

Prevalence by Region based on the reconstructed Danish Diabetes Register

SDC

August 2019

<http://bendixcarstensen.com/DMreg/NewAna.pdf>

Draft version 1

Compiled Thursday 15th August, 2019, 13:00
from: /home/bendix/sdc/DMreg/NewReg/nyr/RegPrev.tex

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

Contents

0.1	Prevalence of diabetes by region	1
-----	--	---

...now input from regprev.tex

0.1 Prevalence of diabetes by region

We have a prevalence dataset constructed from the diabetes register:

```
> library( Epi )
> library( haven )
> prv <- read_sas( '../data/prv.sas7bdat' )
> names( prv ) <- tolower( names(prv) )
> prv$reg <- factor( prv$reg, levels=81:85,
+                   labels=c("Nord","Midt","Syd","Hov","Sjll") )
> prv$sex <- factor( prv$sex, labels=c("M","W") )
> str( prv )
Classes 'tbl_df', 'tbl' and 'data.frame':      63823 obs. of  7 variables:
 $ pdat : num  1996 1996 1996 1996 1996 ...
 $ 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   2 3 4 5 6 7 8 9 10 11 ...
 $ n    : num   3 3 3 1 6 6 1 3 8 8 ...
 $ a5   : num   0 0 0 5 5 5 5 5 10 10 ...
 - attr(*, "label")= chr "PRV"
```

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 ), c(1,3) )
> ftable( tt )
```

	state	noDM	T1	T2	Sum
sex reg					
M	Nord	281014	1727	15448	298189
	Midt	621284	3785	31063	656132
	Syd	575594	3835	31796	611225
	Hov	852344	4759	41533	898636
	Sjll	389426	2428	24897	416751
W	Nord	277665	1323	12726	291714
	Midt	624381	2900	26050	653331
	Syd	581316	2899	26322	610537
	Hov	887889	3862	35699	927450
	Sjll	396606	1938	20194	418738
Sum	Nord	558679	3050	28174	589903
	Midt	1245665	6685	57113	1309463
	Syd	1156910	6734	58118	1221762
	Hov	1740233	8621	77232	1826086
	Sjll	786032	4366	45091	835489

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

	sex	M	W	Sum
reg				
Nord		5.2	4.4	4.8
Midt		4.7	4.0	4.4
Syd		5.2	4.3	4.8
Hov		4.6	3.8	4.2
Sjll		6.0	4.8	5.4

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 ...
 $ reg: Factor w/ 5 levels "Nord","Midt",...: 4 2 1 5 3 4 2 1 5 3 ...
 $ x  : num  0 0 0 0 0 0 0 0 0 0 ...
 $ n  : num  11364 7419 3020 3931 6124 ...

> 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  Midt 0  7419
3  M 0.5  Nord 0  3020
4  M 0.5  Sjll 0  3931
5  M 0.5  Syd  0  6124
6  M 1.5  Hov  0 10668
```

However, we first take a look at the age-distributions in each region

```
> par( mfrow=c(5,2), mar=c(3,3,0.3,0.1), oma=c(0,0,1,0), mgp=c(3,1,0)/1.6, bty="n" )
> for( rg in c("Hov","Sjll","Syd" ,"Midt","Nord") )
+ for( sx in c("M","W") ) {
+ zz <- barplot( nn[sx,,rg,"Sum"]/1000, ylim=c(0,17),
+               col=if(sx=="M") "blue" else "red",
+               border=if(sx=="M") "blue" else "red" )
+ text( 5, 2, rg, font=2, col="white", cex=2, adj=0 )
+ abline( v=zz[quantile( rep(0:99,nn[sx,,rg,"Sum"]) ) [2:4]], col="white",lwd=2 )
+ abline( v=zz[quantile( rep(0:99,nn[sx,,rg,"Sum"]) ) [2:4]] )
+ }
```

Then at the age-distributions *relative* to Region H:

```
> par( mfrow=c(5,2), mar=c(3,3,0.1,0.1), oma=c(0,0,1,0), mgp=c(3,1,0)/1.6, bty="n" )
> for( rg in c("Hov","Sjll","Syd" ,"Midt","Nord") )
+ for( sx in c("M","W") ) {
+ barplot( nn[sx,,rg,"Sum"]/nn[sx,, "Hov","Sum"], ylim=c(0,1),
+         col=if(sx=="M") "blue" else "red",
+         border=if(sx=="M") "blue" else "red" )
+ text( 5, 0.2, rg, font=2, col="white", cex=2, adj=0 )
+ }
```

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

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"

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

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

```
> clr <- rainbow(6)[-2]
> clr <- c("red","forestgreen","blue","orange","magenta")
> par( mfrow=c(1,2), mar=c(3,3,0.1,0.1), oma=c(0,0,1,0), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( apt, cbind(apr[, "M", "Hov" ],
+                     apr[, "M", "Sjll" ],
+                     apr[, "M", "Syd" ],
+                     apr[, "M", "Midt" ],
+                     apr[, "M", "Nord" ]), plot=TRUE, yaxs="i",
+         col=clr, lwd=2, xlab="Age", ylab="Prevalence (%)", ylim=c(0,20) )
> text( 5, 18:14, c("Hov" ,
+                 "Sjll" ,
+                 "Syd" ,
+                 "Midt" ,
+                 "Nord"), col=clr, font=2, adj=0, cex=1.5 )
> text( 5, 19, "Men", adj=0, cex=1.5 )
> matshade( apt, cbind(apr[, "W", "Hov" ],
+                     apr[, "W", "Sjll" ],
+                     apr[, "W", "Syd" ],
+                     apr[, "W", "Midt" ],
+                     apr[, "W", "Nord" ]), plot=TRUE, yaxs="i",
+         col=clr, lwd=2, xlab="Age", ylab="Prevalence (%)", ylim=c(0,20) )
> text( 5, 18:14, c("Hov" ,
+                 "Sjll" ,
+                 "Syd" ,
+                 "Midt" ,
+                 "Nord"), col=clr, font=2, adj=0, cex=1.5 )
> text( 5, 19, "Women", adj=0, cex=1.5 )
```

```
> # clr <- rainbow(6)[-2]
> # clr <- terrain.colors(5)
> par( mfrow=c(1,1), mar=c(3,3,0.1,2), oma=c(0,0,1,0), mgp=c(3,1,0)/1.6, las=1, bty="n" )
> matshade( apt, cbind(apr[, "M", "Hov" ],
+                     apr[, "M", "Sjll" ],
+                     apr[, "M", "Syd" ],
+                     apr[, "M", "Midt" ],
+                     apr[, "M", "Nord" ],
+                     apr[, "W", "Hov" ],
+                     apr[, "W", "Sjll" ],
+                     apr[, "W", "Syd" ],
+                     apr[, "W", "Midt" ],
+                     apr[, "W", "Nord" ]), plot=TRUE, yaxs="i", alpha=0,
+       col=clr, lwd=3, lty=rep(c("solid", "21"), each=5), lend="butt",
+       xlab="Age", ylab="Prevalence (%)", ylim=c(0,20), xlim=c(20,90) )
> text( 30, 20:16-3, c("Hov" ,
+                     "Sjll",
+                     "Syd" ,
+                     "Midt",
+                     "Nord"), col=clr, font=2, adj=0, cex=1.5 )
> axis( side=4 )
> axis( side=4, at=1:19, labels=NA, tcl=-0.3 )
> axis( side=2, at=1:19, labels=NA, tcl=-0.3 )
> text( 30, 19, "Men" , adj=0, cex=1.5 )
> text( 30, 18, "Women", adj=0, cex=1.5 )
> segments( c(21,21), 19:18, c(29,29), 19:18, lwd=3, lty=c("solid", "21") )
```

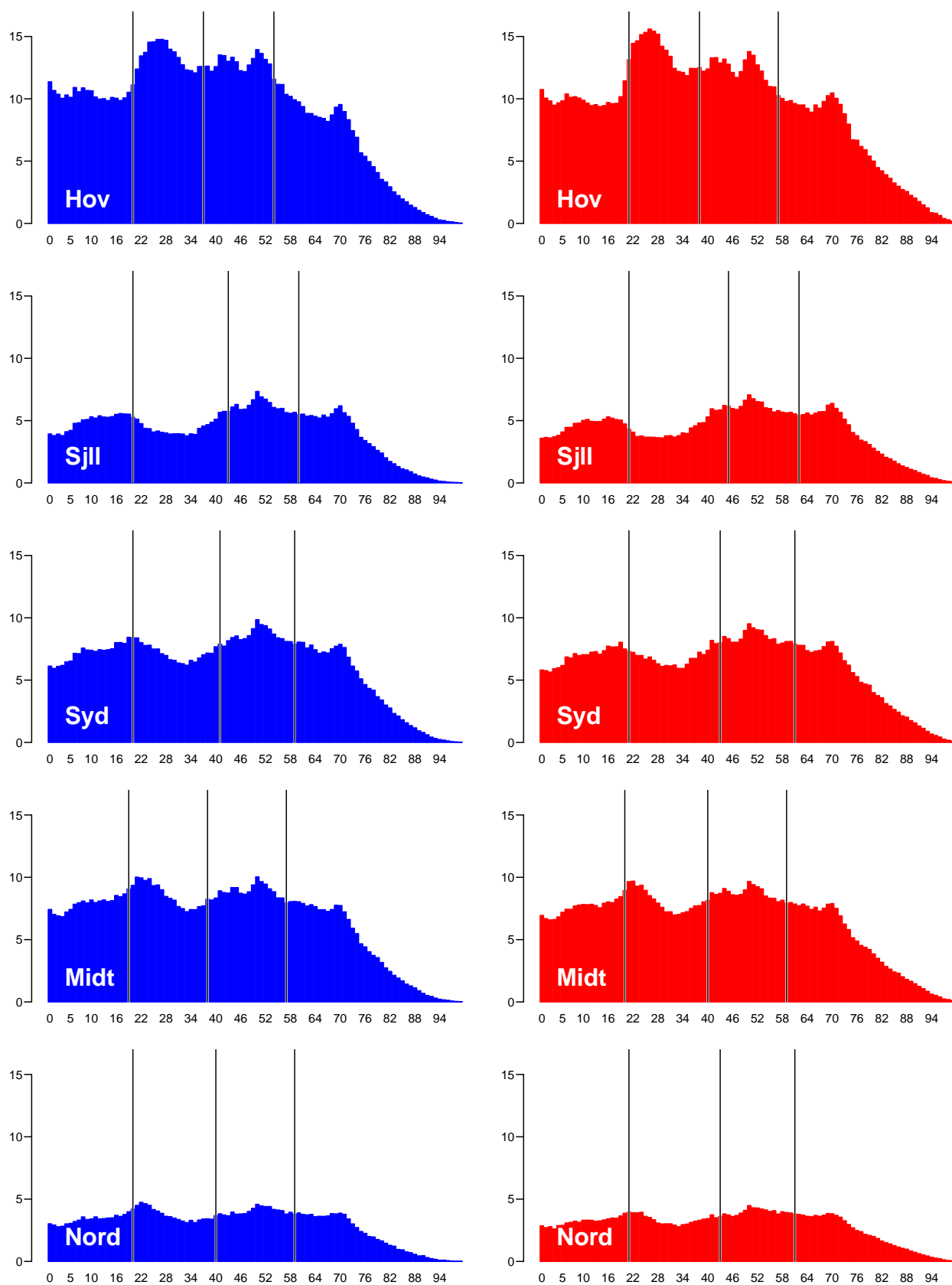


Figure 1: Absolute number of persons in the 100 age-classes for each sex and region. Vertical lines divide population in quarters.

`./graph/regprev-abs`

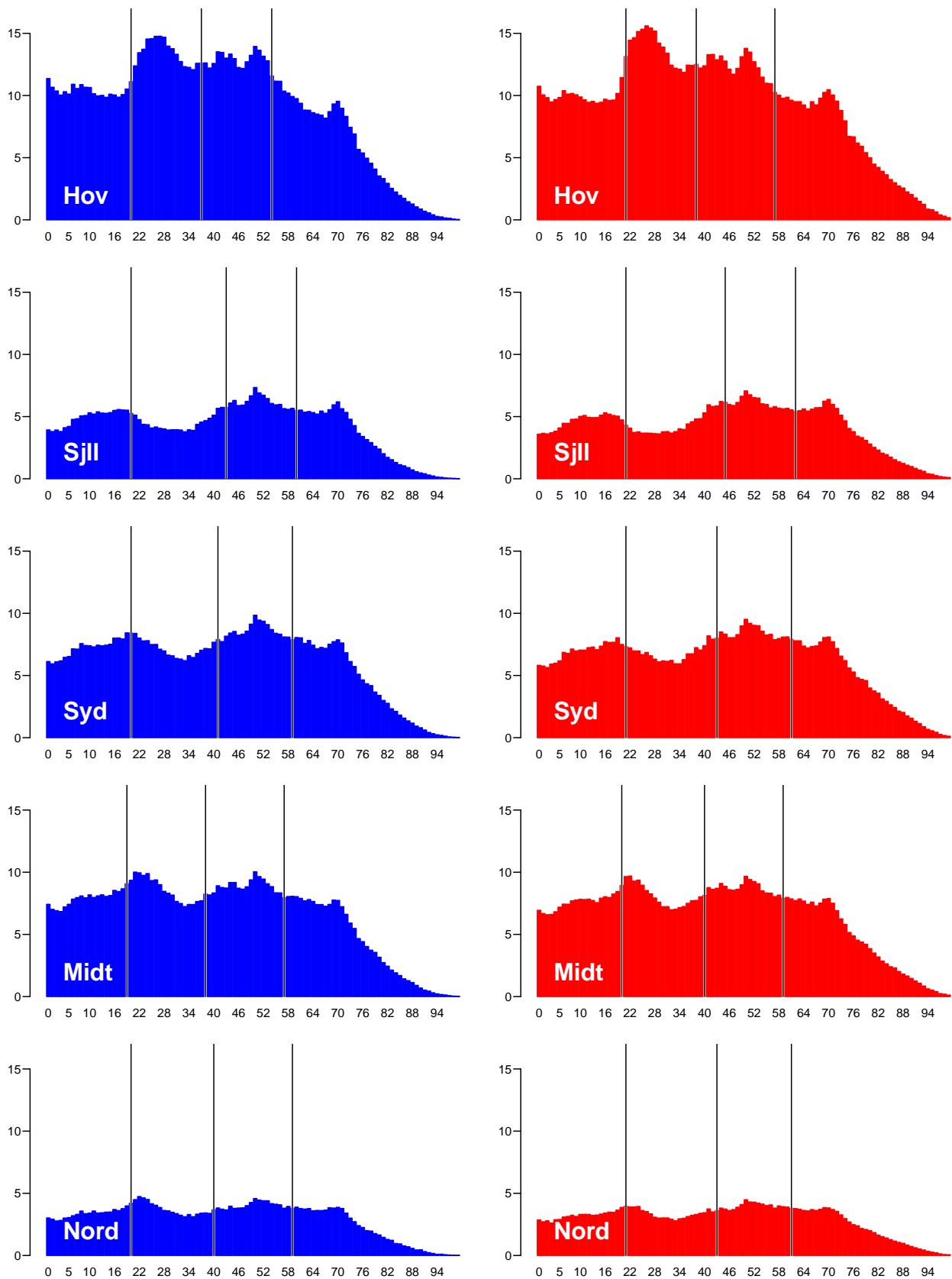


Figure 2: *Relative size of the of population relative to the population in Hov in the 100 age-classes for each sex and region*

`./graph/regprev-abs`

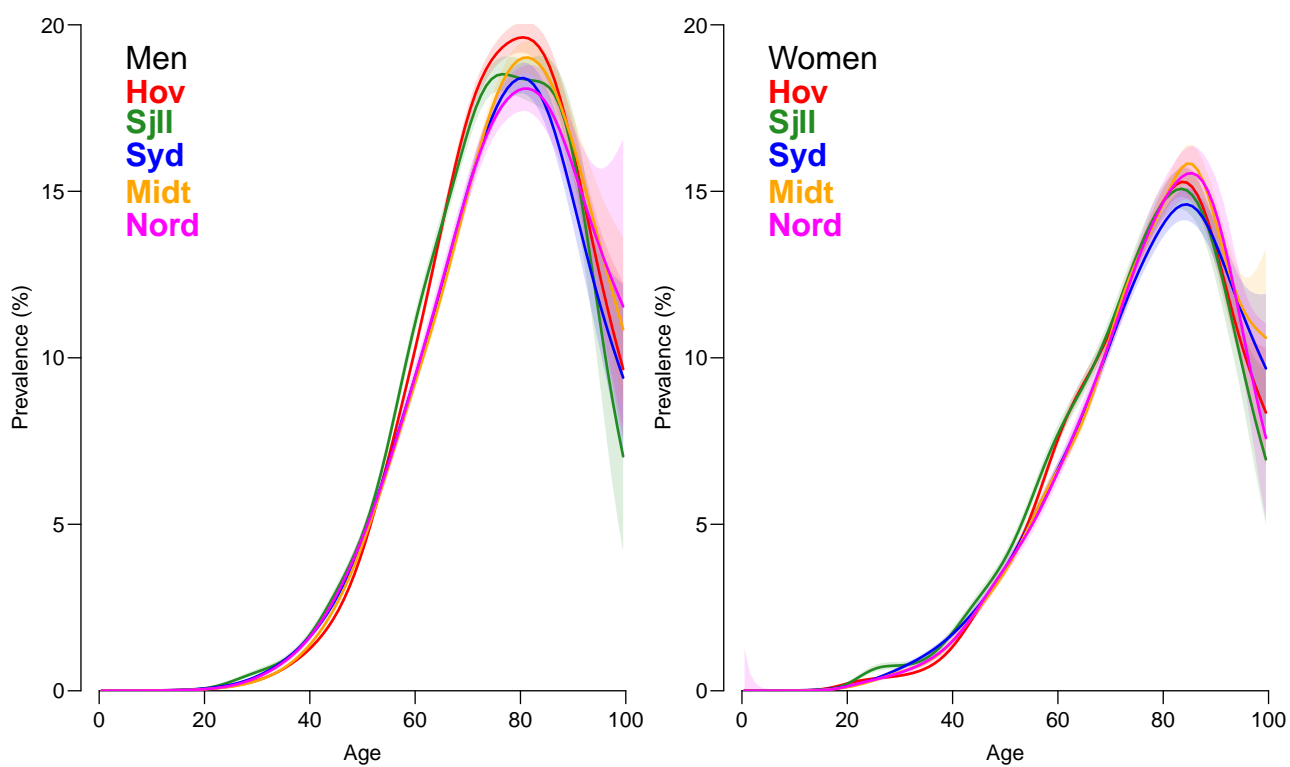


Figure 3: Age-specific prevalences by sex and region. Shaded areas are 95% confidence intervals.
./graph/regprev-aprev

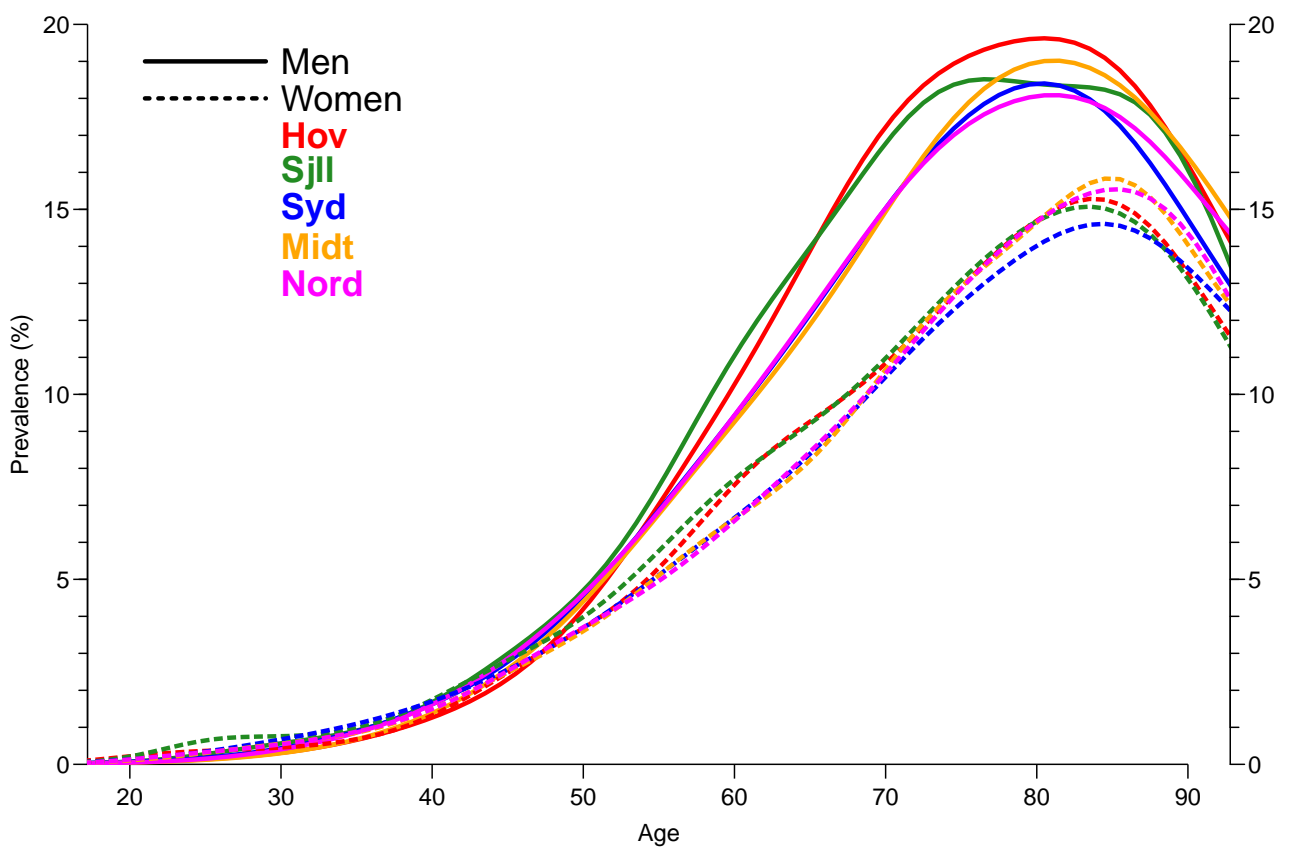


Figure 4: Age-specific prevalences by sex and region.

./graph/regprev-prev