

Notat: Patient populations in SDC Prevalence and trends

SDCC

November 2019

<http://bendixcarstensen.com/>

Version 2

Compiled Thursday 14th November, 2019, 00:39
from: /home/bendix/sdc/proj/NNF/notatNov19.tex

Bendix Carstensen Steno Diabetes Center Copenhagen, Gentofte, Denmark
& Department of Biostatistics, University of Copenhagen
bcar0029@regionh.dk b@bxc.dk
<http://BendixCarstensen.com>

Contents

1	Introduction	1
2	The diabetes population by region and SDC	2
2.1	Region of residence	2
2.1.1	SDC affiliation	3
2.1.2	Region and SDC	5
3	Prevalence of diabetes by regions	6
3.1	Reading the prevalence data	6
3.2	Prevalence of diabetes by region as of 2017	7
3.3	Age-specific prevalences	9
3.4	Modeling prevalences	11
3.5	Trends in prevalence by region	16
3.5.1	Overall number of cases	16
3.5.2	Proper analysis of prevalence trends	17
3.5.3	Non-linear trends between regions	19
4	DMreg: det rekonstruerede diabetes register hos DST	21
4.1	Steno-algoritme	21
4.1.1	Registre	21
	Sygesikringsregistret: ydelse for fodterapi	22
	Dansk Voksendiabetesdatabase (DVDD)	22
	DiaBasen	23
4.1.2	Datoer i diabetes registeret	23
4.1.3	Inklusionskriterium	25
4.1.4	Registerets dækningsperiode	25
4.1.5	Diabetes type	26
4.1.6	Debutdato og diabetes varighed	26
4.1.7	Anvendte parametre	26
4.2	Antal patienter	27
4.3	Region	27
4.4	Forskelle til RUKS	27
4.5	Fremtid	28

Chapter 1

Introduction

This is the first skeleton to a report for the Directors of the SDCs as well as Board of NNF.

There are still missing a section on incidence trends in regions, due to technical problems.

The more interesting section on complications prevalence is not completed because the definitions of the complication entities to consider are not yet finalized.

The register is based on a long standing and still ongoing endeavour at Clinical Epidemiology to reconstruct a proper diabetes register with 1) a reliable type 1/type 2 classification and 2) high sensitivity and specificity.

The register is hosted in our research project(s) at DST's researcher service, and is derived from existing registers. A detailed (Danish) account is given in chapter 4.

Chapter 2

The diabetes population by region and SDC

2.1 Region of residence

In the DMreg is recorded the person's region of residence at the date of inclusion (at the beginning of the year), and the prevalence calculations in the next chapters are based on this information. The population information is likewise based on the region of residence. But the patients's region of residence is *not* updated durin the life after DM diagnosis

2.1.1 SDC affiliation

It is a somewhat other matter if a person belongs to a given SDC center, primarily because this is much more volatile than the region of residence.

In order to assign persons to clinics we have used the earliest record from DVDD which is after the date of inclusion in DMreg, and classified clinics as SDC or other on the basis of the variable *afdeling* for which we have devised a grouping. This is shown in the following table — the classification is entirely made up at SDCC-KE, and we would be happy to be updated.

	SDCA	SDCC	SDCN	SDCO	SDCS	notSDC
	Midt	Hov	Nord	Syd	Sj11	DK
Ambl						
Alb Endokrin. amb.	.	.	2,500	.	.	.
Alb Endokrinologisk Område	*
Diabetes og stofskifte Ambulatorium Herning	308
Diabetes og stofskifte Ambulatorium Holstebro	316
Diabetesambulatorium	337
Endokrinologi ambulant	3,430
Endokrinologi senge	598
Endokrinologisk Amb., AMH	237
Endokrinologisk Ambulatorium	8
Endokrinologisk Ambulatorium I	3,195
Endokrinologisk Ambulatorium Silkeborg	344
Endokrinologisk Ambulatorium Viborg	343
Endokrinologisk ambulatorium Br	91
Endokrinologisk klinik - Randers	333
FRH Medicinsk Ambulatorium	2,024
FS Endokrinologisk Afd., amb.	75
FS Medicinsk Afd., amb.	868
Far Medicinsk Ambulatorium	86
Far Medicinsk amb.	524
Fælles Medicinsk ambulatorie M - Randers	1,667
HI Endokrinologisk Afd., amb.	2,649
HJR Medicinsk Ambulatorium	604
HL Medicinsk Afd., amb.	963
HOL Endokrinologisk Amb.	1,164	.
HOL Medicinsk Amb. Kalundborg	99
Hjerte-Endo. med amb1	126
Hob Medicinsk Amb.	282
Intern medicinsk klinik	1,771
KAL Endokrinologisk Amb.	428	.
KAL Medicinsk Amb.	990
KOE Endokrinologisk Amb.	1,224	.
KOE Medicinsk Amb.	1,886
Kardiologisk-endokrinologisk ambulatorium E	908
MEDICINSK KLINIK CA, GE	154
Med. endokrin. amb. GLO	104
Med. endokrinologisk amb.	2,182
Med.- Endokrinologisk Amb.C	1,664
Med.-endokrinologisk amb. C	1,561
Medicinsk Ambulatorium - HEH	3,072
Medicinsk Ambulatorium Ringkøbing	14
Medicinsk Ambulatorium, FRH	134
Medicinsk Endokrin. Amb. M	3,547
Medicinsk Endokrinologisk Ambulatorium MEA NBG	556
Medicinsk Endokrinologisk Ambulatorium MEA THG	453
Medicinsk afd., Sygehus Fyn	1,199

Medicinsk afdeling	53
Medicinsk afdeling, Fredericia	90
Medicinsk afdeling, amb.	686
Medicinsk amb., diabetes	859
Medicinsk ambulatorklinik, Sygehus Fyn	1,236
Medicinsk ambulatorium	1,409
Medicinsk ambulatorium M	921
Medicinsk ambulatorium MA	1,546
Medicinsk ambulatorium ita	363
Medicinsk ambulatorium, Fredericia	2,934
Medicinsk ambulatorium, Kolding	1,848
Medicinsk ambulatorium, Skive	388
Medicinsk ambulatorium, Viborg	1,831
Medicinsk dagafsnit	1,736
Medicinsk endokrinologisk afdeling J, ambulatorium	2,337
Medicinsk endokrinologisk klinik PE, amb.	970
Medicinsk klinik FA	1,326
NAE Endokrinologisk Amb.	914	.
NAE Medicinsk Amb.	989
NAE Øjen Amb.	172
NFS Endokrinologisk Amb.	877	.
NFS Medicinsk Amb. Nakskov	211
OUH Medicinsk Amb. (Svendborg)	515
Od Endokrinologisk afdeling M	.	.	.	2,027	.	.
Od Endokrinologisk ambulatorium M	.	.	.	2,539	.	.
ROS Øjen Amb.	8
SH Endokrinologisk Afd., amb.	22
SHS Medicinsk Daghospital Ambulatorier (Tønder)	152
SHS Medicinsk Klinik COM Ambulatorier (Aabenraa)	213
SHS Medicinsk Klinik COM Ambulatorier (Haderslev)	52
SHS Medicinsk Klinik MC Ambulatorier (Sønderborg)	453
SLA Endokrinologisk Amb.	1,159	.
SLB Diabetes/Hormonsygdomme, Medicinsk Ambulatorium (Kolding)	64
SLB Endokrinologisk Ambulatorium (Vejle)	1,445
SLB Gastroenterologisk Ambulatorium (Vejle)	476
SVS Endokrinologisk ambulatorium (Esbjerg)	3,416
SVS Endokrinologisk ambulatorium (Grindsted)	679
Steno Diabetes Center, RH-enheden	.	199
Steno Diabetes center, ambulant	.	7,986
Steno diabetes center,sengeafd	.	*
Thy Med. Amb.	59
Thy Med. sengeafdelinger	954
----
Prak	347781
-----	Midt	Hov	Nord	Syd	Sjll	DK
-----	SDCA	SDCC	SDCN	SDCO	SDCS	notSDC

The '*'s in the table represent table entries smaller than 5 that cannot be exported from DST

2.1.2 Region and SDC

The fraction of patients with a given region of residents is shown in the two following tables — one with all patients alive at some point after 2015-01-01, and one with only those included in the register after this date:

Alive after 2015-1-1		Region i Danmark									
		Nord		Midt		Syd		Hov		Sjll	
		All	%	%	%	%	%	%			
Ambl											
SDCN	2,364	2,249	7.1	50	0.1	29	0.0	17	0.0	19	0.0
SDCA	5,877	75	0.2	5,602	8.8	132	0.2	46	0.1	22	0.0
SDCO	3,928	20	0.1	42	0.1	3,768	5.8	59	0.1	39	0.1
SDCC	6,912	32	0.1	69	0.1	87	0.1	6,300	7.2	424	0.9
SDCS	5,219	10	0.0	26	0.0	19	0.0	259	0.3	4,905	10.1
notSDC	54,440	4,308	13.5	12,190	19.2	14,690	22.7	18,662	21.4	4,590	9.5
All	78,740	6,694	21.0	17,979	28.4	18,725	28.9	25,343	29.1	9,999	20.6
Prak											
notSDC	216,906	25,122	79.0	45,359	71.6	46,095	71.1	61,786	70.9	38,544	79.4
All	295,646	31,816	100	63,338	100	64,820	100	87,129	100	48,543	100

Diagnosed after 2015-1-1		Region i Danmark									
		Nord		Midt		Syd		Hov		Sjll	
		All	%	%	%	%	%				
Ambl											
SDCN	168	165	4.2	*	0.0	*	0.0
SDCA	0.0
SDCO	105	105	1.3
SDCC	170	167	1.7	*	0.0
SDCS	364	*	0.0	*	0.0	360	5.8
notSDC	2,738	149	3.8	1,025	13.5	751	9.3	771	8.0	42	0.7
All	3,545	314	8.0	1,027	14.5	857	10.6	941	9.7	406	6.5
Prak											
notSDC	31,938	3,608	92.0	6,563	86.5	7,226	89.4	8,735	90.3	5,806	93.5
All	35,483	3,922	100	7,590	100	8,083	100	9,676	100	6,212	100

The '*'s in the table represent table entries smaller than 5 that cannot be exported from DST

Chapter 3

Prevalence of diabetes by regions

The following text needs to be groomed; for now it is just a verbatim copy from technical report chapters, but it does include the relevant tables and figures—as well as an unfortunate lot of technicalities.

Figures and tables will appear in a more civilized environment in the presentation.

```
> source( file = "E:/workdata/705093/util/elapsed.R" )
> setwd( "E:/workdata/705093/BXC/demoDM/nyr/" )
> start()
```

```
-----
Home: E:/workdata/705093/BXC/demoDM/nyr
Time: 2019-11-13 19:46:22
-----
```

We also define a couple of functions to enhance readability of large numbers (> 5 digits), by formatting these with commas:

```
> fC <- function( x, d=0, w=9, z=NULL ) formatC( x,
+                                               format = "f",
+                                               big.mark = ",",
+                                               digits = d,
+                                               width = w,
+                                               zero.print = z )
> fCp    <- function( x, d=0, w=9, z=".", ... )
+       noquote( fC( x, d=d, w=w, z=z ), ... )
> fCtable <- function( x, d=0, w=9, z=".", ... )
+       ftable( fC( x, d=d, w=w, z=z ), ... )
```

3.1 Reading the prevalence data

First we read the data set and groom it a bit

```
> library( haven )
> library( Epi )
> prv <- read_sas( "../nydata/prv.sas7bdat" )
> str( prv )
Classes 'tbl_df', 'tbl' and 'data.frame':      31840 obs. of  6 variables:
 $ pdat : num  2007 2007 2007 2007 2007 ...
 $ REG  : num  81 81 81 81 81 81 81 81 81 81 ...
```

```

..- attr(*, "label")= chr "Region i Danmark"
$ state: chr "T1" "T1" "T1" "T1" ...
$ sex : num 1 1 1 1 1 1 1 1 1 1 ...
$ age : num 3 4 5 6 7 8 9 10 11 12 ...
$ n : num 1 3 3 2 5 3 6 10 5 9 ...
- attr(*, "label")= chr "PRV"

> names( prv ) <- tolower( names(prv) )
> prv$reg <- factor( prv$reg,
+ levels=80+1:5,
+ labels=c("Nord","Midt"," Syd"," Hov","Sjll") )
> prv$sex <- factor( prv$sex, labels=c("M","W") )
> # library(viridis)
> # rcol <- rainbow(6)[-2]
> # rcol <- viridis(5)[c(2:5,1)]
> rcol <- c("red","forestgreen","blue","orange","magenta")
> names(rcol) <- levels( prv$reg )
> str( prv )

Classes 'tbl_df', 'tbl' and 'data.frame': 31840 obs. of 6 variables:
 $ pdat : num 2007 2007 2007 2007 2007 ...
 $ reg : Factor w/ 5 levels "Nord","Midt",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ state: chr "T1" "T1" "T1" "T1" ...
 $ sex : Factor w/ 2 levels "M","W": 1 1 1 1 1 1 1 1 1 1 ...
 $ age : num 3 4 5 6 7 8 9 10 11 12 ...
 $ n : num 1 3 3 2 5 3 6 10 5 9 ...
- attr(*, "label")= chr "PRV"

> rcol

```

	Nord	Midt	Syd	Hov	Sjll
	"red"	"forestgreen"	"blue"	"orange"	"magenta"

Note that for nice printing of tables the level names have 4 characters, Syd and Hov are preceded by a space:

```

> levels(prv$reg)
[1] "Nord" "Midt" " Syd" " Hov" "Sjll"
> rperm <- c(" Hov", "Sjll", " Syd", "Midt", "Nord")

```

Finally we save the prevalence data for other analyses:

```

> save( prv, rcol, rperm, fC, fCp, fCtable, file="../nydata/prv.Rda" )

```

3.2 Prevalence of diabetes by region as of 2017

Initially we are only interested in prevalences as of 2017 and persons under 100

```

> pr <- subset( prv, pdat==2017 & age<100 )
> tt <- addmargins( xtabs( n ~ sex + reg + state, data=pr ) )
> fCtable( tt )

```

	state	noDM	T1	T2	Sum
sex reg					
M Nord		281,555	1,664	14,970	298,189
Midt		622,545	3,646	29,941	656,132
Syd		576,881	3,649	30,697	611,227
Hov		854,658	4,437	39,540	898,635

```

      Sjll      390,420      2,288      24,044      416,752
      Sum      2,726,059     15,684     139,192     2,880,935
W   Nord      278,378      1,265      12,071      291,714
      Midt      626,115      2,714      24,503      653,332
      Syd      583,107      2,678      24,753      610,538
      Hov      891,012      3,472      32,965      927,449
      Sjll      397,996      1,801      18,940      418,737
      Sum      2,776,608     11,930     113,232     2,901,770
Sum Nord      559,933      2,929      27,041      589,903
      Midt      1,248,660      6,360      54,444     1,309,464
      Syd      1,159,988      6,327      55,450     1,221,765
      Hov      1,745,670      7,909      72,505     1,826,084
      Sjll      788,416      4,089      42,984      835,489
      Sum      5,502,667     27,614     252,424     5,782,705

```

```
> str(tt)
```

```
'table' num [1:3, 1:6, 1:4] 281555 278378 559933 622545 626115 ...
- attr(*, "dimnames")=List of 3
..$ sex : chr [1:3] "M" "W" "Sum"
..$ reg : chr [1:6] "Nord" "Midt" " Syd" " Hov" ...
..$ state: chr [1:4] "noDM" "T1" "T2" "Sum"
```

```
> prpct <- sweep( tt[,2:3], 1:2, tt[,4], "/" ) * 100
> round( ftable( prpct, col.vars=c(3,1) ), 2 )
```

reg	state T1			T2		
	sex M	W	Sum	M	W	Sum
Nord	0.56	0.43	0.50	5.02	4.14	4.58
Midt	0.56	0.42	0.49	4.56	3.75	4.16
Syd	0.60	0.44	0.52	5.02	4.05	4.54
Hov	0.49	0.37	0.43	4.40	3.55	3.97
Sjll	0.55	0.43	0.49	5.77	4.52	5.14
Sum	0.54	0.41	0.48	4.83	3.90	4.37

```
> round( ftable( tt[,3]/tt[,4]*100, row.vars=2 ), 1 )
```

reg	sex M	W	Sum
Nord	5.0	4.1	4.6
Midt	4.6	3.8	4.2
Syd	5.0	4.1	4.5
Hov	4.4	3.6	4.0
Sjll	5.8	4.5	5.1
Sum	4.8	3.9	4.4

So we see that the overall prevalence is smaller in Region H than other places. A slightly more detailed picture can be obtained by looking at the prevalences by age in the different regions:

```
> nn <- addmargins( xtabs( n ~ sex + age + reg + state, data=pr ), 4 )
> str( nn )
```

```
'table' num [1:2, 1:100, 1:5, 1:4] 3020 2859 2920 2712 2778 ...
- attr(*, "dimnames")=List of 4
..$ sex : chr [1:2] "M" "W"
..$ age : chr [1:100] "0" "1" "2" "3" ...
..$ reg : chr [1:5] "Nord" "Midt" " Syd" " Hov" ...
..$ state: chr [1:4] "noDM" "T1" "T2" "Sum"
```

```
> t2 <- as.data.frame( nn[,,"T2" ] ) ; names(t2)[4] <- "x"
> N <- as.data.frame( nn[,,"Sum" ] ) ; names(N)[4] <- "n"
> aa <- merge( t2, N )
> str( aa )
```

```
'data.frame':      1000 obs. of  5 variables:
 $ sex: Factor w/ 2 levels "M","W": 1 1 1 1 1 1 1 1 1 1 ...
 $ age: Factor w/ 100 levels "0","1","2","3",...: 1 1 1 1 1 2 2 2 2 2 ...
 $ reg: Factor w/ 5 levels "Nord","Midt",...: 4 3 2 1 5 4 3 2 1 5 ...
 $ x  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ n  : num  11364 6124 7419 3020 3931 ...

> aa$age <- as.numeric( as.character(aa$age) ) + 0.5
> head(aa)

  sex age reg x    n
1  M 0.5 Hov 0 11364
2  M 0.5 Syd 0  6124
3  M 0.5 Midt 0  7419
4  M 0.5 Nord 0  3020
5  M 0.5 Sjl1 0  3931
6  M 1.5 Hov 0 10668
```

However, we first take a look at the age-distributions in each region. Then at the age-distributions *relative* to Region H:

The age-distributions are quite different in the regions, so the absolute prevalence of DM may not be a relevant measure.

3.3 Age-specific prevalences

The age-specific prevalences for men and women for the different regions, is modeled by penalized splines:

```
> library( mgcv )
> apt <- 0:99+0.5
> prf <- data.frame( age=apt )
> apr <- NArray( list( age = apt,
+                     sex = levels(aa$sex),
+                     reg = levels(aa$reg),
+                     est = c("est","lo","hi") ) )
> str( apr )
logi [1:100, 1:2, 1:5, 1:3] NA NA NA NA NA NA ...
- attr(*, "dimnames")=List of 4
..$ age: chr [1:100] "0.5" "1.5" "2.5" "3.5" ...
..$ sex: chr [1:2] "M" "W"
..$ reg: chr [1:5] "Nord" "Midt" " Syd" " Hov" ...
..$ est: chr [1:3] "est" "lo" "hi"

> app <- apr
> totdev <- 0
> for( sx in levels(aa$sex) )
+ {
+   mp <- gam( cbind(x,n-x) ~ s(age,k=15) + reg, family=binomial(link=log),
+             data=subset(aa, sex==sx ) )
+   for( rg in levels(aa$reg) )
+   {
+     mm <- gam( cbind(x,n-x) ~ s(age,k=15), family=binomial(link=log),
+               data=subset(aa, sex==sx & reg==rg ) )
+     totdev <- totdev + deviance( mm )
+     apr[,sx,rg,] <- ci.pred( mm, prf )*100
+     app[,sx,rg,] <- ci.pred( mp, data.frame(prf,reg=rg) )*100
+   }
+ }
```

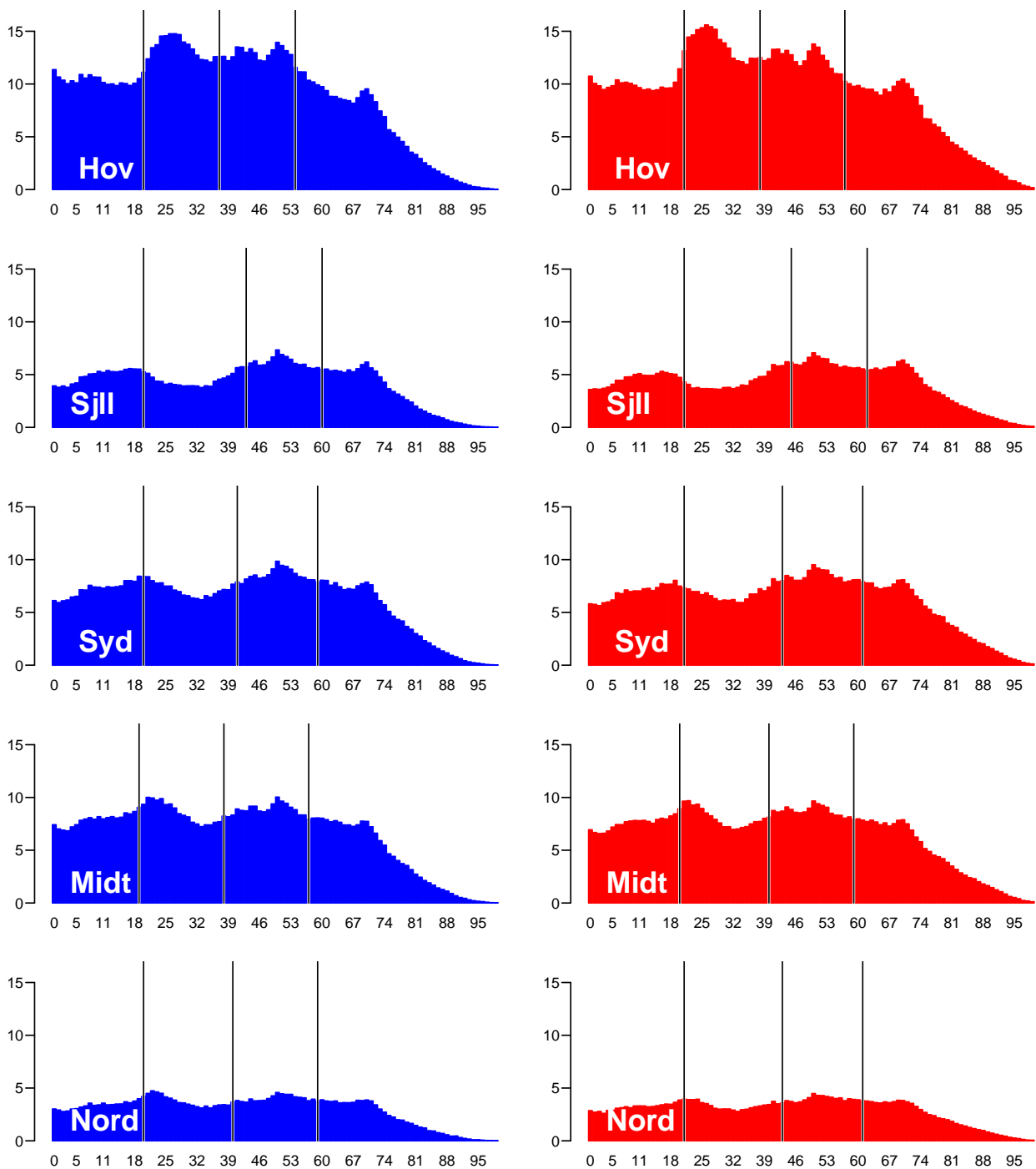


Figure 3.1: Absolute number of persons in the 100 age-classes for each sex and region. Vertical lines divide population in quarters. ./graph/regPrev-abs

```
+ }
> totdev
[1] 862.7574
```

With the predicted age-specific prevalences for man, resp women we can show them by region.

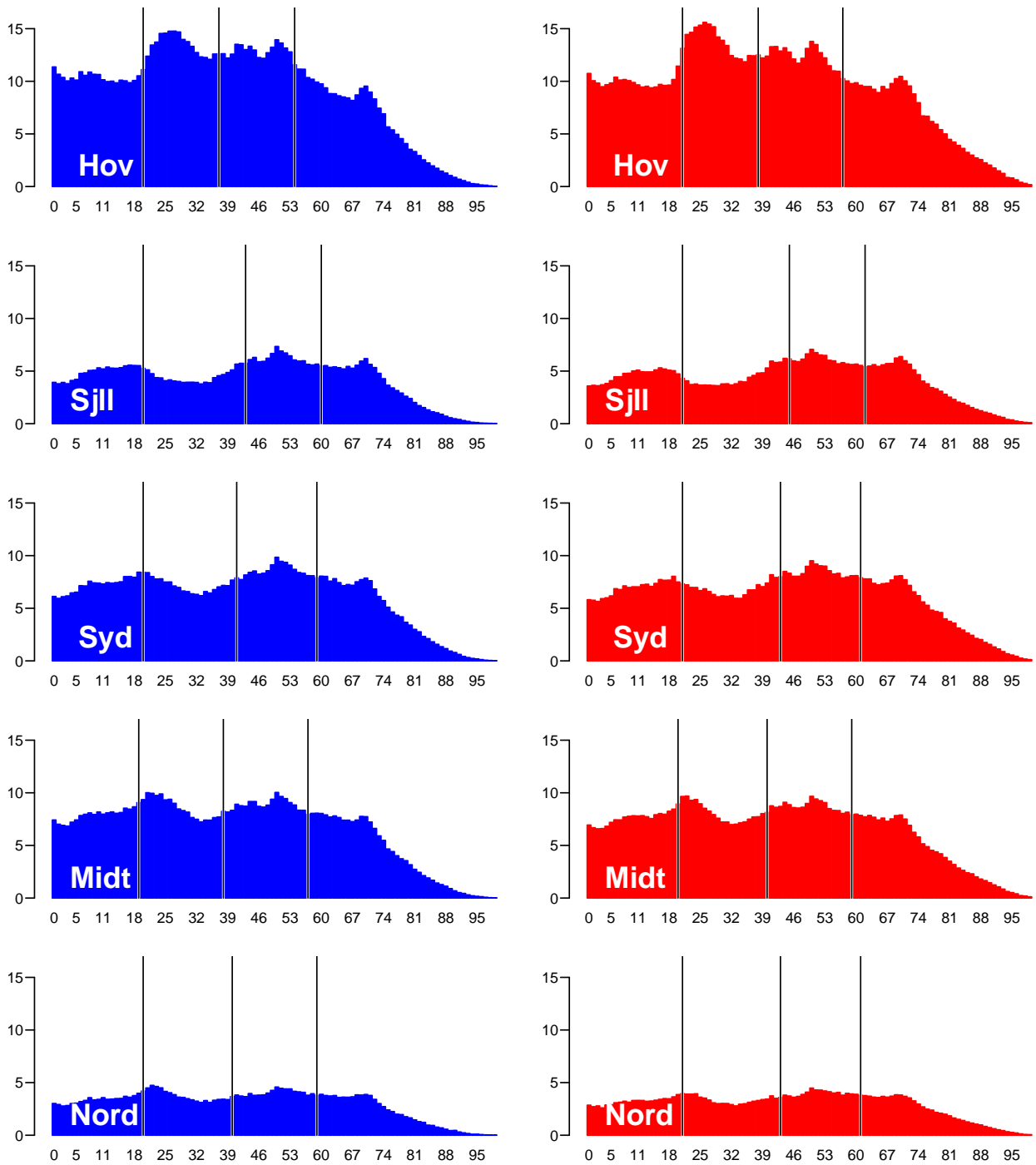


Figure 3.2: Relative size of the of population relative to the population in Hov in the 100 age-classes for each sex and region ./graph/regPrev-abs

3.4 Modeling prevalences

From figure 3.5 it is pretty clear that any proportionality assumption for prevalences among men and women would be pretty far fetched. A little less so for a proportionality assumption between regions *within* sex.

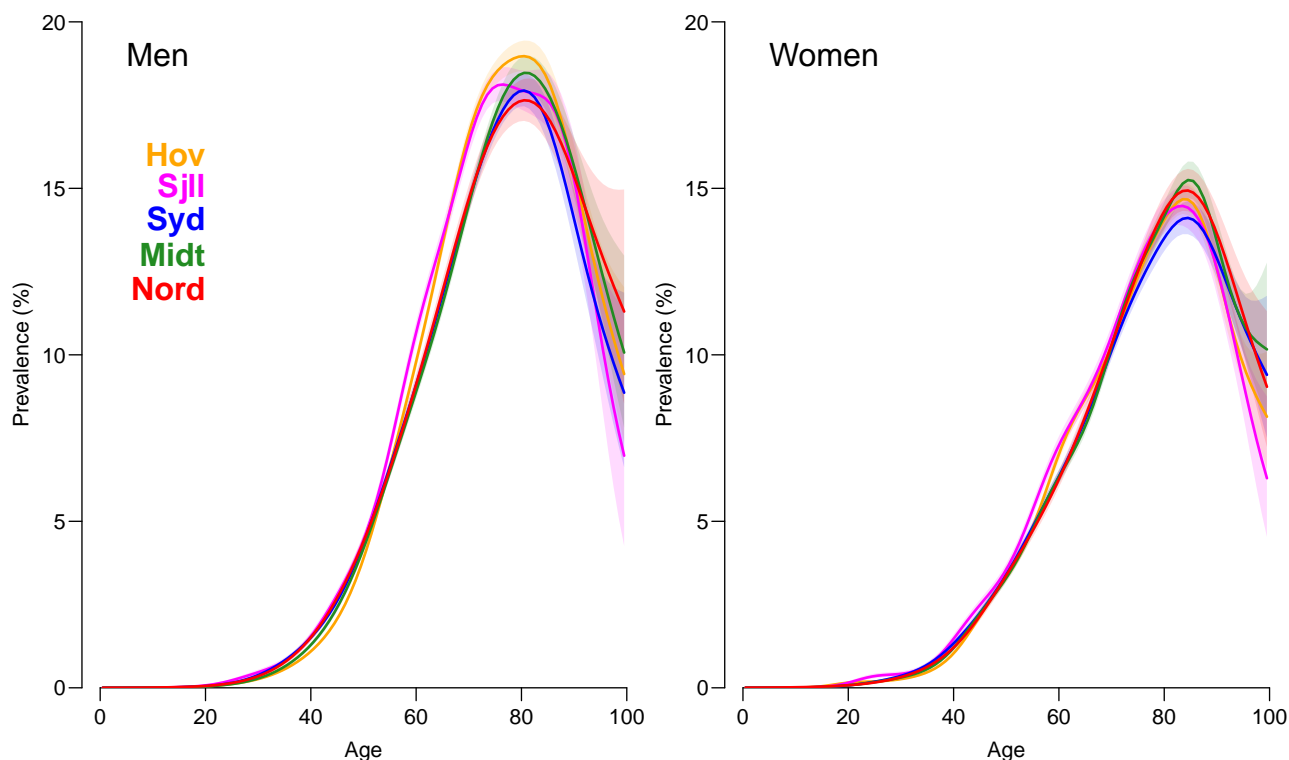


Figure 3.3: Age-specific prevalences by sex and region for T2D, separate curves for region and sex. Shaded areas are 95% confidence intervals.

./graph/regPrev-aprev

We do both, (and and extract the relative prevalences with region H as the reference:

```
> mf <- gam( cbind(x,n-x) ~ s(age,k=15,by=interaction(sex,reg)),
+           family = binomial(link=log),
+           data = transform(aa,reg=relevel(reg,4)) )
> mis <- gam( cbind(x,n-x) ~ s(age,k=15,by=sex) + sex + reg:sex,
+           family = binomial(link=log),
+           data = transform(aa,reg=relevel(reg,4)) )
> mrs <- gam( cbind(x,n-x) ~ s(age,k=15,by=sex) + sex + reg,
+           family = binomial(link=log),
+           data = transform(aa,reg=relevel(reg,4)) )
> totdev # sum of deviances from 10 models for each sex/region
[1] 862.7574
> anova( mf, mis, mrs, test="Chisq" )
Analysis of Deviance Table

Model 1: cbind(x, n - x) ~ s(age, k = 15, by = interaction(sex, reg))
Model 2: cbind(x, n - x) ~ s(age, k = 15, by = sex) + sex + reg:sex
Model 3: cbind(x, n - x) ~ s(age, k = 15, by = sex) + sex + reg
  Resid. Df Resid. Dev      Df Deviance Pr(>Chi)
1     887.82     874.16
2     964.30    1486.39 -76.4838  -612.23 < 2.2e-16 ***
3     968.30    1518.91  -4.0012  -32.52 1.502e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary(mis)
```

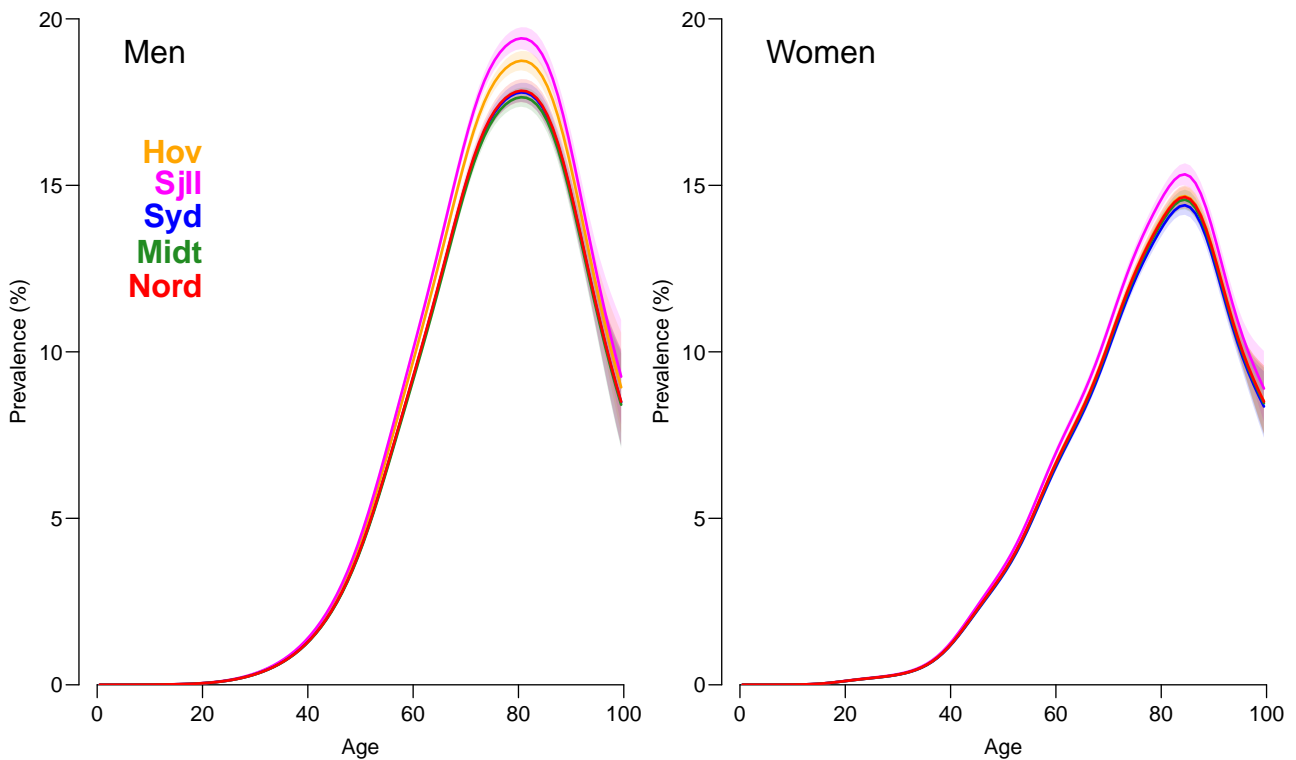


Figure 3.4: Age-specific prevalences of T2D by sex and region assuming proportionality between regions within each sex. Shaded areas are 95% confidence intervals. `./graph/regPrev-aprop`

Family: binomial
Link function: log

Formula:
`cbind(x, n - x) ~ s(age, k = 15, by = sex) + sex + reg:sex`

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.538275	0.037237	-121.876	< 2e-16 ***
sexW	-0.042753	0.054090	-0.790	0.4293
sexM:regNord	-0.049223	0.008973	-5.486	4.12e-08 ***
sexW:regNord	-0.001818	0.010121	-0.180	0.8575
sexM:regMidt	-0.060235	0.007163	-8.409	< 2e-16 ***
sexW:regMidt	-0.007218	0.008032	-0.899	0.3688
sexM:reg Syd	-0.052022	0.007112	-7.314	2.59e-13 ***
sexW:reg Syd	-0.018843	0.008013	-2.352	0.0187 *
sexM:regSjll	0.035312	0.007617	4.636	3.55e-06 ***
sexW:regSjll	0.043684	0.008675	5.036	4.76e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(age):sexM	10.66	11.72	77146	<2e-16 ***
s(age):sexW	13.82	13.98	61779	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

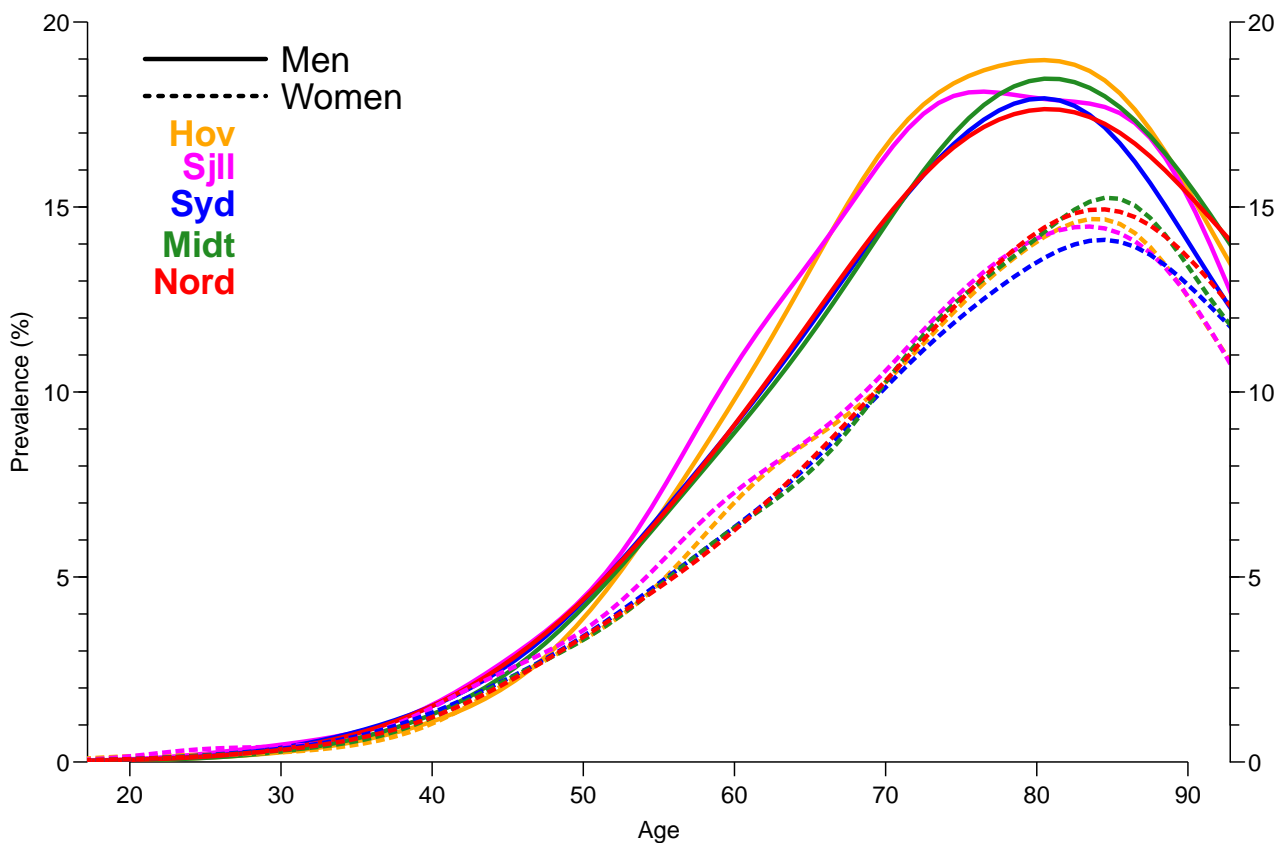


Figure 3.5: Age-specific prevalences by sex and region.

./graph/regPrev-prev

R-sq.(adj) = 0.995 Deviance explained = 99.6%
 UBRE = 0.55536 Scale est. = 1 n = 1000

```
> round( cbind( ci.exp( mis, subset="M:" ),
+             ci.exp( mis, subset="W:" ) ), 2 )
```

	exp(Est.)	2.5%	97.5%	exp(Est.)	2.5%	97.5%
sexM:regNord	0.95	0.94	0.97	1.00	0.98	1.02
sexM:regMidt	0.94	0.93	0.95	0.99	0.98	1.01
sexM:reg Syd	0.95	0.94	0.96	0.98	0.97	1.00
sexM:regSjll	1.04	1.02	1.05	1.04	1.03	1.06

```
> summary(mrs)
```

Family: binomial
 Link function: log

Formula:

```
cbind(x, n - x) ~ s(age, k = 15, by = sex) + sex + reg
```

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.549392	0.037149	-122.464	< 2e-16 ***
sexW	-0.018022	0.053781	-0.335	0.738
regNord	-0.028261	0.006714	-4.209	2.57e-05 ***
regMidt	-0.036749	0.005346	-6.874	6.26e-12 ***
reg Syd	-0.037223	0.005320	-6.997	2.61e-12 ***
regSjll	0.039356	0.005724	6.876	6.17e-12 ***

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:
```

```
      edf Ref.df Chi.sq p-value
s(age):sexM 10.66  11.72  77173 <2e-16 ***
s(age):sexW 13.82  13.98  61850 <2e-16 ***
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.995  Deviance explained = 99.6%
```

```
UBRE = 0.57987  Scale est. = 1          n = 1000
```

```
> round( ci.exp( mrs, subset=2:6 ), 2 )
```

```
      exp(Est.) 2.5% 97.5%
sexW      0.98 0.88  1.09
regNord   0.97 0.96  0.99
regMidt   0.96 0.95  0.97
reg Syd   0.96 0.95  0.97
regSjll   1.04 1.03  1.05
```

The deviance for the separate models is not quite the same as the deviance for the interaction model, but sufficiently close to be credible.

We see that region Sjælland is higher than region H whereas the other regions are below. That pattern is also seen in the (obviously very inadequate) model where sex is used as main effect too.

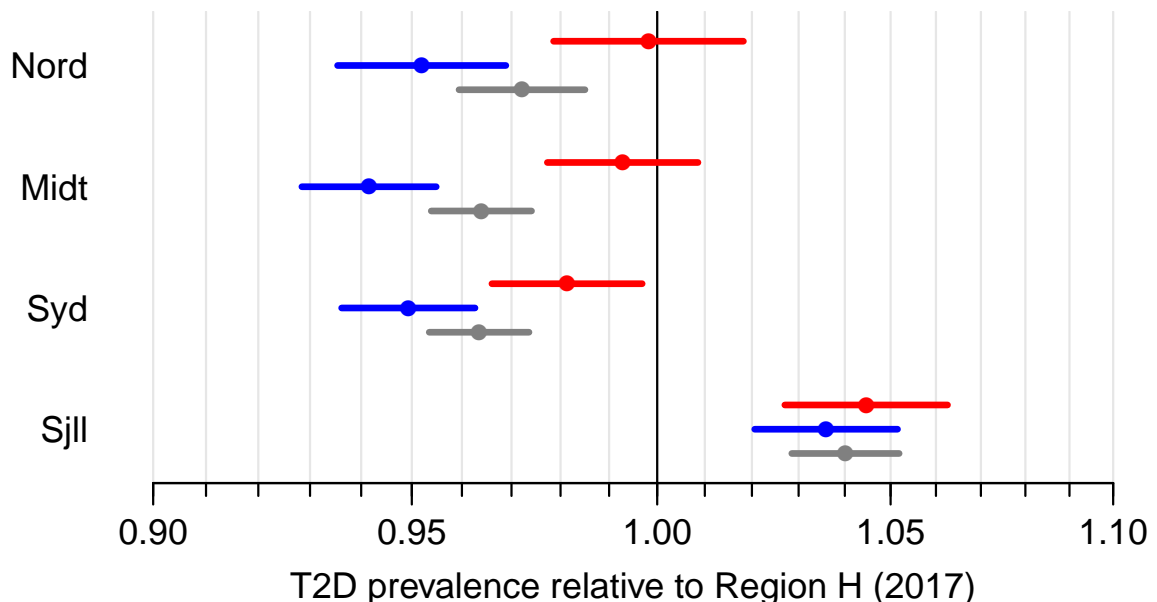


Figure 3.6: *Prevalences of diabetes relative to region H as of 2017. Red:Women, blue:Men, gray:M+W (main effects model).*

`./graph/regPrev-foret`

```
> source( file = "E:/workdata/705093/util/elapsed.R" )
> setwd( "E:/workdata/705093/BXC/demoDM/nyr/" )
> start()
```

```
-----
Home: E:/workdata/705093/BXC/demoDM/nyr
Time: 2019-11-13 20:32:26
-----
```

```
> library( Epi )
> library( mgcv )
> load( file="./nydata/prv.Rda", verbose=TRUE )
```

Loading objects:

```
prv
rcol
rperm
fC
fCp
fCtable
```

3.5 Trends in prevalence by region

```
> tt <- addmargins( xtabs( n ~ pdat + reg + state, data=prv ), 3 )
> # Type1 diabetes:
> round( cbind( tt[, ,2], tt[, ,2]/tt[, ,"Sum"]*100 ), 2 )
```

	Nord	Midt	Syd	Hov	Sjll	Nord	Midt	Syd	Hov	Sjll
2007	2607	5618	5748	6994	3957	0.45	0.46	0.48	0.43	0.48
2008	2660	5671	5804	7036	3969	0.46	0.46	0.49	0.42	0.48
2009	2680	5721	5911	7111	3963	0.46	0.46	0.49	0.42	0.48
2010	2711	5810	5940	7194	3972	0.47	0.46	0.50	0.42	0.48
2011	2724	5901	5978	7285	3935	0.47	0.47	0.50	0.43	0.48
2012	2747	5978	5994	7400	3918	0.47	0.47	0.50	0.43	0.48
2013	2778	6005	6072	7504	3918	0.48	0.47	0.51	0.43	0.48
2014	2815	6088	6137	7585	3949	0.48	0.48	0.51	0.43	0.48
2015	2855	6167	6198	7710	3988	0.49	0.48	0.51	0.43	0.48
2016	2903	6277	6268	7821	4069	0.49	0.48	0.52	0.43	0.49
2017	2929	6360	6327	7909	4089	0.50	0.49	0.52	0.43	0.49

```
> # Type2 diabetes:
> round( cbind( tt[, ,3], tt[, ,3]/tt[, ,"Sum"]*100 ), 1 )
```

	Nord	Midt	Syd	Hov	Sjll	Nord	Midt	Syd	Hov	Sjll
2007	15170	28618	31160	44382	24188	2.6	2.3	2.6	2.7	2.9
2008	16139	30589	33102	46321	25558	2.8	2.5	2.8	2.8	3.1
2009	17265	32903	35428	49035	27339	3.0	2.6	3.0	2.9	3.3
2010	18304	35637	37868	51687	29078	3.2	2.8	3.2	3.1	3.5
2011	19410	38616	40497	55219	31241	3.3	3.1	3.4	3.2	3.8
2012	21746	43754	44608	60827	34283	3.7	3.5	3.7	3.5	4.2
2013	23513	46864	47374	64719	36829	4.0	3.7	3.9	3.7	4.5
2014	24481	48828	49666	66795	38645	4.2	3.8	4.1	3.8	4.7
2015	25393	50681	51483	68755	39859	4.4	3.9	4.3	3.9	4.8
2016	26298	52417	53274	70719	41308	4.5	4.0	4.4	3.9	5.0
2017	27051	54461	55471	72533	43000	4.6	4.2	4.5	4.0	5.1

3.5.1 Overall number of cases

We will first just plot the overall number of cases by date of prevalence (1 January 2007–2017):

3.5.2 Proper analysis of prevalence trends

As we saw in the previous section there is a substantial difference in age composition between the different regions, but the data set allows us to take different age and sex-composition into account when evaluating trends in prevalence. To this end we need a wide dataset—cases of T1, T2 and noDM must be side by side:

```
> df1 <- subset(prv,state=="T1")[,-grep('state',names(prv))]
> df2 <- subset(prv,state=="T2")[,-grep('state',names(prv))]
> dfn <- subset(prv,state=="noDM")[,-grep('state',names(prv))]
> names(df1)[grep("n",names(df1))]<-"T1"
> names(df2)[grep("n",names(df2))]<-"T2"
> names(dfn)[grep("n",names(dfn))]<-"noDM"
> prw <- merge( df1, merge( df2, dfn, all=T ), all=T )
> str( prw )
'data.frame':      11825 obs. of  7 variables:
 $ pdat: num  2007 2007 2007 2007 2007 ...
 $ reg  : Factor w/ 5 levels "Nord","Midt",...: 1 1 1 1 1 1 1 1 1 1 ...
 $ sex  : Factor w/ 2 levels "M","W": 1 1 1 1 1 1 1 1 1 1 ...
 $ age  : num   0  1  2  3  4  5  6  7  8  9 ...
 $ T1   : num  NA NA NA  1  3  3  2  5  3  6 ...
 $ T2   : num  NA NA NA NA NA NA NA NA NA NA ...
 $ noDM: num  3225 3452 3274 3321 3409 ...
> for( i in 5:7 ) prw[,i] <- ifelse( is.na(prw[,i]), 0, prw[,i] )
> prw$N <- prw$T1 + prw$T2 + prw$noDM
> summary(prw)
```

	pdat	reg	sex	age	T1	T2
Min.	:2007	Nord:2326	M:5866	Min. : 0.0	Min. : 0.00	Min. : 0.0
1st Qu.:	:2009	Midt:2375	W:5959	1st Qu.: 26.0	1st Qu.: 4.00	1st Qu.: 3.0
Median	:2012	Syd:2361		Median : 53.0	Median : 20.00	Median : 60.0
Mean	:2012	Hov:2416		Mean : 53.3	Mean : 24.33	Mean : 186.1
3rd Qu.:	:2015	Sjll:2347		3rd Qu.: 80.0	3rd Qu.: 38.00	3rd Qu.: 302.0
Max.	:2017			Max. :119.0	Max. :119.00	Max. :1599.0
	noDM	N				
Min.	: 0	Min. : 1				
1st Qu.:	: 2234	1st Qu.: 2556				
Median	: 4667	Median : 4965				
Mean	: 5005	Mean : 5215				
3rd Qu.:	: 7704	3rd Qu.: 7972				
Max.	:15523	Max. :15599				

So for trends in prevalence we can fit log-link binomial models that quantify how prevalence has changed between the regions

Two types models will be fitted here for each combination of sex and type of DM:

1. the first ones are separate smooth functions by time for each region, controlling for age, these will be used to show how predicted trends in prevalence develop over time.
2. the second ones are linear versions that summarizes the increase in prevalence over the last 10 yrs in each region.

```
> ms1 <- gam( cbind(T1,N) ~ reg+s(pdat,by=reg)+s(pdat)+s(age),
+           family=binomial(link=log), data=subset(prw,sex=="M") )
> ml1 <- update( ms1, . ~ reg + pdat:reg + s(age) )
```

```

> ws1 <- gam( cbind(T1,N) ~ reg+s(pdat,by=reg)+s(pdat)+s(age),
+           family=binomial(link=log), data=subset(prw,sex=="W") )
> wl1 <- update( ws1, . ~ reg + pdat:reg + s(age) )
> ms2 <- gam( cbind(T2,N) ~ reg+s(pdat,by=reg)+s(pdat)+s(age),
+           family=binomial(link=log), data=subset(prw,sex=="M") )
> ml2 <- update( ms2, . ~ reg + pdat:reg + s(age) )
> ws2 <- gam( cbind(T2,N) ~ reg+s(pdat,by=reg)+s(pdat)+s(age),
+           family=binomial(link=log), data=subset(prw,sex=="W") )
> wl2 <- update( ws2, . ~ reg + pdat:reg + s(age) )

```

We can formally test whether the trends are non-linear:

```

> anova( ms1, ml1, test="Chisq" )
Analysis of Deviance Table

Model 1: cbind(T1, N) ~ reg + s(pdat, by = reg) + s(pdat) + s(age)
Model 2: cbind(T1, N) ~ reg + s(age) + reg:pdat
  Resid. Df Resid. Dev      Df  Deviance Pr(>Chi)
1      5847      6944.8
2      5847      6944.8 -0.0019122 -0.0001339 0.008601 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova( ws1, wl1, test="Chisq" )
Analysis of Deviance Table

Model 1: cbind(T1, N) ~ reg + s(pdat, by = reg) + s(pdat) + s(age)
Model 2: cbind(T1, N) ~ reg + s(age) + reg:pdat
  Resid. Df Resid. Dev      Df  Deviance Pr(>Chi)
1      5939.4      6627.5
2      5940.0      6628.3 -0.60023 -0.76628  0.2311

> anova( ms2, ml2, test="Chisq" )
Analysis of Deviance Table

Model 1: cbind(T2, N) ~ reg + s(pdat, by = reg) + s(pdat) + s(age)
Model 2: cbind(T2, N) ~ reg + s(age) + reg:pdat
  Resid. Df Resid. Dev      Df  Deviance Pr(>Chi)
1      5834.9      7896.6
2      5847.0      9117.9 -12.096  -1221.3 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> anova( ws2, wl2, test="Chisq" )
Analysis of Deviance Table

Model 1: cbind(T2, N) ~ reg + s(pdat, by = reg) + s(pdat) + s(age)
Model 2: cbind(T2, N) ~ reg + s(age) + reg:pdat
  Resid. Df Resid. Dev      Df  Deviance Pr(>Chi)
1      5930.4      8064.2
2      5940.0      8964.5 -9.5997  -900.32 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

We see very large deviations from linear trends for T2 but not for T1.

The linear models gives the following annual trends

```

> trend <- NArray( list( type=c("T1","T2"),
+                          sex=c("M","W"),
+                          reg=levels(prw$reg),
+                          est=c("chg","lo","hi") ) )
> e2p <- function(x) (x-1)*100
> trend["T1","M",,] <- e2p( ci.exp( ml1, subset='pdat' ) )
> trend["T1","W",,] <- e2p( ci.exp( wl1, subset='pdat' ) )
> trend["T2","M",,] <- e2p( ci.exp( ml2, subset='pdat' ) )
> trend["T2","W",,] <- e2p( ci.exp( wl2, subset='pdat' ) )
> # Annual trend in prevalence over the period 2007-17
> round( ftable( trend[,rperm,], row.vars=2:3 ), 2 )

```

	sex	reg	type	est	T1	lo	hi	T2	lo	hi
	M	Hov	chg	0.08	-0.20	0.37	3.26	3.16	3.36	
		Sjll	chg	-0.17	-0.56	0.22	3.97	3.84	4.11	
		Syd	chg	0.65	0.33	0.97	4.19	4.07	4.31	
		Midt	chg	0.49	0.17	0.81	4.69	4.56	4.81	
		Nord	chg	0.99	0.52	1.47	4.57	4.40	4.75	
	W	Hov	chg	0.45	0.12	0.78	3.60	3.49	3.72	
		Sjll	chg	0.28	-0.17	0.73	3.94	3.78	4.09	
		Syd	chg	0.71	0.34	1.09	4.02	3.89	4.16	
		Midt	chg	0.77	0.40	1.15	4.38	4.24	4.51	
		Nord	chg	0.78	0.24	1.33	4.23	4.04	4.42	

3.5.3 Non-linear trends between regions

We can look at the non-linear trends for the prevalences, but making a prediction for 35 year old T1D and 75 old T2 persons for each region:

```

> pm1 <- pw1 <- pm2 <- pw2 <- NULL
> for( rg in rperm )
+   {
+     nd1 <- data.frame( pdat=2007:2017, reg=rg, age=35 )
+     pm1 <- cbind( pm1, ci.pred(ms1,nd1) )
+     pw1 <- cbind( pw1, ci.pred(ws1,nd1) )
+     nd2 <- data.frame( pdat=2007:2017, reg=rg, age=75 )
+     pm2 <- cbind( pm2, ci.pred(ms2,nd2) )
+     pw2 <- cbind( pw2, ci.pred(ws2,nd2) )
+   }
> str( pm2 )

```

```

num [1:11, 1:15] 0.112 0.116 0.12 0.125 0.132 ...
- attr(*, "dimnames")=List of 2
..$ : chr [1:11] "1" "2" "3" "4" ...
..$ : chr [1:15] "Estimate" "2.5%" "97.5%" "Estimate" ...

```

With these predictions we can now plot them in separate panels: It is clear that there is not much non-linearity in the T1D trends, and the non-linearity is mainly attributable to the administrative bumps around 2012, pretty similar between regions.

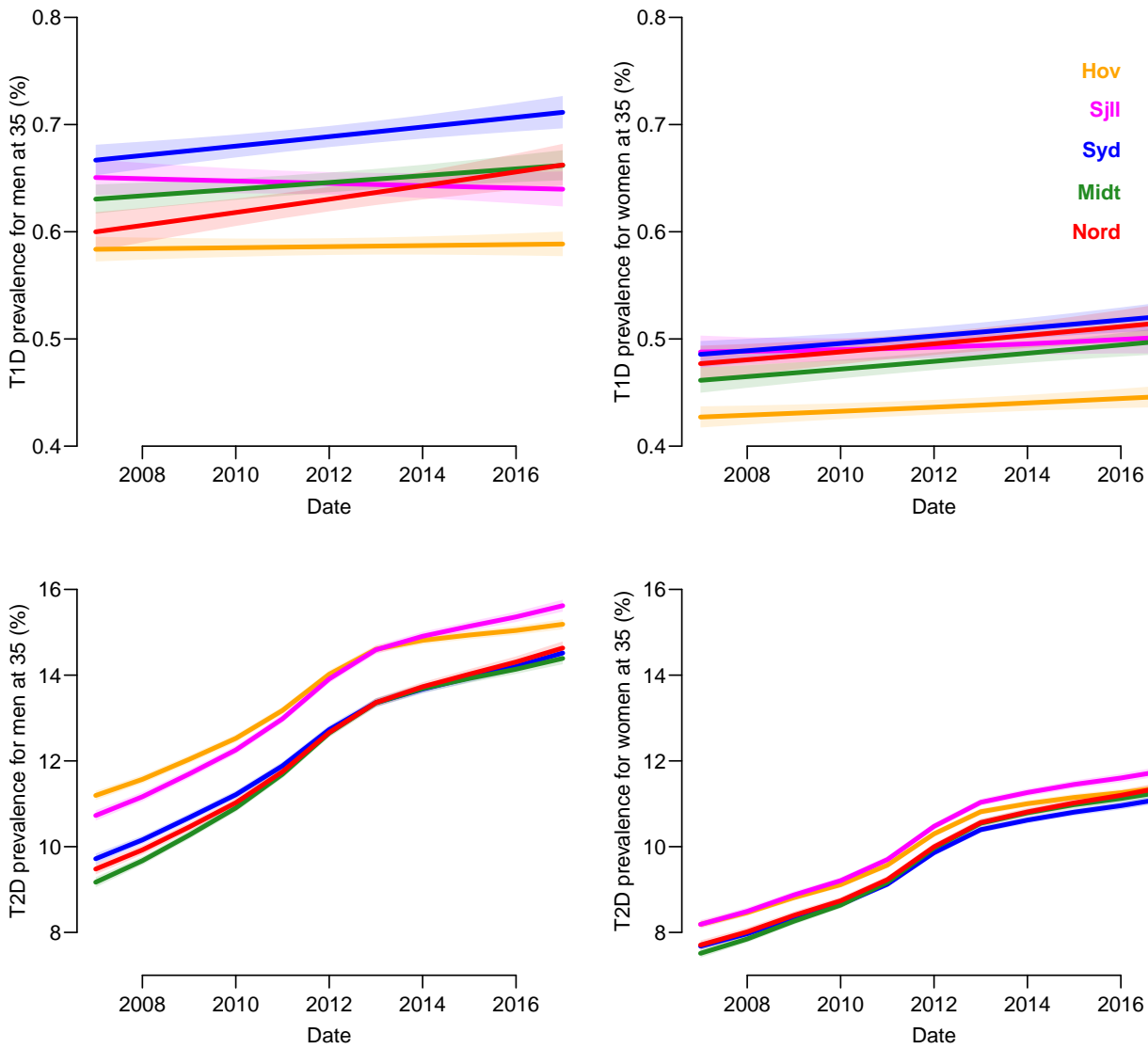


Figure 3.7: Trends in prevalence 2007–17 at ages 35 for T1D and 75 for T2D.
 ./graph/regPrev-Rtrends

Chapter 4

DMreg: det rekonstruerede diabetes register hos DST

4.1 Steno-algoritme

Her beskrives den algoritme til konstruktion af et diabetes register som vi (projektgruppen i afdelingen for Klinisk Epidemiologi, BxC & MaEJ) har brugt i forbindelse med et projekt (Daffodil) finansieret af Astra-Zeneca.

Den er blevet revideret et par gange, senest i december 2018 efter møde med Sundhedsdatastyrelsen (SDS), især med henblik på at få en algoritme som lå nærmere på den SDS anvender.

Det dannede diabetes register ligger som et SAS-datasæt under vores projekt under Danmarks Statistiks forskerordning (projektnummer 705093), og er således ikke tilgængeligt for andre end os. Den SAS-kode der danner registeret er dog naturligvis tilgængelig, således at enhver der har adgang til de samme registre (og SAS) kan danne in kopi af det reviderede register.

I det følgende omtales dette register som DMreg.

4.1.1 Registre

Algoritmen der danner registeret er baseret på data fra:

- Landspatientregisteret, LPR (1977–2016)
- Lægemedelstatistikregisteret, LMS (1995–2016)
- Sygesikringsregisteret, SSR (1990–2016)
- Dansk Voksen Diabetes Database, DVDD (2005–2016)
- Diabasen (øjenscreeningsdatabase for diabetikere), DiaB (2009–2016)

I forhold til RUKS anvender Stenos algoritme altså yderligere oplysninger om fodterapiydelse fra sygesikringsregisteret, oplysninger fra Dansk Voksendiabetesdatabase (DVDD) samt oplysninger fra DiaBasen. Disse tre beskrives derfor nærmere her:

Sygesikringsregistret: ydelse for fodterapi

For at få registreret en dato for ydelse for diabetisk fodterapi kræves følgende:

1. Elektronisk henvisning fra praktiserende læge eller diabetesambulatorium til fodterapeut med ydernummer til diabetesspecifik fodterapi
2. Risikostratificering (inklusive vurdering af diabetisk øjenssygdom) hos fodterapeut mhp. at fastlægge antallet af årlige fodbehandlinger:
 - Gruppe 1 - Lav risiko for fodsår: Ingen tilskudsberettigede behandlinger udover årlig fodstatus/risikovurdering.
 - Gruppe 2 - Mellem risiko for fodsår: Maksimalt fire tilskudsberettigede behandlingsydelse udover årlig fodstatus/risikovurdering.
 - Gruppe 3 - Mellem risiko for fodsår med særlige behov: Maksimalt ni tilskudsberettigede behandlingsydelse udover årlig fodstatus/risikovurdering.
 - Gruppe 4 - Høj risiko for fodsår: Tilskudsberettigede behandlingsydelse efter behov.
3. Gennemført biotesiometrisk undersøgelse specifik for diabetisk perifer neuropati.

Risikoen for fejlagtig registrering af personer uden diabetes som diabetikere anses at være ikke eksisterende (kræver fejl i henvisning + gennemført diabetesspecifik undersøgelse og risikostratifikation samt betalingsopkrævning hos den undersøgte person som diabetiker).

Datoer Der er omkring 8,8 mio. ydelser registreret, fordelt på ca. 263 000 personer. Den tidligste dato for hver person er udtrukket. Af rent administrative årsager (uenigheder mellem fodterapeuter og regionerne) er disse første datoer pr. patient *meget* ujævnt fordelt over årene; i årene 1991–2005 voksede det årlige antal nye patienter fra ca. 5000 til omkring 15 000, fra 2006–2010 var der omkring 1500 årlige første datoer, i 2011 35 000, og derefter falder tallet til omkring 12 000 pr. år.

Dansk Voksendiabetesdatabase (DVDD)

DVDD er en landsdækkende klinisk kvalitetsdatabase for behandling af diabetes hos personer over 18 år. Databasen har eksisteret siden 2005 og indgår som en af tre diabetesdatabaser under regionernes kliniske kvalitetsdatabaser (RKKP). For den enkelte patient registreres dato for undersøgelse, diabetestype, debutår og årlig klinisk status.

Almen praksis har siden lukningen af Dansk Almen Medicinsk Database i september 2014 ikke indberettet data, og databasen er således kun komplet for hospitalsbehandlede diabetespatienter. Al type 1 diabetes i Danmark behandles imidlertid i hospitalsregi og defineres ud fra guidelines ved måling af antistoffer og residual insulinproduktion, og dermed udgør DVDD den mest valide kilde til identifikation af type 1 diabetes i Danmark. Se nedenfor for beskrivelse hvordan oplysingerne bruges i algoritmen.

Risikoen for fejlagtig registrering af personer uden diabetes som diabetikere i DVDD må anses for minimal.

Registeret indeholder pr. april 2018 ca. 740 000 records fra ca. 200 000 personer.

Datoer Indberetningerne til DVDD ligger fra 2005 og fremefter, men diagnose datoerne registreret for patienterne rækker noget længere tilbage; den ældste til 1890 (som nok er en kodefejl), men der er f.eks. 62 diagnose datoer i år 1900 (!).

Diagnose-datoerne (`diag_dato`) er notorisk upræcise; 1. januar og 15. juni udgør 86.6% af alle registrerede diagnosedatoer. Dette må antages i realiteten blot at være en angivelse årstallet, hvor der tydeligvis forekommer en præference for årstallet 2000. I konstruktionen af indgangs-dato og -kriterium tages der højde for dette, se diskussionen nedenfor.

DiaBasen

DiaBasen er den landsdækkende kliniske kvalitetsdatabase for screening af diabetisk retinopati og maculopati for såvel praktiserende oftalmologer som øjenafdelinger. DiaBasen har eksisteret siden 2009 og indgår som en af tre diabetesdatabaser under regionernes kliniske kvalitetsdatabaser (RKKP). For den enkelte patient registreres dato for undersøgelse og detaljeret beskrivelse af diabetiske øjenforandringer.

DiaBasen indeholder pr. april 2018 ca. 425 000 registreringer på ca. 170 000 patienter og er således ikke fuldstændig, men risikoen for fejlregistrering (dvs. inklusion af ikke-diabetikere) må også her anses for minimal.

4.1.2 Datoer i diabetes registeret

For hver person er defineret følgende datoer som anvendes i definition af personens eventuelle inklusion i registeret (en del af disse vil naturligvis være uoplyst for en given person):

- Datoer for hospitalsindlæggelse/ambulant besøg med aktionsdiagnose diabetes (ICD-10: E10, E11, E12, E13, E14; ICD-8: 24900, 24901, 24902, 24903, 24904, 24905, 24906, 24907, 24908, 24909, 25000, 25001, 25002, 25003, 25004, 25005, 25006, 25007, 25008, 25009)
- Datoer for hospitalsindlæggelse/ambulant besøg med aktionsdiagnose gestationel diabetes (ICD-10: O24, ICD-8: 64474, Y6449)
- Datoer for hospitalsindlæggelse/ambulant besøg med aktionsdiagnose PCOS (ICD-10: E282, ICD-8: 61520, 61521).
- Datoer for indløsning af metformin (ATC: A10BAx).
- Datoer for indløsning af andre ikke-insulin antidiabetika (ATC: A10Bxx bortset fra A10BAx).
- Datoer for indløsning af insulinpræparater (ATC: A10Axx)
- Datoer for ydelsen fodterapi for diabetikere (speciale=54xx).
- Registrerede diagnose datoer fra DVDD (subsidiært rapporteringsdatoen).
- Registrerede screeningsdatoer fra DiaBasen,

Rationalet er at inkludere personer som diabetikere med en inklusionsdato som den tidligste af ovenstående datoer (med undtagelse af GDM og PCOS diagnose datoer), dog med to undtagelser:

1. Gestationel diabetes (GDM):

Såfremt der foreligger flere LPR registreringer med gestationel diabetes (ICD-10: O24; ICD-8: 63474, Y6449). For at identificere den første GDM dato i hver graviditet fjernes GDM registreringer som ligger mindre end 200 dage efter den nærmest foregående. For de resterende GDM registreringer for en kvinde dannes et vindue fra 30 dage før til 365 dage efter GDM registreringsdatoen. Datoer fra LMS og LPR som falder i disse vinduer tælles ikke med i beregningen af inklusionsdato.

2. Polycystisk ovarie syndrom (PCOS):

Kvinder der lider af PCOS behandles ofte med metformin. Hvis der derfor foreligger LPR registreringer med PCOS (ICD-10: E282, ICD-8: 61520, 61521) tages den tidligste af disse datoer som PCOS debut dato.

For personer med PCOS ses bort fra metformin køb som ligger fra 30 dage før PCOS debut dato og indtil personens 40 års dag.

Hvis en kvinde udelukkende har indløst metformin (og ingen andre antidiabetika, herunder insulin) mellem alder 18 og alder 40 regnes personen som sandsynligvis værende PCOS i fertilitetsbehandling med metformin. I praksis betyder dette at metformin indløsninger hos kvinder mellem 18 og 40 år ikke medregnes i konstruktionen af datoen doOAD.

Med disse to undtagelser defineres så datoerne:

doNPR — dato for første LPR registrering med en diabetes diagnose (refererer til “National Patient Register”)

doNPR2 — dato for anden LPR registrering

doOAD — dato for første indløsning af OAD

doOAD2 — dato for anden indløsning af OAD

doIns — dato for første indløsning af insulin

doIns2 — dato for anden indløsning af insulin

doPod — dato for første fodterapiydelse (“podiatry”)

doDiaB — dato for tidligste øjenundersøgelse i diabasen

doDVD — tidligste diagnose dato eller indberetnings dato rapporteret i DVDD, dog kun for personer der ikke har nogen anden dato oplyst. 83% af de rapporterede diagnose datoer i DVDD er enten 1 januar eller 15 juni, og er derfor ikke pålidelige.

I den officielle version defineres inklusionsdatoen doDM som den mindste af datoerne doPod, doDVD, doDiaB samt den næst-mindste af datoerne doIns, doOAD, doNPR, doIns2, doOAD2 og doNPR2. Det sidste svarer til RUKSs definition hvor patienter ikke inkluderes på baggrund af alene en recept eller en registrering i LPR.

Bemærk at den sidstnævnte dato (altså den næst-mindste af datoerne doIns, doOAD, doNPR, doIns2, doOAD2 og doNPR2) er datoen for opfyldelsen af kriteriet, og *ikke* som i RUKS, hver den *første* af datoerne for de to kriterier anvendes som inklusions dato.

DMreg anvender den anden af datoerne, altså den første data hvor det er verificeret at kriteriet er opfyldt, for at undgå hvad der kaldes “immortal time bias” ved dødelighedsopgørelser. Hvis inklusionsdatoen tilbagedateres til den første af datoerne vil personer pr. definition være udødelige i tidsrummet fra den første til den anden dato, og dødeligheden i det første stykke tid efter inklusion således være underestimeret. Hvis afstanden mellem de to datoer sædvanligvis er lille betyder dette næppe meget. Men så fald betyder der heller ikke meget om man vælger den første eller den anden dato. Hvis omvendt afstanden er stor vil man få en substantiel *underestimation* af dødeligheden umiddelbart efter inklusion ved brug af RUKS definitionen.

En mere liberal diabetes inklusionsdato (doDM1) kan defineres som den mindste af datoerne doNPR, doOAD, doIns, doPod, doDiaB og doDVD. Dette vil dels resultere i flere personer inkluderet og dels at nogen personer inkluderes på en tidligere dato, muligvis også på et andet kriterium. Dette udvidede register findes som datasættet DMxreg.

Endelig suppleres diabetes registeret med personernes fødselsdato og eventuelle dødsdato.

4.1.3 Inklusionskriterium

Der er for datoen doDM 12 mulige inklusionskriterier afhængig af hvilket (sæt af) inklusionskriterium der først opfyldes, nemlig: Pod, DVD, Dia, I-I, I-O, I-N, O-I, O-O, O-N, N-I, N-O hhv. N-N. Bogstaverne referer til dato for køb af insulin (I) køb af andre antidiabetika (O) samt inklusion i landspatientregisteret (N); således betyder f.eks. O-N at de to første registreringer er 1) indlæsning af OAD og 2) LPR-diagnose, og inklusionsdatoen vil være LPR-datoen. I RUKS ville det have været OAD-datoen.

I de fleste tabelleringer vil vi gruppere inklusionskriterier efter det andet kriterium der opfyldes, således at datoen doDM svarer til datoen for dette.

4.1.4 Registerets dækningsperiode

Da en væsentlig del af personerne inkluderes på baggrund af medicinkøb er registeret ikke anvendeligt til incidens- og prævalens opgørelser før efter 1 januar 1996 — et år efter Lægemiddelstatistikregisterets start, seks år efter sygesikrings registerets start og 19 år efter landspatientregisterets start. Inklusionsdatoer der ligger før 1. januar 1996 må anses at være behæftet med så stor usikkerhed at de ikke bør anvendes som datoer for første diabetesrelaterede kontakt med sundhedsvæsenet.

Det dannede register er således anvendeligt fra 1 januar 1996 og fremefter; og dækker p.t. en 21 års periode frem til d. 31. december 2016, både hvad angår prævalens, incidens og mortalitet.

DVDD starter i 2005, men oplysningerne vedrører patienter som er prævalente efter dette tidspunkt, og data fra DVDD bruges i alt væsentligt til at forfine klassifikation af T1 patienter; således er der ikke et væsentligt databrud omkring 2005 f.s.a. incidens eller prævalens.

DiaBasen bidrager med forholdsvis få nye inklusionsdatoer i begyndelsen da den lige som DVDD omfatter både nye og eksisterende diabetes patienter, så heller ikke denne giver anledning til væsentligt databrud selv om den først starter fra 1 januar 2009.

4.1.5 Diabetes type

Der findes et antal forskellige undertyper af diabetes (T2, T1, LADA, MODY, ...), med T2 og T1 som de dominerende; i dette register defineres T1 patienter så præcist som muligt og alle andre kodes som T2.

Definitionen af T1 patienter baseres primært på registreringen i DVDD: En patient kan optræde flere gange i DVDD, stort set en gang om året, og hver registrering indholder en type-definition af personen. Hvis en person optræder med *over* halvdelen af sine registreringer som T1 (dvs. 4 ud af 8 er ikke nok — 4 ud af 7 eller 5 ud af 9 ville være nok) klassificeres patienten som T1 i variabelen `dvdtyp`. Tilsvarende for T2. Således efterlades nogen patienter uklassificerede fra DVDD, de kodes 'NA' i variabelen `dvdtyp`.

Yderligere har vi klassificeret patienternes enkelte *records* fra LPR som LPR-T1 (ICD-8: 249, ICD-10: E10) hhv. som LPR-T2 (ICD-8: 250, ICD-10: E11) eller ukendt (øvrige diagnoser inkluderet fra LPR). Ligesom for DVDD er patienterne blevet klassificeret som T1 i variabelen `nprtyp` hvis over halvdelen af NPR-records er klassificeret som T1, og tilsvarende for T2. Således efterlades nogen patienter uklassificerede fra LPR, de kodes NA. Vi er nødt til at bruge LPR-klassifikationen selv om den ikke er imponerende præcis, eftersom DVDD først er startet 2005, så personer som er døde før 2005 indgår ikke i DVDD.

Personer klassificeres som T1 i DMreg hvis mindst et af følgende kriterier er opfyldt:

- personen er klassificeret som T1 i DVDD
- personen er klassificeret som T1 i LPR, men *ikke* som T2 i DVDD
- personen har indløst antidiabetisk medicin før 15 års alderen eller har indløst insulin før 30 års alderen, og er *ikke* klassificeret som T2 i DVDD.

Personer der ikke efter disse kriterier klassificeres som T1, eller som aldrig har indløst insulin, vil blive klassificeret som T2. Bemærk at der faktisk er enkelte personer som efter de nævnte kriterier klassificeres som T1, men som aldrig har indløst insulin; disse omklassificeres til T2, da man ikke kan have T1 uden insulinbehandling.

Denne definition af type af diabetes har den konsekvens at personer i registeret kan ændre status ved opdateringer af de bagvedliggende registre. Tabelleringer af faktisk forekommende klassifikationer af enkelte patienter tyder dog på at dette vil være et forholdsvis begrænset fænomen

4.1.6 Debutdato og diabetes varighed

Som nævnt ovenfor er debutdatoer registreret før 1 januar 1996 næppe pålidelige, men datoer efter 1996 må anses for at være anvendelige i epidemiologiske opgørelser. Det betyder at epidemiologiske analyser hvor diabetes-varigheden indgår bør begrænses til personer med debutdato efter 1 januar 1996.

4.1.7 Anvendte parametre

Der er i definitionen af registeret anvendt følgende (i realiteten arbitrært fastsatte) konstanter i definitionerne:

- Interval mellem GDM diagnoser for at behandle dem som forskellige: 200 dage

- Interval omkring GDM til eksklusion: $[-30, +365]$ dage
- Interval før PCOS diagnose: -30 dage
- Alders interval for PCOS / metformin $[18, 40]$ år
- T1 aldergrænse for antidiabetisk medicin: 15 år
- T1 aldergrænse for insulin: 30 år

Registerindholdet vil naturligvis ændre sig en smule hvis man regulerer på disse parametre.

Aktuelt betyder disse grænser at der er et mindre hop i antallet af kvindelige diabetikere inkluderet i 40 års alderen, og et mindre dyk i antallet af T1 diabetikere ved 15 hhv. 30 års alderen. Begge dele er naturligvis artefakter, men sådanne vil være til stede uanset hvordan konstanterne defineres.

4.2 Antal patienter

Dannelsen af registeret ud fra de enkelte komponenter foretages i programmet `06-define`, som også laver forskellige oversigts-tabeller hvori nedstående tal er hentet, se s. ?? ff.

Det totale antal personer i registeret er 474 700, hvoraf 98 711 har en debutdato før 1 januar 1996; dette er altså antallet af prævalente diabetikere pr 1 januar 1996.

Såfremt kun DVDD kriteriet anvendes vil der ud af de 474 700 registrerede være 30 337 (6,4%) registreret som T1. Hvis yderligere registreringerne fra LPR anvendes vil yderligere 15 293 (3,2%) blive klassificeret som T1, i alt 45 630 (9,6%) for hele perioden 1996-2016.

Hvis man slækker kriterierne og inkluderer personer med kun en recept, resp. en registrering i LPR, vil man få inkluderet i alt xxx xxx i registeret, altså xx xxx yderligere personer.

4.3 Region

I registeret findes variable `kom` og `reg` som angiver personens bopÅls-kommune hhv. -region pr. den 1 januar inden inklusionsdatoen. Det er denne er bruges som regions variabel.

4.4 Forskelle til RUKS

Steno algoritmen adskiller sig fra algoritmen brugt til RUKS ved:

- der inkluderes flere diagnosekoder fra LPR.
- type klassifikationen er baseret på DVDD og derfor mere pålidelig.
- eksklusion for PCOS er mindre restriktiv (flere ekskluderes) — ud over LPR-registrering som bruges både af RUKS og DMreg, kræver RUKS at personer som udelukkende er behandlet med meformin også har indløst clomifen eller antiandrogen + østrogen, mens DMreg alene baserer sig på metformin. DMreg tillader imidlertid kvinder med PCOS at blive inkluderet som diabetikere på andre kriterier, herunder indløsning af metformin efter 40 års alderen, idet PCOS også er en velkendt risikofaktor for diabetes.

- eksklusion for GDM er mere restriktiv (færre ekskluderes) idet DMreg opererer med et vindue på $[-30, 365]$ dage, mens RUKS bruger et interval på $[-280, 280]$ dage ($=[-40, 40]$ uger, svarende til en normal graviditetslængde) på hver side af en GDM dato.
- Herudover opererer RUKS med en “forældelsesfrist” på 10 år; personer fjernes fra registeret hvis de i en periode på 10 år ikke er registreret som opfyldende et af inklusionkriterierne.

Det sidste er enten temmelig betydningsløst eller også giver det kunstigt lave tal i perioden før 2007 (10 år før den aktuelle opgørelse). Det er forståeligt at ville ekskludere eventuelle falsk positive registreringer, men det bør ikke gøres ved en eksklusion fra registeret som gør tid-trends upålidelige. I stedet burde man som supplement til variablene med den *tidligste* dato for et kriterium (doIns, doOAD osv.) definere den *senest* registrerede dato for hvert kriterium, således at man muliggør en mere nuanceret analyse af det eventuelle problem med falsk positive.

4.5 Fremtid

På kortere sigt (primo 2020) vil Klinisk Epidemiologi som et led i et større projekt, “Trends i incidens, behandling og komplikationer af diabetes i Danmark” få opdateret DMreg t.o.m. 2017 (muligvis 2018 incl.). På det tidspunkt vil det således være muligt at lave analyser som vedrører patient historier *efter* SDC centrenes etablering.