# $HbA_{1c}$ in the SDC clinic

SDC
http://bendixcarstensen.com/SDC/stbb/hba1c.pdf
July 2016
Version 2.0

Compiled Thursday 21st July, 2016, 14:39
from: /home/bendix/sdc/coll/stbb/r/hba1c.tex


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

Stine Byberg — Scientific assistant
Steno Diabetes Center, Gentofte, Denmark

# Contents

# 1 Data extraction

The datasets used for mixed models consists of the following datasets extracted from the Steno EPR:

- `Labka` (all laboratory results at Steno — several records per patient)
  - Key: `pnr_id`, `labka_dato_Sas`, `labka_type`
  - `pnr_id` ID of patient
  - `labka_type` type of measurement
  - `labka_dato_sas` date of measurement
  - `labka_vaerdi_num` measurement value
- `Patientoplysninger` (core information on patients — one record per patient)
  - Key: `pnr_id`
  - `pnr_id` ID of patient
  - `Diabetes_debut_aar` year of diabetes diagnosis (precise date is not recorded)
  - `Diabetes` ICD10 coding of diagnosis
  - `Afsluttet_beh_sas` date of treatment termination
  - `Foedselsdato_sas` date of birth
  - `Koen` sex
- `Behandlingsforloeb` (information on treatment periods at Steno — one record per patient):
  - Key: `pnr_id`
  - `pnr_id` (ID of patient)
  - enrolment date 1-6 (a patient can be enrolled up to 6 times)
  - finish date 1-6 (a patient can finish treatment at steno up to 6 times)

`Labka` database: We deleted all records that were not $HbA_{1c}$ measurements, conducted before 2000-01-01 and $HbA_{1c}$ values that were missing or 0. Data was complete till 2016-06-28. Duplicate measurements by `pnr_id` and `labka_dato_sas` were deleted so each person only has one $HbA_{1c}$ measurement per date.

We merged the datasets `patientoplysninger` and `behandlingsforloeb` by `pnr_id`. Only patients with a diabetes diagnosis (DE10, DE11, DE13, DE14) were kept.

We merged this and the `Labka` dataset. All measurements on patients younger than 18 and older than 75 years at time of $HbA_{1c}$ measurement were excluded and only $HbA_{1c}$ measurements taken more than one year after enrolment at Steno were included.

The final dataset thus consisted of all $HbA_{1c}$ measurements from 1/1-2000-01-01 to 2016-06-28, among patients with a diabetes diagnosis, taken at least one year after latest enrolment date at Steno and between 18 and 75 years of age.

# 2 Hba1c in the SDC clinic

The data we are looking at here are measurements of $HbA_{1c}$ in the SDC-clinic in the period from 2000-01-01 through 2016-06-28. There are many measurements per person and each person may have undergone several treatment periods. The key to the dataset (that is the unique identifier of records) is (patient id, $HbA_{1c}$date) = `(id,d.hba1c)` as we shall use.

First we read and groom the data from the supplied `.xpt`-file:

```
> library( Epi )
> library( foreign )
> hb <- read.xport( "../data/hba1cR.xpt" )
> names( hb ) <- tolower( names(hb) )
> names( hb ) <- gsub("_",".", names(hb) )
> names( hb )
 [1] "start1"  "slut1"   "start2"  "slut2"   "start3"  "slut3"   "start4"  "slut4"
 [9] "start5"  "slut5"   "start6"  "slut6"   "pnr.id"  "dob"     "koen"    "d.hba1c"
[17] "hba1c"   "dmtype"  "dmdiagn"

> names( hb )[c(13,15)]

[1] "pnr.id" "koen"

> names( hb )[c(13,15)] <- c("id","sex")
> names( hb )[c(1:12,14,16)]

 [1] "start1"  "slut1"   "start2"  "slut2"   "start3"  "slut3"   "start4"  "slut4"
 [9] "start5"  "slut5"   "start6"  "slut6"   "dob"     "d.hba1c"

> for( i in c(1:12,14,16) ) hb[,i] <- cal.yr( as.Date(hb[,i], origin="1960-01-01") )
> apply( hb[,1:12], 2, function(x) sum(!is.na(x)) )

start1  slut1 start2  slut2 start3  slut3 start4  slut4 start5  slut5 start6  slut6
145186  11266  11266   1656   1656    216    216     20     20      5      5      0

> hb <- transform( hb, id = factor(id),
+                      sex = factor( sex, labels=c("F","M") ),
+                   dmtype = factor( dmtype, levels=1:2, labels=c("T1D","T2D") ),
+                      doy = d.hba1c - floor(d.hba1c),
+                      age = d.hba1c - dob,
+                    dmdur = d.hba1c - dmdiagn )
> # Check only one record on a given date for each person
> addmargins( table( table(paste(hb$id,hb$d.hba1c) ) ) )

     1    Sum
145186 145186
```

Time since latest start before measurement is computed, and (graphed below)

```
> names(hb)[1:6*2-1]
[1] "start1" "start2" "start3" "start4" "start5" "start6"

> hb$tfe <- apply( hb[,"d.hba1c"]-hb[,1:6*2-1],
+                  1,
+                  function(x) min(x[x>0],na.rm=TRUE) )
> summary( hb[,-(1:12)] )

       id              dob             sex          d.hba1c          hba1c          dmtype
 1265722:   184   Min.   :1925   F:69578   Min.   :2000   Min.   : 17.00   T1D:99393
 1261758:   151   1st Qu.:1947   M:75608   1st Qu.:2006   1st Qu.: 55.00   T2D:45793
 1265633:   132   Median :1956             Median :2010   Median : 63.00
 1267980:   129   Mean   :1958             Mean   :2010   Mean   : 64.44
 1261333:   121   3rd Qu.:1968             3rd Qu.:2014   3rd Qu.: 73.00
 1259995:   116   Max.   :1998             Max.   :2016   Max.   :178.00
 (Other):144353
    dmdiagn           doy                age             dmdur             tfe
 Min.   :1939   Min.   :0.0000   Min.   :18.01   Min.   : 1.025   Min.   : 1.002
 1st Qu.:1979   1st Qu.:0.2190   1st Qu.:40.87   1st Qu.:11.319   1st Qu.: 4.613
```

```
 Median :1991   Median :0.4565   Median :53.12   Median :19.038   Median : 8.851
 Mean   :1988   Mean   :0.4796   Mean   :51.28   Mean   :21.464   Mean   : 9.654
 3rd Qu.:1998   3rd Qu.:0.7413   3rd Qu.:62.84   3rd Qu.:29.723   3rd Qu.:14.064
 Max.   :2015   Max.   :0.9993   Max.   :75.00   Max.   :73.023   Max.   :26.856
 NA's   :71                                     NA's   :71
```

```
> head( hb )
```

```
    start1 slut1 start2 slut2 start3 slut3 start4 slut4 start5 slut5 start6 slut6      id
1 2001.272    NA     NA    NA     NA    NA     NA    NA     NA    NA     NA    NA 1256423
2 2001.272    NA     NA    NA     NA    NA     NA    NA     NA    NA     NA    NA 1256423
3 2001.272    NA     NA    NA     NA    NA     NA    NA     NA    NA     NA    NA 1256423
4 2001.272    NA     NA    NA     NA    NA     NA    NA     NA    NA     NA    NA 1256423
5 2001.272    NA     NA    NA     NA    NA     NA    NA     NA    NA     NA    NA 1256423
6 2001.272    NA     NA    NA     NA    NA     NA    NA     NA    NA     NA    NA 1256423
       dob sex  d.hba1c hba1c dmtype dmdiagn        doy      age    dmdur      tfe
1 1934.999   M 2002.808    87    T2D    1992 0.80766598 67.80835 10.80767 1.535934
2 1934.999   M 2003.062    97    T2D    1992 0.06228611 68.06297 11.06229 1.790554
3 1934.999   M 2003.169   103    T2D    1992 0.16906229 68.16975 11.16906 1.897331
4 1934.999   M 2003.350    81    T2D    1992 0.34976044 68.35044 11.34976 2.078029
5 1934.999   M 2003.760    84    T2D    1992 0.76043806 68.76112 11.76044 2.488706
6 1934.999   M 2004.240    96    T2D    1992 0.23956194 69.24025 12.23956 2.967830
```

Now we have a groomed data frame we can analyse for trends and cycles in $\mathrm{HbA_{1c}}$, and we can list how many observations, resp.patients we have for each combination of sex and diabetes type:

```
> rr <- with( hb                      , table( sex, dmtype ) )
> pp <- with( hb[!duplicated(hb$id),], table( sex, dmtype ) )
> cbind( rr, pp, round(rr/pp,1) )
```

```
   T1D   T2D  T1D  T2D  T1D  T2D
F 51897 17681 1441  680 36.0 26.0
M 47496 28112 1588 1114 29.9 25.2
```

We can illustrate the location of the measurement dates relative to the starting date for the persons for those with a first starting date (all, that is), those with a second, etc.

```
> par( mfrow=c(2,2), oma=c(2,2,0,0), mar=c(2,3,1,1), mgp=c(3,1,0)/1.6 )
> with( subset(hb,                  d.hba1c<start2|is.na(start2) ),
+       plot( tfe, start1, pch=16, cex=0.2, ylim=c(1994,2016), xlim=c(0,23), ylab="", xaxs="i" ) )
>       text( 10, 2014, "1st", adj=c(1,1) )
>       abline( v=1 )
>       for( y in 2000:2017) abline( y, -1, col="red" )
> with( subset(hb,d.hba1c>start2 & (d.hba1c<start3|is.na(start3))),
+       plot( tfe, start2, pch=16, cex=0.3, ylim=c(1993,2016), xlim=c(0,23), ylab="", xaxs="i" ) )
>       text( 10, 2014, "2nd", adj=c(1,1) )
>       abline( v=1 )
>       for( y in 2000:2017) abline( y, -1, col="red" )
> with( subset(hb,d.hba1c>start3 & (d.hba1c<start4|is.na(start4))),
+       plot( tfe, start3, pch=16, cex=0.4, ylim=c(1993,2016), xlim=c(0,23), ylab="", xaxs="i" ) )
>       text( 10, 2014, "3rd", adj=c(1,1) )
>       abline( v=1 )
>       for( y in 2000:2017) abline( y, -1, col="red" )
> with( subset(hb,d.hba1c>start4 & (d.hba1c<start5|is.na(start5))),
+       plot( tfe, start4, pch=16, cex=0.5, ylim=c(1993,2016), xlim=c(0,23), ylab="", xaxs="i" ) )
>       text( 10, 2014, "4th", adj=c(1,1) )
>       abline( v=1 )
>       for( y in 2000:2017) abline( y, -1, col="red" )
> mtext( "Time since entry to SDC", side=1, outer=TRUE )
> mtext( "Date of entry to SDC", side=2, outer=TRUE, las=0 )
```

Figure 1 show that measurements taken during the first year after entry or re-entry are omitted, so for modeling purposes we subtract 1 (variable `tf1`), which means that the

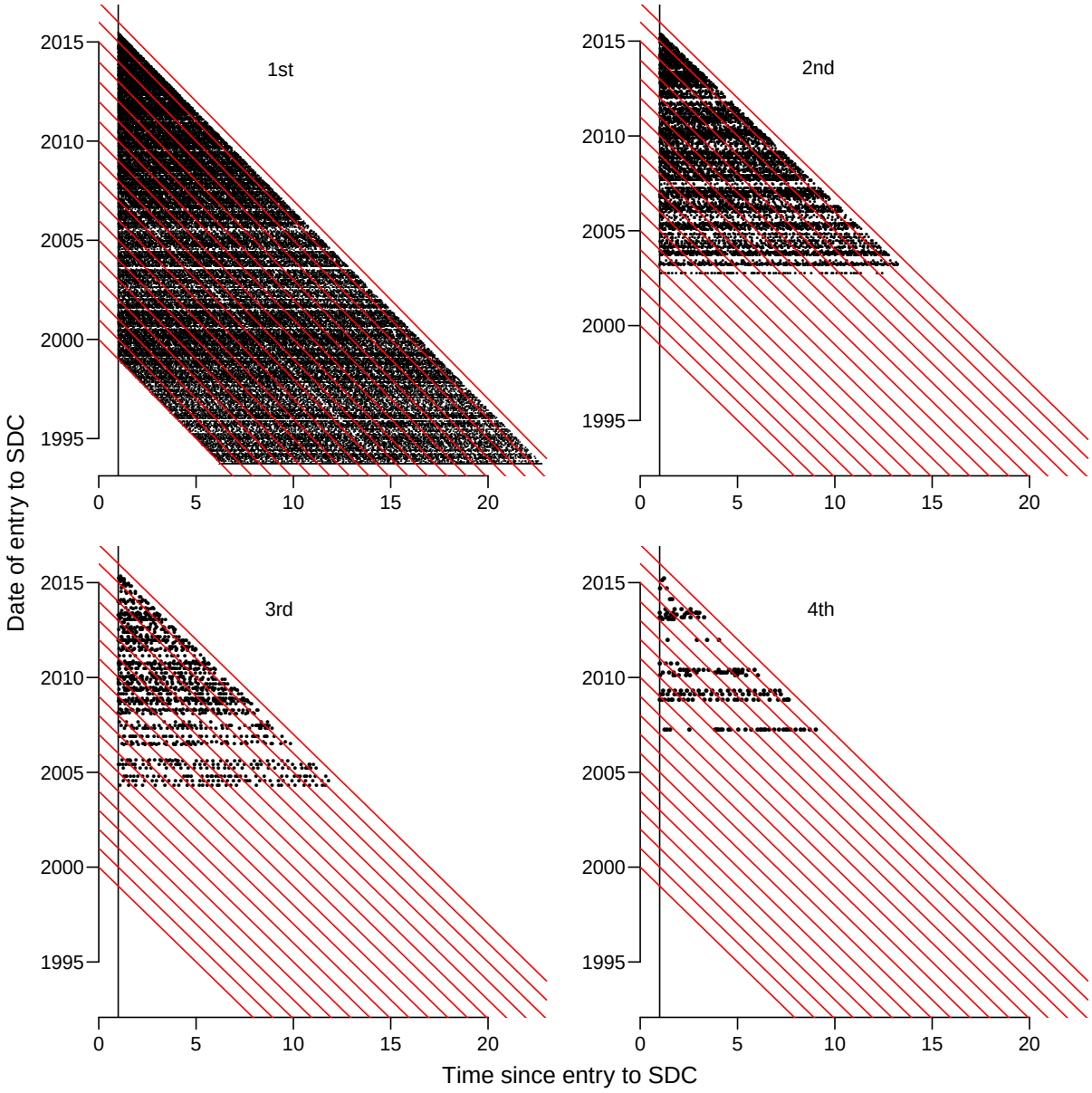


Figure 1: *Time of examination since latest entry to SDC, after 1st, 2nd, 3rd and 4th entry to SDC during the period. The red lines indicate date of measurement.*

between-person variation for the random effect associated with the intercept in the random slope model will refer to the between-patient variation at 1 year after entry to SDC.

```
> summary( hb$tfe )
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.002   4.613   8.851   9.654  14.060  26.860
> hb$tf1 <- hb$tfe - 1
```

## 2.1 Empirical age means over time

Here are the colors used for men and women (well, vice versa):

```
> levels( hb$sex )
[1] "F" "M"
> clr <- c("red","blue")
```

and here is the plot of the mean age at measurement for all measurements:

```
> mag <- with( hb, tapply( d.hba1c-dob,
+                          list( sex,
+                             dmtype,
+                  mna=floor(d.hba1c) ),
+                          mean ) )
> sag <- with( hb, tapply( d.hba1c-dob,
+                          list( sex,
+                             dmtype,
+                  mna=floor(d.hba1c) ),
+                          sd ) )
> matplot( as.numeric( dimnames(mag)[[3]] )+0.5,
+          cbind( t(mag[1,,]), t(mag[2,,]) ),
+          type = "l", lwd=c(4,2), col=rep(clr,each=2), lty=1,
+          xlab = paste( "Date of sample" ),
+          ylab = "Age (years)", ylim=c(0,70), xlim=c(1998,2017) )
> matlines( as.numeric( dimnames(sag)[[3]] )+0.5,
+           cbind( t(sag[1,,]), t(sag[2,,]) ),
+           type = "l", lwd=c(4,2), col=rep(clr,each=2), lty=2 )
> text( rep(2016.9,2), 34:35+0:1*2, dimnames(mag)[[1]], col=clr,
+       cex=1.2, font=2, adj=1 )
>     text( rep(2016.9,2), 30:29-0:1*2, dimnames(mag)[[2]], cex=1.2, font=2, adj=1 )
> segments( rep(2014.0,2), 30:29-0:1*2,
+           rep(2015.5,2), 30:29-0:1*2, lwd=c(4,2), lty=1 )
```

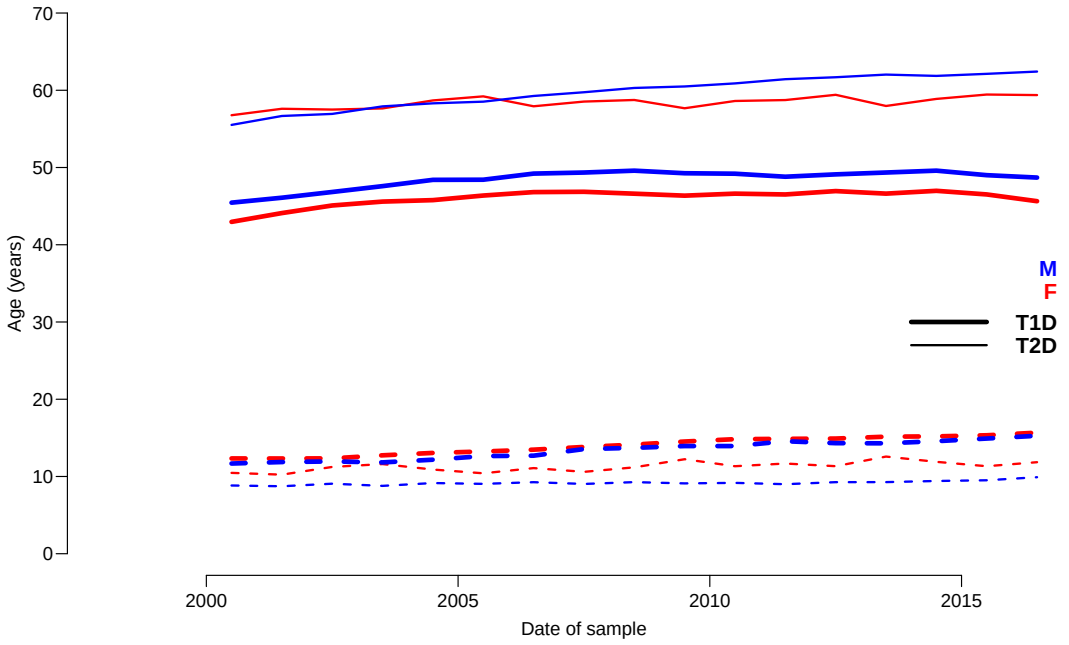


Figure 2: *Mean age by sex and diabetes type (full lines). The broken lines at the bottom shows the standard deviation of ages in the 4 groups of measurements.*

Figure 2 shows that there is not a lot of drift in the age-distribution of those who have measurements at different times.

## 2.2 Empirical means of $HbA_{1c}$ by calendar time

We make a little film showing the effect of grouping data, it will be in the folder http://bendixcarstensen.com/SDC/stbb/ as `hb-mean.pdf`:

```
> plh <-
+ function(ii)
+     {
+ mhb <- with( hb, tapply( hba1c,
+                          list( sex,
+                             dmtype,
+             mnh=floor(d.hba1c*ii)/ii ),
+                          mean ) )
+ matplot( as.numeric( dimnames(mhb)[[3]] )+1/(ii*2),
+          cbind( t(mhb["F",,]), t(mhb["M",,]) ),
+          type = "l", lwd=c(4,2), col=rep(clr,each=2), lty=1,
+          xlab = paste( "Date of sample (",ii,"intervals/year)" ),
+          ylab = "HbA1C (mmol/mol)", ylim=c(58,78), xlim=c(1998,2017) )
+ abline( v=1990:2020 )
+ text( rep(2016.9,2), 75:74-1:0/3, c("Men","Women"), col=c("blue","red"),
+       cex=1.2, font=2, adj=1 )
+ text( rep(2016.9,2), 73:72+0:1/3, c("T1D","T2D"), cex=1.2, font=2, adj=1 )
+ segments( rep(2014,2), 73:72+0:1/3,
+           rep(2016,2), 73:72+0:1/3, lwd=c(4,2), lty=1 )
+      }
> plh( 12 )
> plh(  6 )
> plh(  4 )
> plh(  3 )
> plh(  2 )
> plh(  1 )


> plh( 12 )


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

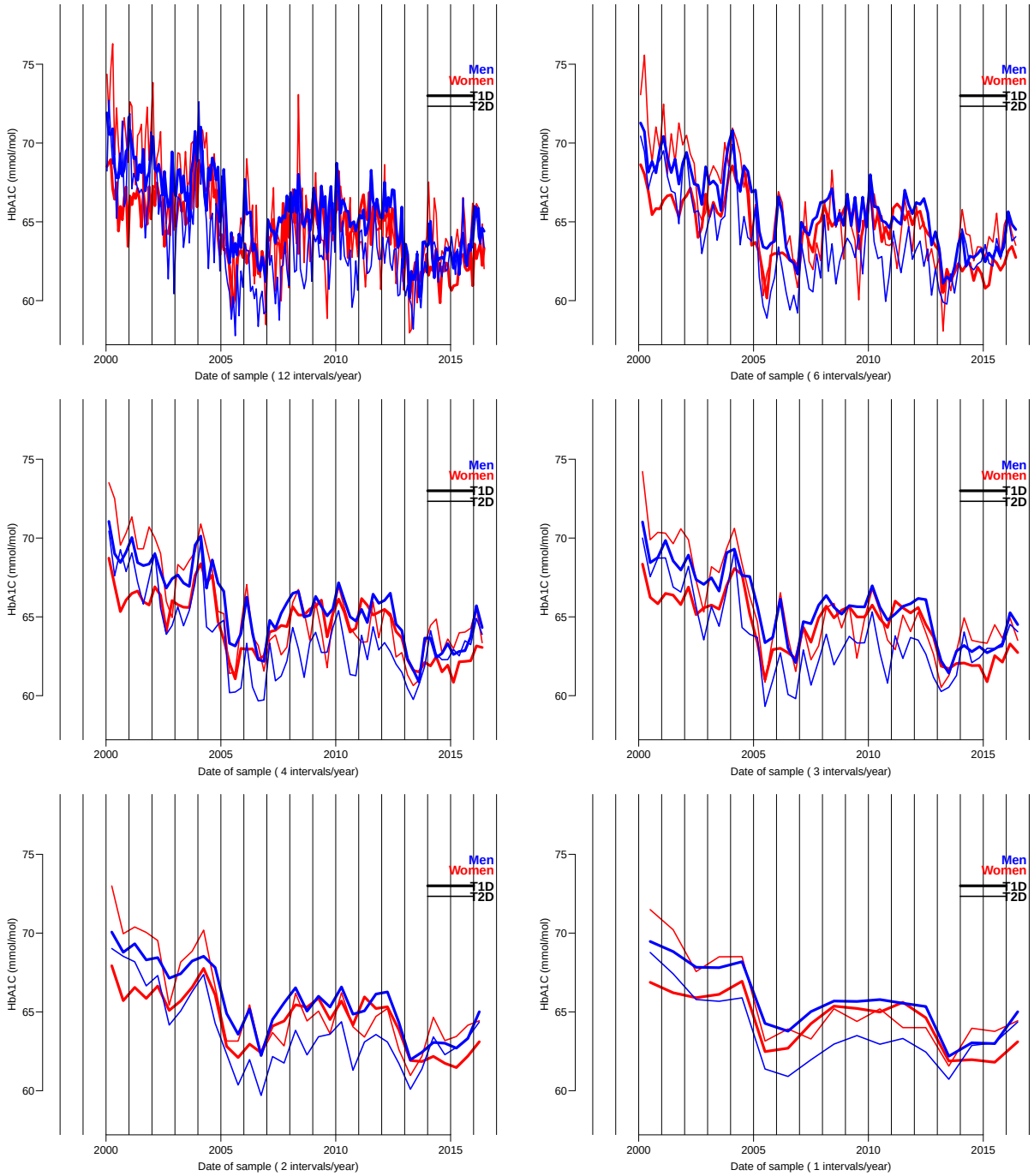


Figure 3: *Mean $HbA_{1c}$ in different time intervals by sex and diabetes type.*

# 3 Modeling

To summarize the effects of age (at measurement), diabetes duration, time of year (seasonality) and calendar time, we model the $HbA_{1c}$-levels using a random person-effect and systematic effects of:

- date
- season (date in year)
- time since (last) enrollment to SDC
- age
- diabetes duration
- sex
- DM-type

## 3.1 Model structure

We stratify the analyses by sex and DM-type, so specifically for each combination of these we fit the following model for person $p$'s $HbA_{1c}$ measurement, $y_{pt}$ at date $t$, at which point the season is $s_{pt}$, the duration of diabetes is $d_{pt}$, the person's age is $a_{pt}$, and the time since entry to SDC is $v_{pt}$

$$y_{pt} = \mu + \gamma(v_{pt} - 1) + h(s_{pt}) + f(t) + g(a_{pt}) + l(d_{pt}) + u_p + w_p(v_{pt} - 1) + e_{pt},$$
$$(u_p, w_p) \sim \mathcal{N}_2(0, \Gamma), \quad e_{pt} \sim (0, \Sigma)$$

The error-terms $(u_p, w_p)$ have an unspecified (freely varying) covariance matrix, $\Gamma$, whereas the terms $e_{pt}$ are independent between persons, but correlated *within* persons with a correlation between points at a distance of $T = |t_i - t_j|$ of $\exp(-\zeta T)$ — this function is 1 for $T = 0$ and 0 for $T = \infty$.

Note that we are using the variable $v_{pt} - 1$ (in the dataset called `tf1`) in order to render the variation in the random intercept interpretable as the between patient variation at 1 year after inclusion at SDC.

We are making the implicit assumption that there are no interactions between the variables of interest — and the assumption that *all* of the variables (including the random effects) have different effects for different values of sex and diabetes type.

## 3.2 Fitting the models

First we read in a simple function to generate harmonics:

```
> source("Hm.R")
> Hm
```

```
function (x, o)
{
    if (any(x < 0) | any(x > 1))
        warning("x not in the range 0:1\nNOTE: harmonic cycle has period 1 on x-scale")
    x <- x * 2 * pi
    res <- NULL
    for (i in 1:o) res <- cbind(res, sin(x * i), cos(x * i))
    colnames(res) <- paste(c("sin", "cos"), rep(1:o, each = 2),
        sep = "")
    res
}
```

Then we can set up a random effects model, but due to the fishy scoping of the `lme` function we need to define the spline terms separately

```
> load( "../data/hb.Rda")
> hb <- subset( hb, !is.na(dmdur) )
> library(Epi)
> library(nlme)
> ( a.kn <- with( hb, quantile( age    , probs = (1: 4-0.5)/ 4 ) ) )
   12.5%    37.5%    62.5%    87.5%
32.52567 47.39493 58.16290 67.86585

> ( d.kn <- with( hb, quantile( dmdur  , probs = (1: 4-0.5)/ 4 ) ) )

    12.5%     37.5%     62.5%     87.5%
 7.233402 15.134668 23.811773 37.873888

> ( t.kn <- with( hb, quantile( d.hba1c, probs = (1:11-0.5)/11 ) ) )

4.545455% 13.63636% 22.72727% 31.81818% 40.90909%       50% 59.09091% 68.18182% 77.27273%
 2001.242  2003.561  2005.436  2007.248  2008.940  2010.372  2011.687  2012.902  2014.063
86.36364% 95.45455%
 2015.068  2016.064

> a.Spl <- function(x) Ns( x, knots=a.kn )
> d.Spl <- function(x) Ns( x, knots=d.kn )
> t.Spl <- function(x) Ns( x, knots=t.kn )
> mods <- list( NULL )
> length( mods ) <- 4
> dim( mods ) <- c(2,2)
> dimnames( mods ) <- list( sex = levels(hb$sex),
+                           dmtype = levels(hb$dmtype) )
> mods

   dmtype
sex T1D  T2D
  F NULL NULL
  M NULL NULL

> lapply( mods, class )

[[1]]
[1] "NULL"

[[2]]
[1] "NULL"

[[3]]
[1] "NULL"

[[4]]
[1] "NULL"
```

```
> for( sx in dimnames(mods)[[1]] )
+ for( tp in dimnames(mods)[[2]] )
+    {
+ cat( "sex =", sx, ", DM-type =", tp, " took " )
+ st <- Sys.time()
+ mods[[sx,tp]] <- lme( hba1c ~ tf1 + Hm( doy, 2 ) +
+                                     a.Spl( age ) +
+                                     d.Spl( dmdur ) +
+                                     t.Spl( d.hba1c ),
+                     random = ~ tf1 | id,
+                       corr = corExp( form = ~ d.hba1c | id ),
+                       data = subset( hb, sex==sx & dmtype==tp ),
+                    control = list( niterEM=100 ) )
+ cat( formatC(as.numeric(Sys.time()-st),format="f",digits=2), "minutes \n" )
+ flush.console()
+    }
sex = F , DM-type = T1D  took 7.13 minutes
sex = F , DM-type = T2D  took 1.36 minutes
sex = M , DM-type = T1D  took 4.48 minutes
sex = M , DM-type = T2D  took 1.59 minutes
> save( mods, a.kn, d.kn, t.kn, a.Spl, d.Spl, t.Spl, Hm, file="../data/lme-mods.Rda")
```

## 3.3 Reporting the model results

The random effects model is fitted separately for the 4 possible combinations of sex and diabetes type.

The model has the following effects influencing the level of $\mathrm{HbA_{1c}}$:

- time since entry to SDC ($v$) — we only estimate a linear effect of this.
- season (time of year) ($h(s)$)
- date of measurement ($f(t)$) — this is the effect of primary interest, particularly if there is tendency to change after
- age ($g(a)$)
- duration of diabetes ($l(d)$)

The natural way to report these effects is by showing the average $\mathrm{HbA_{1c}}$ level by age and duration — a curve starting at say age 50 showing the $\mathrm{HbA_{1c}}$level as a function of *current age* for a person diagnosed at age 50, that is also with increasing duration.

This will be the predicted values for $\mathrm{HbA_{1c}}$, of course for a given sex and diabetes type and for a chosen reference value of date of measurement, say 1 January 2012, time since entry to SDC equal to 1 year (we only included measurements from one year after entry to SDC).

The seasonal component is estimated as a separate term which is constrained to be 0 on average over the year - that is with no pre-specified reference point in the year. But since the terms is included in the model we must include it with a value that corresponds to the date we are using as a reference point.

These predicted values and effects for the two sexes and the two types of DM should be shown together for all effects.

First we show the results of the model fits:

```
> library( nlme )
> load( file="../data/lme-mods.Rda")
> for( sx in dimnames(mods)[[1]] )
+ for( tp in dimnames(mods)[[2]] )
+    {
+ cat( "\n------------------------------------------------------\n",
+      "Sex =", sx, ", type =", tp, "\n" )
+ print( mods[[sx,tp]] )
+    }
------------------------------------------------------
 Sex = F , type = T1D
Linear mixed-effects model fit by REML
  Data: subset(hb, sex == sx & dmtype == tp)
  Log-restricted-likelihood: -164686.1
  Fixed: hba1c ~ tf1 + Hm(doy, 2) + a.Spl(age) + d.Spl(dmdur) + t.Spl(d.hba1c)
     (Intercept)              tf1   Hm(doy, 2)sin1   Hm(doy, 2)cos1   Hm(doy, 2)sin2
     69.17480532      -0.19265607       0.04043430       0.34604151       0.01559146
  Hm(doy, 2)cos2      a.Spl(age)1      a.Spl(age)2      a.Spl(age)3    d.Spl(dmdur)1
      0.09709970       0.52568261      -1.73168243       0.73680235       0.23663506
   d.Spl(dmdur)2    d.Spl(dmdur)3  t.Spl(d.hba1c)1  t.Spl(d.hba1c)2  t.Spl(d.hba1c)3
      2.14217096      -0.58569737      -2.29284183      -4.19575608      -0.17954320
 t.Spl(d.hba1c)4  t.Spl(d.hba1c)5  t.Spl(d.hba1c)6  t.Spl(d.hba1c)7  t.Spl(d.hba1c)8
     -2.76748331       0.44469882      -4.03784965      -4.25709225      -4.92014223
 t.Spl(d.hba1c)9 t.Spl(d.hba1c)10
     -4.59976065      -3.70446875

Random effects:
 Formula: ~tf1 | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 13.5658769 (Intr)
tf1          0.8840922 -0.689
Residual     7.8209595

Correlation Structure: Exponential spatial correlation
 Formula: ~d.hba1c | id
 Parameter estimate(s):
    range
0.5250154
Number of Observations: 51896
Number of Groups: 1440

------------------------------------------------------
 Sex = F , type = T2D
Linear mixed-effects model fit by REML
  Data: subset(hb, sex == sx & dmtype == tp)
  Log-restricted-likelihood: -60519.56
  Fixed: hba1c ~ tf1 + Hm(doy, 2) + a.Spl(age) + d.Spl(dmdur) + t.Spl(d.hba1c)
     (Intercept)              tf1   Hm(doy, 2)sin1   Hm(doy, 2)cos1   Hm(doy, 2)sin2
      65.1030226        0.1216043        0.2147684        0.3336391        0.1690627
  Hm(doy, 2)cos2      a.Spl(age)1      a.Spl(age)2      a.Spl(age)3    d.Spl(dmdur)1
       0.1351404        2.3158024        5.1663263       -2.1402164        4.3201149
   d.Spl(dmdur)2    d.Spl(dmdur)3  t.Spl(d.hba1c)1  t.Spl(d.hba1c)2  t.Spl(d.hba1c)3
      10.8747239        2.2170921       -6.6745441       -8.4578912       -5.4748802
 t.Spl(d.hba1c)4  t.Spl(d.hba1c)5  t.Spl(d.hba1c)6  t.Spl(d.hba1c)7  t.Spl(d.hba1c)8
      -7.4775616       -5.2385322      -10.3095085       -7.8070032       -8.5469022
 t.Spl(d.hba1c)9 t.Spl(d.hba1c)10
      -9.6896126       -7.2426762

Random effects:
 Formula: ~tf1 | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 12.1788254 (Intr)
tf1          0.9155121 -0.49
```

```
Residual     9.8107545

Correlation Structure: Exponential spatial correlation
 Formula: ~d.hba1c | id
 Parameter estimate(s):
    range
0.5839186
Number of Observations: 17680
Number of Groups: 681

------------------------------------------------------------
 Sex = M , type = T1D
Linear mixed-effects model fit by REML
  Data: subset(hb, sex == sx & dmtype == tp)
  Log-restricted-likelihood: -152051
  Fixed: hba1c ~ tf1 + Hm(doy, 2) + a.Spl(age) + d.Spl(dmdur) + t.Spl(d.hba1c)
     (Intercept)              tf1   Hm(doy, 2)sin1   Hm(doy, 2)cos1   Hm(doy, 2)sin2
     68.41043996      -0.08149855       0.08949382       0.53672963       0.14130698
  Hm(doy, 2)cos2      a.Spl(age)1      a.Spl(age)2      a.Spl(age)3    d.Spl(dmdur)1
      0.24475974      -1.50690723      -3.91737370      -1.84575566       0.90562573
   d.Spl(dmdur)2    d.Spl(dmdur)3  t.Spl(d.hba1c)1  t.Spl(d.hba1c)2  t.Spl(d.hba1c)3
      5.29562799      -0.91801061      -2.62756724      -5.84358830      -0.76503064
 t.Spl(d.hba1c)4  t.Spl(d.hba1c)5  t.Spl(d.hba1c)6  t.Spl(d.hba1c)7  t.Spl(d.hba1c)8
     -4.19506335      -1.24736079      -5.20974868      -5.42922653      -5.67745247
 t.Spl(d.hba1c)9 t.Spl(d.hba1c)10
     -5.13658657      -3.74641920

Random effects:
 Formula: ~tf1 | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 13.3285895 (Intr)
tf1          0.7794011 -0.617
Residual     6.6104003

Correlation Structure: Exponential spatial correlation
 Formula: ~d.hba1c | id
 Parameter estimate(s):
   range
0.374688
Number of Observations: 47496
Number of Groups: 1589

------------------------------------------------------------
 Sex = M , type = T2D
Linear mixed-effects model fit by REML
  Data: subset(hb, sex == sx & dmtype == tp)
  Log-restricted-likelihood: -97382.69
  Fixed: hba1c ~ tf1 + Hm(doy, 2) + a.Spl(age) + d.Spl(dmdur) + t.Spl(d.hba1c)
     (Intercept)              tf1   Hm(doy, 2)sin1   Hm(doy, 2)cos1   Hm(doy, 2)sin2
     69.02699904       0.03703666       0.43651573       0.68518959       0.38809675
  Hm(doy, 2)cos2      a.Spl(age)1      a.Spl(age)2      a.Spl(age)3    d.Spl(dmdur)1
      0.36453650      -4.03843331      -8.88716536      -6.61903905       1.72703693
   d.Spl(dmdur)2    d.Spl(dmdur)3  t.Spl(d.hba1c)1  t.Spl(d.hba1c)2  t.Spl(d.hba1c)3
     11.70043648       2.27924307      -4.84851942      -6.23473839      -2.73821932
 t.Spl(d.hba1c)4  t.Spl(d.hba1c)5  t.Spl(d.hba1c)6  t.Spl(d.hba1c)7  t.Spl(d.hba1c)8
     -4.44289212      -2.44878824      -6.46921893      -4.42079908      -5.30426074
 t.Spl(d.hba1c)9 t.Spl(d.hba1c)10
     -5.14067920      -4.06407952

Random effects:
 Formula: ~tf1 | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev     Corr
(Intercept) 11.7296829 (Intr)
```

```
tf1          0.8613414 -0.524
Residual     9.6160078

Correlation Structure: Exponential spatial correlation
 Formula: ~d.hba1c | id
 Parameter estimate(s):
    range
0.4824872
Number of Observations: 28043
Number of Groups: 1113
```

To clarify what we extract from the model we set up arrays to collect the relevant effects — first we define the ranges of the effects we want to see:

```
> # Dates of measurement (time)
> tpr <- seq(2000,2016.5,,500)
> # Seasonal values (december and january included twice)
> hpr <- seq(-31,365+31,,500)
> # Diabetes duration
> dpr <- seq(0,40,,500)
> # Matrix to fish out time effects relative to 2012-01-01 / 2014-01-01
> S12 <- Ns( tpr, knots=t.kn, ref=2012 )
> S14 <- Ns( tpr, knots=t.kn, ref=2014 )
> H.0 <- Hm( rep(0,500), 2 )
> Hpr <- Hm( hpr/365, 2 )
> # Joint effects of age and duration, also including prediction intervals
> agedur <- NArray( c( dimnames(mods),
+                      list(  age = seq(20,70,5),
+                             dur = dpr,
+                            what = c("Est","lo","up","pr.l","pr.u") ) ) )
> str( agedur )
 logi [1:2, 1:2, 1:11, 1:500, 1:5] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ sex   : chr [1:2] "F" "M"
  ..$ dmtype: chr [1:2] "T1D" "T2D"
  ..$ age   : chr [1:11] "20" "25" "30" "35" ...
  ..$ dur   : chr [1:500] "0" "0.0801603206412826" "0.160320641282565" "0.240480961923848" ...
  ..$ what  : chr [1:5] "Est" "lo" "up" "pr.l" ...

> # Seasonal effects
> season <- NArray( c( dimnames(mods),
+                      list( time = hpr,
+                            what = c("Est","lo","up") ) ) )
> str( season )
 logi [1:2, 1:2, 1:500, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ sex   : chr [1:2] "F" "M"
  ..$ dmtype: chr [1:2] "T1D" "T2D"
  ..$ time  : chr [1:500] "-31" "-30.1442885771543" "-29.2885771543086" "-28.4328657314629" ...
  ..$ what  : chr [1:3] "Est" "lo" "up"

> # Time trajectories
> trends <- NArray( c( dimnames(mods),
+                      list( ref = 2012+0:1*2,
+                           time = tpr,
+                           what = c("Est","lo","up") ) ) )
> str( trends )
 logi [1:2, 1:2, 1:2, 1:500, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 5
  ..$ sex   : chr [1:2] "F" "M"
  ..$ dmtype: chr [1:2] "T1D" "T2D"
  ..$ ref   : chr [1:2] "2012" "2014"
  ..$ time  : chr [1:500] "2000" "2000.03306613226" "2000.06613226453" "2000.09919839679" ...
  ..$ what  : chr [1:3] "Est" "lo" "up"
```

```
> # Random effects and fixed slope by individual time since start
> meanch <- NArray( c( dimnames(mods),
+                      list( parm = c("pop.sd","pop.sl","cor","fix.sl","res.sd",
+                                     paste("cor0.",cval<-seq(9,1,-2),sep="")),
+                            what = c("Est","lo","up") ) ) )
> str( meanch )
 logi [1:2, 1:2, 1:10, 1:3] NA NA NA NA NA NA ...
 - attr(*, "dimnames")=List of 4
  ..$ sex   : chr [1:2] "F" "M"
  ..$ dmtype: chr [1:2] "T1D" "T2D"
  ..$ parm  : chr [1:10] "pop.sd" "pop.sl" "cor" "fix.sl" ...
  ..$ what  : chr [1:3] "Est" "lo" "up"
> # Loop to put values into the arrays
> for( sx in dimnames(mods)[[1]] )
+ for( tp in dimnames(mods)[[2]] )
+    {
+ season[sx,tp,,] <- ci.exp( mods[[sx,tp]], subset="Hm" , ctr.mat=Hpr, Exp=FALSE )
+ trends[sx,tp,"2012",,] <- ci.exp( mods[[sx,tp]], subset="t.S", ctr.mat=S12, Exp=FALSE )
+ trends[sx,tp,"2014",,] <- ci.exp( mods[[sx,tp]], subset="t.S", ctr.mat=S14, Exp=FALSE )
+ # extract random effects with c.i.s
+ ii <- try( intervals( mods[[sx,tp]] ) )
+ if( inherits( ii, "intervals.lme") )
+   {
+   cat( "random effects OK for", sx, tp, "\n" )
+ meanch[sx,tp,"fix.sl",] <- ii$fixed["tf1",c(2,1,3)]
+ meanch[sx,tp,1:3     ,] <- as.matrix(ii$reStruct$id)[,c(2,1,3)]
+ meanch[sx,tp,"res.sd",] <- ii$sigma[c(2,1,3)]
+ for(id in cval) meanch[sx,tp,paste("cor0.",id,sep=""),] <- -log(id/10) * ii$corStruct[c(2,1,3)]
+   }
+ # For the different fixed ages at diagnosis (aa) compute the current
+ # ages along the duration points. This will be a prediction at some
+ # reference values for the a) individual duration (1 year), b) season
+ # (approx 1 sept) and c) date of measurement (1.1.2012):
+ for( aa in dimnames(agedur)[[3]] )
+    {
+     ad <- as.numeric(aa)
+    Cad <- cbind( 1, 0,
+                  Hm(    rep(0,length(dpr)), 2 ),
+               a.Spl(              ad+dpr ),
+               d.Spl(                dpr ),
+               t.Spl( rep(2012,length(dpr)) ) )
+    # estimate and se
+    ES <- ci.lin( mods[[sx,tp]], Exp=FALSE, ctr.mat=Cad )[,1:2]
+    # prediction sd
+    PS <- sqrt( ES[,2]^2 + meanch[sx,tp,"res.sd",1]^2 )
+    # mean, ci and pi
+    agedur[sx,tp,aa,,] <- cbind( ES %*% ci.mat(),
+                  cbind(ES[,1],PS) %*% ci.mat() )[,-4]
+    }
+    }
random effects OK for F T2D
random effects OK for M T1D
random effects OK for M T2D
```

Thus we have the trends and the seasonal effects as well as the estimated individual mean change within persons — we print the numbers in different layout:

```
> round( ftable( meanch, col.vars=c(2,1,4) ), 3 )
       dmtype    T1D                                    T2D
       sex         F                   M                F                   M
       what      Est     lo     up    Est     lo     up   Est     lo     up   Est     lo     up
parm
```

```
pop.sd            NA     NA     NA 13.329 12.836 13.840 12.179 11.363 13.053 11.730 11.118 12.374
pop.sl            NA     NA     NA  0.779  0.735  0.826  0.916  0.827  1.014  0.861  0.789  0.940
cor               NA     NA     NA -0.617 -0.660 -0.570 -0.490 -0.588 -0.378 -0.524 -0.598 -0.441
fix.sl            NA     NA     NA -0.081 -0.183  0.020  0.122 -0.075  0.318  0.037 -0.113  0.187
res.sd            NA     NA     NA  6.610  6.548  6.673  9.811  9.618 10.008  9.616  9.481  9.753
cor0.9            NA     NA     NA  0.039  0.038  0.041  0.062  0.059  0.065  0.051  0.049  0.053
cor0.7            NA     NA     NA  0.134  0.130  0.137  0.208  0.198  0.219  0.172  0.166  0.179
cor0.5            NA     NA     NA  0.260  0.253  0.267  0.405  0.385  0.426  0.334  0.322  0.347
cor0.3            NA     NA     NA  0.451  0.439  0.464  0.703  0.669  0.739  0.581  0.560  0.603
cor0.1            NA     NA     NA  0.863  0.839  0.887  1.345  1.279  1.414  1.111  1.070  1.153
```

### 3.3.1 Time since entry to SDC, and random effects

The average annual change in $HbA_{1c}$(`fix.sl`) is negative for T1D patients and positive for T2D patients; the absolute size larger for women than for men, but quite small relative to the population variation of the slopes (`pop.sl`). The population SD of the $HbA_{1c}$ level (`pop.sd`) is in the vicinity of 12 mmol/mol and the residual variation (`res.sd`) is between 50 and 80% of this, so the total standard deviation is in the range 15–20 mmol/mol.

The last 9 lines gives the distance between measurement points (in years) needed to get to a correlation of 0.9,. . . ,0.1, respectively. We see that beyond one year's difference in measurements there is very little correlation between observations.

### 3.3.2 Age- and duration effects

We then plot the age-specific mean $HbA_{1c}$in two different guises:

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1,6, las=1, bty="n" )
> plot( NA, xlim = c(30,90), ylim = c(45,80),
+           xlab = "Age"   , ylab = expression("Hb"*A[1][c]) )
> for( aa in 3:7*10 )
+ matlines( aa+dpr, cbind( agedur["F","T1D",paste(aa),,1:3],
+                           agedur["F","T2D",paste(aa),,1:3],
+                           agedur["M","T1D",paste(aa),,1:3],
+                           agedur["M","T2D",paste(aa),,1:3] ),
+           lty=c(rep("solid",3),"21","63","63"),
+           col=rep(clr,each=6), lwd=c(3,1,1) )
```

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1,6, las=1, bty="n" )
> plot( NA, xlim = c(30,90), ylim = c(45,80),
+           xlab = "Age"   , ylab = expression("Hb"*A[1][c]) )
> for( aa in seq(20,70,5) )
+ matlines( aa+dpr, cbind( agedur["F","T1D",paste(aa),,1],
+                           agedur["F","T2D",paste(aa),,1],
+                           agedur["M","T1D",paste(aa),,1],
+                           agedur["M","T2D",paste(aa),,1] ),
+           lty=c("solid","21"), lend=1, col=rep(clr,each=2), lwd=3 )
```

The figures 4 and 5 indicate that there is an increase in $HbA_{1c}$ in the decade after diagnosis, and subsequently a flattening for T1D patients and even a decline for T2D patients. Moreover there seem to be very little influence of age at diagnosis after some 10 years of DM duration; all curves are largely coincident, with a weak downward slope by age for T1D patients and a somewhat stronger slop among T2D patients.

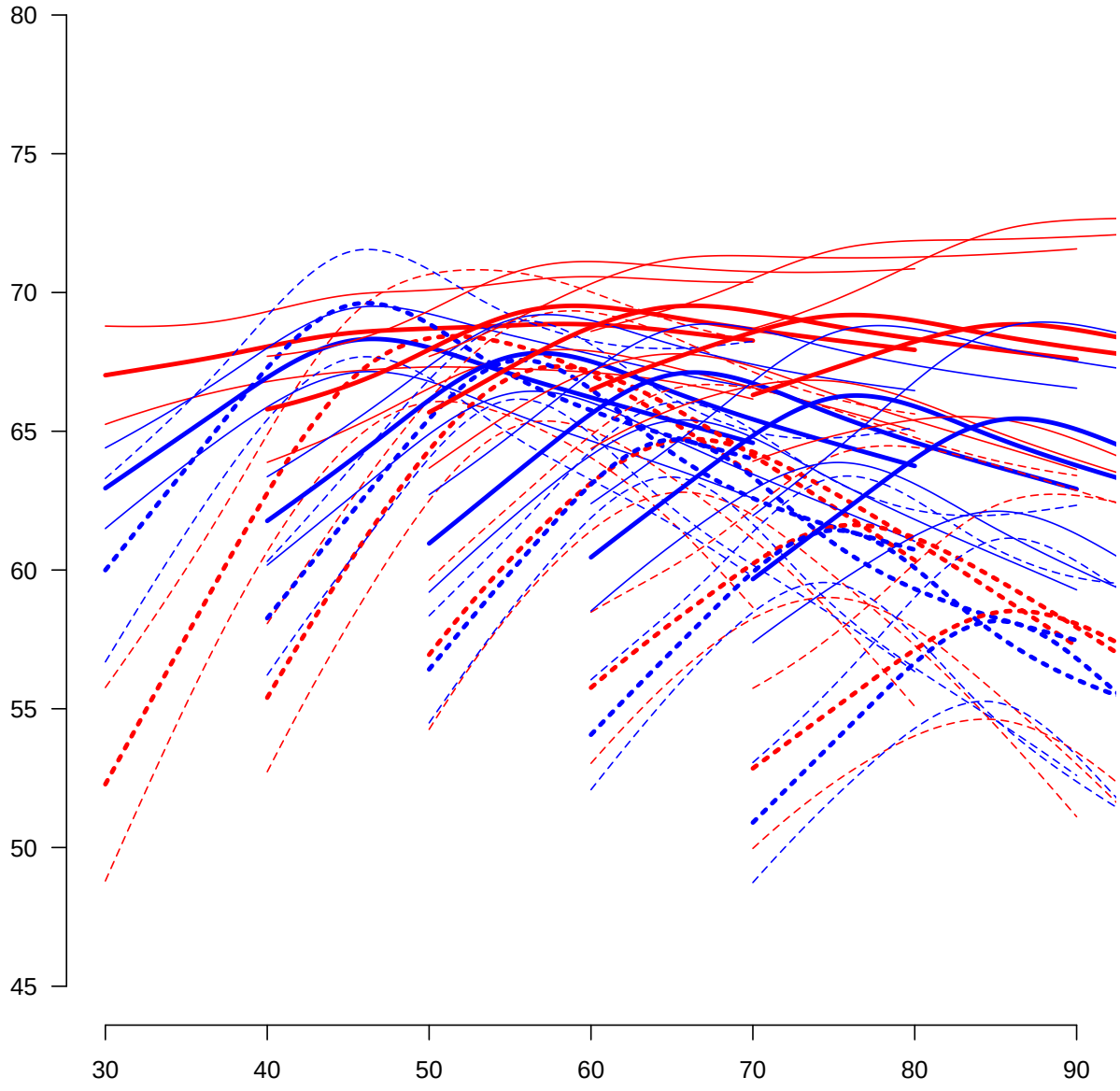


Figure 4: *Effect of age and duration on the mean $HbA_{1c}$in SDC clinic, for ages st onset 30, 40,. . . ,70. Men blue, women red, T1D full lines, T2D broken lines, thin lines are 95% c.i.s.*

### 3.3.3 Seasonal effects

We then plot the estimated seasonal effects — note that January and December both appear twice (the effects in these are the same)

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matplot( hpr, cbind( season["F","T1D",,],
+                      season["F","T2D",,],
+                      season["M","T1D",,],
+                      season["M","T2D",,] ),
+          type="l", lty=c(rep("solid",3),"21","63","63"), lend=1,
+          col=rep(clr,each=6), lwd=c(3,1,1),
+          xaxt="n", xlab="", ylim=c(-3,3),
```

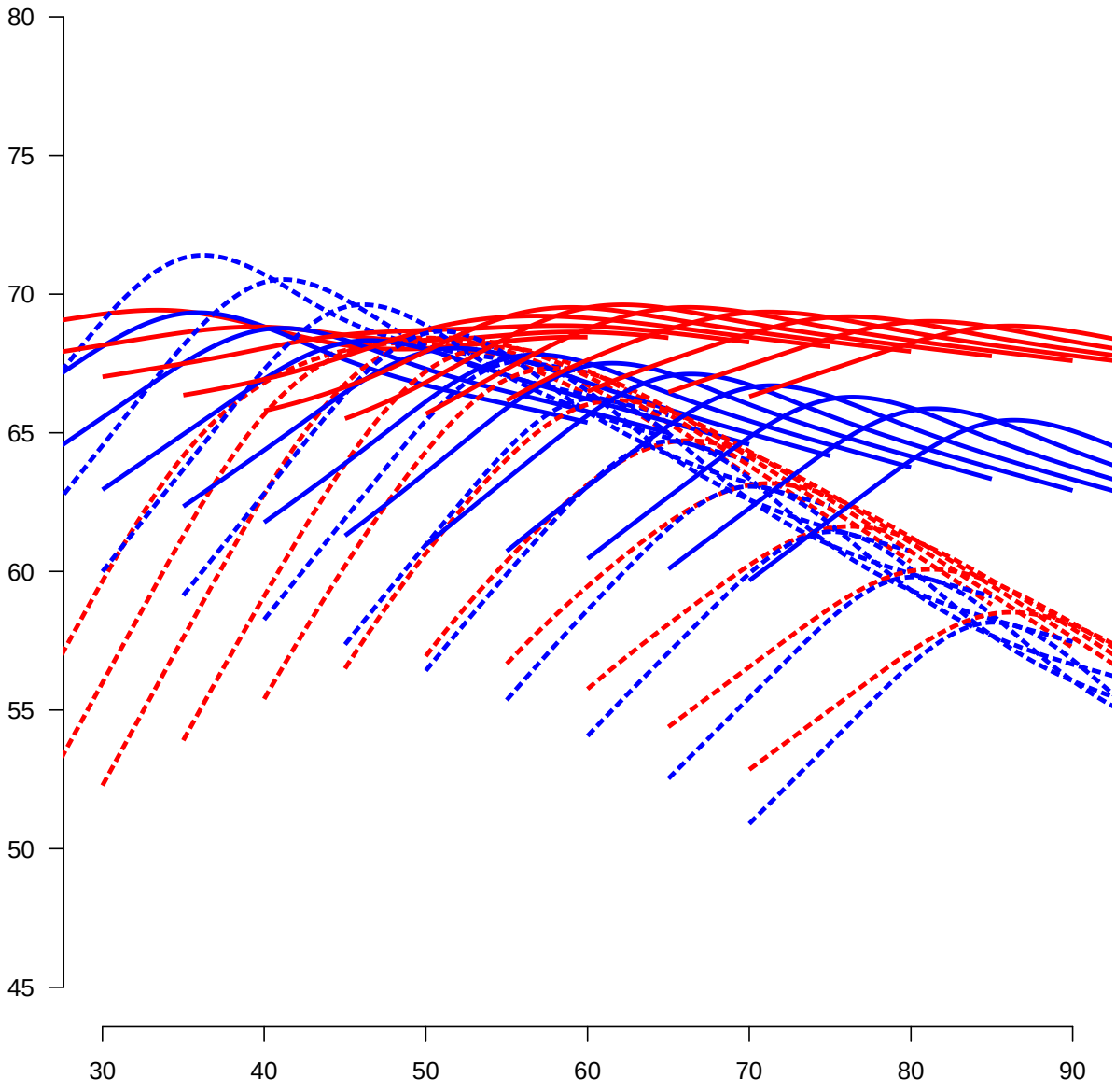


Figure 5: *Effect of age and duration on the mean $HbA_{1c}$in SDC clinic, for ages at onset 20,25,...,70. men blue, women red, T1D full lines, T2D broken lines.*

```
+          ylab=expression("Seasonal variation in Hb"*A[1][c]*" (mmol/mol)") )
> m.len <- c(31,31,28,31,30,31,30,31,31,30,31,30,31,31)
> m.div <- cumsum( c(-31,m.len) )
> m.mid <- m.div[-1] - m.len/2
> m.nam <- format( ISOdate(2004,c(12,1:12,1),1), "%b" )
> axis( side=1, at=m.div, labels=NA )
> axis( side=1, at=m.mid, labels=m.nam, tck=0 )
> abline( v=c(0,365) )
```

The seasonal variation seen in figure 6 is modeled by a 2nd order harmonic, that is with 4 parameters for each combination of sex and diabetes type. There is a remarkably unanimous peak for all 4 groups around 1 February, and a low-plateau from May to

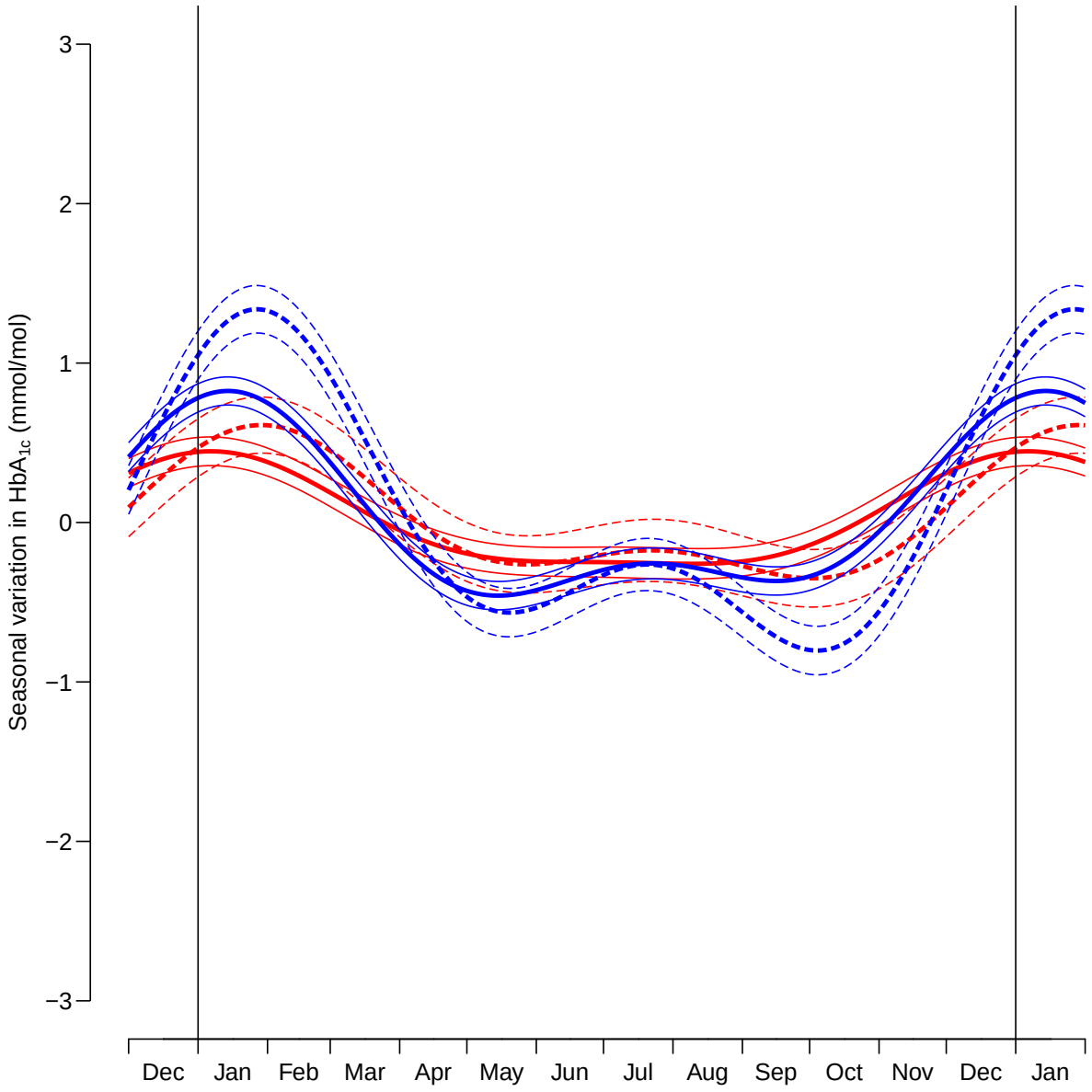


Figure 6: *Seasonal effects in $HbA_{1c}$in the period 1997–2015, men blue, women red, T1D full lines, T2D broken lines, thin lines are 95% c.i.s.*

October; the difference between the two being approximately 1 mmol/mol — possibly slightly larger for men with T2D (broken blue line).

### 3.3.4 Time trends

The overall time trend in $HbA_{1c}$ is of major interest, particularly regarding any changes occurring after 2015-04-01 (which in `cal.yr` terms is 2015.25). So we plot the estimated mean difference in $HbA_{1c}$ relative to 2012-01-01 for the 4 groups — thus we have 4 curves all passing through (2012,0). This is the $HbA_{1c}$ difference

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matplot( tpr, cbind( trends["F","T1D","2012",,],
+                      trends["F","T2D","2012",,],
+                      trends["M","T1D","2012",,],
+                      trends["M","T2D","2012",,] ),
+          type="l", lty=c(rep("solid",3),"21","63","63"), lend=1,
+          col=rep(clr,each=6), lwd=c(3,1,1),
+          xlab="Date of measurement", ylim=c(-5,10),
+          ylab=expression("Hb"*A[1][c]*" (mmol/mol)") )
> abline( h=0, v=c(2012,2015.25) )
> axis( side=1, at=1999:2016, labels=NA )
> rug( t.kn, lwd=3 )
```

```
> par( mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matplot( tpr, cbind( trends["F","T1D","2014",,],
+                      trends["F","T2D","2014",,],
+                      trends["M","T1D","2014",,],
+                      trends["M","T2D","2014",,] ),
+          type="l", lty=c(rep("solid",3),"21","63","63"), lend=1,
+          col=rep(clr,each=6), lwd=c(3,1,1),
+          xlab="Date of measurement", ylim=c(-5,10),
+          ylab=expression("Hb"*A[1][c]*" (mmol/mol)") )
> abline( h=0, v=c(2014,2015.25) )
> axis( side=1, at=1999:2016, labels=NA )
> rug( t.kn, lwd=3 )
```

From figures 7 and 8 which only differ in reference point for the curves, we see that there is an upward trend in the $HbA_{1c}$ measurements after 2015-04-01 (the second vertical line) — particularly for the T1D patients.

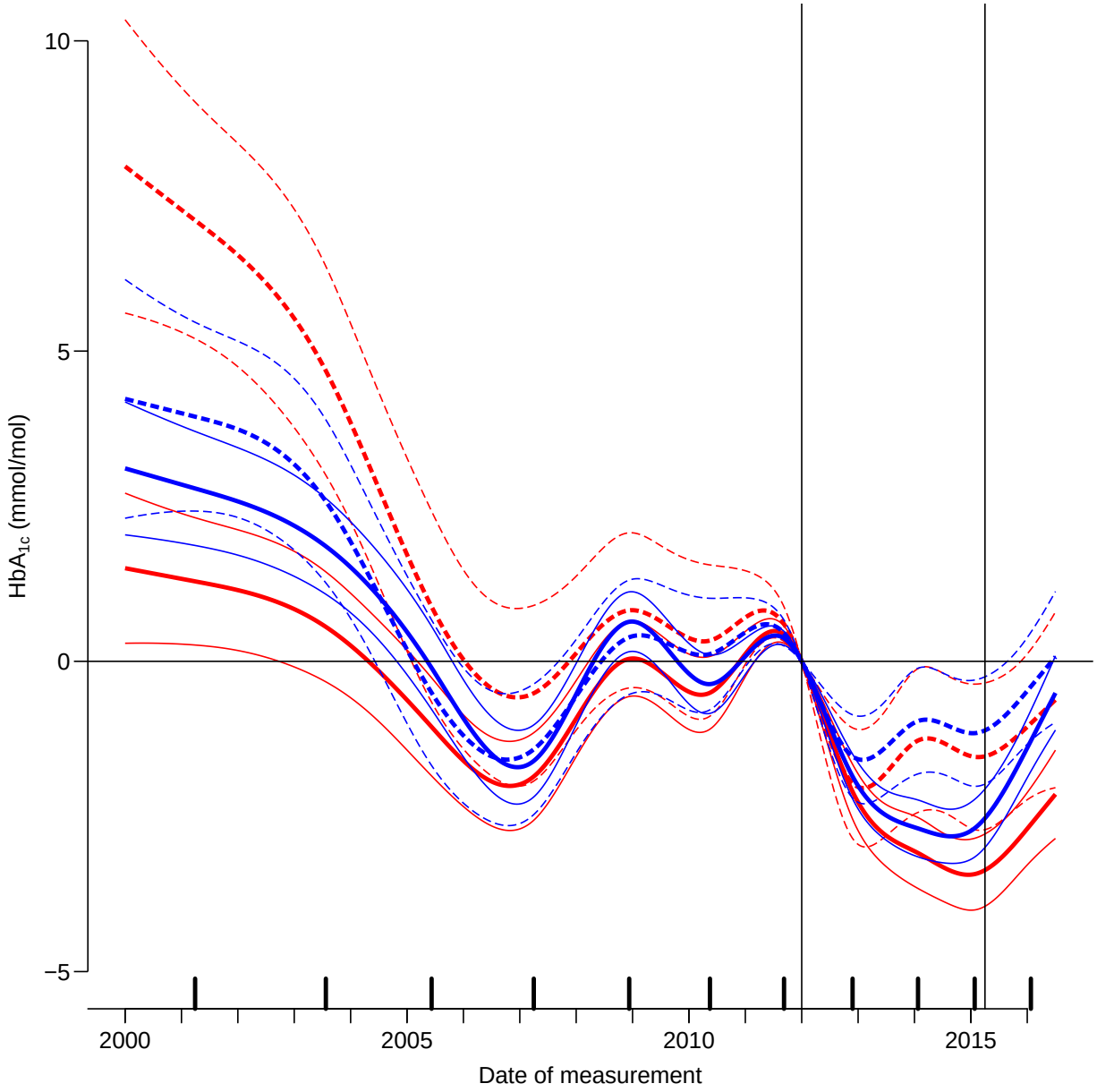


Figure 7: *Trend effects in $HbA_{1c}$relative to 1 January 2012; tick marks represent the start of each calendar year. Men blue, women red, T1D full lines, T2D broken lines, thin lines are 95% c.i.s. The rug at the bottom indicate the knots used in the spline model.*

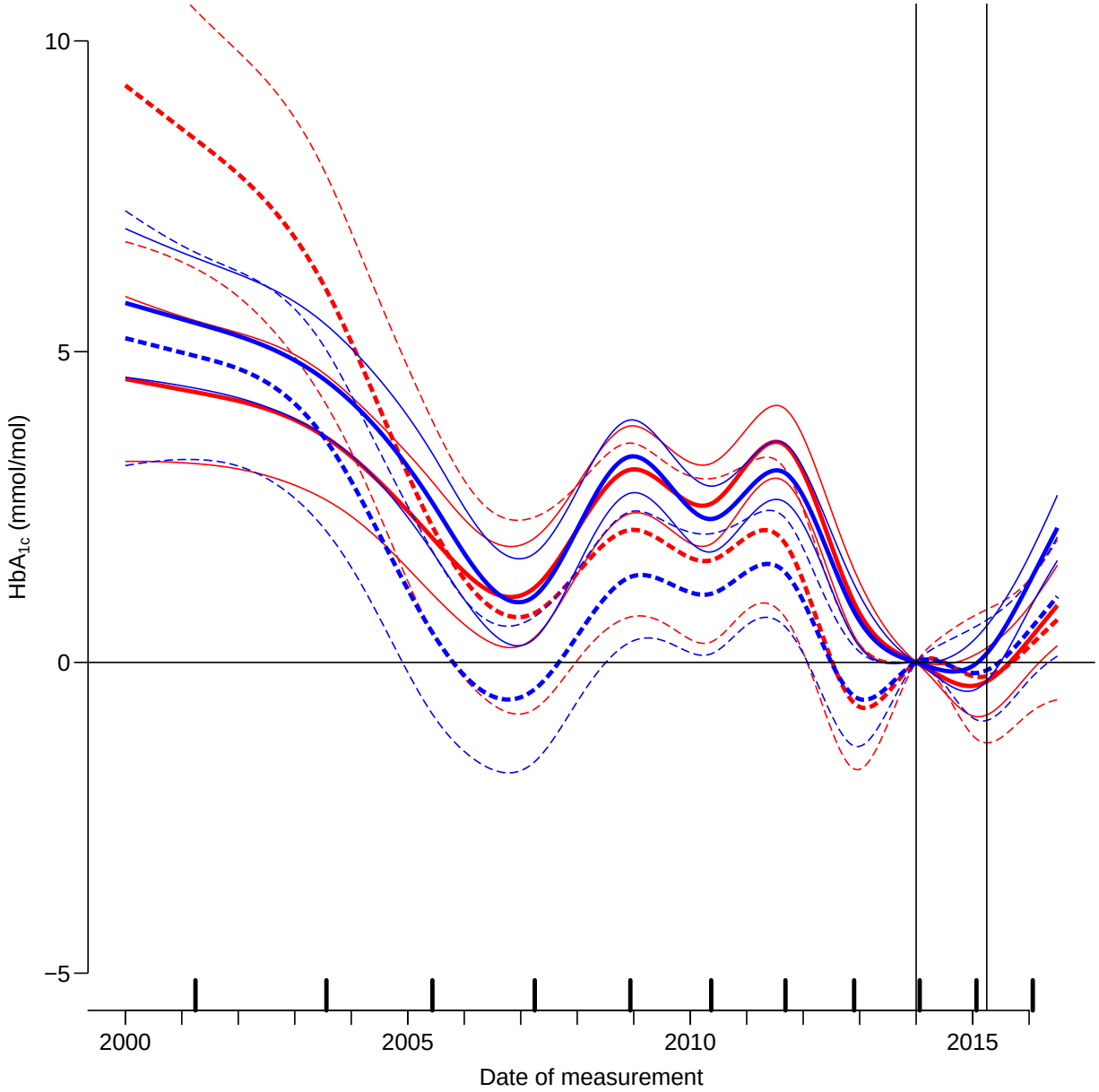


Figure 8: *Trend effects in $HbA_{1c}$relative to 1 January 2014; tick marks represent the start of each calendar year. Men blue, women red, T1D full lines, T2D broken lines, thin lines are 95% c.i.s. The rug at the bottom indicate the knots used in the spline model.*