=" " )
                     M: DK   FI   SE   SC   AU  Sum    F: DK   FI   SE   SC   AU  Sum    Total
All sites              401 1000 1882  253  504 4040      641 1408 2180  280  600 5109     9149
Non-sex-specific       352  832 1480  222  443 3329      351  680 1122  146  364 2663     5992
Oesophagus               4   16   29    9    9   67        1    5   14    7    3   30       97
Stomach                  6   47   64    5   12  134       14   50   39    4   13  120      254
Colon                   26   56  173   18       273       16   66  119    9       210      483
Rectum                  13   46   84   15       158       10   41   57    6       114      272
Colorectal              39  102  257   33   61  492       26  107  176   15   60  384      876
Liver                    9   34   44   12   14  113        6    8   17    3    7   41      154
Pancreas                15   54   52    7   19  147        9   41   30    5    8   93      240
Lung                    39  119  134   37   41  370       38   58  128   19   29  272      642
Melanoma of skin        21   60  122   16   86  305       59   59  103   15   72  308      613
Breast                                                   184  546  710   99  184 1723     1723
Cervix uteri                                              48   36   85   14   11  194      194
Corpus uteri                                              40   97  149   12   25  323      323
Ovary                                                     25   80  114   11   22  252      252
Prostate                12  148  341   11   41  553                                        553
Testis                  37   23   57   20   22  159                                        159
Kidney                  23   67   62   14   21  187       15   47   37    4   15  118      305
Bladder                 17   49  124   11   12  213        7   16   35    6    2   66      279
Brain, CNS              31   35   79   18   17  180       42   32   98   28   13  213      393
Thyroid                  6   18   15    4   15   58       29  104   51   12   45  241      299
Non-Hodgkin lymphoma    26   56  111   21   30  244       14   46   56    8   19  143      387
Hodgkin lymphoma        11   14   20        11   56        4    9    9         2   24       80
Multiple myeloma         4   14   26    3    6   53        5    8   13         2   28       81
Leukaemia               14   39   38   13       104       16   33   28    9        86      190
```

A similarly laid out table for the two versions of person-years:

```
> YY <- xtabs( cbind(y0,y)/1000 ~ Cnt + sex,
+              data = subset(all.ana,T1=="T1DM") )
> YY <- addmargins( YY, 1 )
> YY <- cbind( t(YY[,"M",]), t(YY[,"F",]), YY["Sum","M",]+YY["Sum","F",] )
> colnames( YY )[13] <- "Total"
> round( YY, 1 )
      DK    FI    SE    SC    AU    Sum    DK    FI    SE    SC    AU    Sum  Total
y0 255.6 547.9 737.5 178.7 255.5 1975.1 289.0 636.8 631.1 145.5 255.4 1957.7 3932.9
y  258.9 553.8 746.0 179.5 255.5 1993.7 293.9 648.0 645.2 146.8 255.4 1989.3 3983.0
```

8.3 Baseline splines

8.3.1 Case-distribution

In order to have country-specific splines for the underlying population rates, we explore the period-range for each of the countries:

```
> load( file = "../data/ALLana.Rda" )
> print(
+ ftable( xtabs( d0 ~ floor(P) + Cnt + T1,
+                data = transform( all.ana,
+                                  T1=Relevel(T1D,list(1,T1D=2:6)) ) ),
+        col.vars=c(3,2) ), zero.print="." )
         T1    NoDM                                       T1D
         Cnt     DK      FI      SE      SC      AU       DK      FI      SE      SC      AU
floor(P)
1972              .   11487       .       .       .        .       4       .       .       .
1973              .   11559       .       .       .        .       2       .       .       .
1974              .   11773       .       .       .        .       3       .       .       .
1975              .   12060       .       .       .        .       4       .       .       .
1976              .   12587       .       .       .        .       .       .       .       .
1977              .   12903       .       .       .        .       4       .       .       .
1978              .   13158       .       .       .        .      11       .       .       .
1979              .   13500       .       .       .        .       9       .       .       .
1980              .   13892       .       .       .        .       9       .       .       .
1981              .   13961       .       .       .        .      20       .       .       .
1982              .   14320       .       .       .        .      23       .       .       .
1983              .   14546       .       .       .        .      19       .       .       .
1984              .   14648       .       .       .        .      15       .       .       .
1985              .   14978       .       .       .        .      15       .       .       .
1986              .   15404       .       .       .        .      24       .       .       .
1987              .   15707   34084       .       .        .      22      49       .       .
1988              .   15718   34000       .       .        .      30      62       .       .
1989              .   15658   34023       .       .        .      47      63       .       .
1990              .   16034   34493       .       .        .      33      51       .       .
1991              .   16489   34756       .       .        .      20      77       .       .
1992              .   17184   34844       .       .        .      57      66       .       .
1993              .   17196   35392       .       .        .      44      91       .       .
1994              .   17886   35835       .       .        .      50      94       .       .
1995          20980   18267   35354   22048       .       15      62     108      14       .
1996          19233   19296   35577   23468       .       19      56     108      15       .
1997          21484   19312   35743   22772       .       16      85     130      16       .
1998          23752   19747   36503   22509       .       25      71     137      15       .
1999          22490   19959   37827   22444       .       34      73     138      26       .
2000          20531   20389   37960   22829   80927       21      86     172      21      87
2001          22708   20715   38615   22863   83093       43      87     181      18     102
2002          25185   21495   39395   23572   86480       46      88     177      31     108
2003          21605   22003   40918   23578   88023       41      97     205      37     107
2004          24628   23410   42262   24104   92575       56     110     243      35     125
2005          25198   23905   42019   23743   94913       47     142     230      28     141
2006          26202   23697   41662   24169   97165       71     140     248      37     133
2007          26515   23140   41710   24693  100209       69     142     231      44     134
2008          27705   23689   42289   25248  103537       85     160     261      40     167
2009          29491   24873   44700   25523       .      113     184     323      46       .
2010          28147   24879   44640   25422       .       97     177     303      58       .
2011          28622   25602   44855   25722       .      104     183     314      52       .
2012          28054       .       .       .       .      140       .       .       .       .
```

In order to illustrate how cancers among T1D patients are distributed by age and calendar time in the different countries, we extract these from the database:

```
> with( all.ana, table( T1,T1D ) )
      T1D
T1      NoDM    30    35    40
  NoDM  8466     0     0     0
  T1DM     0 55430 16170 15793
> ca.all <- subset( all.ana,
+                   d0>0 & T1D %in% c("30","35","40"),
+                   select=c("d0","A","P","Cnt") )
> ncnt <- xtabs( d0 ~ Cnt, data=ca.all )
> ca.ind <- ca.all[rep(1:nrow(ca.all),ca.all$d0),]
> ca.ind <- transform( ca.ind, A = A + runif(nrow(ca.ind),-0.5,0.5),
+                              P = P + runif(nrow(ca.ind),-0.5,0.5) )
```

Then we can plot the cancer cases in a Lexis diagram:

```
> ca.ind <- ca.ind[order(runif(nrow(ca.ind))),]
> cbind( cclr, levels( ca.ind$Cnt ), lcnt )
     cclr                lcnt
[1,] "red"         "DK" "Denmark"
[2,] "mediumblue"  "FI" "Finland"
[3,] "darkorange"  "SE" "Sweden"
[4,] "forestgreen" "SC" "Scotland"
[5,] "black"       "AU" "Australia"
> # parameters for the graphs
> ypi <- 12
> xli <- c(1970,2013.5)
> yli <- c(0,85)
> # the graphs, 0 corresponds to all countries together
> xtn <- c(".pdf",".eps")
> for( i in 0:5 ) for( j in 1:2 )
+    {
+ if( j==1 ) pdf(
+ paste("../graph/Joint-Lexis-",levels(ca.ind$Cnt)[i], xtn[j], sep=""),
+        height=diff(yli)/ypi+1,
+         width=diff(xli)/ypi+1 )
+ else postscript(
+ paste("../graph/Joint-Lexis-",levels(ca.ind$Cnt)[i], xtn[j], sep=""),
+        height=diff(yli)/ypi+1,
+         width=diff(xli)/ypi+1 )
+ tclr <- cclr
+ tclr[-i] <- "transparent"
+ par( mai=c(3,3,1,1)/4, mgp=c(3,1,0)/1.6, las=1 )
+ plot( NA,    xlim=xli, xaxs="i", ylim=yli, yaxs="i",
+              xlab="Date of cancer diagnosis",
+              ylab="Age at cancer diagnosis" )
+ abline( v=seq(1975,2010,5), h=seq(5,80,5), col=gray(0.8) )
+ box()
+ with( ca.ind, points( P, A, pch=16, cex=0.8, col=tclr[Cnt] ) )
+ text( rep(1979.5,5), 80+5/4-0:4*5/2, lcnt, col=tclr, font=2, adj=1 )
+ text( rep(1984.5,5), 80+5/4-0:4*5/2, ncnt, col=tclr, font=2, adj=1 )
+ if( i == 0 ){
+     text( 1979.5   , 80+5/4-  5*5/2,   "Total", col=gray(0.4), font=2, adj=1 )
+     text( 1984.5   , 80+5/4-  5*5/2, sum(ncnt), col=gray(0.4), font=2, adj=1 )
+            }
+ dev.off()
+     }
```

From figures 8.2 and 8.3 we see that the majority of the cancers are in the age-bracket 40-60 years of age.

8.3.2 Spline knots

From this table it is seen that we have T1D data in different time periods, so we use different sets of splines for period and cohort for different countries, but the same age-splines:

```
> a.kn <- seq(10,80,,8)
> # Period knots
> p.dk <- seq(1996,2011,,3)
> p.fi <- seq(1975,2007,,5)
> p.se <- seq(1988,2007,,4)
> p.sc <- seq(1996,2010,,2)
> p.au <- seq(1997,2007,,2)
> # Cohort knots
> c.dk <- seq(1920,1985,,5)
> c.fi <- seq(1900,1985,,7)
```

```
> c.se <- seq(1910,1985,,6)
> c.sc <- seq(1920,1985,,4)
> c.au <- seq(1920,1985,,5)
```

With these knots for the splines we can set up the design matrices for the baseline-effects for each county:

```
> rnam <- function(M,pre){colnames(M)<-paste(pre,colnames(M),sep="");M}
> # Denmark
> M.dk <- with( DK.an <- subset(all.ana,Cnt=="DK"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.dk." ),
+                    rnam( Ns(P   ,knots=p.dk)                 , "P.dk." ),
+                    rnam( Ns(P-A,knots=c.dk)                  , "C.dk." )))
> # Finland
> M.fi <- with( FI.an <- subset(all.ana,Cnt=="FI"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.fi." ),
+                    rnam( Ns(P   ,knots=p.fi)                 , "P.fi." ),
+                    rnam( Ns(P-A,knots=c.fi)                  , "C.fi." )))
> # Sweden
> M.se <- with( SE.an <- subset(all.ana,Cnt=="SE"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.se." ),
+                    rnam( Ns(P   ,knots=p.se)                 , "P.se." ),
+                    rnam( Ns(P-A,knots=c.se)                  , "C.se." )))
> # Scotland
> M.sc <- with( SC.an <- subset(all.ana,Cnt=="SC"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.sc." ),
+                    rnam( Ns(P   ,knots=p.sc)                 , "P.sc." ),
+                    rnam( Ns(P-A,knots=c.sc)                  , "C.sc." )))
> # Australia
> M.au <- with( AU.an <- subset(all.ana,Cnt=="AU"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.au." ),
+                    rnam( Ns(P   ,knots=p.au)                 , "P.au." ),
+                    rnam( Ns(P-A,knots=c.au)                  , "C.au." )))
> # Overview
> c( colnames(M.dk),
+    colnames(M.fi),
+    colnames(M.se),
+    colnames(M.sc),
+    colnames(M.au) )
 [1] "A.dk.1" "A.dk.2" "A.dk.3" "A.dk.4" "A.dk.5" "A.dk.6" "A.dk.7" "A.dk.8" "P.dk.1" "P.dk.2"
[11] "C.dk.1" "C.dk.2" "C.dk.3" "C.dk.4" "A.fi.1" "A.fi.2" "A.fi.3" "A.fi.4" "A.fi.5" "A.fi.6"
[21] "A.fi.7" "A.fi.8" "P.fi.1" "P.fi.2" "P.fi.3" "P.fi.4" "C.fi.1" "C.fi.2" "C.fi.3" "C.fi.4"
[31] "C.fi.5" "C.fi.6" "A.se.1" "A.se.2" "A.se.3" "A.se.4" "A.se.5" "A.se.6" "A.se.7" "A.se.8"
[41] "P.se.1" "P.se.2" "P.se.3" "C.se.1" "C.se.2" "C.se.3" "C.se.4" "C.se.5" "A.sc.1" "A.sc.2"
[51] "A.sc.3" "A.sc.4" "A.sc.5" "A.sc.6" "A.sc.7" "A.sc.8" "P.sc.1" "C.sc.1" "C.sc.2" "C.sc.3"
[61] "A.au.1" "A.au.2" "A.au.3" "A.au.4" "A.au.5" "A.au.6" "A.au.7" "A.au.8" "P.au.1" "C.au.1"
[71] "C.au.2" "C.au.3" "C.au.4"
```

```
> addmargins( rbind( DK=dim(M.dk),
+                    FI=dim(M.fi),
+                    SE=dim(M.se),
+                    SC=dim(M.sc),
+                    AU=dim(M.au) ), margin=1 )
     [,1] [,2]
DK  15816   14
FI  30538   18
SE  24345   16
SC  14706   12
AU  10454   13
Sum 95859   73
```

With these model matrices in place we can now set up the total model matrix for the baseline rates, basically putting the 5 model matrices diagonally as matrices surrounded by 0s:

```
> c.dk <- ncol(M.dk) ; r.dk <- nrow(M.dk)
> c.fi <- ncol(M.fi) ; r.fi <- nrow(M.fi)
> c.se <- ncol(M.se) ; r.se <- nrow(M.se)
> c.sc <- ncol(M.sc) ; r.sc <- nrow(M.sc)
> c.au <- ncol(M.au) ; r.au <- nrow(M.au)
> MB <- rbind( cbind(                M.dk, matrix(0,r.dk,c.fi+c.se+c.sc+c.au) ),
+              cbind( matrix(0,r.fi,c.dk), M.fi, matrix(0,r.fi,c.se+c.sc+c.au) ),
+              cbind( matrix(0,r.se,c.dk+c.fi), M.se, matrix(0,r.se,c.sc+c.au) ),
+              cbind( matrix(0,r.sc,c.dk+c.fi+c.se), M.sc, matrix(0,r.sc,c.au) ),
+              cbind( matrix(0,r.au,c.dk+c.fi+c.se+c.sc), M.au                ) )
> dim( MB )
[1] 95859    73
> colnames( MB ) <- c( colnames(M.dk),
+                      colnames(M.fi),
+                      colnames(M.se),
+                      colnames(M.sc),
+                      colnames(M.au) )
```

Since we have the base model matrix set up as one, but ultimately we will be doing analyses by sex, so we must also have it available subdivided by sex:

```
> MB.m <- MB[all.ana$sex=="M",]
> MB.f <- MB[all.ana$sex=="F",]
```

8.4 Simple models for the HR

8.4.1 All cancer

We now model the effect of T1D for all patients:

```
> system.time(
+ m0.i <- glm( d0 ~ -1 + MB.m + T1:Cnt,
+              family = poisson,
+              offset = log(y0),
+              data = subset(all.ana,sex=="M") ) )
   user  system elapsed
  2.511   0.012   2.523
> m0.j <- update( m0.i, . ~ . - T1:Cnt + T1 )
> f0.i <- update( m0.i, . ~ . - MB.m + MB.f,
+              data = subset(all.ana,sex=="F") )
> f0.j <- update( f0.i, . ~ . - T1:Cnt + T1 )
> anova( m0.i, m0.j, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ -1 + MB.m + T1:Cnt
Model 2: d0 ~ MB.m + T1 - 1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     47755      21751
2     47759      21803 -4  -51.893 1.452e-10
> anova( f0.i, f0.j, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ MB.f + T1:Cnt - 1
Model 2: d0 ~ MB.f + T1 - 1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     47958      22187
2     47962      22202 -4  -15.108 0.004483
```

We see there is a considerable inhomogeneity between countries for the male cancers, this is clearly attributable to the Swedish data:

```
> HR.m <-
+ rbind( ci.exp( m0.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( m0.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> HR.f <-
+ rbind( ci.exp( f0.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( f0.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> rownames( HR.m ) <-
+ rownames( HR.f ) <- c(levels(all.ana$Cnt),"Joint")
> rownames( HR.m ) <-
+ rownames( HR.f ) <- c("Denmark","Finland","Sweden","Scotland","Australia","Joint")
> round( cbind( HR.m, HR.f ), 3 )
          exp(Est.)  2.5% 97.5% exp(Est.)  2.5% 97.5%
Denmark       1.101 0.998 1.215     1.003 0.928 1.084
Finland       1.179 1.108 1.255     1.005 0.953 1.059
Sweden        0.900 0.860 0.941     1.131 1.084 1.179
Scotland      1.061 0.938 1.201     1.054 0.937 1.186
Australia     1.025 0.940 1.119     1.093 1.009 1.185
Joint         1.001 0.971 1.033     1.068 1.039 1.098
```

We can plot the results for men and women together, showing this even more clearly:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( HR.m, txtpos=c(7:3,1), y=c(7:3,1)+0.15,
+          lwd=4, cex=1.5, xlog=TRUE, col="blue",
+          grid=8:15/10, xtic=c(8:15)/10, vref=1,
+          xlab="HR of total cancer, T1D vs Pop" )
> linesEst( HR.f, y=c(7:3,1)-0.15, lwd=4, cex=1.5, xlog=TRUE, col="red" )
```

8.4.1.1 Excluding follow-up in young ages: Interaction

It could be argued that cancers occurring in young ages (under 20, say) are substantially different from cancers in older ages, and these should be excluded from analyses overall. Essentially this is an interaction model with interaction between T1D effect and age dichotomized (at age 20), where only the effect for ages over 20 is reported.

To this there are two objections; one is that the dichotomization point is completely arbitrary, and the other that it is substantially more credible to expect a T1D effect that varies *continuously* with age.

The logical approach would therefore be to include the relevant interaction in the model; just as we subsequently shall include the interaction with T1D duration (although this is slightly different in that the interaction variable is only defined for T1D persons).

We therefore extend the joint models with T1D×age interactions, using a spline with knots at 10,35,50,65 (3 parameters) using age 50 as reference point for convenience:

```
> ia.kn <- c(10,35,50,65)
> m0.a <- update( m0.j, . ~ . + I((T1=="T1DM")*1):Ns(A,knots=ia.kn) )
> f0.a <- update( f0.j, . ~ . + I((T1=="T1DM")*1):Ns(A,knots=ia.kn) )
> anova( m0.j, m0.a, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ MB.m + T1 - 1
Model 2: d0 ~ MB.m + T1 + I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn) -
    1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     47759      21803
2     47756      21719  3   84.228 < 2.2e-16
> anova( f0.j, f0.a, test="Chisq" )
Analysis of Deviance Table

Model 1: d0 ~ MB.f + T1 - 1
```

```
Model 2: d0 ~ MB.f + T1 + I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn) -
    1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     47962      22202
2     47959      22184  3   18.444 0.0003562
```

We see there is a significant T1×age interaction, much stronger for man than for women. We can have a look at the estimated shape of the HR by age as for all cancers:

```
> a.pt <- 10:70
> CA <- cbind(-1,1,Ns(a.pt,knots=ia.kn))
> ci.exp( m0.a, subset="T1" )
                                            exp(Est.)      2.5%     97.5%
T1NoDM                                      1.0983425 0.8178511 1.4750316
T1T1DM                                      1.0000000 1.0000000 1.0000000
I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn)1 1.0449488 0.8549407 1.2771857
I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn)2 1.4923567 0.8103207 2.7484534
I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn)3 0.7563009 0.6682469 0.8559578

> ci.exp( m0.j, subset="T1" )

       exp(Est.)      2.5%    97.5%
T1NoDM 0.9985599 0.9681311 1.029945
T1T1DM 1.0000000 1.0000000 1.000000

> mrr.a <- ci.exp( m0.a, subset="T1", ctr.mat=CA )
> frr.a <- ci.exp( f0.a, subset="T1", ctr.mat=CA )
> mrr.c <- ci.exp( m0.j, subset="T1", ctr.mat=CA[,1:2] )
> frr.c <- ci.exp( f0.j, subset="T1", ctr.mat=CA[,1:2] )
> matplot( a.pt, cbind(mrr.a,frr.a,mrr.c,frr.c),
+          type="l", lty=rep(c(1,3), each=6), lwd=c(4,1,1),
+          col=rep(c("blue","red"),each=3),
+          ylim=c(0.5,2), log="y",
+          xlab="Age at follow-up",
+          ylab="All cancer HR T1D vs NoDM" )
> abline( h=1 )
```

We see that the interaction is substantially different between man and women.

8.4.2 Sensitivity analysis of PY-approximation

For convenience we will use all time alive as the person-years denominator in the analysis of the specific sites, thus including follow-up after cancer diagnosis as risk time for cancers which is formally wrong. By far the largest error this can induce is in the analysis of all cancers, where the post-diagnosis follow-up constitutes 0.03% of all follow-up among T1D patients:

```
> dy <- xtabs( cbind("      D"=d0,PY=y0/1000,pmil=y0) ~ I(floor(A/10)*10) + Cnt, data=all.ana )
> round( ftable( dy <- addmargins(swpct(dy,2:3),1), col.vars=2:3 ), 1 )
                    Cnt       DK                FI                SE                SC
                              D    PY  pmil       D    PY  pmil       D    PY  pmil       D    PY  p
I(floor(A/10) * 10)
0                           0.5  12.5  12.5     0.6  12.6  12.6     0.5  12.0  12.0     0.4  11.6  1
10                          0.5  12.0  12.0     0.6  13.7  13.7     0.4  12.6  12.6     0.4  12.6  1
20                          1.5  12.9  12.9     1.3  14.6  14.6     1.1  13.3  13.3     1.2  13.4  1
30                          3.6  14.9  14.9     3.1  14.7  14.7     2.7  14.1  14.1     3.1  14.7  1
40                          8.3  14.7  14.7     8.1  14.1  14.1     6.9  14.2  14.2     7.4  14.8  1
50                         18.2  13.6  13.6    17.4  12.4  12.4    15.1  12.6  12.6    16.1  12.8  1
60                         29.2  10.5  10.5    28.2   9.7   9.7    27.8  10.5  10.5    27.6  10.4  1
70                         28.4   6.9   6.9    30.5   6.5   6.5    32.9   8.0   8.0    32.1   7.4
80                          9.8   2.2   2.2    10.3   1.7   1.7    12.6   2.7   2.7    11.6   2.2
Sum                       100.0 100.0 100.0   100.0 100.0 100.0   100.0 100.0 100.0   100.0 100.0 10
```

```
> aa0 <- addmargins( xtabs( y0 ~ I(floor(A/10)*10) + Cnt, data=all.ana ), 1 )
> aa  <- addmargins( xtabs( y  ~ I(floor(A/10)*10) + Cnt, data=all.ana ), 1 )
> dy[,,3] <- 1000*(aa-aa0)/aa0
> names(dimnames( dy ))[1] <- "Agr"
> round( ftable( dy, col.vars=2:3 ), 1 )
    Cnt      DK                  FI                  SE                  SC                  AU
                 D    PY  pmil       D    PY  pmil       D    PY  pmil       D    PY  pmil       D    P
Agr
0              0.5  12.5   0.0     0.6  12.6   0.0     0.5  12.0   0.0     0.4  11.6   0.1     0.4  13.
10             0.5  12.0   0.0     0.6  13.7   0.0     0.4  12.6   0.0     0.4  12.6   0.1     0.5  13.
20             1.5  12.9   0.0     1.3  14.6   0.0     1.1  13.3   0.0     1.2  13.4   0.2     1.6  14.
30             3.6  14.9   0.2     3.1  14.7   0.1     2.7  14.1   0.0     3.1  14.7   0.5     3.9  15.
40             8.3  14.7   0.2     8.1  14.1   0.2     6.9  14.2   0.1     7.4  14.8   1.2     9.2  14.
50            18.2  13.6   0.1    17.4  12.4   0.2    15.1  12.6   0.2    16.1  12.8   3.0    18.6  12.
60            29.2  10.5   0.0    28.2   9.7   0.2    27.8  10.5   0.3    27.6  10.4   6.3    26.3   8.
70            28.4   6.9   0.0    30.5   6.5   0.1    32.9   8.0   0.2    32.1   7.4  10.4    28.5   5.
80             9.8   2.2   0.0    10.3   1.7   0.0    12.6   2.7   0.1    11.6   2.2  12.9    10.9   1.
Sum          100.0 100.0   0.1   100.0 100.0   0.1   100.0 100.0   0.1   100.0 100.0   2.4   100.0 100.
```

The above tabulations shows that the largest relative difference (`pmil` is the fraction follow-up (per 1000) after cancer diagnoses) is in the 90–99 age class where the relative difference in PY is 7% in Sweden, an age-class with 0.1% of the follow-up. For single sites this fraction is of course substantially lower.

In order to substantiate the magnitude of the possible error, we re-do the analysis of all cancers using the follow-up, post cancer too, and compare the results. The differences for the single site analyses will all be substantially smaller, as the amount of follow-up after specific cancers is substantially smaller than for .

```
> system.time(
+ xm0.i <- glm( d0 ~ -1 + MB.m + T1:Cnt,
+               family = poisson,
+               offset = log(y),
+               data = subset(all.ana,sex=="M") ) )
   user  system elapsed
  2.351   0.016   2.366
> xm0.j <- update( xm0.i, . ~ . - T1:Cnt + T1 )
> xf0.i <- update( xm0.i, . ~ . - MB.m + MB.f,
+               data = subset(all.ana,sex=="F") )
> xf0.j <- update( xf0.i, . ~ . - T1:Cnt + T1 )
> xHR.m <-
+ rbind( ci.exp( xm0.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( xm0.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> xHR.f <-
+ rbind( ci.exp( xf0.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( xf0.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> rownames( xHR.m ) <-
+ rownames( xHR.f ) <- c(levels(all.ana$Cnt),"Joint")
> rownames( xHR.m ) <-
+ rownames( xHR.f ) <- c("Denmark","Finland","Sweden","Scotland","Australia","Joint")
> round( cbind(  HR.m,  HR.f ), 3 )
          exp(Est.)  2.5% 97.5% exp(Est.)  2.5% 97.5%
Denmark       1.101 0.998 1.215     1.003 0.928 1.084
Finland       1.179 1.108 1.255     1.005 0.953 1.059
Sweden        0.900 0.860 0.941     1.131 1.084 1.179
Scotland      1.061 0.938 1.201     1.054 0.937 1.186
Australia     1.025 0.940 1.119     1.093 1.009 1.185
Joint         1.001 0.971 1.033     1.068 1.039 1.098
> round( cbind( xHR.m, xHR.f ), 3 )
          exp(Est.)  2.5% 97.5% exp(Est.)  2.5% 97.5%
Denmark       1.080 0.979 1.192     0.977 0.903 1.056
Finland       1.136 1.068 1.209     0.966 0.916 1.018
Sweden        0.861 0.823 0.901     1.074 1.029 1.120
```

```
Scotland      1.049 0.927 1.187     1.036 0.921 1.165
Australia     1.025 0.940 1.119     1.093 1.009 1.185
Joint         0.969 0.939 0.999     1.029 1.001 1.058
> cbind(
+ round( rbind(  cbind(  HR.m,   HR.f )/
+                cbind( xHR.m, xHR.f ) ), 2 ),
+ round( rbind(  cbind(  HR.m,   HR.f )-
+                cbind( xHR.m, xHR.f ) ), 2 ),
+ round( rbind( (cbind(  HR.m,   HR.f )/
+                cbind( xHR.m, xHR.f )-1)*100 ), 1 ) )
          exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
Denmark        1.02 1.02  1.02      1.03 1.03  1.03      0.02 0.02  0.02      0.03 0.02  0.03
Finland        1.04 1.04  1.04      1.04 1.04  1.04      0.04 0.04  0.05      0.04 0.04  0.04
Sweden         1.05 1.05  1.05      1.05 1.05  1.05      0.04 0.04  0.04      0.06 0.05  0.06
Scotland       1.01 1.01  1.01      1.02 1.02  1.02      0.01 0.01  0.01      0.02 0.02  0.02
Australia      1.00 1.00  1.00      1.00 1.00  1.00      0.00 0.00  0.00      0.00 0.00  0.00
Joint          1.03 1.03  1.03      1.04 1.04  1.04      0.03 0.03  0.03      0.04 0.04  0.04
          exp(Est.) 2.5% 97.5% exp(Est.) 2.5% 97.5%
Denmark         2.0  2.0   2.0       2.7  2.7   2.7
Finland         3.8  3.8   3.8       4.0  4.0   4.0
Sweden          4.5  4.5   4.5       5.3  5.3   5.3
Scotland        1.2  1.2   1.2       1.8  1.8   1.8
Australia       0.0  0.0   0.0       0.0  0.0   0.0
Joint           3.4  3.4   3.4       3.8  3.8   3.8
```

We see that the HRs is about 1–5% smaller if we use the incorrect follow-up in the case of *all* cancers. For the specific sites, the deflation of the HR due to the extra person-years used will therefore be smaller than 1% since none of the specific sites constitute more than a third of all cases in any country as seen in the table in page 130.

```
> tt <- sapply( all.ana[,9:32], FUN=function(x) tapply(x,list(all.ana$Cnt,all.ana$sex),sum) )
> tt <- t(100*sweep( tt[,-1], 1, tt[,1], "/" ))
> dim(tt)
[1] 23 10
> snam <- conv$Clab[match( dimnames(tt)[[1]], conv$NCnam )]
> dim(tt) <- c(dim(tt)[1],5,2)
> dimnames(tt) <- list( site = snam,
+                    country = levels(all.ana$Cnt),
+                        sex = levels(all.ana$sex) )
> round(ftable(tt,col.vars=3:2),1)
                     sex        M                         F
                     country   DK   FI   SE   SC   AU   DK   FI   SE   SC   AU
site
Oesophagus                    2.3  1.3  1.2  4.0  1.4  0.8  1.0  0.5  2.2  0.7
Stomach                       2.6  5.6  3.1  4.0  2.2  1.4  4.6  2.1  2.4  1.4
Colon                         9.2  5.0  7.1  9.1   NA  9.4  6.2  7.9  8.2   NA
Rectum                        5.7  4.1  4.9  5.8   NA  3.7  3.7  4.1  3.8   NA
Liver                         1.5  1.6  1.5  1.6  1.4  0.7  1.1  1.1  0.8  0.7
Pancreas                      3.1  3.6  2.3  2.4  2.0  3.0  3.9  2.6  2.4  2.1
Lung                         16.7 19.2  8.5 20.4 10.3 13.9  4.8  6.6 16.3  7.7
Breast                        0.2  0.1  0.2  0.2  0.2 31.9 29.3 29.3 28.6 29.1
Cervix uteri                  0.0  0.0  0.0  0.0  0.0  3.2  2.0  2.4  2.6  1.8
Corpus uteri                  0.0  0.0  0.0  0.0  0.0  5.3  6.5  6.2  4.1  4.2
Ovary                         0.0  0.0  0.0  0.0  0.0  4.6  4.8  4.5  4.9  2.8
Prostate                     22.5 24.5 33.6 19.2 27.9  0.0  0.0  0.0  0.0  0.0
Testis                        2.4  0.8  1.3  1.7  1.3  0.0  0.0  0.0  0.0  0.0
Kidney                        2.9  3.5  2.7  3.0  2.7  1.7  2.8  2.1  1.9  1.9
Bladder                       9.4  5.4  7.4  8.3  2.8  3.2  1.8  2.8  3.6  1.1
Melanoma of skin              4.7  3.0  4.2  2.9 10.5  5.7  3.0  4.6  3.7 10.0
Brain, CNS                    3.2  2.9  2.8  3.0  1.5  2.7  4.2  3.6  3.6  1.4
Thyroid                       0.4  0.7  0.4  0.4  0.7  1.0  2.5  1.2  0.9  2.8
Non-Hodgkin lymphoma          3.6  3.7  3.4  3.6  3.8  2.9  3.7  3.0  3.4  3.9
Hodgkin lymphoma              0.6  0.8  0.5   NA  0.5  0.4  0.6  0.4   NA  0.5
Multiple myeloma              1.6  1.3  1.4  1.4  1.2  1.2  1.4  1.2  1.2  1.2
Leukaemia                     3.4  2.8  2.7  3.1   NA  2.5  2.5  2.1  2.2   NA
Colorectal                   14.9  9.0 12.0 14.7 13.0 13.1  9.9 12.0 11.9 12.9
```

8.4.2.1 Specific sites

Breast cancer constitutes about 40% of the female cases of cancer, and the second largest of female cancers (endometrial and melanoma) only 7.4%. Among men, prostate cancer constitutes 18% and the second largest of male cancer lung cancer about 10%. Therefore clearly the main concern in terms of inaccuracy is the breast cancer, so we analyze breast cancer using the two extremes, one over- and one under-estimating the HR versus the population:

```
> system.time(
+ b0.i <- glm( d18 ~ -1 + MB.f + T1:Cnt,
+               family = poisson,
+               offset = log(y),
+               data = subset(all.ana,sex=="F") ) )
   user  system elapsed
  5.885   0.012   5.895
> b0.j <- update( b0.i, . ~ . - T1:Cnt + T1 )
> xb0.i <- glm( d18 ~ -1 + MB.f + T1:Cnt,
+               family = poisson,
+               offset = log(y0),
+               data = subset(all.ana,sex=="F") )
> xb0.j <- update( xb0.i, . ~ . - T1:Cnt + T1 )
> bHR <-
+ rbind( ci.exp( b0.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( b0.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> xbHR <-
+ rbind( ci.exp( xb0.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( xb0.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> rownames(  bHR ) <-
+ rownames( xbHR ) <- c(levels(all.ana$Cnt),"Joint")
> rownames(  bHR ) <-
+ rownames( xbHR ) <- c("Denmark","Finland","Sweden","Scotland","Australia","Joint")
> round( cbind(  bHR,  xbHR ), 3 )
          exp(Est.)  2.5% 97.5% exp(Est.)  2.5% 97.5%
Denmark       0.743 0.642 0.859     0.765 0.661 0.884
Finland       0.881 0.810 0.959     0.916 0.842 0.996
Sweden        0.952 0.884 1.025     0.997 0.926 1.074
Scotland      0.911 0.748 1.110     0.928 0.762 1.131
Australia     0.892 0.772 1.031     0.892 0.772 1.031
Joint         0.894 0.852 0.937     0.926 0.883 0.971
> round( cbind(  bHR/xbHR ), 3 )
          exp(Est.)  2.5% 97.5%
Denmark       0.971 0.971 0.971
Finland       0.962 0.962 0.962
Sweden        0.954 0.954 0.954
Scotland      0.982 0.982 0.982
Australia     1.000 1.000 1.000
Joint         0.966 0.966 0.966
> round( cbind(  bHR-xbHR ), 3 )
          exp(Est.)   2.5%  97.5%
Denmark      -0.022 -0.019 -0.025
Finland      -0.035 -0.032 -0.038
Sweden       -0.045 -0.042 -0.049
Scotland     -0.017 -0.014 -0.020
Australia     0.000  0.000  0.000
Joint        -0.032 -0.030 -0.033
```

Thus the difference in HR between the value over and under is less than 0.05, which an upper bound of the bias introduced by the convenience follow-up. There is no difference between the estimates for the Australian data because follow-up has only been computed to death / end of FU in Australian data.

8.4.3 Non-sex-specific cancers

We now estimate the HR associated with T1D diagnosis for the non-sex-specific cancers:

```
> system.time(
+ m00.i <- glm( d00 ~ -1 + MB.m + T1:Cnt,
+               family = poisson,
+               offset = log(y0),
+               data = subset(all.ana,sex=="M") ) )
   user  system elapsed
  2.339   0.008   2.346
> m00.j <- update( m00.i, . ~ . - T1:Cnt + T1 )
> f00.i <- update( m00.i, . ~ . - MB.m + MB.f,
+              data = subset(all.ana,sex=="F") )
> f00.j <- update( f00.i, . ~ . - T1:Cnt + T1 )
> anova( m00.i, m00.j, test="Chisq" )
Analysis of Deviance Table

Model 1: d00 ~ -1 + MB.m + T1:Cnt
Model 2: d00 ~ MB.m + T1 - 1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     47755      17717
2     47759      17739 -4   -22.61 0.0001515
> anova( f00.i, f00.j, test="Chisq" )
Analysis of Deviance Table

Model 1: d00 ~ MB.f + T1:Cnt - 1
Model 2: d00 ~ MB.f + T1 - 1
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1     47958      17483
2     47962      17490 -4  -7.5492   0.1096
```

We see there still is a considerable inhomogeneity between countries for the male cancers, this is clearly attributable to the Swedish data:

```
> HR.m <-
+ rbind( ci.exp( m00.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( m00.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> HR.f <-
+ rbind( ci.exp( f00.i, subset="T1", ctr.mat=cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)] ),
+        ci.exp( f00.j, subset="T1", ctr.mat=rbind(c(-1,1)) ) )
> rownames( HR.m ) <-
+ rownames( HR.f ) <- c(levels(all.ana$Cnt),"Joint")
> rownames( HR.m ) <-
+ rownames( HR.f ) <- c("Denmark","Finland","Sweden","Scotland","Australia","Joint")
> round( cbind( HR.m, HR.f ), 3 )
          exp(Est.)  2.5% 97.5% exp(Est.)  2.5% 97.5%
Denmark       1.198 1.079 1.331     1.162 1.046 1.291
Finland       1.298 1.213 1.390     1.075 0.997 1.160
Sweden        1.058 1.006 1.114     1.181 1.114 1.253
Scotland      1.140 0.999 1.300     1.254 1.065 1.475
Australia     1.162 1.058 1.275     1.259 1.136 1.396
Joint         1.144 1.106 1.184     1.163 1.120 1.208
```

We can plot the results for men and women together, showing this even more clearly:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6 )
> plotEst( HR.m, y=c(7:3,1)+0.15,
+          lwd=4, cex=1.5, xlog=TRUE, col="blue",
+          grid=8:15/10, xtic=c(8:15)/10, vref=1,
+          xlab="HR of total cancer, T1D vs population" )
> linesEst( HR.f, y=c(7:3,1)-0.15, lwd=4, cex=1.5, xlog=TRUE, col="red" )
```

8.4.3.1 Age-interaction

We repeat the interaction model for the non sex-specific cancers, using the same approach as for all cancers:

```
> ia.kn <- c(10,35,50,65)
> m00.a <- update( m00.j, . ~ . + I((T1=="T1DM")*1):Ns(A,knots=ia.kn) )
> f00.a <- update( f00.j, . ~ . + I((T1=="T1DM")*1):Ns(A,knots=ia.kn) )
> anova( m00.j, m00.a, test="Chisq" )
Analysis of Deviance Table

Model 1: d00 ~ MB.m + T1 - 1
Model 2: d00 ~ MB.m + T1 + I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn) -
    1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     47759      17739
2     47756      17698  3   41.366 5.468e-09
> anova( f00.j, f00.a, test="Chisq" )
Analysis of Deviance Table

Model 1: d00 ~ MB.f + T1 - 1
Model 2: d00 ~ MB.f + T1 + I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn) -
    1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1     47962      17490
2     47959      17470  3    20.35 0.0001436
```

We see there is a significant T1×age interaction, and we can have a look at the estimated shape of the HR by age:

```
> a.pt <- 10:70
> CA <- cbind(-1,1,Ns(a.pt,knots=ia.kn))
> ci.exp( m00.a, subset="T1" )
                                          exp(Est.)      2.5%     97.5%
T1NoDM                                    0.9590926 0.7107818 1.2941503
T1T1DM                                    1.0000000 1.0000000 1.0000000
I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn)1 1.1086814 0.8992080 1.3669523
I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn)2 1.3149780 0.7075255 2.4439645
I((T1 == "T1DM") * 1):Ns(A, knots = ia.kn)3 0.8392869 0.7334365 0.9604139
> ci.exp( m00.j, subset="T1" )
       exp(Est.)      2.5%     97.5%
T1NoDM 0.8738695 0.8445618 0.9041942
T1T1DM 1.0000000 1.0000000 1.0000000
> mrr.a <- ci.exp( m00.a, subset="T1", ctr.mat=CA )
> frr.a <- ci.exp( f00.a, subset="T1", ctr.mat=CA )
> mrr.c <- ci.exp( m00.j, subset="T1", ctr.mat=CA[,1:2] )
> frr.c <- ci.exp( f00.j, subset="T1", ctr.mat=CA[,1:2] )
> matplot( a.pt, cbind(mrr.a,frr.a,mrr.c,frr.c),
+          type="l", lty=rep(c(1,3), each=6), lwd=c(4,1,1),
+          col=rep(c("blue","red"),each=3),
+          ylim=c(0.5,2), log="y",
+          xlab="Age at follow-up",
+          ylab="Non sex-specific cancer HR T1D vs NoDM" )
> abline( h=1 )
```

We see (not surprisingly) that the interaction has approximately the same form as for all cancers (see fig. 8.5), and in particular very different between man and women.

8.4.4 All sub-sites

In order to do the analyses for all sub-sites available we set up an array to hold the results

```
> Earr <- NArray( list( site = wh.ca,
+                        sex = levels( all.ana$sex ),
+                    country = c(levels( all.ana$Cnt ),"Joint"),
+                         wh = c("HR","lo","up") ) )
> Tarr <- NArray( c( dimnames(Earr)[1:2],
+                    list( wh = c("Chisq","P","conv S","conv J") ) ) )
> str( Earr )
 logi [1:25, 1:2, 1:6, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ site   : chr [1:25] "d0" "d00" "d6" "d7" ...
  ..$ sex    : chr [1:2] "M" "F"
  ..$ country: chr [1:6] "DK" "FI" "SE" "SC" ...
  ..$ wh     : chr [1:3] "HR" "lo" "up"
> str( Tarr )
 logi [1:25, 1:2, 1:4] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ site: chr [1:25] "d0" "d00" "d6" "d7" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ wh  : chr [1:4] "Chisq" "P" "conv S" "conv J"
```

With these arrays in place we can fit models for each site separately, but we first fill in the values from the analyses of all cancers and non-sex-specific cancers:

```
> Tarr["d0","M",] <- c( as.numeric( anova( m0.j, m0.i, test="Chisq" )[2,4:5] ),
+                       m0.i$converged, m0.j$converged )
> Tarr["d0","F",] <- c( as.numeric( anova( f0.j, f0.i, test="Chisq" )[2,4:5] ),
+                       f0.i$converged, f0.j$converged )
> Tarr["d00","M",] <- c( as.numeric( anova( m00.j, m00.i, test="Chisq" )[2,4:5] ),
+                        m00.i$converged, m00.j$converged )
> Tarr["d00","F",] <- c( as.numeric( anova( f00.j, f00.i, test="Chisq" )[2,4:5] ),
+                        f00.i$converged, f00.j$converged )
> # HR estimates
> CI <- cbind(-diag(5),diag(5))[,c(1,6,2,7,3,8,4,9,5,10)]
> CM <- rbind(c(-1,1))
> Earr["d0","M",,] <- rbind( ci.exp( m0.i, subset="T1", ctr.mat=CI ),
+                            ci.exp( m0.j, subset="T1", ctr.mat=CM ) )
> Earr["d0","F",,] <- rbind( ci.exp( f0.i, subset="T1", ctr.mat=CI ),
+                            ci.exp( f0.j, subset="T1", ctr.mat=CM ) )
> Earr["d00","M",,] <- rbind( ci.exp( m00.i, subset="T1", ctr.mat=CI ),
+                             ci.exp( m00.j, subset="T1", ctr.mat=CM ) )
> Earr["d00","F",,] <- rbind( ci.exp( f00.i, subset="T1", ctr.mat=CI ),
+                             ci.exp( f00.j, subset="T1", ctr.mat=CM ) )
```

Then we do a loop over the rest of the cancer sites; note that we now use `y` instead of `y0` in the offset expression:

```
> # id <-  dimnames(Earr)[[1]][3]
> for( id in dimnames(Earr)[[1]][-(1:2)] )
+    {
+    cat( id, "started:", format(Sys.time()) )
+    all.ana$D <- all.ana[,id]
+ m0.i <- glm( D ~ -1 + MB.m + T1:Cnt,
+              family = poisson,
+              offset = log(y),
+              data = subset(all.ana,sex=="M") )
+ m0.j <- update( m0.i, . ~ . - T1:Cnt + T1 )
+ f0.i <- update( m0.i, . ~ . - MB.m + MB.f,
+              data = subset(all.ana,sex=="F") )
+ f0.j <- update( f0.i, . ~ . - T1:Cnt + T1 )
+ # Test for homogeneity of the HR
+ Tarr[id,"M",] <- c( as.numeric( anova( m0.j, m0.i, test="Chisq" )[2,4:5] ),
+                     m0.i$converged, m0.j$converged )
+ Tarr[id,"F",] <- c( as.numeric( anova( f0.j, f0.i, test="Chisq" )[2,4:5] ),
+                     f0.i$converged, f0.j$converged )
+ # HR estimates - first allow for fewer countrie in the interaction
```

```
+ wh <- unique( substr( rownames(ci.exp(m0.i,subset="T1")), 11, 12 ) )
+ nn <- length( wh )
+ wh <- c(wh,"Joint")
+ Ci <- CI[1:nn,1:(2*nn)]
+ Earr[id,"M",wh,] <- rbind( ci.exp( m0.i, subset="T1", ctr.mat=Ci ),
+                            ci.exp( m0.j, subset="T1", ctr.mat=CM ) )
+ Earr[id,"F",wh,] <- rbind( ci.exp( f0.i, subset="T1", ctr.mat=Ci ),
+                            ci.exp( f0.j, subset="T1", ctr.mat=CM ) )
+ cat( " and ended:", format( Sys.time(), format="%H:%M:%S" ), "\n" )
+ flush.console()
+     }
> round( ftable( Tarr[1:4,,] ), 3 )
> save( Earr, Tarr, conv, file="../data/Arr0.Rda" )
```

8.4.4.1 Plotting the results

We can now show the results from the single countries together with the joint results in a forest plot, which is somewhat busy:

```
> library( Epi )
> load( file="../data/Arr0.Rda" )
> dimnames( Tarr )[[1]] <-
+ dimnames( Earr )[[1]] <- conv[match(dimnames( Earr )[[1]],conv$NCnam),"Clab"]
> Earr["Breast"      ,"M",,] <-
+ Earr["Cervix uteri","M",,] <-
+ Earr["Corpus uteri","M",,] <-
+ Earr["Ovary"       ,"M",,] <-
+ Earr["Prostate"    ,"F",,] <-
+ Earr["Testis"      ,"F",,] <- NA
> Tarr["Breast"      ,"M",] <-
+ Tarr["Cervix uteri","M",] <-
+ Tarr["Corpus uteri","M",] <-
+ Tarr["Ovary"       ,"M",] <-
+ Tarr["Prostate"    ,"F",] <-
+ Tarr["Testis"      ,"F",] <- NA
> wh <- grep("Corpus",dimnames(Earr)$site)
> dimnames(Earr)$site[wh] <-
+ dimnames(Tarr)$site[wh] <- "Endometrium"
```

```
> y <- (dim(Earr)[1]:1)-1
> rg <- 0.3
> par( mar=c(3,1,1,0.5), cex=1.2 )
> plotEst( Earr[,"M","Joint",], y=y-rg, txtpos=y, ylim=c(0,max(y)+0.5),
+          lwd=2, cex=1.0, xlog=TRUE, col="blue", xlim=c(0.2,6),
+          xtic=c(2:15/10,2:5), grid=c(2:15/10,2:5), vref=1,
+          xlab="HR cancer, T1D vs population" )
> linesEst( Earr[,"M","Joint",], y=y-rg   , lwd=2, cex=1.0, col="blue" )
> linesEst( Earr[,"F","Joint",], y=y+rg/6, lwd=2, cex=1.0, col="red"  )
> for( i in 1:5 ) {
+ linesEst( Earr[,"M",i,], y=y-rg  +i*rg/6, lwd=1, pch=3, cex=0.6, col="#7777FF" )
+ linesEst( Earr[,"F",i,], y=y+rg/6+i*rg/6, lwd=1, pch=3, cex=0.6, col="#FF7777" ) }
> rhs <- 10^par("usr")[2]
> text( rep(rhs,length(y)), y-rg*3/4,
+       gsub("NA","",formatC(Tarr[,"M","P"],format="f",digits=3)),
+       col="blue", adj=1, cex=0.6 )
> text( rep(rhs,length(y)), y+rg*3/4,
+       gsub("NA","",formatC(Tarr[,"F","P"],format="f",digits=3)),
+       col="red", adj=1, cex=0.6 )
> text( rhs, max(y)+1, "Homogeneity of HR", cex=0.6, adj=1 )
```

```
> y <- ((dim(Earr)[1]):1)-1
> par( mar=c(3.5,1,1,0), mgp=c(3,1,0)/1.4, cex=1.2 )
> plotEst( Earr[,"M","Joint",], y=y-0.15, txtpos=y,
+          lwd=2, cex=1.0, xlog=TRUE, col="blue", ylim=c(0,max(y)+1/2),
+          xtic=c(5:15/10,2,2.5,3), grid=c(5:15/10,2,2.5,3), vref=1,
+          xlab="HR of cancer, T1D vs population", xlim=c(0.5,4.2) )
> # linesEst( Earr[,"M","Joint",], y=y-0.15, lwd=2, cex=1.0, col="blue" )
> linesEst( Earr[,"F","Joint",], y=y+0.15, lwd=2, cex=1.0, col="red" )
> rrF <- paste( formatC(Earr[,"F","Joint",1],format="f",digits=2), " (",
+               formatC(Earr[,"F","Joint",2],format="f",digits=2), ",",
+               formatC(Earr[,"F","Joint",3],format="f",digits=2), ")", sep="")
> rrF[grep("NA",rrF)] <- ""
> rrM <- paste( formatC(Earr[,"M","Joint",1],format="f",digits=2), " (",
+               formatC(Earr[,"M","Joint",2],format="f",digits=2), ",",
+               formatC(Earr[,"M","Joint",3],format="f",digits=2), ")", sep="")
> rrM[grep("NA",rrM)] <- ""
> text( rep(3.1,dim(Earr)[1]), y+0.2, rrF, adj=0, col="red" , cex=0.6 )
> text( rep(3.1,dim(Earr)[1]), y-0.2, rrM, adj=0, col="blue", cex=0.6 )
> text( 3.1, max(y)+1, "Hazard ratio", cex=0.6, adj=0 )
```

The numbers in the plot are here in print:

```
> round( ftable( Earr[,,1:3,], row.vars=1:2 ), 2 )
                          country  DK             FI             SE
                          wh       RR   lo   up   RR   lo   up   RR   lo   up
site                     sex
All sites                M       1.10 1.00 1.22 1.18 1.11 1.25 0.90 0.86 0.94
                         F       1.00 0.93 1.08 1.00 0.95 1.06 1.13 1.08 1.18
Non-sex-specific         M       1.20 1.08 1.33 1.30 1.21 1.39 1.06 1.00 1.11
                         F       1.16 1.05 1.29 1.08 1.00 1.16 1.18 1.11 1.25
Oesophagus               M       0.58 0.22 1.56 1.33 0.81 2.18 1.04 0.72 1.49
                         F       0.42 0.06 3.01 1.20 0.50 2.90 2.01 1.19 3.41
Stomach                  M       0.72 0.32 1.60 1.49 1.11 1.98 1.21 0.95 1.55
                         F       2.65 1.55 4.51 1.84 1.39 2.43 1.47 1.07 2.02
Colon                    M       1.30 0.88 1.92 1.23 0.95 1.60 1.26 1.08 1.46
                         F       0.73 0.45 1.19 1.14 0.90 1.46 1.10 0.92 1.32
Rectum                   M       0.90 0.52 1.55 1.31 0.98 1.75 0.83 0.67 1.02
                         F       0.82 0.44 1.53 1.17 0.86 1.59 0.91 0.71 1.18
Colorectal               M       1.13 0.83 1.55 1.26 1.04 1.54 1.07 0.95 1.21
                         F       0.76 0.52 1.12 1.15 0.95 1.40 1.03 0.89 1.20
Liver                    M       2.12 1.10 4.11 2.64 1.88 3.71 1.45 1.07 1.95
                         F       2.52 1.12 5.69 1.05 0.52 2.11 1.23 0.77 1.99
Pancreas                 M       1.84 1.10 3.07 1.89 1.44 2.47 1.08 0.82 1.42
                         F       1.37 0.71 2.65 1.58 1.16 2.16 0.84 0.59 1.20
Lung                     M       1.27 0.92 1.74 1.22 1.02 1.46 0.79 0.67 0.94
                         F       1.00 0.73 1.38 1.15 0.89 1.49 1.01 0.85 1.20
Melanoma of skin         M       0.50 0.32 0.76 1.16 0.90 1.50 0.92 0.77 1.10
                         F       0.69 0.53 0.89 0.86 0.66 1.11 0.80 0.66 0.97
Breast                   M         NA   NA   NA   NA   NA   NA   NA   NA   NA
                         F       0.74 0.64 0.86 0.88 0.81 0.96 0.95 0.89 1.03
Cervix uteri             M         NA   NA   NA   NA   NA   NA   NA   NA   NA
                         F       0.85 0.64 1.13 0.99 0.72 1.38 1.13 0.91 1.40
Endometrium              M         NA   NA   NA   NA   NA   NA   NA   NA   NA
                         F       2.25 1.65 3.09 1.28 1.05 1.57 1.37 1.16 1.60
Ovary                    M         NA   NA   NA   NA   NA   NA   NA   NA   NA
                         F       1.02 0.69 1.51 1.16 0.93 1.45 1.19 0.99 1.44
Prostate                 M       0.53 0.30 0.93 0.79 0.67 0.92 0.51 0.46 0.56
                         F         NA   NA   NA   NA   NA   NA   NA   NA   NA
Testis                   M       0.78 0.56 1.07 0.89 0.59 1.34 0.98 0.76 1.28
                         F         NA   NA   NA   NA   NA   NA   NA   NA   NA
Kidney                   M       1.65 1.09 2.50 1.75 1.37 2.22 0.96 0.75 1.23
                         F       2.27 1.36 3.80 1.68 1.26 2.24 1.17 0.85 1.61
Bladder                  M       0.94 0.58 1.51 1.34 1.01 1.77 0.87 0.73 1.04
                         F       0.99 0.47 2.09 1.34 0.82 2.20 0.87 0.62 1.21
Brain, CNS               M       1.15 0.81 1.64 0.69 0.50 0.96 0.86 0.69 1.08
                         F       1.49 1.10 2.02 0.38 0.27 0.54 1.09 0.90 1.33
```

```
Thyroid             M         1.14 0.51 2.56 1.28 0.80 2.03 1.01 0.61 1.69
                    F         1.64 1.13 2.36 1.54 1.27 1.87 1.45 1.10 1.91
Non-Hodgkin lymphoma M        1.32 0.89 1.94 1.08 0.83 1.41 1.29 1.07 1.55
                    F         0.93 0.55 1.57 0.92 0.69 1.23 1.07 0.82 1.39
Hodgkin lymphoma    M         1.42 0.78 2.58 0.83 0.49 1.40 1.17 0.75 1.82
                    F         0.59 0.22 1.57 0.63 0.33 1.22 0.82 0.43 1.58
Multiple myeloma    M         0.99 0.37 2.65 1.35 0.80 2.29 0.86 0.58 1.26
                    F         1.50 0.62 3.63 0.79 0.39 1.58 0.72 0.42 1.25
Leukaemia           M         1.02 0.60 1.73 1.39 1.01 1.91 0.61 0.44 0.84
                    F         1.43 0.87 2.35 1.30 0.92 1.83 0.76 0.53 1.10
```

```
> round( ftable( Earr[,,4:6,], row.vars=1:2 ), 2 )
```

```
                         country   SC             AU              Joint
                         wh          RR   lo   up   RR   lo   up    RR   lo   up
site                     sex
All sites                M         1.06 0.94 1.20 1.06 0.98 1.15  1.01 0.98 1.04
                         F         1.05 0.94 1.19 1.13 1.05 1.21  1.07 1.04 1.10
Non-sex-specific         M         1.14 1.00 1.30 1.21 1.11 1.31  1.15 1.11 1.19
                         F         1.26 1.07 1.48 1.31 1.20 1.44  1.17 1.13 1.22
Oesophagus               M         1.07 0.56 2.06 1.30 0.68 2.50  1.08 0.85 1.37
                         F         3.86 1.83 8.14 1.96 0.63 6.09  1.79 1.25 2.56
Stomach                  M         0.85 0.35 2.06 1.15 0.65 2.03  1.23 1.04 1.46
                         F         1.51 0.56 4.04 2.28 1.32 3.93  1.78 1.49 2.13
Colon                    M         1.19 0.75 1.88   NA   NA   NA  1.25 1.11 1.41
                         F         0.86 0.45 1.66   NA   NA   NA  1.06 0.93 1.22
Rectum                   M         1.19 0.71 1.97   NA   NA   NA  0.96 0.82 1.12
                         F         0.85 0.38 1.90   NA   NA   NA  0.97 0.81 1.17
Colorectal               M         1.20 0.85 1.68 1.20 0.95 1.52  1.14 1.04 1.24
                         F         0.86 0.52 1.44 1.54 1.21 1.95  1.09 0.99 1.21
Liver                    M         3.49 1.97 6.17 2.35 1.52 3.66  2.00 1.67 2.40
                         F         2.89 0.93 9.04 3.51 1.75 7.05  1.55 1.14 2.10
Pancreas                 M         1.50 0.71 3.15 2.55 1.68 3.88  1.53 1.30 1.79
                         F         1.95 0.81 4.71 1.96 1.05 3.66  1.25 1.02 1.53
Lung                     M         1.42 1.02 1.96 1.45 1.10 1.92  1.06 0.96 1.17
                         F         1.14 0.73 1.79 1.32 0.92 1.89  1.07 0.95 1.21
Melanoma of skin         M         0.94 0.57 1.54 1.03 0.86 1.25  0.94 0.84 1.05
                         F         0.69 0.42 1.15 0.91 0.75 1.12  0.81 0.73 0.90
Breast                   M           NA   NA   NA   NA   NA   NA    NA   NA   NA
                         F         0.91 0.75 1.11 0.90 0.79 1.03  0.90 0.85 0.94
Cervix uteri             M           NA   NA   NA   NA   NA   NA    NA   NA   NA
                         F         0.71 0.42 1.20 0.54 0.32 0.90  0.92 0.80 1.06
Endometrium              M           NA   NA   NA   NA   NA   NA    NA   NA   NA
                         F         1.42 0.80 2.50 1.49 1.05 2.12  1.42 1.27 1.58
Ovary                    M           NA   NA   NA   NA   NA   NA    NA   NA   NA
                         F         0.75 0.41 1.35 1.40 0.95 2.08  1.15 1.02 1.30
Prostate                 M         0.53 0.29 0.96 0.48 0.36 0.65  0.56 0.51 0.61
                         F           NA   NA   NA   NA   NA   NA    NA   NA   NA
Testis                   M         0.88 0.56 1.36 0.82 0.55 1.21  0.88 0.75 1.02
                         F           NA   NA   NA   NA   NA   NA    NA   NA   NA
Kidney                   M         1.49 0.88 2.51 1.17 0.79 1.75  1.30 1.12 1.49
                         F         1.01 0.38 2.71 1.52 0.92 2.53  1.47 1.23 1.77
Bladder                  M         0.84 0.46 1.51 1.37 0.78 2.42  0.97 0.85 1.11
                         F         1.54 0.69 3.45 0.82 0.20 3.27  1.01 0.79 1.28
Brain, CNS               M         1.05 0.66 1.68 1.44 0.96 2.17  0.92 0.80 1.06
                         F         1.60 1.10 2.32 2.23 1.48 3.36  0.97 0.85 1.11
Thyroid                  M         1.53 0.57 4.11 1.52 0.93 2.48  1.25 0.97 1.61
                         F         1.77 1.00 3.13 1.43 1.09 1.86  1.51 1.34 1.72
Non-Hodgkin lymphoma     M         1.48 0.96 2.27 1.21 0.88 1.66  1.24 1.09 1.40
                         F         1.02 0.51 2.05 1.32 0.90 1.92  1.04 0.88 1.21
Hodgkin lymphoma         M           NA   NA   NA 1.24 0.69 2.25  1.11 0.85 1.44
                         F           NA   NA   NA 0.27 0.07 1.08  0.61 0.41 0.91
Multiple myeloma         M         1.01 0.33 3.15 1.03 0.46 2.29  0.99 0.76 1.30
                         F         0.00 0.00  Inf 0.56 0.14 2.26  0.77 0.53 1.12
Leukaemia                M         1.47 0.85 2.53   NA   NA   NA  0.92 0.76 1.12
                         F         1.81 0.94 3.50   NA   NA   NA  1.10 0.89 1.36
```

We want to test whether HRs are the same between men and women, and to this end we construct a small R-function that computes the P-value from the confidence interval:

```
> ci2p <-
+ function( l1, u1, l2, u2, Exp=TRUE, alpha=0.05, df=Inf )
+ {
+ if( Exp )
+   {
+   l1 <- log(l1)
+   u1 <- log(u1)
+   l2 <- log(l2)
+   u2 <- log(u2)
+   }
+ se1 <- (u1-l1)/(qt(1-alpha/2,df)*2)
+ se2 <- (u2-l2)/(qt(1-alpha/2,df)*2)
+ chi <- ((l1+u1)/2-(l2+u2)/2)^2/(se1^2+se2^2)
+ cbind( l1=if(Exp) exp(l1) else l1,
+        u1=if(Exp) exp(u1) else u1,
+        l2=if(Exp) exp(l2) else l2,
+        u2=if(Exp) exp(u2) else u2,
+        Chisq=chi, P=1-pchisq(chi,1) )
+ }
```

And here are the tests for equality of the HR between men and women:

```
> round(
+ mfP <-ci2p( Earr[,"M","Joint","lo"],
+             Earr[,"M","Joint","up"],
+             Earr[,"F","Joint","lo"],
+             Earr[,"F","Joint","up"] ), 3 )
                      l1    u1    l2    u2 Chisq     P
All sites          0.976 1.037 1.044 1.102 9.530 0.002
Non-sex-specific   1.112 1.189 1.128 1.216 0.504 0.478
Oesophagus         0.848 1.371 1.247 2.562 5.234 0.022
Stomach            1.036 1.456 1.487 2.132 8.621 0.003
Colon              1.110 1.408 0.927 1.216 3.138 0.077
Rectum             0.823 1.125 0.812 1.170 0.010 0.919
Colorectal         1.042 1.241 0.989 1.206 0.363 0.547
Liver              1.670 2.398 1.144 2.103 1.989 0.158
Pancreas           1.300 1.794 1.024 1.534 2.250 0.134
Lung               0.957 1.171 0.951 1.207 0.023 0.878
Melanoma of skin   0.843 1.047 0.727 0.903 3.596 0.058
Breast                NA    NA 0.854 0.938    NA    NA
Cervix uteri          NA    NA 0.801 1.061    NA    NA
Endometrium           NA    NA 1.274 1.582    NA    NA
Ovary                 NA    NA 1.019 1.304    NA    NA
Prostate           0.514 0.607    NA    NA    NA    NA
Testis             0.750 1.022    NA    NA    NA    NA
Kidney             1.123 1.494 1.228 1.765 1.181 0.277
Bladder            0.851 1.113 0.789 1.281 0.053 0.818
Brain, CNS         0.797 1.063 0.851 1.107 0.283 0.595
Thyroid            0.966 1.613 1.337 1.715 1.757 0.185
Non-Hodgkin lymphoma 1.095 1.403 0.882 1.215 3.031 0.082
Hodgkin lymphoma   0.850 1.439 0.408 0.909 5.948 0.015
Multiple myeloma   0.756 1.298 0.531 1.116 1.157 0.282
Leukaemia          0.759 1.117 0.887 1.357 1.435 0.231
```

From the plot in figure 8.8 we see that some of the Swedish HRs are among the lowest ones, but not consistently throughout, though.

We can also see that the tests for interaction (different HRs between countries) are significant for all sites, liver, pancreas, lung(M), cervix uteri, prostate, kidney(M), melanoma(M), brain and multiple myeloma(F). For all these sites it is predominantly the Swedish figures that stand out as having a lower HR associated with T1D, the second table here contains the same numbers as above, but differently arranged for comparison:

```
> round( Tarr[,,2], 3 )
                     sex
site                      M     F
  All sites           0.000 0.003
  Non-sex-specific    0.000 0.023
  Oesophagus          0.581 0.097
  Stomach             0.347 0.373
  Colon               0.990 0.326
  Rectum              0.083 0.590
  Colorectal          0.681 0.013
  Liver               0.026 0.073
  Pancreas            0.006 0.038
  Lung                0.000 0.698
  Melanoma of skin    0.007 0.488
  Breast                 NA 0.048
  Cervix uteri           NA 0.043
  Endometrium            NA 0.075
  Ovary                  NA 0.433
  Prostate            0.001    NA
  Testis              0.844    NA
  Kidney              0.010 0.206
  Bladder             0.112 0.552
  Brain, CNS          0.057 0.000
  Thyroid             0.833 0.941
  Non-Hodgkin lymphoma 0.754 0.675
  Hodgkin lymphoma    0.543 0.471
  Multiple myeloma    0.777 0.532
  Leukaemia           0.001 0.046
> round( ftable( Earr, row.vars=c(1,3) ), 2 )
                                sex    M                 F
                                wh    RR   lo   up   RR   lo   up
site                 country
All sites            DK             1.10 1.00 1.22 1.00 0.93 1.08
                     FI             1.18 1.11 1.25 1.00 0.95 1.06
                     SE             0.90 0.86 0.94 1.13 1.08 1.18
                     SC             1.06 0.94 1.20 1.05 0.94 1.19
                     AU             1.06 0.98 1.15 1.13 1.05 1.21
                     Joint          1.01 0.98 1.04 1.07 1.04 1.10
Non-sex-specific     DK             1.20 1.08 1.33 1.16 1.05 1.29
                     FI             1.30 1.21 1.39 1.08 1.00 1.16
                     SE             1.06 1.00 1.11 1.18 1.11 1.25
                     SC             1.14 1.00 1.30 1.26 1.07 1.48
                     AU             1.21 1.11 1.31 1.31 1.20 1.44
                     Joint          1.15 1.11 1.19 1.17 1.13 1.22
Oesophagus           DK             0.58 0.22 1.56 0.42 0.06 3.01
                     FI             1.33 0.81 2.18 1.20 0.50 2.90
                     SE             1.04 0.72 1.49 2.01 1.19 3.41
                     SC             1.07 0.56 2.06 3.86 1.83 8.14
                     AU             1.30 0.68 2.50 1.96 0.63 6.09
                     Joint          1.08 0.85 1.37 1.79 1.25 2.56
Stomach              DK             0.72 0.32 1.60 2.65 1.55 4.51
                     FI             1.49 1.11 1.98 1.84 1.39 2.43
                     SE             1.21 0.95 1.55 1.47 1.07 2.02
                     SC             0.85 0.35 2.06 1.51 0.56 4.04
                     AU             1.15 0.65 2.03 2.28 1.32 3.93
                     Joint          1.23 1.04 1.46 1.78 1.49 2.13
Colon                DK             1.30 0.88 1.92 0.73 0.45 1.19
                     FI             1.23 0.95 1.60 1.14 0.90 1.46
                     SE             1.26 1.08 1.46 1.10 0.92 1.32
                     SC             1.19 0.75 1.88 0.86 0.45 1.66
                     AU               NA   NA   NA   NA   NA   NA
                     Joint          1.25 1.11 1.41 1.06 0.93 1.22
Rectum               DK             0.90 0.52 1.55 0.82 0.44 1.53
                     FI             1.31 0.98 1.75 1.17 0.86 1.59
                     SE             0.83 0.67 1.02 0.91 0.71 1.18
                     SC             1.19 0.71 1.97 0.85 0.38 1.90
```

	AU	NA	NA	NA	NA	NA	NA
	Joint	0.96	0.82	1.12	0.97	0.81	1.17
Colorectal	DK	1.13	0.83	1.55	0.76	0.52	1.12
	FI	1.26	1.04	1.54	1.15	0.95	1.40
	SE	1.07	0.95	1.21	1.03	0.89	1.20
	SC	1.20	0.85	1.68	0.86	0.52	1.44
	AU	1.20	0.95	1.52	1.54	1.21	1.95
	Joint	1.14	1.04	1.24	1.09	0.99	1.21
Liver	DK	2.12	1.10	4.11	2.52	1.12	5.69
	FI	2.64	1.88	3.71	1.05	0.52	2.11
	SE	1.45	1.07	1.95	1.23	0.77	1.99
	SC	3.49	1.97	6.17	2.89	0.93	9.04
	AU	2.35	1.52	3.66	3.51	1.75	7.05
	Joint	2.00	1.67	2.40	1.55	1.14	2.10
Pancreas	DK	1.84	1.10	3.07	1.37	0.71	2.65
	FI	1.89	1.44	2.47	1.58	1.16	2.16
	SE	1.08	0.82	1.42	0.84	0.59	1.20
	SC	1.50	0.71	3.15	1.95	0.81	4.71
	AU	2.55	1.68	3.88	1.96	1.05	3.66
	Joint	1.53	1.30	1.79	1.25	1.02	1.53
Lung	DK	1.27	0.92	1.74	1.00	0.73	1.38
	FI	1.22	1.02	1.46	1.15	0.89	1.49
	SE	0.79	0.67	0.94	1.01	0.85	1.20
	SC	1.42	1.02	1.96	1.14	0.73	1.79
	AU	1.45	1.10	1.92	1.32	0.92	1.89
	Joint	1.06	0.96	1.17	1.07	0.95	1.21
Melanoma of skin	DK	0.50	0.32	0.76	0.69	0.53	0.89
	FI	1.16	0.90	1.50	0.86	0.66	1.11
	SE	0.92	0.77	1.10	0.80	0.66	0.97
	SC	0.94	0.57	1.54	0.69	0.42	1.15
	AU	1.03	0.86	1.25	0.91	0.75	1.12
	Joint	0.94	0.84	1.05	0.81	0.73	0.90
Breast	DK	NA	NA	NA	0.74	0.64	0.86
	FI	NA	NA	NA	0.88	0.81	0.96
	SE	NA	NA	NA	0.95	0.89	1.03
	SC	NA	NA	NA	0.91	0.75	1.11
	AU	NA	NA	NA	0.90	0.79	1.03
	Joint	NA	NA	NA	0.90	0.85	0.94
Cervix uteri	DK	NA	NA	NA	0.85	0.64	1.13
	FI	NA	NA	NA	0.99	0.72	1.38
	SE	NA	NA	NA	1.13	0.91	1.40
	SC	NA	NA	NA	0.71	0.42	1.20
	AU	NA	NA	NA	0.54	0.32	0.90
	Joint	NA	NA	NA	0.92	0.80	1.06
Endometrium	DK	NA	NA	NA	2.25	1.65	3.09
	FI	NA	NA	NA	1.28	1.05	1.57
	SE	NA	NA	NA	1.37	1.16	1.60
	SC	NA	NA	NA	1.42	0.80	2.50
	AU	NA	NA	NA	1.49	1.05	2.12
	Joint	NA	NA	NA	1.42	1.27	1.58
Ovary	DK	NA	NA	NA	1.02	0.69	1.51
	FI	NA	NA	NA	1.16	0.93	1.45
	SE	NA	NA	NA	1.19	0.99	1.44
	SC	NA	NA	NA	0.75	0.41	1.35
	AU	NA	NA	NA	1.40	0.95	2.08
	Joint	NA	NA	NA	1.15	1.02	1.30
Prostate	DK	0.53	0.30	0.93	NA	NA	NA
	FI	0.79	0.67	0.92	NA	NA	NA
	SE	0.51	0.46	0.56	NA	NA	NA
	SC	0.53	0.29	0.96	NA	NA	NA
	AU	0.48	0.36	0.65	NA	NA	NA
	Joint	0.56	0.51	0.61	NA	NA	NA
Testis	DK	0.78	0.56	1.07	NA	NA	NA
	FI	0.89	0.59	1.34	NA	NA	NA
	SE	0.98	0.76	1.28	NA	NA	NA
	SC	0.88	0.56	1.36	NA	NA	NA

```
                    AU        0.82 0.55 1.21   NA   NA   NA
                    Joint     0.88 0.75 1.02   NA   NA   NA
Kidney              DK        1.65 1.09 2.50 2.27 1.36 3.80
                    FI        1.75 1.37 2.22 1.68 1.26 2.24
                    SE        0.96 0.75 1.23 1.17 0.85 1.61
                    SC        1.49 0.88 2.51 1.01 0.38 2.71
                    AU        1.17 0.79 1.75 1.52 0.92 2.53
                    Joint     1.30 1.12 1.49 1.47 1.23 1.77
Bladder             DK        0.94 0.58 1.51 0.99 0.47 2.09
                    FI        1.34 1.01 1.77 1.34 0.82 2.20
                    SE        0.87 0.73 1.04 0.87 0.62 1.21
                    SC        0.84 0.46 1.51 1.54 0.69 3.45
                    AU        1.37 0.78 2.42 0.82 0.20 3.27
                    Joint     0.97 0.85 1.11 1.01 0.79 1.28
Brain, CNS          DK        1.15 0.81 1.64 1.49 1.10 2.02
                    FI        0.69 0.50 0.96 0.38 0.27 0.54
                    SE        0.86 0.69 1.08 1.09 0.90 1.33
                    SC        1.05 0.66 1.68 1.60 1.10 2.32
                    AU        1.44 0.96 2.17 2.23 1.48 3.36
                    Joint     0.92 0.80 1.06 0.97 0.85 1.11
Thyroid             DK        1.14 0.51 2.56 1.64 1.13 2.36
                    FI        1.28 0.80 2.03 1.54 1.27 1.87
                    SE        1.01 0.61 1.69 1.45 1.10 1.91
                    SC        1.53 0.57 4.11 1.77 1.00 3.13
                    AU        1.52 0.93 2.48 1.43 1.09 1.86
                    Joint     1.25 0.97 1.61 1.51 1.34 1.72
Non-Hodgkin lymphoma DK       1.32 0.89 1.94 0.93 0.55 1.57
                    FI        1.08 0.83 1.41 0.92 0.69 1.23
                    SE        1.29 1.07 1.55 1.07 0.82 1.39
                    SC        1.48 0.96 2.27 1.02 0.51 2.05
                    AU        1.21 0.88 1.66 1.32 0.90 1.92
                    Joint     1.24 1.09 1.40 1.04 0.88 1.21
Hodgkin lymphoma    DK        1.42 0.78 2.58 0.59 0.22 1.57
                    FI        0.83 0.49 1.40 0.63 0.33 1.22
                    SE        1.17 0.75 1.82 0.82 0.43 1.58
                    SC          NA   NA   NA   NA   NA   NA
                    AU        1.24 0.69 2.25 0.27 0.07 1.08
                    Joint     1.11 0.85 1.44 0.61 0.41 0.91
Multiple myeloma    DK        0.99 0.37 2.65 1.50 0.62 3.63
                    FI        1.35 0.80 2.29 0.79 0.39 1.58
                    SE        0.86 0.58 1.26 0.72 0.42 1.25
                    SC        1.01 0.33 3.15 0.00 0.00  Inf
                    AU        1.03 0.46 2.29 0.56 0.14 2.26
                    Joint     0.99 0.76 1.30 0.77 0.53 1.12
Leukaemia           DK        1.02 0.60 1.73 1.43 0.87 2.35
                    FI        1.39 1.01 1.91 1.30 0.92 1.83
                    SE        0.61 0.44 0.84 0.76 0.53 1.10
                    SC        1.47 0.85 2.53 1.81 0.94 3.50
                    AU          NA   NA   NA   NA   NA   NA
                    Joint     0.92 0.76 1.12 1.10 0.89 1.36
```

8.5 Age-interactions

Since T1D is the dominant type of diabetes in younger ages, those at risk below age 30, 35 or 20 are less likely to be T2D cases, and it might therefore be expected that the cancer pattern is very different for persons in these age-classes. However the number of cases for specific types of cancer is tiny among the T1D population. Note the fishy behaviour of `xtabs` — the 0s for Scotland and Australia is because the Hodgkin lymphomas are missing for Scotland and Leukaemias are missing for Australia, and therefore must be resurrected separately and added in the appropriate place of the array:

```
> load( file = "../data/ALLana.Rda" )
> lls()
   name        mode      class        size
1  aa          numeric   table matrix 10 5
2  aa0         numeric   table matrix 10 5
3  af          numeric   table array  85 6 2
4  Af          numeric   table array  85 6 2
5  a.kn        numeric   numeric      8
6  all.ana     list      data.frame   95859 35
7  a.pt        numeric   integer      61
8  AU.an       list      data.frame   10454 35
9  au.ana      list      data.frame   14759 34
10 au.ca       character character    41
11 au.wh       numeric   integer      21
12 b0.i        list      glm lm       30
13 b0.j        list      glm lm       30
14 bHR         numeric   matrix       6 3
15 CA          numeric   matrix       61 5
16 ca.all      list      data.frame   7159 4
17 ca.ind      list      data.frame   9149 4
18 c.au        numeric   integer      1
19 cclr        character character    5
20 c.dk        numeric   integer      1
21 c.fi        numeric   integer      1
22 CI          numeric   matrix       5 10
23 ci2p        function  function     1
24 CM          numeric   matrix       1 2
25 cn          character character    1
26 conv        list      data.frame   42 3
27 c.sc        numeric   integer      1
28 c.se        numeric   integer      1
29 dd          numeric   matrix       7 1
30 df          numeric   table array  9 6 2
31 DK.an       list      data.frame   15816 35
32 dk.ana      list      data.frame   15877 33
33 dk.ca       character character    41
34 dk.wh       numeric   integer      24
35 dn          character character    25
36 dy          numeric   table array  10 5 3
37 Earr        numeric   array        25 2 6 3
38 f00.a       list      glm lm       30
39 f00.i       list      glm lm       30
40 f00.j       list      glm lm       30
41 f0.a        list      glm lm       30
42 f0.i        list      glm lm       30
43 f0.j        list      glm lm       30
44 FI.an       list      data.frame   30538 35
45 fi.ana      list      data.frame   30538 51
46 fi.ca       character character    41
47 fi.wh       numeric   integer      40
48 frr.a       numeric   matrix       61 3
49 frr.c       numeric   matrix       61 3
50 HR.f        numeric   matrix       6 3
51 HR.m        numeric   matrix       6 3
52 i           numeric   integer      1
53 ia.kn       numeric   numeric      4
54 ic          numeric   integer      1
55 it          numeric   integer      1
56 j           numeric   integer      1
57 lcnt        character character    5
58 m00.a       list      glm lm       30
59 m00.i       list      glm lm       30
60 m00.j       list      glm lm       30
61 m0.a        list      glm lm       30
62 m0.i        list      glm lm       30
63 m0.j        list      glm lm       30
```

```
64  M.au        numeric   matrix       10454 13
65  MB          numeric   matrix       95859 73
66  MB.f        numeric   matrix       48031 73
67  MB.m        numeric   matrix       47828 73
68  M.dk        numeric   matrix       15816 14
69  M.fi        numeric   matrix       30538 18
70  mfP         numeric   matrix       25 6
71  mm          numeric   matrix       95859 25
72  MM          numeric   matrix       13 5
73  mrr.a       numeric   matrix       61 3
74  mrr.c       numeric   matrix       61 3
75  M.sc        numeric   matrix       14706 12
76  M.se        numeric   matrix       24345 16
77  ncnt        numeric   xtabs table  5
78  p.au        numeric   numeric      2
79  p.dk        numeric   numeric      3
80  pf          numeric   table array  41 6 2
81  Pf          numeric   table array  41 6 2
82  p.fi        numeric   numeric      5
83  pp          numeric   array        2 6 18
84  p.sc        numeric   numeric      2
85  p.se        numeric   numeric      4
86  r.au        numeric   integer      1
87  r.dk        numeric   integer      1
88  r.fi        numeric   integer      1
89  rg          numeric   numeric      1
90  rhs         numeric   numeric      1
91  rnam        function  function     1
92  rrF         character character    25
93  rrM         character character    25
94  r.sc        numeric   integer      1
95  r.se        numeric   integer      1
96  S.all       function  function     1
97  SC.an       list      data.frame   14706 35
98  sc.ana      list      data.frame   14910 50
99  sc.ca       character character    41
100 sc.wh       numeric   integer      40
101 SE.an       list      data.frame   24345 35
102 se.ana      list      data.frame   25226 48
103 se.ca       character character    41
104 se.wh       numeric   integer      39
105 S.m1        function  function     1
106 snam        character character    23
107 swpct       function  function     1
108 Tarr        numeric   array        25 2 4
109 tclr        character character    5
110 tt          numeric   array        23 5 2
111 TT          numeric   matrix       25 13
112 vars        character character    33
113 vn          character character    25
114 wd          numeric   numeric      5
115 wh          numeric   integer      1
116 what        character character    2
117 wh.ca       character character    25
118 which.ctrib character character    41
119 wi          numeric   numeric      7
120 xb0.i       list      glm lm       30
121 xb0.j       list      glm lm       30
122 xbHR        numeric   matrix       6 3
123 xcl         numeric   numeric      7
124 xclr        character character    6
125 xcnt        character character    6
126 xf0.i       list      glm lm       30
127 xf0.j       list      glm lm       30
128 xHR.f       numeric   matrix       6 3
129 xHR.m       numeric   matrix       6 3
```

```
130 xli         numeric   numeric      2
131 xm0.i       list      glm lm       30
132 xm0.j       list      glm lm       30
133 xtn         character character    2
134 y           numeric   numeric      25
135 yli         numeric   numeric      2
136 ypi         numeric   numeric      1
137 YY          numeric   matrix       2 13
> tt <- xtabs( cbind(All=d0,
+           "non-Hogkin"=d36,
+               Hodgkin=d37,
+             Leukaemia=d40,
+           "PY (1000s)"=y0/1000) ~ I(floor(A/5)*5) + Cnt,
+             data = subset( all.ana, A<30 & T1=="T1DM" ) )
> str( tt )
 xtabs [1:6, 1:5, 1:5] 0 2 2 7 14 45 4 6 10 17 ...
 - attr(*, "dimnames")=List of 3
  ..$ I(floor(A/5) * 5): chr [1:6] "0" "5" "10" "15" ...
  ..$ Cnt              : chr [1:5] "DK" "FI" "SE" "SC" ...
  ..$                  : chr [1:5] "All" "non-Hogkin" "Hodgkin" "Leukaemia" ...
 - attr(*, "class")= chr [1:2] "xtabs" "table"
 - attr(*, "call")= language xtabs(formula = cbind(All = d0, `non-Hogkin` = d36, Hodgkin = d37, Leuk
> sc <- xtabs( cbind(All=d0,
+           "non-Hogkin"=d36,
+             Leukaemia=d40,
+           "PY (1000s)"=y0/1000) ~ I(floor(A/5)*5) + Cnt,
+             data = subset( all.ana, A<30 & T1=="T1DM" ) )
> au <- xtabs( cbind(All=d0,
+           "non-Hogkin"=d36,
+               Hodgkin=d37,
+           "PY (1000s)"=y0/1000) ~ I(floor(A/5)*5) + Cnt,
+             data = subset( all.ana, A<30 & T1=="T1DM" ) )
> dimnames(tt)
$`I(floor(A/5) * 5)`
[1] "0"  "5"  "10" "15" "20" "25"

$Cnt
[1] "DK" "FI" "SE" "SC" "AU"

[[3]]
[1] "All"        "non-Hogkin" "Hodgkin"    "Leukaemia"  "PY (1000s)"
> dimnames(sc)
$`I(floor(A/5) * 5)`
[1] "0"  "5"  "10" "15" "20" "25"

$Cnt
[1] "DK" "FI" "SE" "SC" "AU"

[[3]]
[1] "All"        "non-Hogkin" "Leukaemia"  "PY (1000s)"
> dimnames(tt[,,-3])
$`I(floor(A/5) * 5)`
[1] "0"  "5"  "10" "15" "20" "25"

$Cnt
[1] "DK" "FI" "SE" "SC" "AU"

[[3]]
[1] "All"        "non-Hogkin" "Leukaemia"  "PY (1000s)"
> tt[,"SC",-3] <- sc[,"SC",]
> dimnames(au)[[3]]
[1] "All"        "non-Hogkin" "Hodgkin"    "PY (1000s)"
> dimnames(tt[,,-4])[[3]]
[1] "All"        "non-Hogkin" "Hodgkin"    "PY (1000s)"
```

```
> tt[,"AU",-4] <- au[,"AU",]
> sum.0.20 <- function(x) sum(x[1:4])
> tt <- addmargins( tt, 1:2, FUN=list(list(sum.0.20,sum),sum) )
Margins computed over dimensions
in the following order:
1: I(floor(A/5) * 5)
2: Cnt
> ftable( round( tt[,,-5]   ), row.vars=c(3,1) )
                              Cnt  DK  FI  SE  SC  AU sum
            I(floor(A/5) * 5)
All         0                       0   4   3   1   1   9
            5                       2   6   5   1   4  18
            10                      2  10  14   3  10  39
            15                      7  17  14   7  10  55
            20                     14  35  30   7  22 108
            25                     45  47  68  18  42 220
            sum.0.20               11  37  36  12  25 121
            sum                    70 119 134  37  89 449
non-Hogkin  0                       0   2   1   0   0   3
            5                       0   0   1   0   0   1
            10                      0   0   3   0   0   3
            15                      0   0   1   2   1   4
            20                      1   3   0   1   1   6
            25                      1   3   2   2   2  10
            sum.0.20                0   2   6   2   1  11
            sum                     2   8   8   5   4  27
Hodgkin     0                       0   0   0   0   0   0
            5                       0   0   0   0   1   1
            10                      2   0   0   0   2   4
            15                      0   5   2   0   1   8
            20                      1   3   2   0   0   6
            25                      4   2   3   0   1  10
            sum.0.20                2   5   2   0   4  13
            sum                     7  10   7   0   5  29
Leukaemia   0                       0   1   0   1   0   2
            5                       0   4   2   0   0   6
            10                      0   3   4   2   0   9
            15                      0   4   3   1   0   8
            20                      1   4   2   0   0   7
            25                      1   2   2   3   0   8
            sum.0.20                0  12   9   4   0  25
            sum                     2  18  13   7   0  40
>          round( tt[,, 5], 1)
                  Cnt
I(floor(A/5) * 5)    DK    FI    SE    SC    AU    sum
         0          1.9   9.1   6.7   1.9   3.4   23.0
         5          7.7  35.2  28.1   9.1  15.1   95.2
         10        15.3  61.3  54.3  19.0  30.3  180.3
         15        23.3  77.3  73.9  25.2  39.8  239.5
         20        32.0  87.8  84.2  29.1  43.5  276.6
         25        49.5 112.5 101.3  34.3  51.3  348.8
         sum.0.20  48.2 183.0 162.9  55.2  88.7  538.0
         sum      129.6 383.3 348.4 118.6 183.5 1163.4
```

Thus we see that the total number of cancers among T1D patients under age 20 is a mere 121 and under 30, 449 cancers. If we focus on lymphomas and leukaemias we see that we have 11 Hodgkin, 13 non-Hodgkin and 25 Leukaemias in the total study. These numbers are very small and unlikely to provide any information on the interaction with age.

However the question raised by a reviewer about a special (different?) effect of T1D on cancer occurrence among persons under 20 is really a special case of an interaction with age (at follow-up). Thus, logically, what would be of interest is analysis of interactions between diabetes status and other variables; that is whether and in particular how the HR between T1D and the general population varies with age (at follow-up) and duration of diabetes. The latter is already addressed in a separate analysis, whereas the former is assumed to be absent without further justification. As is the interaction between T1D status and other variables such as calendar time.

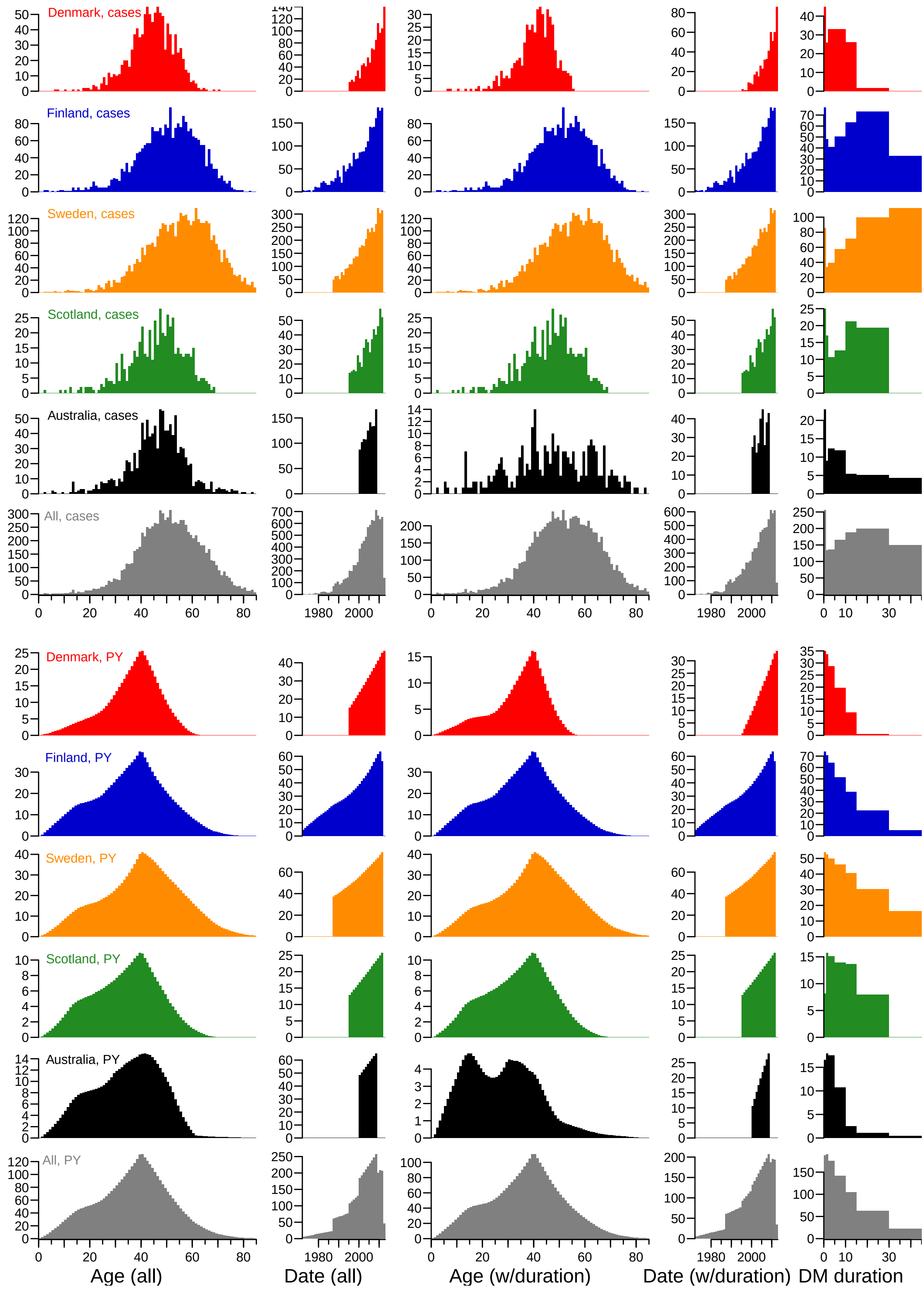

Figure 8.1: *Distribution of cancer cases (top 5 rows) and person-years (in 1000s, bottom 5 rows) among T1D patients in the analysis. Note that only the x-axes of are identical across countries; the y-axis scales are different.*

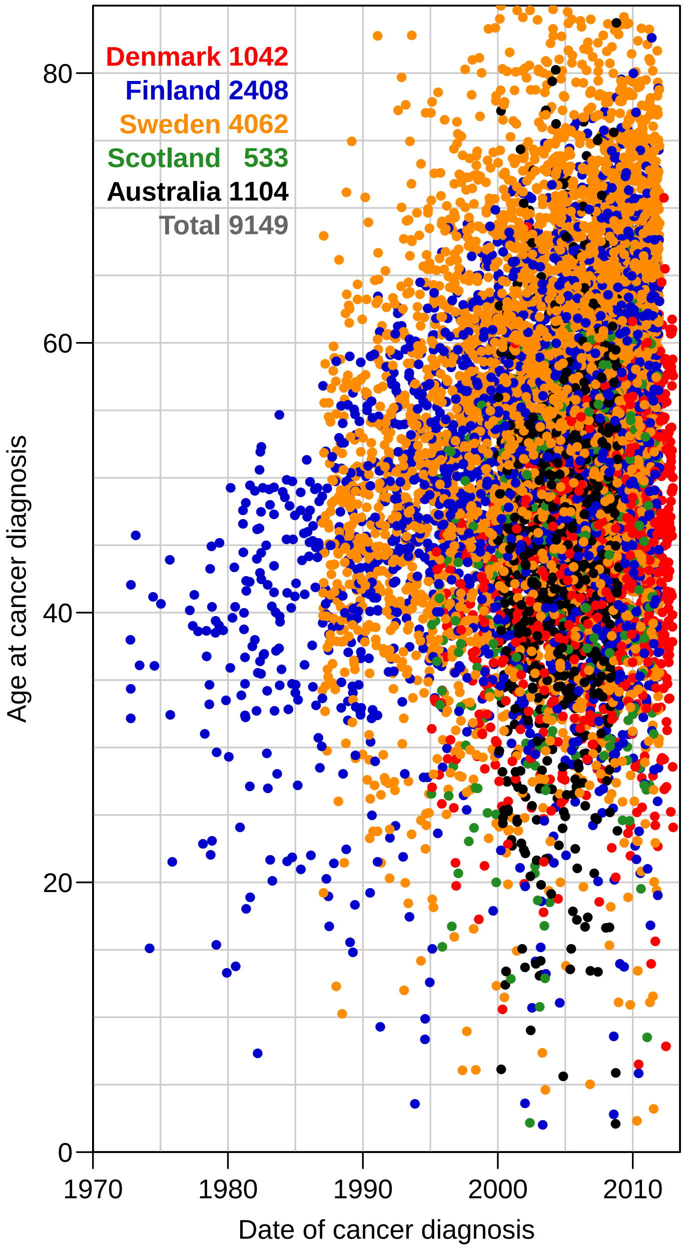

Figure 8.2: *Lexis diagram with a dot for each cancer case in T1D patients by age and date at diagnosis of cancer.*

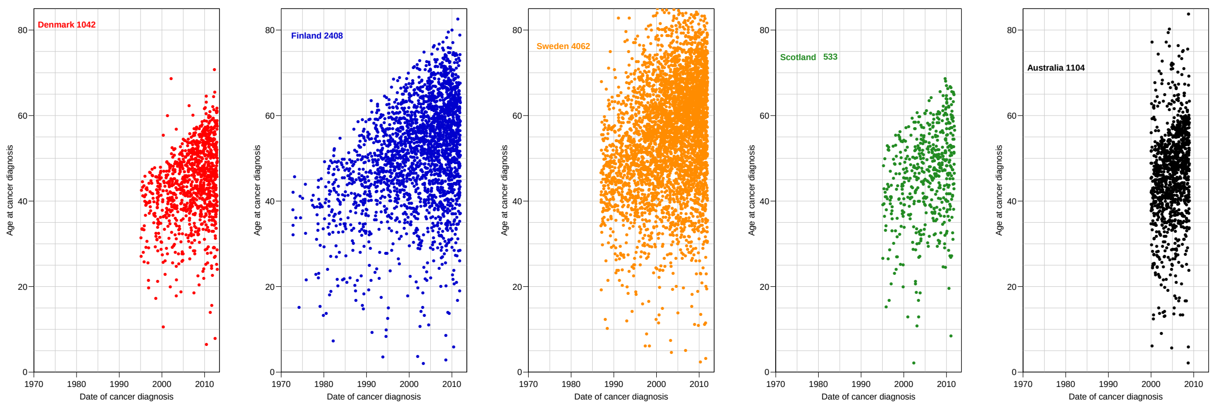

Figure 8.3: *Lexis diagrams for the single countries, showing the cancer occurrences by date and age at diagnosis of cancer.*

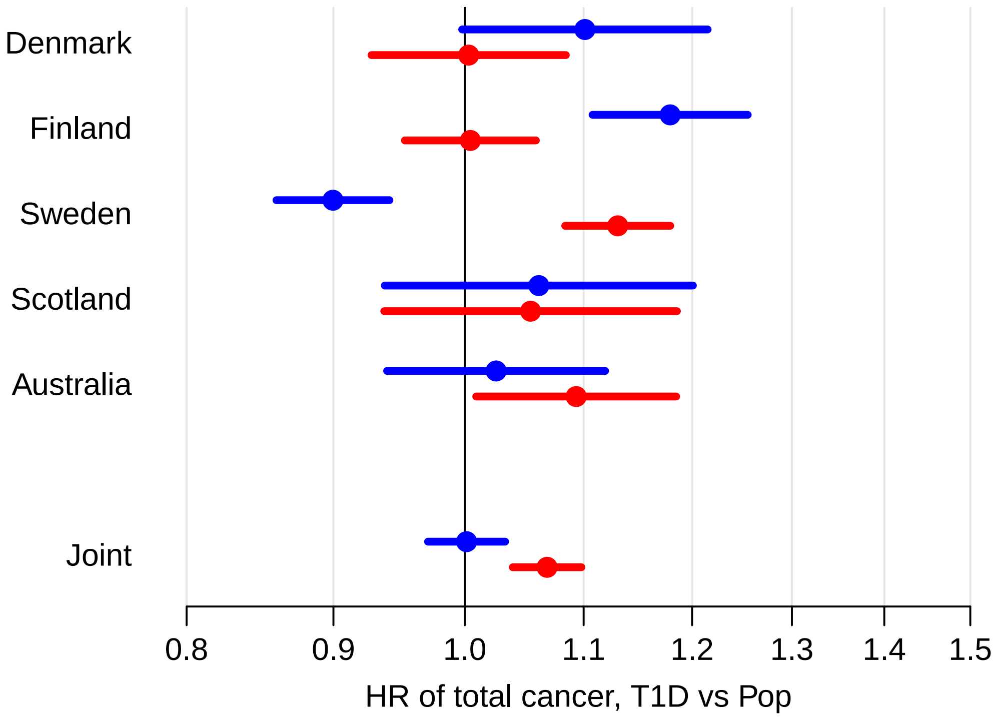

Figure 8.4: *HR of any cancer for men (blue) and women (red), separately for each country, and jointly. There is a strong inhomogeneity between countries for men, but not for women (P=*

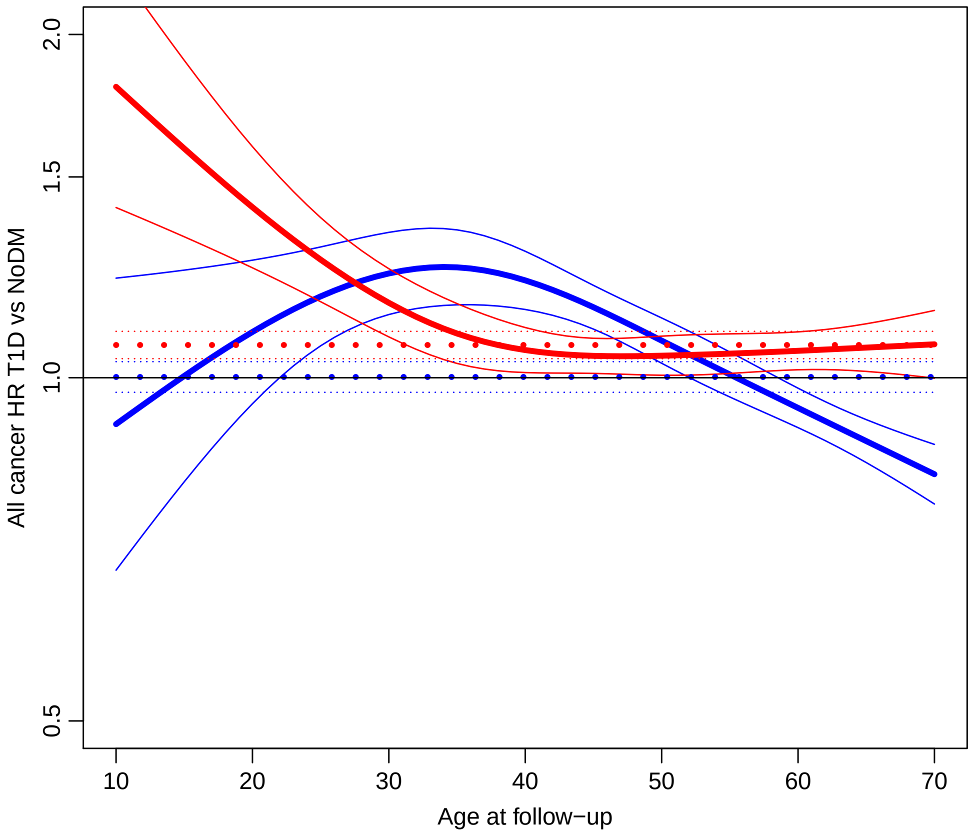

Figure 8.5: *Age-interaction of the T1D associated HR relative to the population; blue lines are men, red women. The dotted lines indicate the estimated HRs from the model without interaction.*

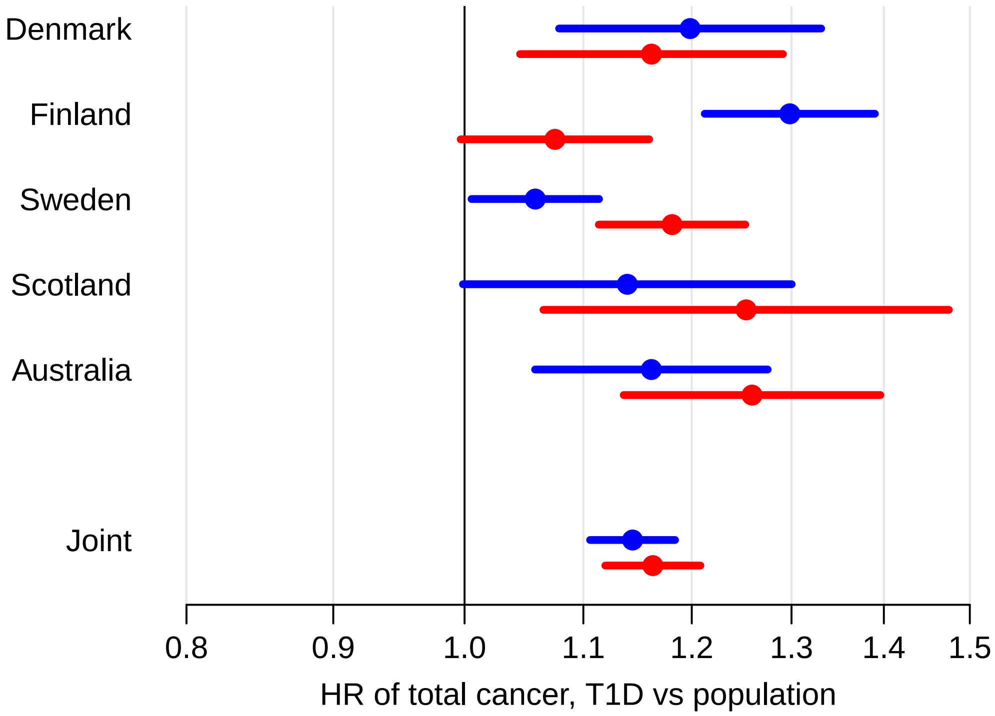

Figure 8.6: *HR of the non-sex-specific cancers for men (blue) and women (red), separately for each country, and jointly. There is a strong inhomogeneity between countries for men, but not for women (P=*

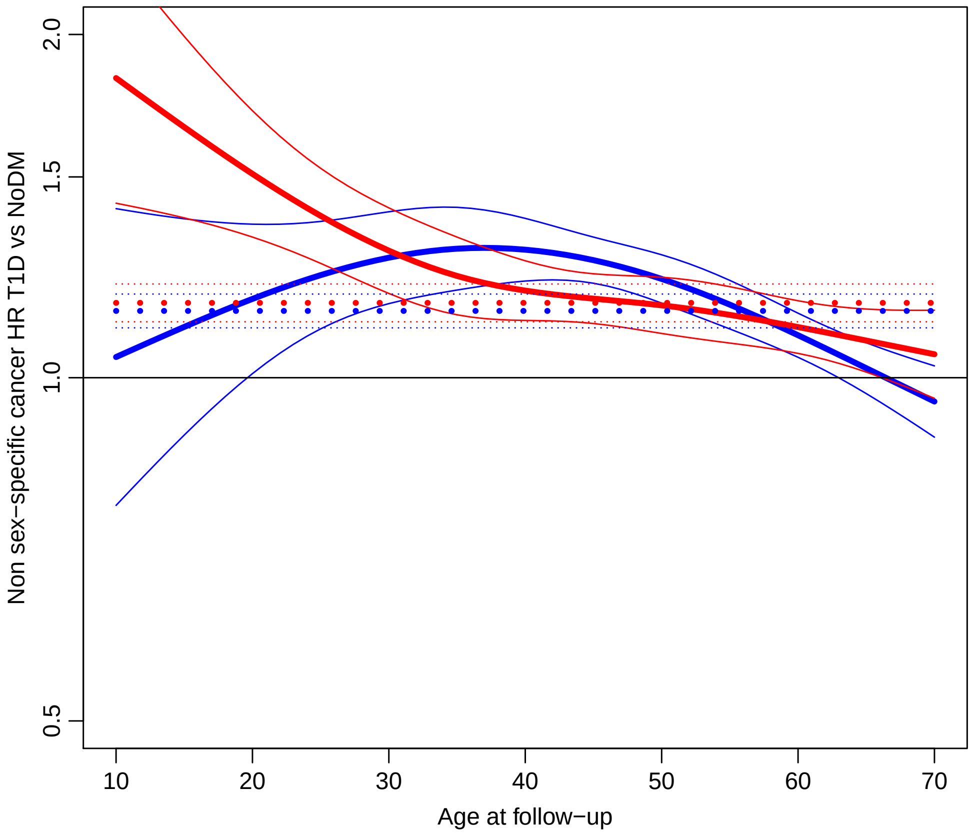

Figure 8.7: *Age-interaction of the T1D associated HR for non-sex-specific cancers relative to the population; blue lines are men, red women. The dotted lines indicate the estimate HRs from the model without interaction.*

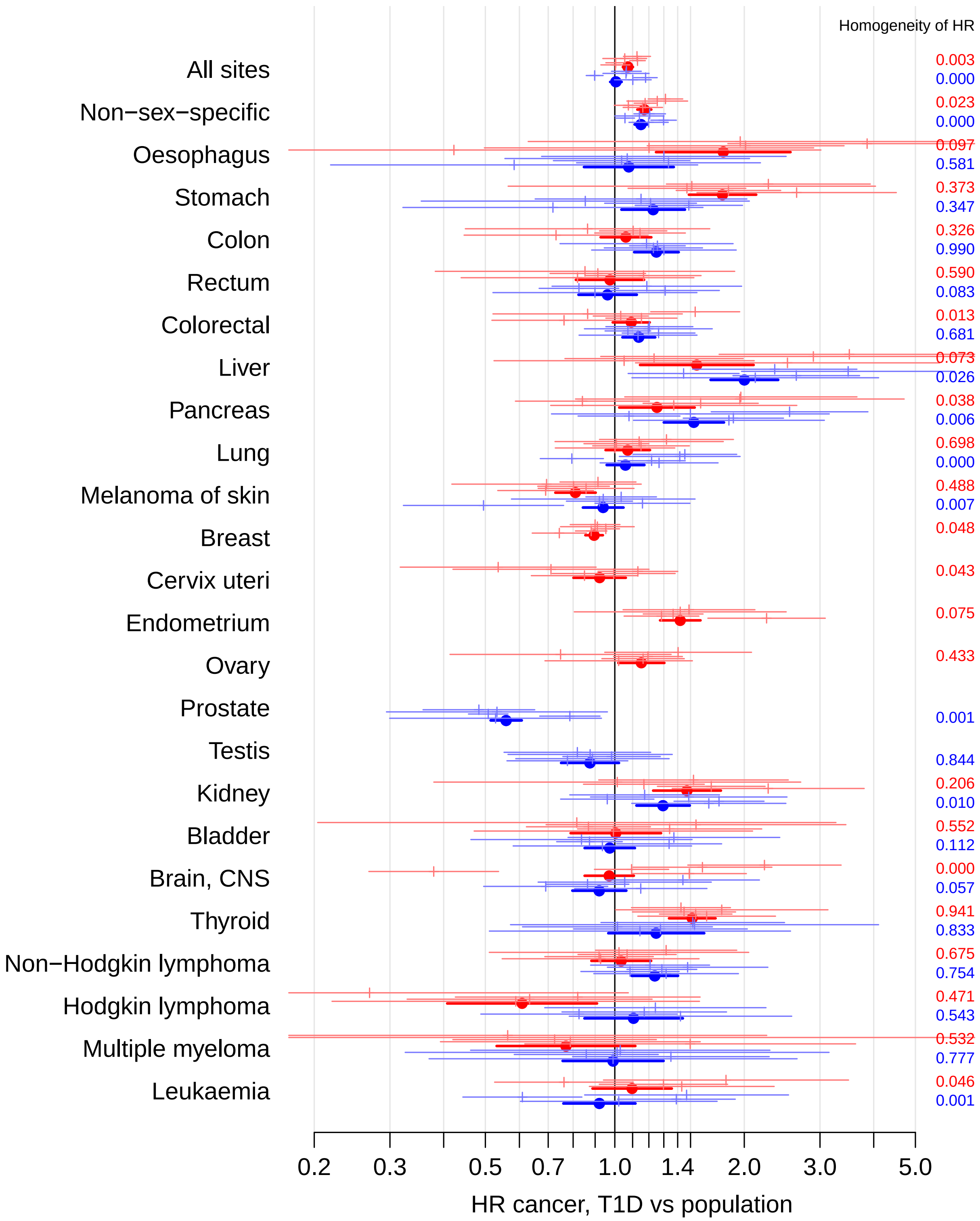

Figure 8.8: *HR of different cancers from the joint analysis (thick, full color), and the interaction with country (thin, pale color, ordered bottom-up as: DK, FI, SE, SC, AU).*

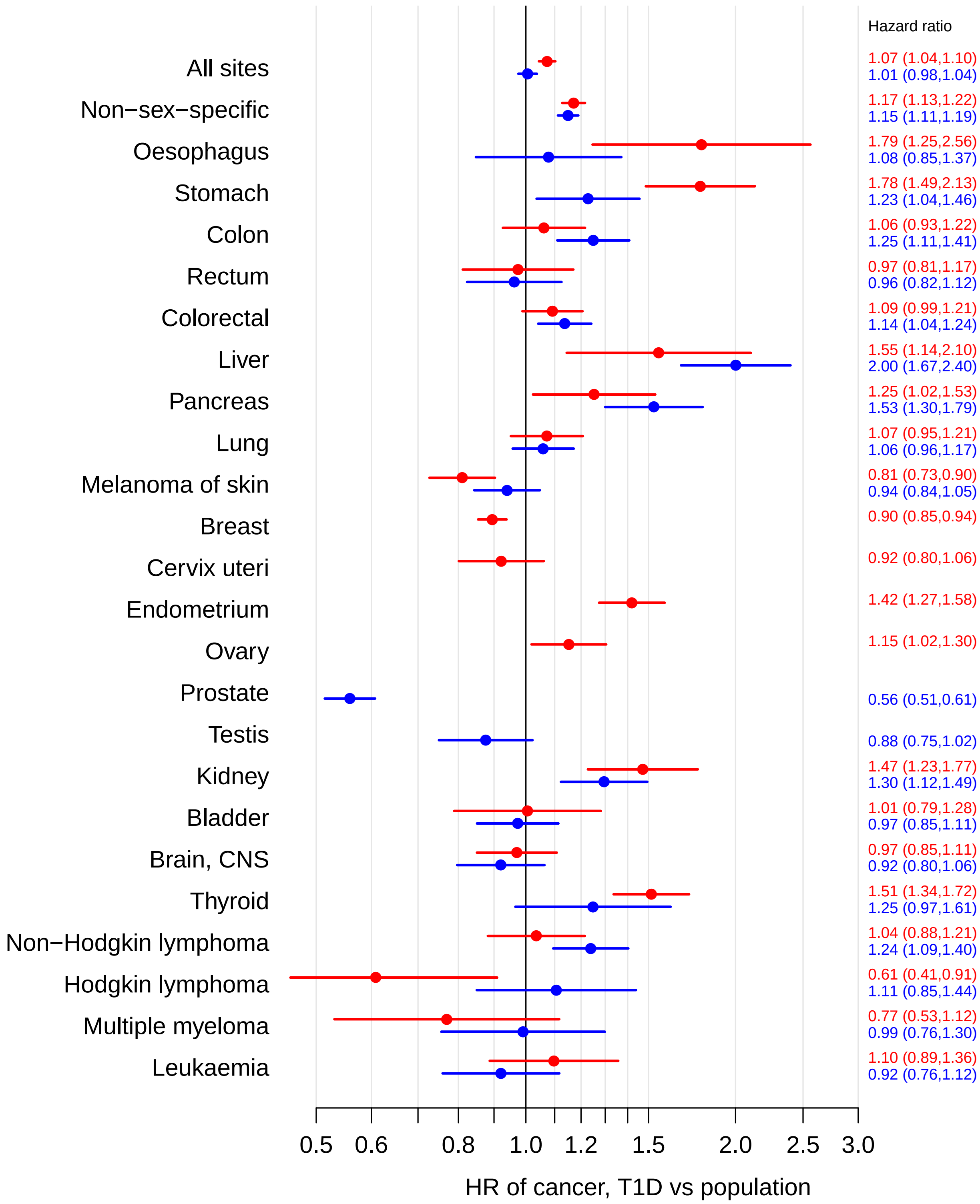

Figure 8.9: *HR of different cancers from the joint analysis.*

8.6 Analysis by diabetes duration

The models used for this are quite similar to those used above, the difference is that the dataset used is a bit smaller and that the exposure is on several levels, namely the duration groups.

First we reload the analysis dataset and subset it to the patients with known duration of DM:

```
> library( Epi )
> library( splines )
> options( width=90 )
> clear()
> load( file = "../data/ALLana.Rda" )
> lls()
  name    mode      class      size
1 all.ana list      data.frame 95859 35
2 conv    list      data.frame 42 3
3 wh.ca   character character  25
>  DM.sum <- function(x) sum(x[-1])
> Dur.sum <- function(x) sum(x[-c(1,length(x))])
> round( ftable( addmargins( xtabs( cbind( All.Ca=d0, PYears=y0/1000 ) ~ DMdur + sex, data=all.ana )
+                            margin=1:2,
+                            FUN=list(list(Dur.sum,DM.sum,sum),sum ) ),
+                 col.vars=3:2 ) )
Margins computed over dimensions
in the following order:
1: DMdur
2: sex
             All.Ca                  PYears
        sex       M       F     sum       M       F     sum
DMdur
NoDM        1754141 1586500 3340641  383045  390712  773757
0                98     158     256      92      95     187
1                57      77     134      94      96     190
2               153     256     409     259     264     523
5               349     477     826     353     352     705
10              374     564     938     264     255     519
15             1340    1651    2991     465     454     919
30             1090    1148    2238     160     155     315
Unkn            579     778    1357     287     288     575
Dur.sum        3461    4331    7792    1688    1670    3358
DM.sum         4040    5109    9149    1975    1958    3933
sum         1758181 1591609 3349790  385020  392670  777690
> # Cases and person years per year of F.U.
> round( ftable( xx <- addmargins( xtabs( cbind( All.Ca=d0, PYears=y0/1000 ) ~ DMdur + Cnt, data=all
+                            margin=1:2,
+                            FUN=list(list(Dur.sum,DM.sum,sum),sum ) ),
+                 col.vars=3:2 ) )
Margins computed over dimensions
in the following order:
1: DMdur
2: Cnt
             All.Ca                                          PYears
        Cnt      DK      FI      SE      SC      AU     sum      DK      FI      SE      SC      AU
DMdur
NoDM         442530  707026  959456  404707  826922 3340641   95492  198320  217267   85184  177493
0                45      77      86      25      23     256      35      73      54       8      17
1                26      48      34      17       9     134      33      70      52      16      18
2                99     123     118      32      37     409      85     191     149      45      53
5               165     252     287      63      59     826      97     255     229      69      54
10              131     317     357     106      27     938      47     191     201      68      12
15               26    1100    1498     290      77    2991       7     330     448     118      16
30                0     491    1682       0      65    2238       0      74     235       0       6
Unkn            550       0       0       0     807    1357     240       0       0       0     335
```

```
Dur.sum         492    2408    4062     533     297    7792     304    1185    1369     324     176
DM.sum         1042    2408    4062     533    1104    9149     545    1185    1369     324     511
sum          443572  709434  963518  405240  828026 3349790   96037  199505  218636   85508  178004
> print.table( round( xx[,,1]/xx[,,2], 2 ), na.print="." )
         Cnt
DMdur       DK    FI    SE    SC    AU   sum
  NoDM    4.63  3.57  4.42  4.75  4.66  4.32
  0       1.30  1.05  1.59  3.06  1.39  1.37
  1       0.78  0.68  0.65  1.08  0.50  0.71
  2       1.17  0.64  0.79  0.71  0.70  0.78
  5       1.69  0.99  1.25  0.91  1.10  1.17
  10      2.79  1.66  1.77  1.56  2.17  1.81
  15      3.59  3.34  3.34  2.46  4.77  3.25
  30         .  6.63  7.17     . 10.26  7.11
  Unkn    2.29     .     .     .  2.41  2.36
  Dur.sum 1.62  2.03  2.97  1.64  1.69  2.32
  DM.sum  1.91  2.03  2.97  1.64  2.16  2.33
  sum     4.62  3.56  4.41  4.74  4.65  4.31
```

We then restrict the dataset to those where `DMdur` is *not* `Unkn`

```
> all.ana <- transform( subset( all.ana, DMdur != "Unkn" ),
+                       T1D = factor( T1D ),
+                     DMdur = factor( DMdur ) )
> # A lot of bells and whistles to get nicely formatted numbers in the table
> suppressWarnings( noquote( formatC(
+ addmargins( xtabs( d0 ~ sex + T1D, data = all.ana ),
+             margin=1:2,
+             FUN=list(sum,list(DM.sum,sum)), quiet=TRUE ),
+          format="f", digits=0, big.mark=",", width=6 ) ) )
     T1D
sex   NoDM      30      35     40     DM.sum sum
  M   1,754,141 1,266     733  1,462  3,461 1,757,602
  F   1,586,500 1,894  1,041  1,396  4,331 1,590,831
  sum 3,340,641 3,160  1,774  2,858  7,792 3,348,433
> ftable( xtabs( d0 ~ DMdur + Cnt + T1D, data = all.ana ),
+         col.vars=1 )
         DMdur   NoDM      0      1      2      5     10     15     30
Cnt T1D
DK  NoDM       442530      0      0      0      0      0      0      0
    30              0      9      4     24     39     29      8      0
    35              0      9      7     21     37     45      5      0
    40              0     27     15     54     89     57     13      0
FI  NoDM       707026      0      0      0      0      0      0      0
    30              0     25     10     30     62     64    402    272
    35              0     14     16     35     59     92    293    103
    40              0     38     22     58    131    161    405    116
SE  NoDM       959456      0      0      0      0      0      0      0
    30              0     27      9     32     73     93    490   1154
    35              0     19      9     27     67     95    386    239
    40              0     40     16     59    147    169    622    289
SC  NoDM       404707      0      0      0      0      0      0      0
    30              0      7      6      8     16     37    113      0
    35              0      7      4      4     13     24     72      0
    40              0     11      7     20     34     45    105      0
AU  NoDM       826922      0      0      0      0      0      0      0
    30              0     17      3     19     19      3     21     35
    35              0      5      1      6     18      6     22     14
    40              0      1      5     12     22     18     34     16
```

With this reduced dataset we can now do the analyses of all cancers by duration of diabetes. We only do the analyses for the major sites, that is those with more than 240 cases among DM patients of either sex:

```
> wh <- names(all.ana)[which( substr(names(all.ana),1,1)=="d" )]
> ad <- all.ana[all.ana$T1D!="NoDM",c("sex","DMdur",wh)]
> ( nc <- xtabs( as.matrix(ad[,-(1:2)]) ~ ad$sex ) )
ad$sex   d0   d6   d7   d9  d10  d11  d13  d16  d18  d19  d20  d22  d24  d25  d27  d28
     M 3052   47  115  239  133   81  112  265    6    0    0    0  492  101  139  178
     F 3910   19   95  191  102   30   74  198 1335  149  266  207    0    0   96   56

ad$sex  d29  d32  d33  d36  d37  d38  d40  d52  d00
     M  191  131   36  174   43   42   83  372 2458
     F  197  158  173  109   22   24   68  293 1987
> # Those with more than 135 in the sex with most PLUS liver and pancreas
> wh <- c( colnames(nc)[apply( nc, 2, max )>135], "d11","d13" )
> wh <- conv[match(wh,conv$NCnam),]    ; dim(wh)
[1] 18  3
> xcl <- match(c("Colon","Rectum"),wh$Clab)
> xcl <- xcl[!is.na(xcl)]
> wh <- wh[-xcl,] ; dim(wh)
[1] 17  3
> res.var <- wh[,"NCnam"]
> res.nam <- wh[,"Clab"]
> res.nam[grep("Cerv",res.nam)] <- "Cervix"
> res.nam[grep("Corp",res.nam)] <- "Endometrium"
> res.nam[grep("Mela",res.nam)] <- "Melanoma"
> nc <- nc[,res.var]
> data.frame( res.var, res.nam, t(nc), wh )
    res.var              res.nam    M    F DKnam NCnam                 Clab
d0       d0            All sites 3052 3910    d0    d0            All sites
d16     d16                 Lung  265  198   d33   d16                 Lung
d18     d18               Breast    6 1335   d70   d18               Breast
d19     d19               Cervix    0  149   d82   d19         Cervix uteri
d20     d20          Endometrium    0  266   d83   d20         Corpus uteri
d22     d22                Ovary    0  207   d84   d22                Ovary
d24     d24             Prostate  492    0   d91   d24             Prostate
d27     d27               Kidney  139   96  d101   d27               Kidney
d28     d28              Bladder  178   56  d103   d28              Bladder
d29     d29             Melanoma  191  197   d51   d29     Melanoma of skin
d32     d32          Brain, CNS  131  158  d113   d32           Brain, CNS
d33     d33              Thyroid   36  173  d121   d33              Thyroid
d36     d36 Non-Hodgkin lymphoma  174  109  d132   d36 Non-Hodgkin lymphoma
d52     d52           Colorectal  372  293  d251   d52           Colorectal
d00     d00     Non-sex-specific 2458 1987         d00     Non-sex-specific
d11     d11                Liver   81   30   d26   d11                Liver
d13     d13             Pancreas  112   74   d28   d13             Pancreas
```

With this selection of sites we can inspect the availability of cancer cases by duration:

```
> dur.cases <- xtabs( as.matrix(ad[,res.var]) ~ DMdur + sex, data=ad )
> dimnames( dur.cases )[[3]] <- res.nam
> ftable( dur.cases, row.vars=3:2, zero.print="." )
                        DMdur NoDM    0    1    2    5   10   15   30
                    sex
All sites           M            0   98   57  153  349  374 1340 1090
                    F            0  158   77  256  477  564 1651 1148
Lung                M            0    1    0    5   30   36  159   86
                    F            0    4    1    3   10   10  112   83
Breast              M            0    0    0    0    0    1    3    3
                    F            0   16   18   68  135  193  609  430
Cervix              M            0    0    0    0    0    0    0    0
                    F            0    9    4   19   37   30   52   13
Endometrium         M            0    0    0    0    0    0    0    0
                    F            0   11    2   15   29   37   97   92
Ovary               M            0    0    0    0    0    0    0    0
                    F            0    5    2   20   28   34   83   53
Prostate            M            0    0    0    1    5   15  177  319
                    F            0    0    0    0    0    0    1    0
```

```
Kidney              M         0    4    3    4   21   28   71   29
                    F         0    2    2    8    9   18   47   17
Bladder             M         0    4    2    4   16   17   83   72
                    F         0    3    0    2    9    4   20   26
Melanoma            M         0    6    4   10   31   32   86   60
                    F         0    6    6   23   33   41   62   51
Brain, CNS          M         0   14    5   12   25   23   46   27
                    F         0   17   13   14   33   22   61   32
Thyroid             M         0    4    1    5   11    7   12    5
                    F         0   16    9   20   42   37   57   20
Non-Hodgkin lymphoma M        0   11    6   13   22   29   65   60
                    F         0    3    0   10   10   22   49   27
Colorectal          M         0    9    3   13   35   33  162  165
                    F         0    7    2    7   21   33  135  113
Non-sex-specific    M         0   96   47  131  322  334 1121  764
                    F         0  117   53  138  251  274  825  569
Liver               M         0    2    1    1    9   13   50   21
                    F         0    0    0    4    3    7   13    9
Pancreas            M         0   13    3    7   14    5   60   29
                    F         0    9    0    2    6   10   31   25
```

As before we set up the splines etc.

```
> a.kn <- seq(10,80,,8)
> # Period knots
> p.dk <- seq(1996,2011,,3)
> p.fi <- seq(1975,2007,,5)
> p.se <- seq(1988,2007,,4)
> p.sc <- seq(1996,2010,,2)
> p.au <- seq(1997,2007,,2)
> # Cohort knots
> c.dk <- seq(1920,1985,,5)
> c.fi <- seq(1900,1985,,7)
> c.se <- seq(1910,1985,,6)
> c.sc <- seq(1920,1985,,4)
> c.au <- seq(1920,1985,,5)
```

With these knots for the splines we can set up the design matrices for the baseline-effects for each county:

```
> rnam <- function(M,pre){colnames(M)<-paste(pre,colnames(M),sep="");M}
> # Denmark
> M.dk <- with( DK.an <- subset(all.ana,Cnt=="DK"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.dk." ),
+                    rnam( Ns(P  ,knots=p.dk)                , "P.dk." ),
+                    rnam( Ns(P-A,knots=c.dk)                , "C.dk." )))
> # Finland
> M.fi <- with( FI.an <- subset(all.ana,Cnt=="FI"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.fi." ),
+                    rnam( Ns(P  ,knots=p.fi)                , "P.fi." ),
+                    rnam( Ns(P-A,knots=c.fi)                , "C.fi." )))
> # Sweden
> M.se <- with( SE.an <- subset(all.ana,Cnt=="SE"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.se." ),
+                    rnam( Ns(P  ,knots=p.se)                , "P.se." ),
+                    rnam( Ns(P-A,knots=c.se)                , "C.se." )))
> # Scotland
> M.sc <- with( SC.an <- subset(all.ana,Cnt=="SC"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.sc." ),
+                    rnam( Ns(P  ,knots=p.sc)                , "P.sc." ),
+                    rnam( Ns(P-A,knots=c.sc)                , "C.sc." )))
> # Australia
> M.au <- with( AU.an <- subset(all.ana,Cnt=="AU"),
+              cbind(rnam( Ns(  A,knots=a.kn, intercept=TRUE), "A.au." ),
+                    rnam( Ns(P  ,knots=p.au)                , "P.au." ),
+                    rnam( Ns(P-A,knots=c.au)                , "C.au." )))
```

```
> # Overview
> addmargins( rbind( DK=dim(M.dk),
+                    FI=dim(M.fi),
+                    SE=dim(M.se),
+                    SC=dim(M.sc),
+                    AU=dim(M.au) ), margin=1 )
     [,1] [,2]
DK  10089   14
FI  30538   18
SE  24345   16
SC  14706   12
AU   8097   13
Sum 87775   73
```

With these model matrices in place we can now set up the total model matrix for the baseline rates, basically putting the 5 model matrices diagonally as sub-matrices surrounded by 0s:

```
> c.dk <- ncol(M.dk) ; r.dk <- nrow(M.dk)
> c.fi <- ncol(M.fi) ; r.fi <- nrow(M.fi)
> c.se <- ncol(M.se) ; r.se <- nrow(M.se)
> c.sc <- ncol(M.sc) ; r.sc <- nrow(M.sc)
> c.au <- ncol(M.au) ; r.au <- nrow(M.au)
> MB <- rbind( cbind(                 M.dk, matrix(0,r.dk,c.fi+c.se+c.sc+c.au) ),
+              cbind( matrix(0,r.fi,c.dk), M.fi, matrix(0,r.fi,c.se+c.sc+c.au) ),
+              cbind( matrix(0,r.se,c.dk+c.fi), M.se, matrix(0,r.se,c.sc+c.au) ),
+              cbind( matrix(0,r.sc,c.dk+c.fi+c.se), M.sc, matrix(0,r.sc,c.au) ),
+              cbind( matrix(0,r.au,c.dk+c.fi+c.se+c.sc), M.au                ) )
> dim( MB )
[1] 87775    73
> colnames( MB ) <- c( colnames(M.dk),
+                      colnames(M.fi),
+                      colnames(M.se),
+                      colnames(M.sc),
+                      colnames(M.au) )
```

Since we have the base model matrix set up as one, but ultimately we will be doing analyses by sex, so we must also have it available subdivided by sex:

```
> MB.m <- MB[all.ana$sex=="M",]
> MB.f <- MB[all.ana$sex=="F",]
```

8.6.1 Statistical models

In order to do the analyses for all cancers and the 13 sub-sites chosen we set up an array to hold the results:

```
> Edrr <- NArray( list( site = res.var,
+                        sex = levels( all.ana$sex ),
+                    country = c(levels( all.ana$Cnt ),"Joint"),
+                        dur = levels( all.ana$DMdur )[-1],
+                         wh = c("HR","lo","up") ) )
> Tdrr <- NArray( c( dimnames(Edrr)[1:2],
+                    list( wh = c("Chisq","P","conv S","conv J") ) ) )
> str( Edrr )
 logi [1:17, 1:2, 1:6, 1:7, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ site   : chr [1:17] "d0" "d16" "d18" "d19" ...
  ..$ sex    : chr [1:2] "M" "F"
  ..$ country: chr [1:6] "DK" "FI" "SE" "SC" ...
  ..$ dur    : chr [1:7] "0" "1" "2" "5" ...
  ..$ wh     : chr [1:3] "HR" "lo" "up"
```

```
> str( Tdrr )
 logi [1:17, 1:2, 1:4] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ site: chr [1:17] "d0" "d16" "d18" "d19" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ wh  : chr [1:4] "Chisq" "P" "conv S" "conv J"
```

With these arrays in place we can fit models for each site separately in a loop over the rest of the cancer sites; note that we now use `y` instead of `y0` in the offset expression:

```
> id <- dimnames(Edrr)[[1]][1]
> for( id in dimnames(Edrr)[[1]] )
+    {
+ cat( id, format( Sys.time(), format="%H:%M:%S" ), "\n" )
+ flush.console()
+    all.ana$D <- all.ana[,id]
+ m0.i <- glm( D ~ MB.m + Cnt + DMdur:Cnt,
+              family = poisson,
+              offset = log(if(id=="d0") y0 else y),
+                data = subset(all.ana,sex=="M") )
+ m0.j <- update( m0.i, . ~ . - DMdur:Cnt + DMdur )
+ f0.i <- glm( D ~ MB.f + Cnt + DMdur:Cnt,
+              family = poisson,
+              offset = log(if(id=="d0") y0 else y),
+                data = subset(all.ana,sex=="F") )
+ f0.j <- update( f0.i, . ~ . - DMdur:Cnt + DMdur )
+ # Test for homogeneity
+ Tdrr[id,"M",] <- c( as.numeric( anova( m0.j, m0.i, test="Chisq" )[2,4:5] ),
+                     m0.i$converged, m0.j$converged )
+ Tdrr[id,"F",] <- c( as.numeric( anova( f0.j, f0.i, test="Chisq" )[2,4:5] ),
+                     f0.i$converged, f0.j$converged )
+ # HR estimates
+ for( cnt in levels(all.ana$Cnt) )
+    {
+ Edrr[id,"M",cnt,,] <- ci.exp( m0.i, subint=c("DMdur",cnt) )
+ Edrr[id,"F",cnt,,] <- ci.exp( f0.i, subint=c("DMdur",cnt) )
+    }
+ Edrr[id,"M","Joint",,] <- ci.exp( m0.j, subset="DMdur" )
+ Edrr[id,"F","Joint",,] <- ci.exp( f0.j, subset="DMdur" )
+ cat( id, format( Sys.time(), format="%H:%M:%S" ), "\n" )
+ flush.console()
+     }
d0 19:12:10
d0 19:12:29
d16 19:12:29
d16 19:13:11
d18 19:13:11
d18 19:14:00
d19 19:14:00
d19 19:14:52
d20 19:14:52
d20 19:15:44
d22 19:15:44
d22 19:16:34
d24 19:16:34
d24 19:17:28
d27 19:17:28
d27 19:18:07
d28 19:18:07
d28 19:18:50
d29 19:18:50
d29 19:19:28
d32 19:19:28
d32 19:20:04
d33 19:20:04
d33 19:20:41
```

```
d36 19:20:41
d36 19:21:22
d52 19:21:22
d52 19:22:00
d00 19:22:00
d00 19:22:19
d11 19:22:19
d11 19:23:02
d13 19:23:02
d13 19:23:48
> Tdrr["d18","M",] <- NA
> Tdrr["d19","M",] <- NA
> Tdrr["d20","M",] <- NA
> Tdrr["d22","M",] <- NA
> Tdrr["d24","F",] <- NA
> round( ftable( Tdrr[,,] ), 3 )
         wh  Chisq      P conv S conv J
site sex
d0   M      56.225  0.001  1.000  1.000
     F      65.053  0.000  1.000  1.000
d16  M      26.668  0.427  1.000  1.000
     F      23.478  0.606  1.000  1.000
d18  M          NA     NA     NA     NA
     F      20.818  0.751  1.000  1.000
d19  M          NA     NA     NA     NA
     F      40.092  0.038  1.000  1.000
d20  M          NA     NA     NA     NA
     F      31.904  0.196  1.000  1.000
d22  M          NA     NA     NA     NA
     F      33.241  0.155  1.000  1.000
d24  M      36.670  0.080  1.000  1.000
     F          NA     NA     NA     NA
d27  M      34.435  0.124  1.000  1.000
     F      25.787  0.475  1.000  1.000
d28  M      37.871  0.062  1.000  1.000
     F      24.075  0.572  1.000  1.000
d29  M      21.951  0.691  1.000  1.000
     F      28.617  0.329  1.000  1.000
d32  M      43.507  0.017  1.000  1.000
     F      81.518  0.000  1.000  1.000
d33  M      29.809  0.276  1.000  1.000
     F      37.111  0.073  1.000  1.000
d36  M      26.303  0.447  1.000  1.000
     F      35.345  0.104  1.000  1.000
d52  M      25.453  0.493  1.000  1.000
     F      34.288  0.128  1.000  1.000
d00  M      38.326  0.056  1.000  1.000
     F      59.598  0.000  1.000  1.000
d11  M      25.328  0.500  1.000  1.000
     F      18.051  0.874  1.000  1.000
d13  M      33.130  0.158  1.000  1.000
     F      25.717  0.479  1.000  1.000
> save( Edrr, Tdrr, file="../data/ArrD.Rda" )
```

```
> load( file="../data/ArrD.Rda" )
> dimnames( Edrr )[[1]] <- res.nam
> str( Edrr )
 num [1:17, 1:2, 1:6, 1:7, 1:3] 1.75 3.43 9.06e-08 2.17e+03 2.17e+03 ...
 - attr(*, "dimnames")=List of 5
  ..$ site   : chr [1:17] "All sites" "Lung" "Breast" "Cervix" ...
  ..$ sex    : chr [1:2] "M" "F"
  ..$ country: chr [1:6] "DK" "FI" "SE" "SC" ...
  ..$ dur    : chr [1:7] "0" "1" "2" "5" ...
  ..$ wh     : chr [1:3] "HR" "lo" "up"
```

```
> str( Tdrr )
 num [1:17, 1:2, 1:4] 56.2 26.7 NA NA NA ...
 - attr(*, "dimnames")=List of 3
  ..$ site: chr [1:17] "d0" "d16" "d18" "d19" ...
  ..$ sex : chr [1:2] "M" "F"
  ..$ wh  : chr [1:4] "Chisq" "P" "conv S" "conv J"
> Edrr[c("Breast","Cervix","Endometrium","Ovary"),"M",,,] <- NA
> Edrr["Prostate","F",,,] <- NA
> Edrr <- ifelse( Edrr>10^2 | Edrr<10^-4, NA, Edrr )
> round( ftable( Edrr[,,"Joint",,], row.vars=c(1,3) ), 2 )
                         sex     M                 F
                         wh     HR    lo    up    HR    lo    up
site                     dur
All sites                0    2.28  1.87  2.78  2.34  2.00  2.74
                         1    1.23  0.95  1.60  1.03  0.83  1.29
                         2    1.07  0.91  1.25  1.04  0.92  1.17
                         5    1.34  1.20  1.49  1.01  0.93  1.11
                         10   1.24  1.12  1.37  1.12  1.03  1.22
                         15   1.02  0.97  1.07  1.03  0.99  1.09
                         30   0.81  0.76  0.86  1.07  1.01  1.14
Lung                     0    0.99  0.14  7.00  4.59  1.72 12.24
                         1      NA    NA    NA  0.94  0.13  6.68
                         2    0.98  0.41  2.36  0.69  0.22  2.14
                         5    2.02  1.41  2.90  0.81  0.43  1.50
                         10   1.43  1.03  1.99  0.53  0.28  0.98
                         15   1.20  1.02  1.40  1.26  1.04  1.51
                         30   0.65  0.53  0.81  0.98  0.79  1.22
Breast                   0      NA    NA    NA  0.89  0.54  1.45
                         1      NA    NA    NA  0.84  0.53  1.33
                         2      NA    NA    NA  0.83  0.66  1.06
                         5      NA    NA    NA  0.72  0.61  0.86
                         10     NA    NA    NA  0.87  0.76  1.01
                         15     NA    NA    NA  0.87  0.81  0.95
                         30     NA    NA    NA  1.08  0.98  1.19
Cervix                   0      NA    NA    NA  1.19  0.62  2.29
                         1      NA    NA    NA  0.50  0.19  1.32
                         2      NA    NA    NA  0.80  0.51  1.26
                         5      NA    NA    NA  1.08  0.78  1.50
                         10     NA    NA    NA  1.16  0.81  1.66
                         15     NA    NA    NA  1.04  0.79  1.37
                         30     NA    NA    NA  0.67  0.39  1.15
Endometrium              0      NA    NA    NA 13.67  7.54 24.78
                         1      NA    NA    NA  2.11  0.53  8.45
                         2      NA    NA    NA  4.03  2.43  6.71
                         5      NA    NA    NA  2.85  1.98  4.10
                         10     NA    NA    NA  2.20  1.59  3.04
                         15     NA    NA    NA  1.05  0.86  1.28
                         30     NA    NA    NA  1.17  0.95  1.43
Ovary                    0      NA    NA    NA  1.89  0.79  4.54
                         1      NA    NA    NA  0.67  0.17  2.69
                         2      NA    NA    NA  1.98  1.28  3.07
                         5      NA    NA    NA  1.35  0.93  1.96
                         10     NA    NA    NA  1.39  0.99  1.94
                         15     NA    NA    NA  1.03  0.83  1.27
                         30     NA    NA    NA  1.05  0.80  1.38
Prostate                 0      NA    NA    NA    NA    NA    NA
                         1      NA    NA    NA    NA    NA    NA
                         2    1.00  0.14  7.11    NA    NA    NA
                         5    0.76  0.32  1.82    NA    NA    NA
                         10   0.65  0.39  1.07    NA    NA    NA
                         15   0.52  0.45  0.60    NA    NA    NA
                         30   0.59  0.53  0.66    NA    NA    NA
Kidney                   0    3.65  1.37  9.75  2.82  0.70 11.27
                         1    2.42  0.78  7.51  2.60  0.65 10.39
                         2    0.91  0.34  2.43  3.11  1.55  6.24
                         5    2.07  1.35  3.18  1.67  0.87  3.22
```

	10	2.09	1.44	3.03	2.68	1.68	4.25
	15	1.40	1.11	1.77	1.65	1.24	2.20
	30	0.71	0.49	1.03	0.72	0.45	1.16
Bladder	0	3.98	1.49	10.63	8.39	2.70	26.14
	1	1.72	0.43	6.87	NA	NA	NA
	2	0.96	0.36	2.57	1.38	0.35	5.54
	5	1.62	0.99	2.64	2.77	1.44	5.34
	10	1.16	0.72	1.87	0.89	0.33	2.37
	15	1.12	0.90	1.39	0.89	0.58	1.38
	30	0.76	0.60	0.96	1.00	0.68	1.47
Melanoma	0	1.08	0.48	2.40	0.63	0.28	1.40
	1	0.65	0.24	1.73	0.58	0.26	1.29
	2	0.51	0.28	0.95	0.71	0.47	1.07
	5	0.93	0.65	1.32	0.63	0.45	0.89
	10	1.04	0.74	1.47	0.97	0.71	1.31
	15	1.01	0.82	1.25	0.68	0.53	0.87
	30	0.93	0.72	1.20	0.98	0.74	1.28
Brain, CNS	0	2.63	1.55	4.44	2.84	1.77	4.57
	1	0.90	0.37	2.16	2.07	1.20	3.57
	2	0.74	0.42	1.31	0.75	0.45	1.27
	5	0.97	0.65	1.43	1.11	0.79	1.57
	10	0.93	0.62	1.41	0.78	0.52	1.19
	15	0.73	0.55	0.98	0.77	0.60	0.99
	30	0.77	0.53	1.12	0.77	0.54	1.09
Thyroid	0	3.56	1.33	9.48	3.29	2.01	5.38
	1	0.83	0.12	5.89	1.77	0.92	3.41
	2	1.35	0.56	3.26	1.30	0.84	2.02
	5	1.82	1.01	3.29	1.73	1.28	2.34
	10	1.31	0.63	2.76	1.85	1.34	2.56
	15	0.96	0.54	1.69	1.34	1.03	1.74
	30	0.77	0.32	1.85	1.29	0.83	2.00
Non-Hodgkin lymphoma	0	3.60	1.99	6.50	1.48	0.48	4.59
	1	1.81	0.81	4.03	NA	NA	NA
	2	1.26	0.73	2.17	1.45	0.78	2.70
	5	1.20	0.79	1.82	0.80	0.43	1.49
	10	1.49	1.04	2.15	1.65	1.09	2.51
	15	1.05	0.82	1.34	1.02	0.77	1.35
	30	1.20	0.93	1.55	0.76	0.52	1.10
Colorectal	0	3.85	2.00	7.41	2.82	1.34	5.93
	1	1.12	0.36	3.47	0.71	0.18	2.85
	2	1.36	0.79	2.34	0.70	0.33	1.47
	5	1.60	1.15	2.23	0.94	0.62	1.45
	10	1.11	0.79	1.56	1.19	0.84	1.67
	15	1.14	0.98	1.33	1.15	0.97	1.36
	30	1.00	0.86	1.16	0.95	0.79	1.14
Non-sex-specific	0	2.81	2.30	3.43	3.01	2.51	3.61
	1	1.26	0.95	1.68	1.28	0.98	1.67
	2	1.11	0.93	1.31	1.07	0.90	1.26
	5	1.44	1.29	1.61	1.12	0.99	1.27
	10	1.30	1.16	1.44	1.22	1.09	1.38
	15	1.16	1.09	1.23	1.11	1.04	1.19
	30	0.85	0.79	0.91	0.94	0.87	1.02
Liver	0	6.39	1.59	25.60	NA	NA	NA
	1	2.92	0.41	20.78	NA	NA	NA
	2	0.87	0.12	6.16	4.91	1.83	13.14
	5	3.33	1.73	6.42	1.84	0.59	5.71
	10	3.23	1.87	5.57	3.48	1.65	7.32
	15	2.51	1.90	3.31	1.44	0.84	2.49
	30	1.00	0.65	1.53	0.93	0.48	1.79
Pancreas	0	33.80	19.50	58.57	29.53	15.22	57.28
	1	6.43	2.07	19.99	NA	NA	NA
	2	3.81	1.81	8.01	1.47	0.37	5.88
	5	2.86	1.69	4.84	1.72	0.77	3.83
	10	0.67	0.28	1.61	1.94	1.04	3.61
	15	1.68	1.31	2.17	1.12	0.79	1.59
	30	0.83	0.58	1.19	0.87	0.59	1.29

```
> # As laid out in the ESM:
> round( ftable( Edrr[c(14, 2,10,12, 8),,"Joint",,], row.vars=c(2,3) ), 2 )
        site Colorectal             Lung             Melanoma             Thyroid             Kidne
        wh             HR    lo    up    HR    lo    up       HR    lo    up    HR    lo    up    H
sex dur
M   0                3.85  2.00  7.41  0.99  0.14  7.00     1.08  0.48  2.40    3.56  1.33  9.48  3.6
    1                1.12  0.36  3.47    NA    NA    NA     0.65  0.24  1.73    0.83  0.12  5.89  2.4
    2                1.36  0.79  2.34  0.98  0.41  2.36     0.51  0.28  0.95    1.35  0.56  3.26  0.9
    5                1.60  1.15  2.23  2.02  1.41  2.90     0.93  0.65  1.32    1.82  1.01  3.29  2.0
    10               1.11  0.79  1.56  1.43  1.03  1.99     1.04  0.74  1.47    1.31  0.63  2.76  2.0
    15               1.14  0.98  1.33  1.20  1.02  1.40     1.01  0.82  1.25    0.96  0.54  1.69  1.4
    30               1.00  0.86  1.16  0.65  0.53  0.81     0.93  0.72  1.20    0.77  0.32  1.85  0.7
F   0                2.82  1.34  5.93  4.59  1.72 12.24     0.63  0.28  1.40    3.29  2.01  5.38  2.8
    1                0.71  0.18  2.85  0.94  0.13  6.68     0.58  0.26  1.29    1.77  0.92  3.41  2.6
    2                0.70  0.33  1.47  0.69  0.22  2.14     0.71  0.47  1.07    1.30  0.84  2.02  3.1
    5                0.94  0.62  1.45  0.81  0.43  1.50     0.63  0.45  0.89    1.73  1.28  2.34  1.6
    10               1.19  0.84  1.67  0.53  0.28  0.98     0.97  0.71  1.31    1.85  1.34  2.56  2.6
    15               1.15  0.97  1.36  1.26  1.04  1.51     0.68  0.53  0.87    1.34  1.03  1.74  1.6
    30               0.95  0.79  1.14  0.98  0.79  1.22     0.98  0.74  1.28    1.29  0.83  2.00  0.7
> round( ftable( Edrr[c( 9,11,13,16,17),,"Joint",,], row.vars=c(2,3) ), 2 )
        site Bladder            Brain, CNS            Non-Hodgkin lymphoma            Liver
        wh          HR    lo    up          HR    lo    up                  HR    lo    up    HR    lo
sex dur
M   0             3.98  1.49 10.63        2.63  1.55  4.44                3.60  1.99  6.50  6.39  1.59
    1             1.72  0.43  6.87        0.90  0.37  2.16                1.81  0.81  4.03  2.92  0.41
    2             0.96  0.36  2.57        0.74  0.42  1.31                1.26  0.73  2.17  0.87  0.12
    5             1.62  0.99  2.64        0.97  0.65  1.43                1.20  0.79  1.82  3.33  1.73
    10            1.16  0.72  1.87        0.93  0.62  1.41                1.49  1.04  2.15  3.23  1.87
    15            1.12  0.90  1.39        0.73  0.55  0.98                1.05  0.82  1.34  2.51  1.90
    30            0.76  0.60  0.96        0.77  0.53  1.12                1.20  0.93  1.55  1.00  0.65
F   0             8.39  2.70 26.14        2.84  1.77  4.57                1.48  0.48  4.59    NA    NA
    1               NA    NA    NA        2.07  1.20  3.57                  NA    NA    NA    NA    NA
    2             1.38  0.35  5.54        0.75  0.45  1.27                1.45  0.78  2.70  4.91  1.83
    5             2.77  1.44  5.34        1.11  0.79  1.57                0.80  0.43  1.49  1.84  0.59
    10            0.89  0.33  2.37        0.78  0.52  1.19                1.65  1.09  2.51  3.48  1.65
    15            0.89  0.58  1.38        0.77  0.60  0.99                1.02  0.77  1.35  1.44  0.84
    30            1.00  0.68  1.47        0.77  0.54  1.09                0.76  0.52  1.10  0.93  0.48
> round( ftable( Edrr[c( 7, 3, 5, 4, 6),,"Joint",,], row.vars=c(2,3) ), 2 )
        site Prostate            Breast             Endometrium             Cervix             Ovar
        wh            HR    lo    up    HR    lo    up         HR    lo    up    HR    lo    up    H
sex dur
M   0                 NA    NA    NA    NA    NA    NA         NA    NA    NA    NA    NA    NA    N
    1                 NA    NA    NA    NA    NA    NA         NA    NA    NA    NA    NA    NA    N
    2               1.00  0.14  7.11    NA    NA    NA         NA    NA    NA    NA    NA    NA    N
    5               0.76  0.32  1.82    NA    NA    NA         NA    NA    NA    NA    NA    NA    N
    10              0.65  0.39  1.07    NA    NA    NA         NA    NA    NA    NA    NA    NA    N
    15              0.52  0.45  0.60    NA    NA    NA         NA    NA    NA    NA    NA    NA    N
    30              0.59  0.53  0.66    NA    NA    NA         NA    NA    NA    NA    NA    NA    N
F   0                 NA    NA    NA  0.89  0.54  1.45      13.67  7.54 24.78  1.19  0.62  2.29  1.8
    1                 NA    NA    NA  0.84  0.53  1.33       2.11  0.53  8.45  0.50  0.19  1.32  0.6
    2                 NA    NA    NA  0.83  0.66  1.06       4.03  2.43  6.71  0.80  0.51  1.26  1.9
    5                 NA    NA    NA  0.72  0.61  0.86       2.85  1.98  4.10  1.08  0.78  1.50  1.3
    10                NA    NA    NA  0.87  0.76  1.01       2.20  1.59  3.04  1.16  0.81  1.66  1.3
    15                NA    NA    NA  0.87  0.81  0.95       1.05  0.86  1.28  1.04  0.79  1.37  1.0
    30                NA    NA    NA  1.08  0.98  1.19       1.17  0.95  1.43  0.67  0.39  1.15  1.0
```

We then plot the results for men and women separately, showing the differences between countries;

```
> dmid <- c(0.5,1.5,3.5,7.5,12.5,22.5,35)
> clr <- c("red","blue","orange","limegreen","maroon","black")
> pclr <- rgb( t(col2rgb( clr )*2/3 + 255/3), max=255 )
> pclr[6] <- "black"
> pl.site <-
+ function( wh = 2:5,   # Which sites to plot
```

```
+           nr = 4,     # How many rows
+          cnt = TRUE,  # Country-specific HRs
+          ann = TRUE,  # Annotate with sex and site
+           yl = c(0.1,10) )
+    {
+ par( mfrow=c(nr,2),
+      mar=c(2,0,1,1), mgp=c(3,1,0)/1.6, oma=c(1,3,3,0),
+      las=1, bty="n" )
+ if( !cnt ) clr[1:5] <- pclr[1:5] <- "transparent"
+ for( tp in dimnames(Edrr)[[1]][wh] )
+ for( sx in c("M","F") )
+    {
+ # If Cervix, Endometrium or Ovary, plot a blank for men
+ # If Breast, dont plot one for men, assuming that prostate is taken just
+ # before breast.
+ if( sx=="M" & tp %in% c("Cervix","Endometrium","Ovary") )
+     plot(NA,xlab="",ylab="", xaxt="n",yaxt="n",
+                              xlim=0:1,ylim=0:1 )
+ if( !(sx=="M" & tp=="Breast"     ) &
+     !(sx=="M" & tp=="Cervix"     ) &
+     !(sx=="M" & tp=="Endometrium") &
+     !(sx=="M" & tp=="Ovary"      ) &
+     !(sx=="F" & tp=="Prostate"   ) )
+     {
+ plot( NA, xlim=c(0,35), ylim=yl*(1+(tp=="Endometrium")*1.5),
+       ylab="", xlab="", log="y", yaxs="i", yaxt="n" )
+ abline( h=c(2:15/10,2:10), col=gray(0.9) )
+ abline( h=1 )
+ if(sx=="M" | tp %in% c("Endometrium","Cervix","Ovary")) axis( side=2 )
+ matlines( dmid, cbind( Edrr[tp,sx,"DK",,],
+                        Edrr[tp,sx,"FI",,],
+                        Edrr[tp,sx,"SE",,],
+                        Edrr[tp,sx,"SC",,],
+                        Edrr[tp,sx,"AU",,],
+                        Edrr[tp,sx,"Joint",,] ),
+           type=c("o","l","l"), lty=1,
+           lwd=c(rep(c(2,1,1),5),c(4,2,2)),
+           pch=16, cex=0.7, log="y", col=rep(clr,each=3))
+ if( ann ) text( 35, yl[2], tp, adj=c(1,1), font=2 )
+      }
+      }
+ mtext( "Time since diagnosis of diabetes (years)", side=1, outer=TRUE,
+        line=0, cex=if(nr>2) 0.8 else 1.0 )
+ mtext( "HR of cancer; T1D vs population", side=2, outer=TRUE, line=2,
+        las=0, cex=if(nr>2) 0.8 else 1.0 )
+ if( ann ){
+ mtext( "Men"  , side=3, outer=TRUE, line=-0.5, at=0.25 )
+ mtext( "Women", side=3, outer=TRUE, line=-0.5, at=0.75 ) }
+ if( cnt )
+ for( i in 1:6 ) mtext( dimnames(Edrr)[[3]][i], col=clr[i], at=i/7, font=2,
+                        side=3, outer=TRUE, line=1.5, las=0 )
+ }
```

```
> pl.site(wh=c(14,2,10,12),nr=4,cnt=FALSE,yl=c(0.2,5))
```

```
> pl.site(wh=c(8,9,11,13),nr=4,cnt=FALSE,yl=c(0.2,5))
```

```
> pl.site(wh=c(7,3,4,5,6),nr=4,cnt=FALSE,yl=c(0.2,5))
```

```
> pl.site(wh=16:17,nr=2,cnt=FALSE,yl=c(0.2,10))
```

```
> pl.site(wh=1,nr=1,yl=c(0.5,4))
```

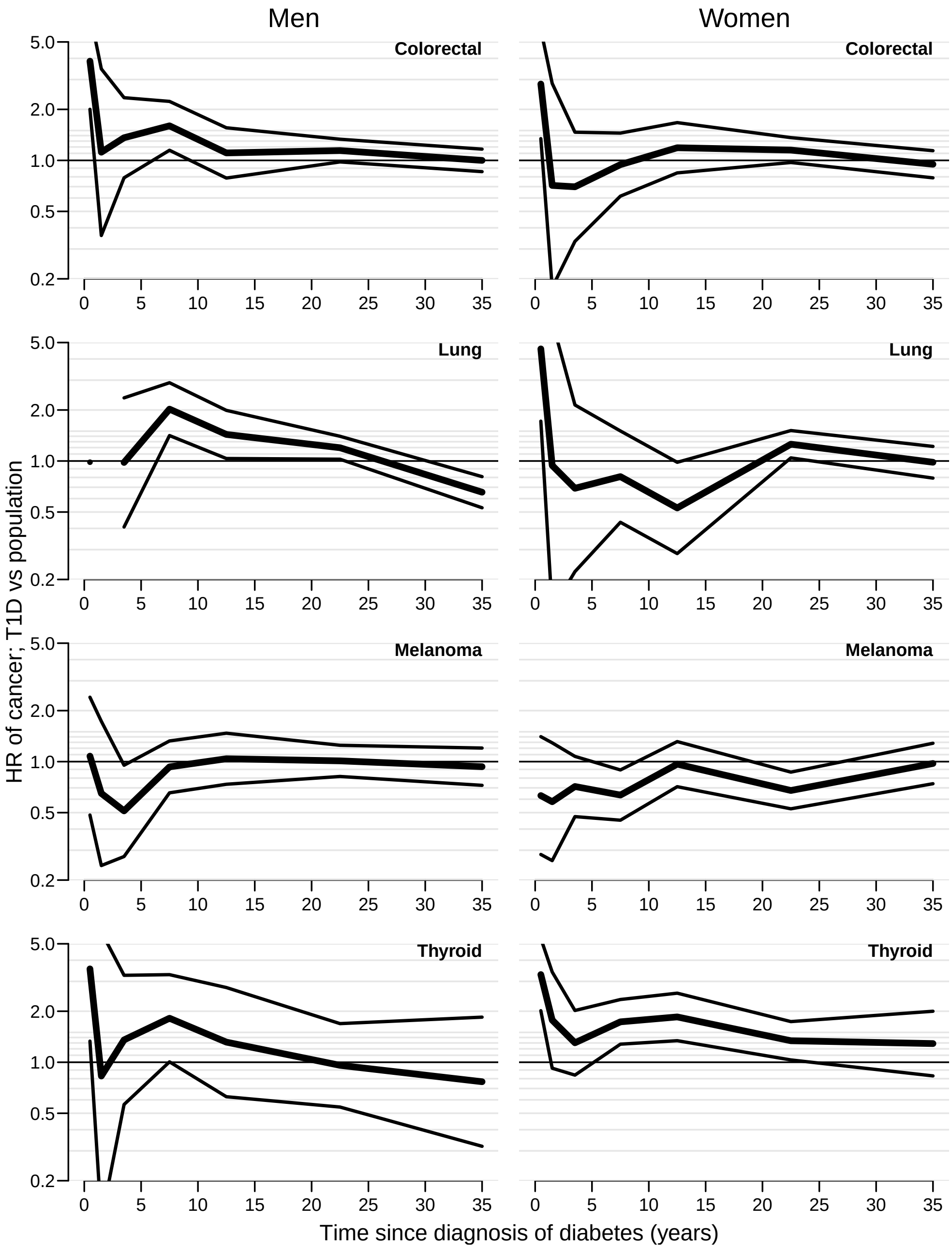

Figure 8.10: *HR of major cancers by duration of DM.*

```
> pl.site(wh=1,nr=1,cnt=FALSE,yl=c(0.8,3))
```

```
> pl.site(wh=1,nr=1,cnt=FALSE,ann=FALSE,yl=c(0.8,3))
> mtext( c("a)","b)"), at=c(5,55)/100, outer=TRUE, cex=1.5, font=2 )
```

```
> pl.site(wh=1,nr=1,cnt=FALSE,ann=FALSE,yl=c(0.8,3))
> mtext( c("a)","b)"), at=c(5,55)/100, outer=TRUE, cex=1.5, font=2 )
> axis( side=2 )
```

```
> pl.site(wh=15,nr=1,yl=c(0.5,4))
```

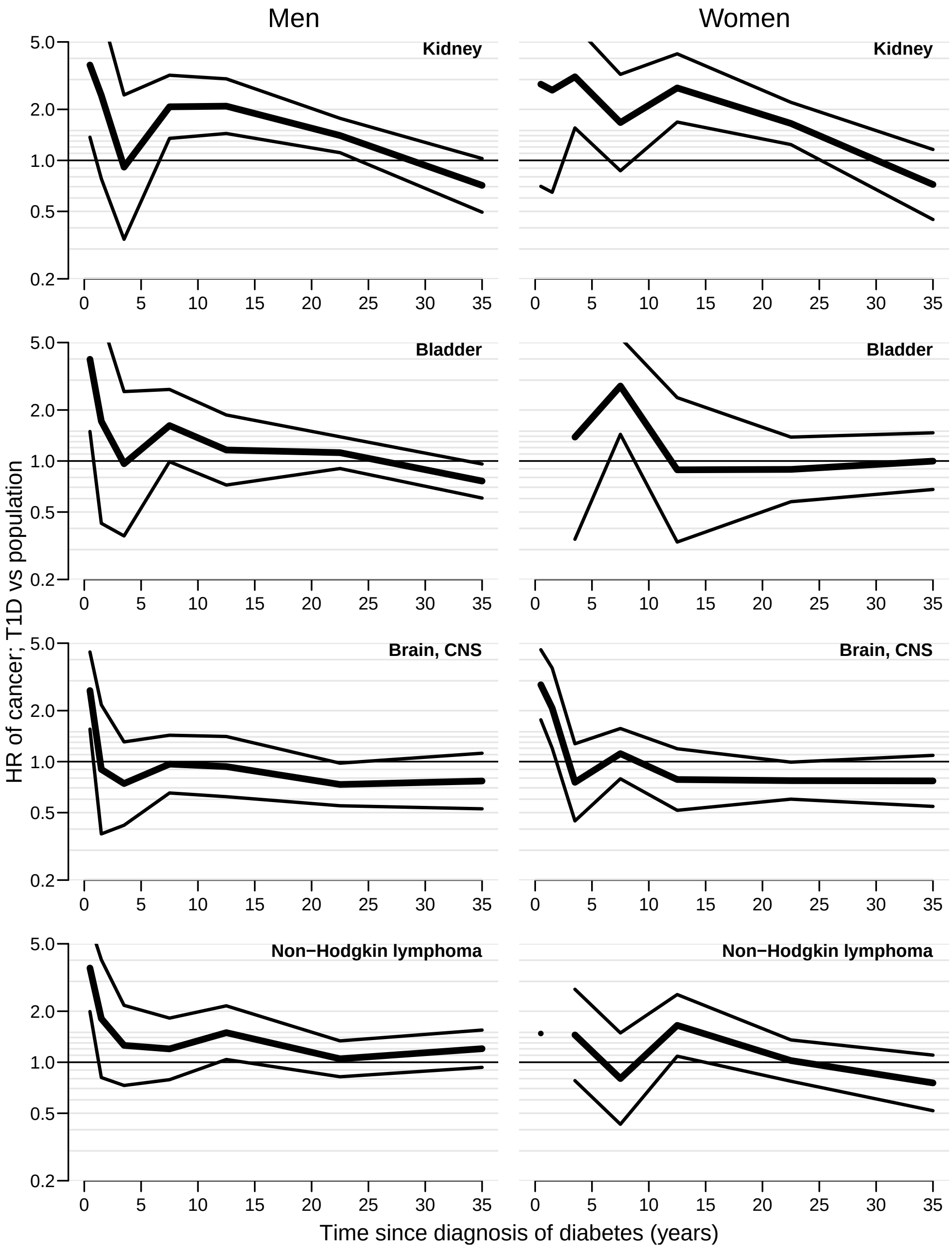

Figure 8.11: *HR of major cancers by duration of DM.*

```
> pl.site(wh=15,nr=1,cnt=FALSE,yl=c(0.8,3))
```

From the figures 8.15, 8.14 and 8.10–8.12 we see that there is a substantial duration effect which seems to be present in all countries; formally there is heterogeneity between the 5 countries, but substantially they look quite similar with respect to time since diagnosis of DM.

It is also seen that for most specific sites there is no significant excess risk after the first few years, but that there seem to be an elevated risk of any cancer of some 20% for men during the first 15 years after diagnosis, but none among women.

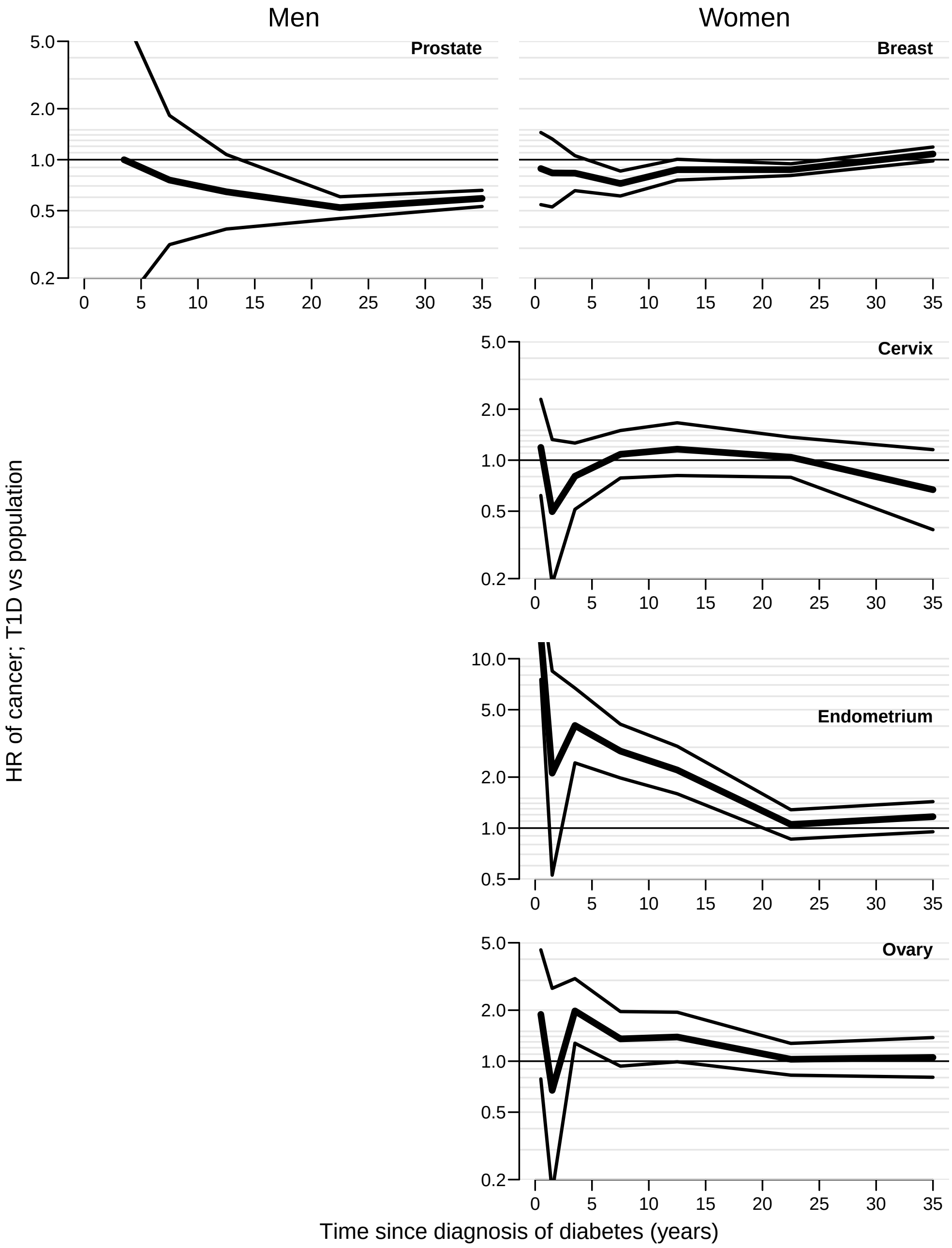

Figure 8.12: *HR of major cancers by duration of DM.*

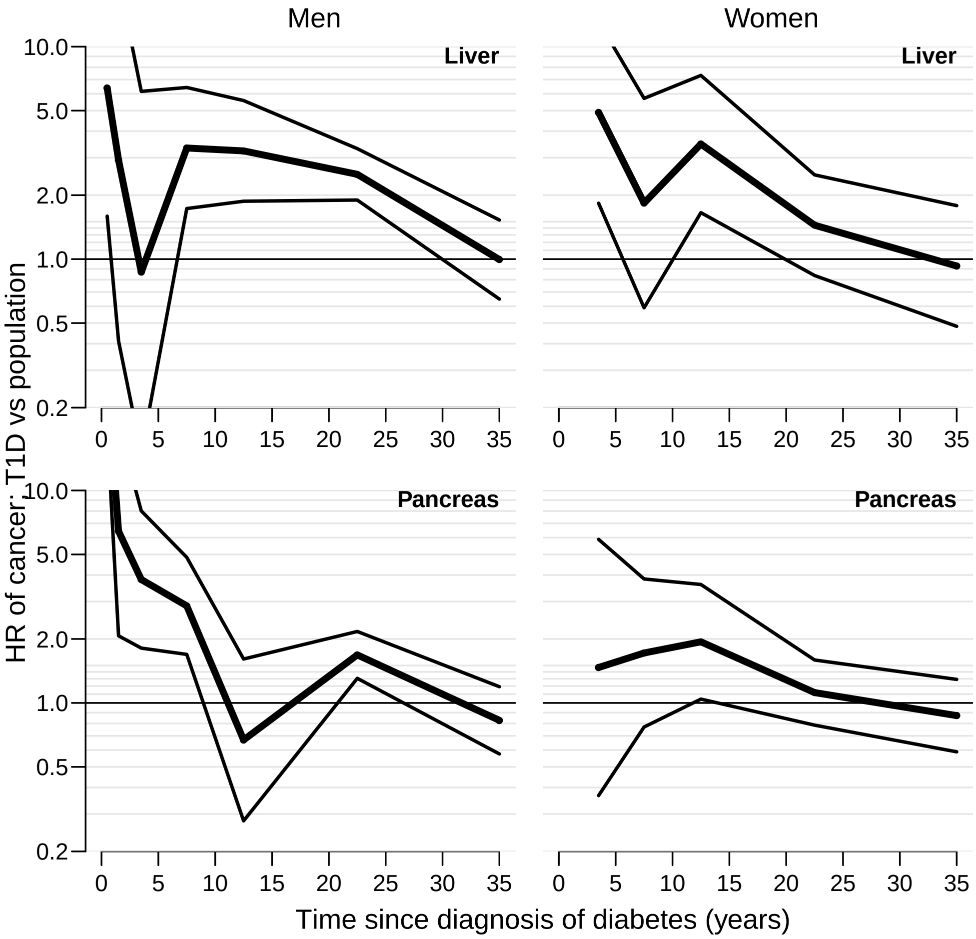

Figure 8.13: *HR of select cancers by duration of DM.*

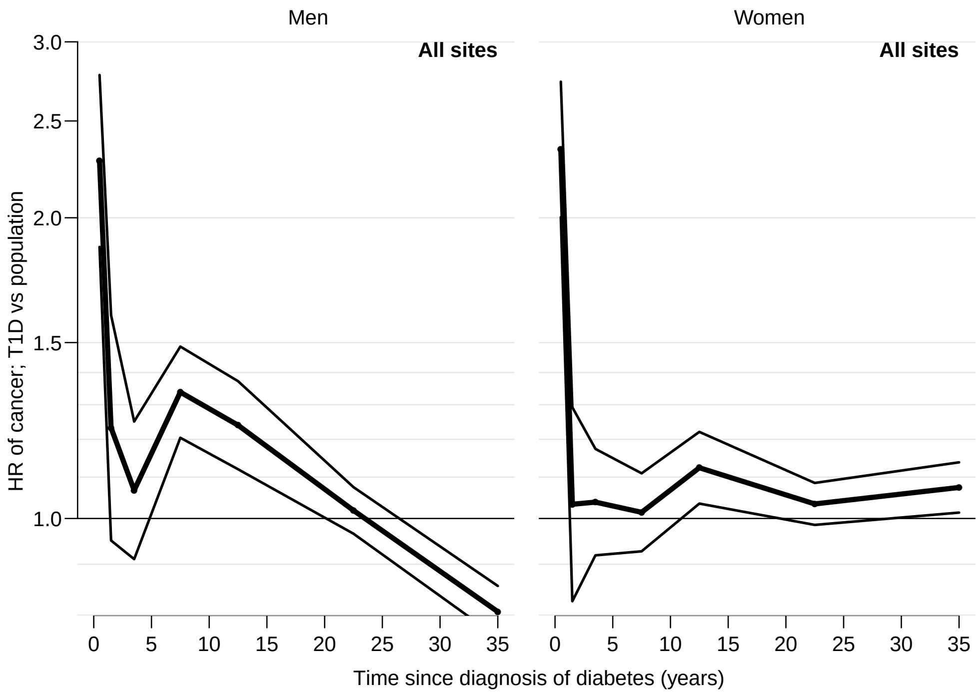

Figure 8.14: *HR of all cancers by duration of DM.*

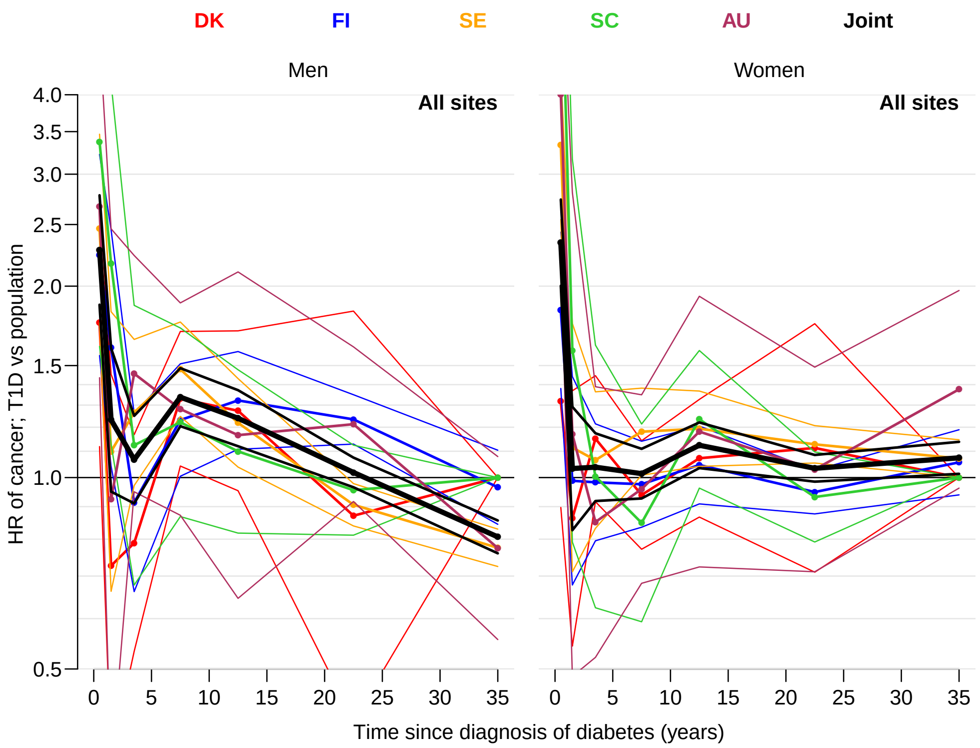

Figure 8.15: *HR of all cancers by duration of DM, separately for each country (coloured), and jointly (black curves).*

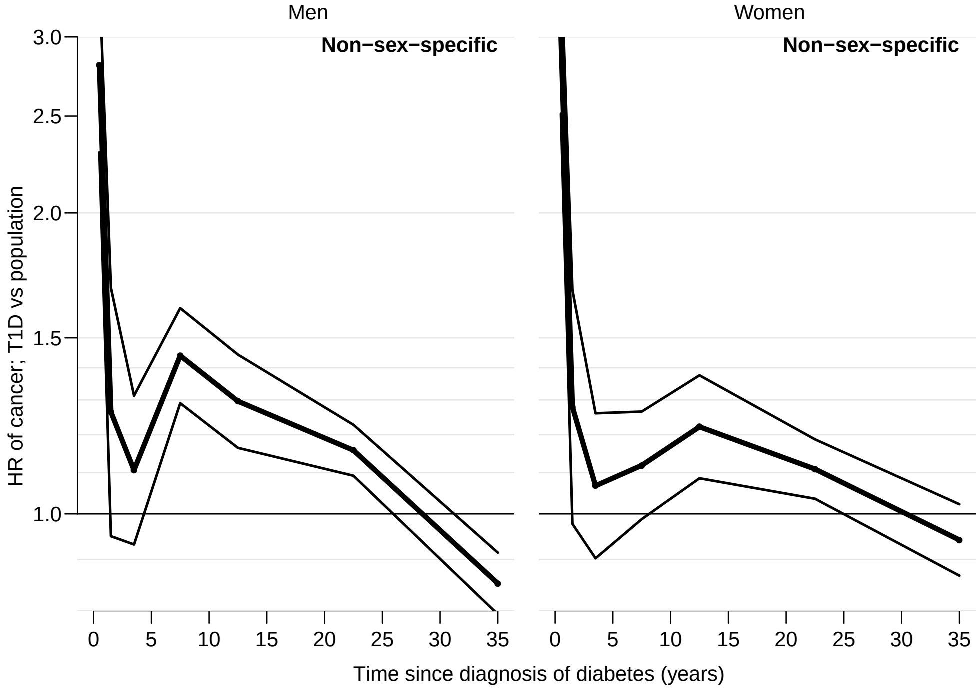

Figure 8.16: *HR of non-sex-specific cancers (all cancers but breast, cervix, endometrium, ovary, prostate and testis) by duration of DM.*

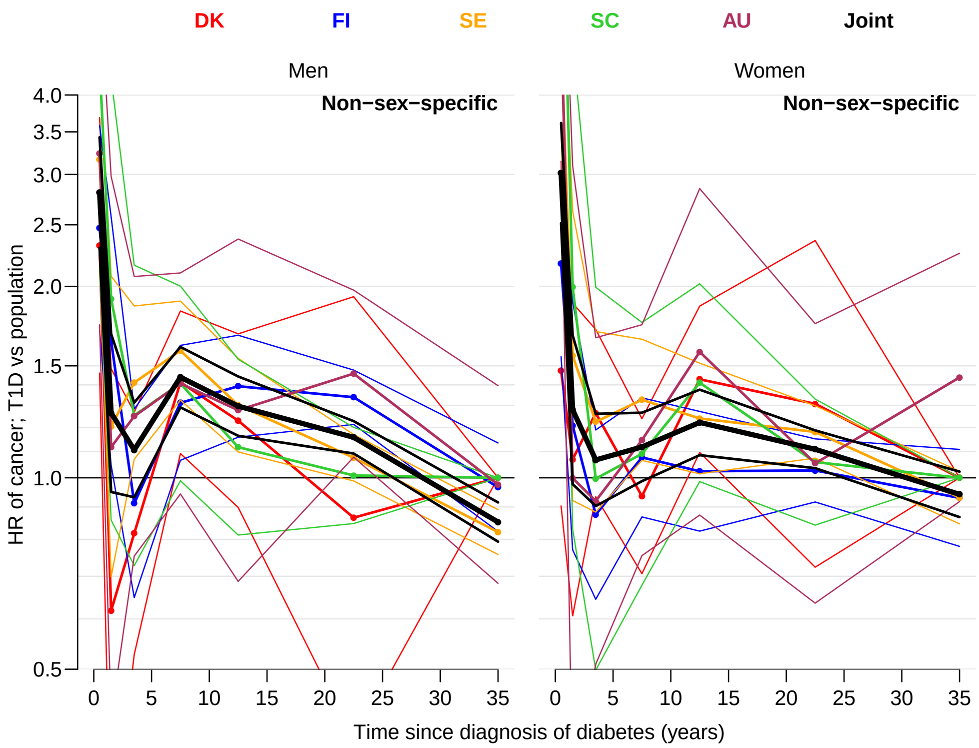

Figure 8.17: *HR of non-sex-specific cancers (all cancers but breast, cervix, endometrium, ovary, prostate and testis) by duration of DM, separately for each country (coloured), and jointly (black curves).*