An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time

Bendix Carstensen Steno Diabetes Center Copenhagen, Gentofte, Denmark
`http://BendixCarstensen.com`

CDC, Atlanta, June 2019

From `/home/bendix/teach/APC/talks/Atlanta.2019/slides.tex` Tuesday 18th June, 2019, 09:37

An overview of APC models

- Data in a Lexis diagram — and where they come from.

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models
- APC models as they usually are

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models
- APC models as they usually are
- APC models as they should be

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models
- APC models as they usually are
- APC models as they should be
- Parameters vs. fitted values

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models
- APC models as they usually are
- APC models as they should be
- Parameters vs. fitted values
- Practical use in forecasting

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models
- APC models as they usually are
- APC models as they should be
- Parameters vs. fitted values
- Practical use in forecasting

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models
- APC models as they usually are
- APC models as they should be
- Parameters vs. fitted values
- Practical use in forecasting

Slides with code in **R** only briefly covered

An overview of APC models

- Data in a Lexis diagram — and where they come from.
- Simple graphs of rates
- Simple AP and AC models
- APC models as they usually are
- APC models as they should be
- Parameters vs. fitted values
- Practical use in forecasting

Slides with code in **R** only briefly covered

Population occurrence rates

- Population rates occur in calendar time

Population occurrence rates

- Population rates occur in calendar time
- . . . depend very strongly on age

Population occurrence rates

- Population rates occur in calendar time
- . . . depend very strongly on age
- describe how rates have evolved

Population occurrence rates

- Population rates occur in calendar time
- . . . depend very strongly on age
- describe how rates have evolved
- predict how they will evolve in the future

Population occurrence rates

- Population rates occur in calendar time
- . . . depend very strongly on age
- describe how rates have evolved
- predict how they will evolve in the future
- Rates as a function of age and calendar time:

Population occurrence rates

- Population rates occur in calendar time
- . . . depend very strongly on age
- describe how rates have evolved
- predict how they will evolve in the future
- Rates as a function of age and calendar time:
    - data representation

Population occurrence rates

- Population rates occur in calendar time
- . . . depend very strongly on age
- describe how rates have evolved
- predict how they will evolve in the future
- Rates as a function of age and calendar time:
    - data representation
    - modeling

Models for tabulated data

Bendix Carstensen

An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time
CDC, Atlanta,June 2019

`http://BendixCarstensen/APC`

Conceptual set-up

Follow-up of the entire (male) population from 1943–2006 w.r.t. occurrence of testis cancer:

- Split follow-up time for all about 4 mil. men in 1-year classes by age and calendar time (y).
- Allocate testis cancer event $(d = 0, 1)$ to each.
- Analyze all $200,000,000$ records by a Poisson model.

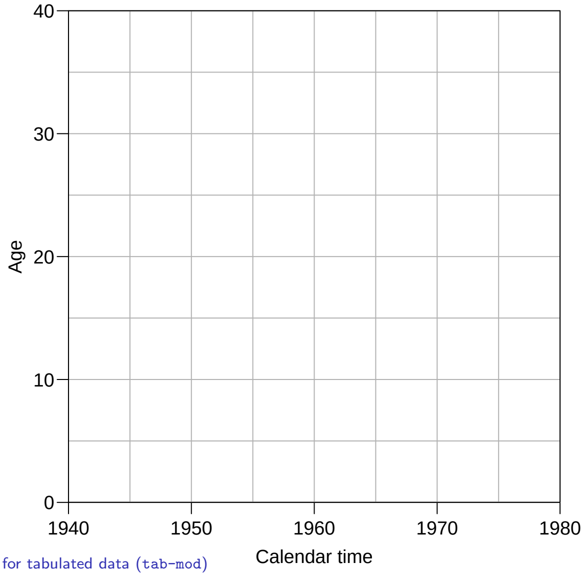
Realistic set-up

- Tabulate the follow-up time and events by age and period.
- 100 age-classes.
- 65 periods (single calendar years).
- 6500 aggregate records of (D, Y).
- Analyze by a Poisson model
- …note: I have not specified how the model looks

Practical set-up

- Tabulate only events (as obtained from the cancer registry) by age and period.
- 100 age-classes.
- 65 periods (single calendar years).
- 6500 aggregate records of D.
- Estimate the population follow-up based on census data from Statistics Denmark (Y_{pop}).
  . . . or get it from the human mortality database.
- If disease is common: tabulate follow-up **after** diagnosis (Y_{dis}), and subtract from population follow-up.
- Analyze (D, Y) by Poisson model.

Lexis diagram [1]

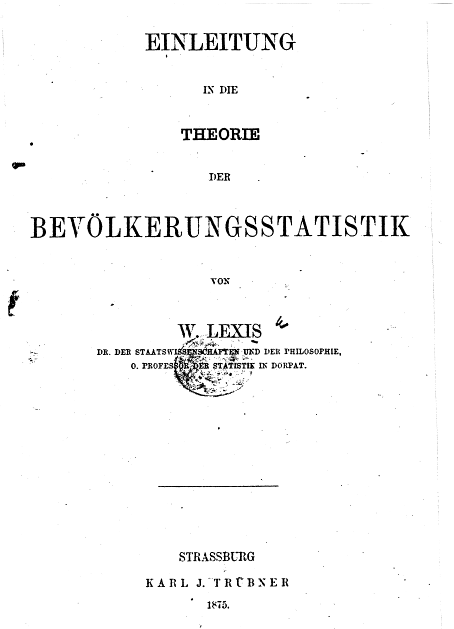


Disease registers record events.

Official statistics collect population data.

---

[1] Named after the German statistician and economist **William Lexis** (1837–1914), who devised this diagram in the book “Einleitung in die Theorie der Bevölkerungsstatistik” (Karl J. Trübner, Strassburg, 1875).



EINLEITUNG

IN DIE

THEORIE

DER

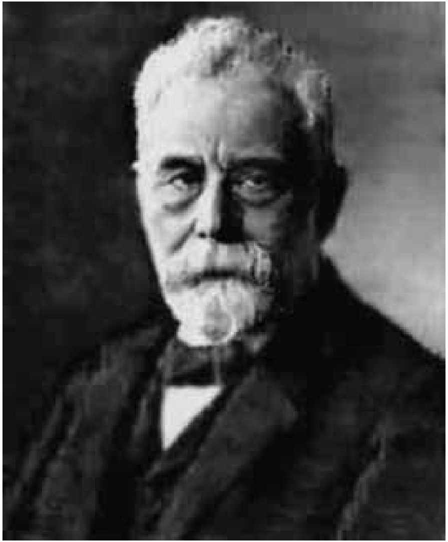
BEVÖLKERUNGSSTATISTIK

VON

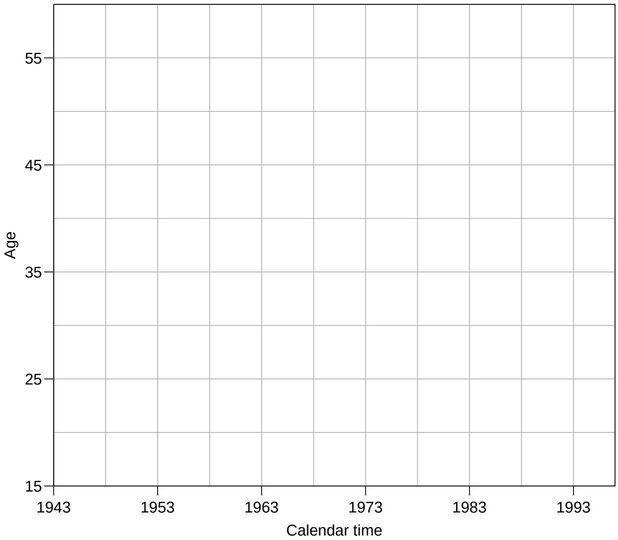
W. LEXIS

DR. DER STAATSWISSENSCHAFTEN UND DER PHILOSOPHIE,
O. PROFESSOR DER STATISTIK IN DORPAT.

STRASSBURG
KARL J. TRÜBNER
1875.



Lexis diagram



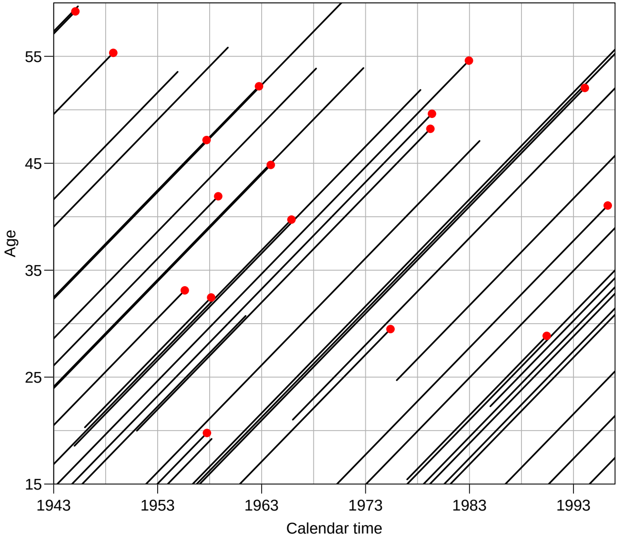
Registration of:

cases (D)

risk time,
person-years (Y)

in subsets of the Lexis diagram.

Lexis diagram



Registration of:

cases (D)

risk time,
person-years (Y)

in subsets of the Lexis diagram.

Rates available in each subset.

Register data

Classification of **cases** (D_{ap}) by age at diagnosis and date of diagnosis, and **population** (Y_{ap}) by age at risk and date at risk, in compartments of the Lexis diagram, e.g.:

```
> fCtable( xtabs( cbind(D,Y) ~ A + P, data=ts ), col.vars=3:2, w=8 )

             D                                       Y
   P      1943     1948     1953     1958     1943     1948     1953     1958
A
15           2        3        4        1  773,812  744,217  794,123  972,853
20           7        7       17        8  813,022  744,706  721,810  770,859
25          28       23       26       35  790,501  781,827  722,968  698,612
30          28       43       49       51  799,293  774,542  769,298  711,596
35          36       42       39       44  769,356  782,893  760,213  760,452
40          24       32       46       53  694,073  754,322  768,471  749,912
```

In analysis format:

```
> ts

     A    P  D      Y
1   15 1943  2 773812
2   20 1943  7 813022
3   25 1943 28 790501
4   30 1943 28 799293
5   35 1943 36 769356
6   40 1943 24 694073
7   15 1948  3 744217
8   20 1948  7 744706
9   25 1948 23 781827
10  30 1948 43 774542
11  35 1948 42 782893
12  40 1948 32 754322
13  15 1953  4 794123
14  20 1953 17 721810
15  25 1953 26 722968
16  30 1953 49 769298
17  35 1953 39 760213
18  40 1953 46 768471
```

Tabulated data

Once data are in tabular form, models are restricted:

- Rates must be assumed constant in each cell of the table / subset of the Lexis diagram.
- With large cells (5×5 years) it is customary to put a separate parameter on each cell or on each levels of classifying factors.
- Output from the model will be rates and rate-ratios.
- Since we use multiplicative Poisson, usually the log rates and the log-RR are reported

Age-Period and Age-Cohort models

Bendix Carstensen

An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time
CDC, Atlanta,June 2019

Register data — rates

Rates in “tiles” of the Lexis diagram:

$$\lambda(a, p) = D_{ap} / Y_{ap}$$

Descriptive epidemiology based on disease registers:
How do the rates vary by age and time:

- Age-specific rates across periods.

Register data — rates

Rates in “tiles” of the Lexis diagram:

$$\lambda(a, p) = D_{ap} / Y_{ap}$$

Descriptive epidemiology based on disease registers:
How do the rates vary by age and time:

- Age-specific rates across periods.
- Age-specific rates across cohorts.

Register data — rates

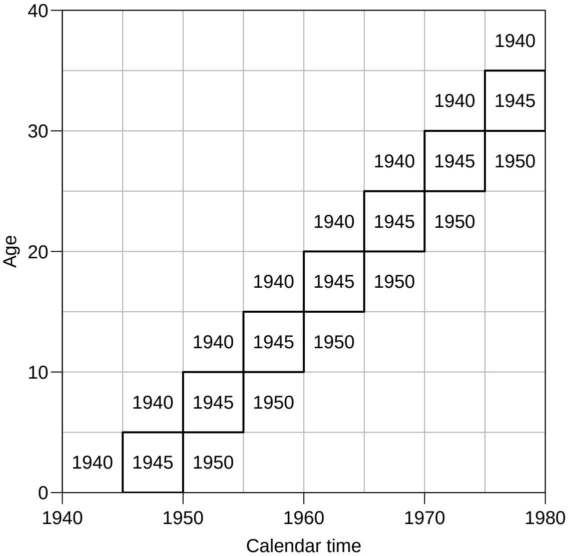
Rates in “tiles” of the Lexis diagram:

$$\lambda(a,p) = D_{ap} / Y_{ap}$$

Descriptive epidemiology based on disease registers:
How do the rates vary by age and time:

- Age-specific rates across periods.
- Age-specific rates across cohorts.
- Age-standardized rates as a function of calendar time. (Weighted averages of the age-specific rates).

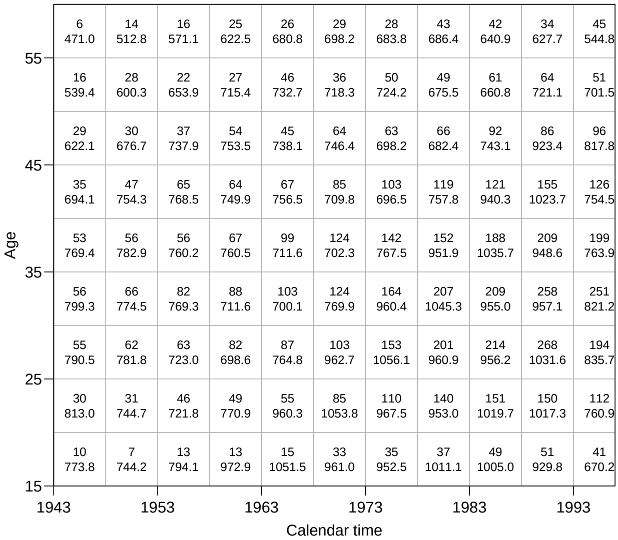
“Synthetic” cohorts



Events and risk time in cells along the diagonals are among persons with roughly same date of birth.

Successively overlapping 10-year periods.

Lexis diagram: data



6 471.0	14 512.8	16 571.1	25 622.5	26 680.8	29 698.2	28 683.8	43 686.4	42 640.9	34 627.7	45 544.8
16 539.4	28 600.3	22 653.9	27 715.4	46 732.7	36 718.3	50 724.2	49 675.5	61 660.8	64 721.1	51 701.5
29 622.1	30 676.7	37 737.9	54 753.5	45 738.1	64 746.4	63 698.2	66 682.4	92 743.1	86 923.4	96 817.8
35 694.1	47 754.3	65 768.5	64 749.9	67 756.5	85 709.8	103 696.5	119 757.8	121 940.3	155 1023.7	126 754.5
53 769.4	56 782.9	56 760.2	67 760.5	99 711.6	124 702.3	142 767.5	152 951.9	188 1035.7	209 948.6	199 763.9
56 799.3	66 774.5	82 769.3	88 711.6	103 700.1	124 769.9	164 960.4	207 1045.3	209 955.0	258 957.1	251 821.2
55 790.5	62 781.8	63 723.0	82 698.6	87 764.8	103 962.7	153 1056.1	201 960.9	214 956.2	268 1031.6	194 835.7
30 813.0	31 744.7	46 721.8	49 770.9	55 960.3	85 1053.8	110 967.5	140 953.0	151 1019.7	150 1017.3	112 760.9
10 773.8	7 744.2	13 794.1	13 972.9	15 1051.5	33 961.0	35 952.5	37 1011.1	49 1005.0	51 929.8	41 670.2

Testis cancer cases in Denmark.

Male person-years in Denmark.

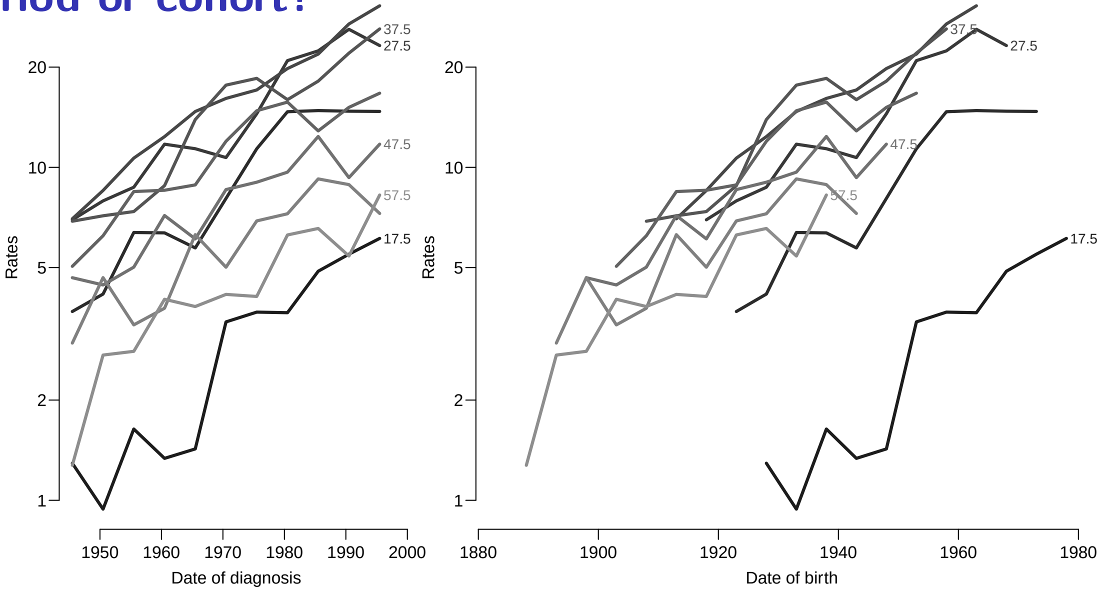
```
> library( Epi )
> data( testisDK )
> head( testisDK )

  A    P D        Y
1 0 1943 1 39649.50
2 1 1943 1 36942.83
3 2 1943 0 34588.33
4 3 1943 1 33267.00
5 4 1943 0 32614.00
6 5 1943 0 32020.33

> ts <- transform( subset( testisDK, A>14 & A<60 ),
+                  A = floor( A      /5)*5     +2.5,
+                  P = floor((P-1943)/5)*5+1943+2.5 )
> ts$C <- ts$P - ts$A
> trate <- xtabs( D ~ A + P, data = ts ) /
+          xtabs( Y ~ A + P, data = ts ) * 100000
> trate[1:5,1:6]

      P
A          1945.5    1950.5    1955.5    1960.5    1965.5    1970.5
  17.5  1.2923036 0.9405857 1.6370257 1.3362759 1.4264867 3.4340862
```

```
22.5  3.6899378  4.1627194  6.3728682  6.3565492  5.7274822  8.0657826
27.5  6.9576174  7.9301414  8.7140826 11.7375624 11.3753792 10.6996275
32.5  7.0061961  8.5211703 10.6590661 12.3665762 14.7122260 16.1068525
37.5  6.8888785  7.1529555  7.3663549  8.8105514 13.9126492 17.6571019
```

```
> par( mfrow=c(2,2) )
> rateplot( trate, col=gray(2:15/18), lwd=3, ann=TRUE )
>  wh = c("ap","ac","pa","ca")
> for( ptp in wh ) {
+    pdf( paste("./graph/AP-AC-",ptp,".pdf",sep=""), height=6, width=8 )
+    par( mar=c(3,3,1,1, mgp=c(3,1,0)/1.6, bty="n", las=1 ))
+    rateplot( trate, which=ptp,
+              col=gray(2:15/18), lwd=3, ann=TRUE, a.lim=c(15,60) )
+    dev.off()
+    }
>
```

```
> library( Epi )
> par( mar=c(3,3,.1,.1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> layout( mat=cbind(1,2),width=c(6,10) )
> for( ptp in c("pa","ca") )
+    rateplot( trate, which=ptp,
+              col=gray(2:15/18), lwd=3, ann=TRUE, a.lim=c(15,60) )
```

Period or cohort?



Age-Period model

Rates are proportional between periods:

$$\lambda(a,p) = a_a \times b_p \qquad \text{or} \qquad \log[\lambda(a,p)] = \alpha_a + \beta_p$$

Choose p_0 as reference period, where $\beta_{p_0} = 0$

$$\log[\lambda(a,p_0)] = \alpha_a + \beta_{p_0} = \alpha_a$$

Fitting the A-P model in R I

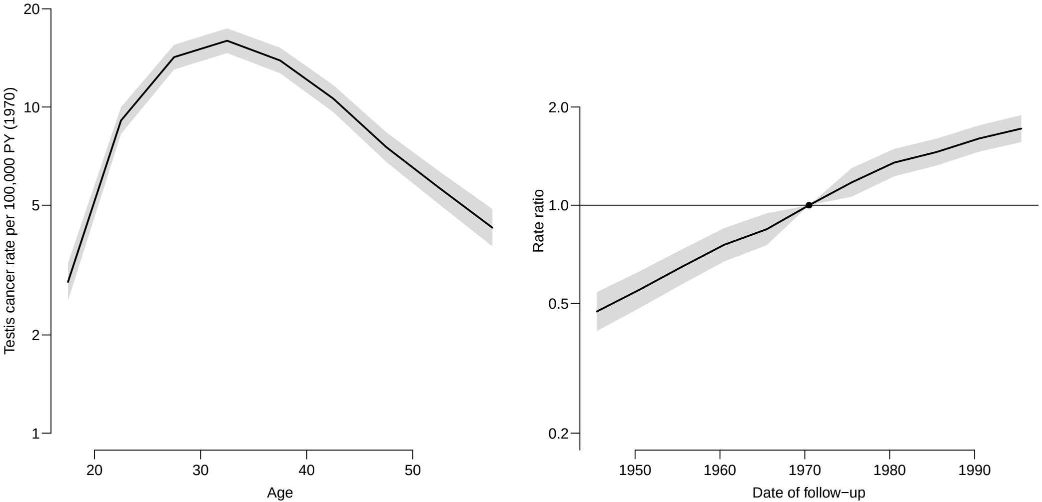
Reference period is the 5th period (1970.5 $\sim$ 1968–72):

```
> ap <- glm( D ~ factor(A) - 1 + relevel( factor(P), "1970.5" ) +
+                 offset( log(Y/10^5) ),
+             family=poisson, data=ts )
> # summary( ap )
```

Estimates with confidence intervals

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> matshade( seq(17.5,57.5,5), ci.exp(ap,subset="A"), plot=TRUE,
+           log="y", lwd=2, ylim=c(1,20), xlab="Age",
+           ylab="Testis cancer rate per 100,000 PY (1970)" )
> matshade( seq(1945.5,1995.5,5),
+           rbind( ci.exp(ap,subset="P")[1:5 ,], 1,
+                  ci.exp(ap,subset="P")[6:10,] ), plot=TRUE,
+           log="y", lwd=2, ylim=c(1,20)/5,
+           xlab="Date of follow-up", ylab="Rate ratio" )
> abline( h = 1)
> points( 1970.5, 1, pch=16 )
```

Estimates from Age-Period model



Age-cohort model

Rates are proportional between cohorts:

$$\lambda(a,c) = a_a \times c_c \qquad \text{or} \qquad \log[\lambda(a,p)] = \alpha_a + \gamma_c$$

Choose c_0 as reference cohort, where $\gamma_{c_0} = 0$

$$\log[\lambda(a,c_0)] = \alpha_a + \gamma_{c_0} = \alpha_a$$

Fitting the A-C model in R I

Reference cohort is the 1933 cohort:

```
> ac <- glm( D ~ factor(A) - 1 + relevel( factor(C), "1933" ) +
+                 offset( log(Y/10^5) ),
+             family=poisson, data=ts )
> summary( ac )

Call:
glm(formula = D ~ factor(A) - 1 + relevel(factor(C), "1933") +
    offset(log(Y/10^5)), family = poisson, data = ts)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.0796  -0.9538  -0.1620   0.5767   3.9525

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
factor(A)17.5                   0.61513    0.07534   8.165 3.23e-16
```

Fitting the A-C model in R II

```
factor(A)22.5                      1.89965    0.05342  35.558  < 2e-16
factor(A)27.5                      2.46911    0.04842  50.990  < 2e-16
factor(A)32.5                      2.70635    0.04695  57.639  < 2e-16
factor(A)37.5                      2.71211    0.04758  57.006  < 2e-16
factor(A)42.5                      2.58676    0.04993  51.803  < 2e-16
factor(A)47.5                      2.36542    0.05459  43.327  < 2e-16
factor(A)52.5                      2.18192    0.06098  35.782  < 2e-16
factor(A)57.5                      2.01519    0.06939  29.041  < 2e-16
relevel(factor(C), "1933")1888 -1.77316    0.41400  -4.283 1.84e-05
relevel(factor(C), "1933")1893 -1.05641    0.19017  -5.555 2.77e-08
relevel(factor(C), "1933")1898 -0.79897    0.12600  -6.341 2.28e-10
relevel(factor(C), "1933")1903 -0.87599    0.10389  -8.432  < 2e-16
relevel(factor(C), "1933")1908 -0.76707    0.08352  -9.184  < 2e-16
relevel(factor(C), "1933")1913 -0.56290    0.07006  -8.035 9.36e-16
relevel(factor(C), "1933")1918 -0.56702    0.06683  -8.484  < 2e-16
relevel(factor(C), "1933")1923 -0.36836    0.06124  -6.015 1.79e-09
relevel(factor(C), "1933")1928 -0.18832    0.05903  -3.190 0.001421
relevel(factor(C), "1933")1938  0.08958    0.05439   1.647 0.099585
relevel(factor(C), "1933")1943 -0.03107    0.05443  -0.571 0.568091
```

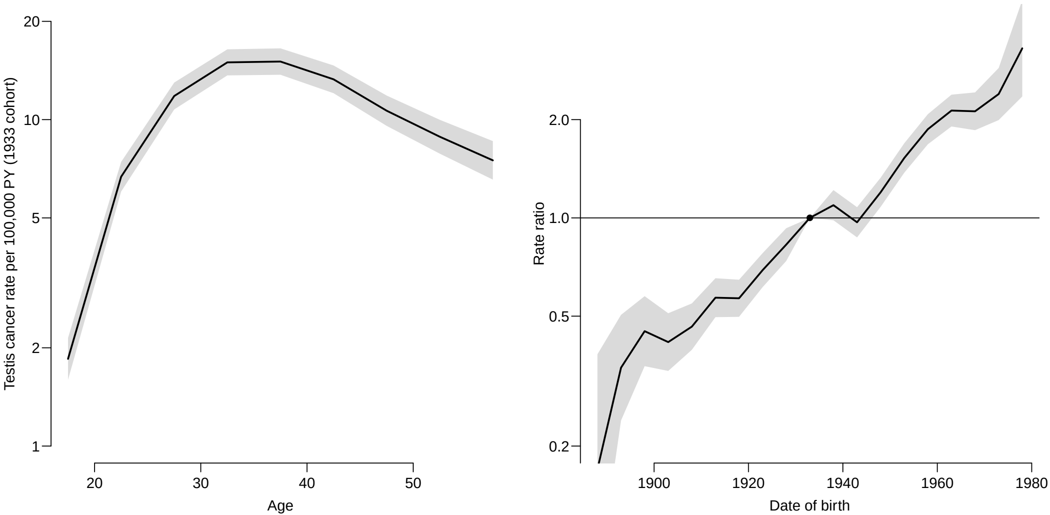
Fitting the A-C model in R III

```
relevel(factor(C), "1933")1948  0.18088    0.05256   3.441 0.000579
relevel(factor(C), "1933")1953  0.42239    0.05309   7.956 1.77e-15
relevel(factor(C), "1933")1958  0.62544    0.05421  11.537  < 2e-16
relevel(factor(C), "1933")1963  0.75687    0.05727  13.215  < 2e-16
relevel(factor(C), "1933")1968  0.75183    0.06799  11.057  < 2e-16
relevel(factor(C), "1933")1973  0.87343    0.09373   9.318  < 2e-16
relevel(factor(C), "1933")1978  1.19601    0.17340   6.898 5.29e-12

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 29193.6  on 2430  degrees of freedom
Residual deviance:  2767.8  on 2403  degrees of freedom
AIC: 8972.2

Number of Fisher Scoring iterations: 5
```

Estimates with confidence intervals

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> matshade( seq(17.5,57.5,5), ci.exp(ac,subset="A"), plot=TRUE,
+           log="y", lwd=2, ylim=c(1,20), xlab="Age",
+           ylab="Testis cancer rate per 100,000 PY (1933 cohort)" )
> matshade( seq(1888,1978,5),
+           rbind( ci.exp(ac,subset="C")[1:9 ,], 1,
+                  ci.exp(ac,subset="C")[10:18,] ), plot=TRUE,
+           log="y", lwd=2, ylim=c(1,20)/5,
+           xlab="Date of birth", ylab="Rate ratio" )
> abline( h = 1)
> points( 1933, 1, pch=16 )
```

Estimates from Age-Cohort model



Hang on:

Age, period and cohort are **quantitative** variables

Hang on:

Age, period and cohort are **quantitative** variables

- but the models we fitted does not use this feature

Hang on:

Age, period and cohort are **quantitative** variables

- but the models we fitted does not use this feature
- they are **exchangeable** models for the A, P and C effects

Hang on:

Age, period and cohort are **quantitative** variables

- but the models we fitted does not use this feature
- they are **exchangeable** models for the A, P and C effects
- meaning that you can exhange the names of two age-classes and still get the same fit

Hang on:

Age, period and cohort are **quantitative** variables

- but the models we fitted does not use this feature
- they are **exchangeable** models for the A, P and C effects
- meaning that you can exhange the names of two age-classes and still get the same fit
- models do not use the fact that $50 < 55 < 60$.

Hang on:

Age, period and cohort are **quantitative** variables

- but the models we fitted does not use this feature
- they are **exchangeable** models for the A, P and C effects
- meaning that you can exhange the names of two age-classes and still get the same fit
- models do not use the fact that $50 < 55 < 60$.
- we need parametric models for the A, P and C effects

$$\log\big(\lambda(a,p)\big) = f(a) + g(p) \qquad \log\big(\lambda(a,p)\big) = f(a) + h(p-a)$$

Parametric models

- f, g and h are **smooth**, **continuous** functions:

$$\log\big(\lambda(a,p)\big) = f(a) + g(p) \qquad \log\big(\lambda(a,p)\big) = f(a) + h(p-a)$$

Parametric models

- f, g and h are **smooth**, **continuous** functions:

$$\log\big(\lambda(a,p)\big) = f(a) + g(p) \qquad \log\big(\lambda(a,p)\big) = f(a) + h(p-a)$$

- **Data** is discrete (1-year, 5-year) intervals

Parametric models

- f, g and h are **smooth**, **continuous** functions:

$$\log\big(\lambda(a,p)\big) = f(a) + g(p) \qquad \log\big(\lambda(a,p)\big) = f(a) + h(p-a)$$

- **Data** is discrete (1-year, 5-year) intervals
- **Models** are continuous, prediction at **any** value for a, p or c

Parametric models

- f, g and h are **smooth**, **continuous** functions:

$$\log\bigl(\lambda(a,p)\bigr) = f(a) + g(p) \qquad \log\bigl(\lambda(a,p)\bigr) = f(a) + h(p-a)$$

- **Data** is discrete (1-year, 5-year) intervals
- **Models** are continuous, prediction at **any** value for a, p or c
- Reference is now to a **specific** age or data — not an age-**band** or **period**

Parametric models

- f, g and h are **smooth**, **continuous** functions:

$$\log\bigl(\lambda(a,p)\bigr) = f(a) + g(p) \qquad \log\bigl(\lambda(a,p)\bigr) = f(a) + h(p-a)$$

- **Data** is discrete (1-year, 5-year) intervals
- **Models** are continuous, prediction at **any** value for a, p or c
- Reference is now to a **specific** age or data — not an age-**band** or **period**
- **Results** are functions to be shown as **curves**

Parametric models

- f, g and h are **smooth**, **continuous** functions:

$$\log\big(\lambda(a,p)\big) = f(a) + g(p) \qquad \log\big(\lambda(a,p)\big) = f(a) + h(p-a)$$

- **Data** is discrete (1-year, 5-year) intervals
- **Models** are continuous, prediction at **any** value for a, p or c
- Reference is now to a **specific** age or data — not an age-**band** or **period**
- **Results** are functions to be shown as **curves**
- in the form of **predictions** and

Parametric models

- f, g and h are **smooth**, **continuous** functions:

$$\log\big(\lambda(a,p)\big) = f(a) + g(p) \qquad \log\big(\lambda(a,p)\big) = f(a) + h(p-a)$$

- **Data** is discrete (1-year, 5-year) intervals
- **Models** are continuous, prediction at **any** value for a, p or c
- Reference is now to a **specific** age or data — not an age-**band** or **period**
- **Results** are functions to be shown as **curves**
- in the form of **predictions** and
- **contrasts** between predictions (RR between p and p_{ref})

Quantitative, natural splines I

```
> library(splines)
> ap <- glm( D ~ Ns(A,knots=seq(15,50,,4)) +
+                Ns(P,knots=seq(1950,1990,,5)),
+            offset = log(Y/10^5),
+            family = poisson, data=ts )
> round( ci.lin(ap), 4 )
```

```
                                     Estimate StdErr       z      P    2.5%   97.5
(Intercept)                            0.0499 0.0712  0.7011 0.4833 -0.0896  0.189
Ns(A, knots = seq(15, 50, , 4))1       1.2480 0.0475 26.2816 0.0000  1.1549  1.341
Ns(A, knots = seq(15, 50, , 4))2       3.5475 0.1394 25.4553 0.0000  3.2743  3.820
Ns(A, knots = seq(15, 50, , 4))3      -0.1530 0.0322 -4.7525 0.0000 -0.2161 -0.089
Ns(P, knots = seq(1950, 1990, , 5))1   0.5795 0.0616  9.4032 0.0000  0.4587  0.700
Ns(P, knots = seq(1950, 1990, , 5))2   0.8348 0.0409 20.4259 0.0000  0.7547  0.914
Ns(P, knots = seq(1950, 1990, , 5))3   1.2830 0.0744 17.2465 0.0000  1.1372  1.428
Ns(P, knots = seq(1950, 1990, , 5))4   0.8935 0.0359 24.8785 0.0000  0.8231  0.963
```

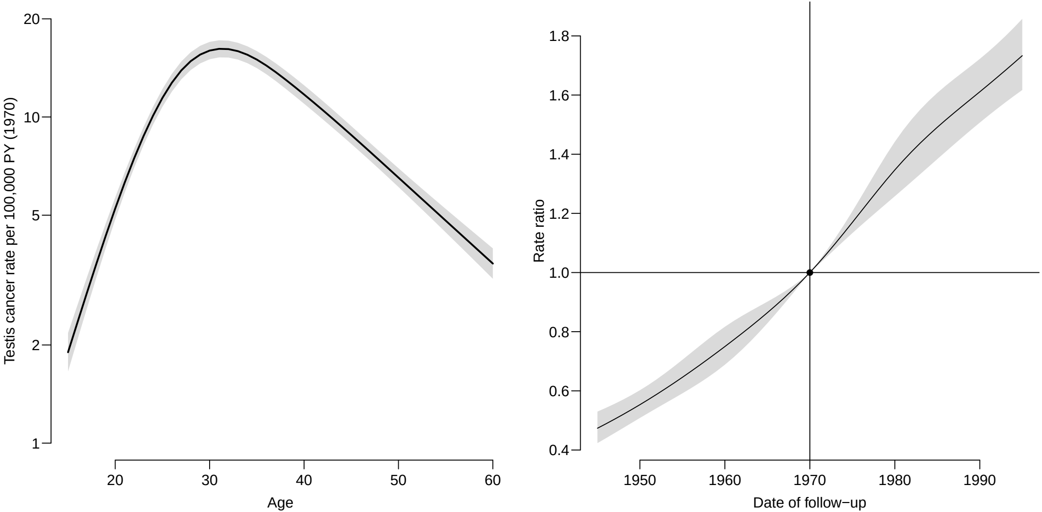
Quantitative, natural splines II

```
> ac <- glm( D ~ Ns(A,knots=seq(15,50,,4)) +
+                Ns(C,knots=seq(1910,1965,,9)),
+             offset = log(Y/10^5),
+             family = poisson, data=ts )
```

Period model predicions I

```
> ndA  <- data.frame( A=15:60, P=1970     , Y=1 )
> ndP  <- data.frame( A=30   , P=1945:1995, Y=1 )
> ndRp <- data.frame( A=30   , P=1970     , Y=1 )
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> matshade( ndA$A,
+           ci.pred(ap,ndA)*10^5, # <- predicted rates using ndA
+           plot=TRUE, log="y", lwd=2, ylim=c(1,20), xlab="Age",
+           ylab="Testis cancer rate per 100,000 PY (1970)" )
> matshade( ndP$P,
+           ci.exp(ap,list(ndP,ndRp)),  # <- RR comparing ndP vs. ndRp
+           plot=TRUE, xlab="Date of follow-up", ylab="Rate ratio" )
> abline( h = 1, v=1970 )
> points( 1970, 1, pch=16 )
```

Estimates from Age-Period model

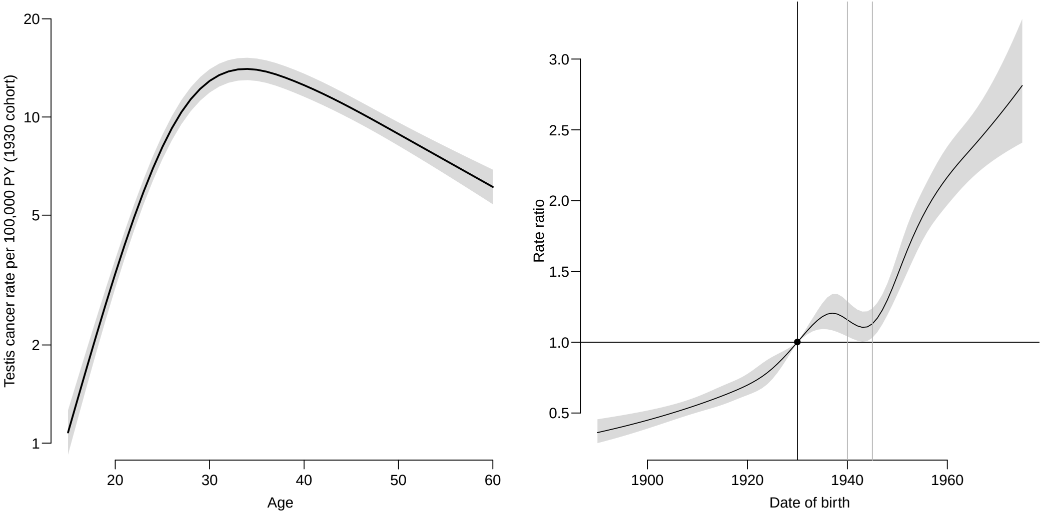


Cohort model I

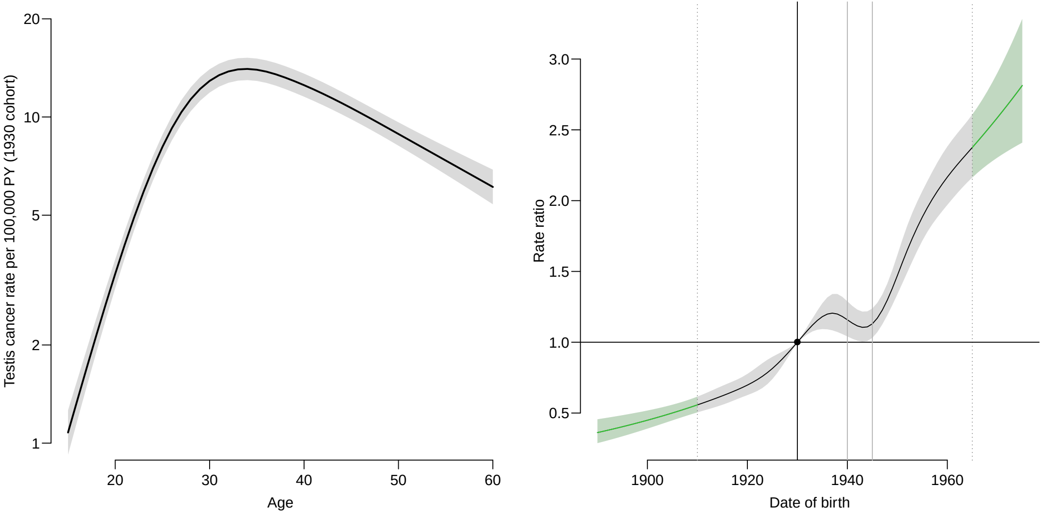
```
> ndA  <- data.frame( A=15:60, C=1930     , Y=1 )
> ndC  <- data.frame( A=30   , C=1890:1975, Y=1 )
> ndRc <- data.frame( A=30   , C=1930     , Y=1 )
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> matshade( ndA$A, ci.pred(ac,ndA)*10^5, plot=TRUE,
+           log="y", lwd=2, ylim=c(1,20), xlab="Age",
+           ylab="Testis cancer rate per 100,000 PY (1930 cohort)" )
> matshade( ndC$C, ci.exp(ac,list(ndC,ndRc)), plot=TRUE,
+           xlab="Date of birth", ylab="Rate ratio" )#, xlim=c(1890,1920), ylim=c
> abline( h = 1, v=1930 )
> abline( v=c(1940,1945), col=gray(0.7) )
> points( 1930, 1, pch=16 )
```

```
> par( mfrow=c(1,2), mar=c(3,3,1,1), mgp=c(3,1,0)/1.6, bty="n", las=1 )
> matshade( ndA$A, ci.pred(ac,ndA)*10^5, plot=TRUE,
+           log="y", lwd=2, ylim=c(1,20), xlab="Age",
+           ylab="Testis cancer rate per 100,000 PY (1930 cohort)" )
> matshade( ndC$C, ci.exp(ac,list(ndC,ndRc)), plot=TRUE,
+           xlab="Date of birth", ylab="Rate ratio" )
> lo <- ndC$C<=1910
> hi <- ndC$C>=1965
> matshade( ndC$C[lo], ci.exp(ac,list(ndC,ndRc))[lo,], col="limegreen" )
> matshade( ndC$C[hi], ci.exp(ac,list(ndC,ndRc))[hi,], col="limegreen" )
> abline(v=c(1910,1965),lty=3,col=gray(0.5))
> abline( h = 1, v=1930 )
> abline( v=c(1940,1945), col=gray(0.7) )
> points( 1930, 1, pch=16 )
```

Estimates from Age-Cohort model



Estimates from Age-Cohort model



Age-drift model

Bendix Carstensen

An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time
CDC, Atlanta,June 2019

`http://BendixCarstensen/APC`

Linear effect of period:

$$\log[\lambda(a,p)] = \alpha_a + \beta_p = \alpha_a + \beta(p - p_0)$$

that is, $\beta_p = \beta(p - p_0)$.

Linear effect of cohort:

$$\log[\lambda(a,p)] = \tilde{\alpha}_a + \gamma_c = \tilde{\alpha}_a + \gamma(c - c_0)$$

that is, $\gamma_c = \gamma(c - c_0)$

Age and linear effect of period:

```
> apd <- glm( D ~ factor( A ) - 1 + I(P-1970.5) +
+             offset( log( Y ) ),
+             family=poisson )
> summary( apd )

Call:
glm(formula = D ~ factor(A) - 1 + I(P - 1970.5) + offset(log(Y)), family = poisso

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.97593  -0.77091   0.02809   0.95914   2.93076

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
factor(A)17.5 -3.58065    0.06306  -56.79   <2e-16
...
factor(A)57.5 -3.17579    0.06256  -50.77   <2e-16
I(P - 1970.5)  0.02653    0.00100   26.52   <2e-16

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 89358.53  on 81  degrees of freedom
Residual deviance:   126.07  on 71  degrees of freedom
```

Age and linear effect of cohort:

```
> acd <- glm( D ~ factor( A ) - 1 + I(C-1933) +
+               offset( log( Y ) ),
+               family=poisson )
> summary( acd )

Call:
glm(formula = D ~ factor(A) - 1 + I(C - 1933) + offset(log(Y)), family = poisson)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.97593  -0.77091   0.02809   0.95914   2.93076

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
factor(A)17.5 -4.11117    0.06760  -60.82   <2e-16
...
factor(A)57.5 -2.64527    0.06423  -41.19   <2e-16
I(C - 1933)    0.02653    0.00100   26.52   <2e-16

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 89358.53  on 81  degrees of freedom
Residual deviance:   126.07  on 71  degrees of freedom
```

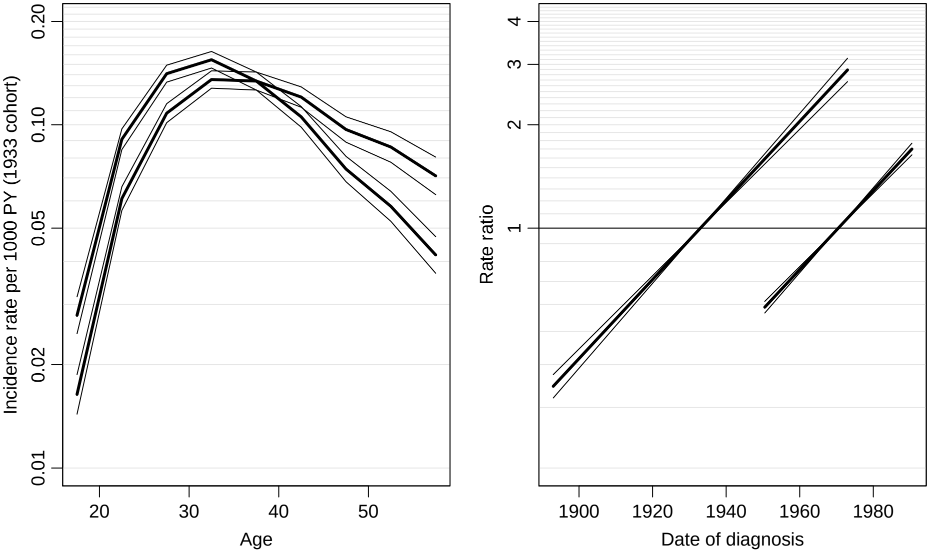
What goes on?

$$p = a + c \qquad p_0 = a_0 + c_0$$

$$\begin{aligned} \alpha_a + \beta(p - p_0) &= \alpha_a + \beta\big(a + c - (a_0 + c_0)\big) \\ &= \underbrace{\alpha_a + \beta(a - a_0)}_{\text{cohort age-effect}} + \beta(c - c_0) \end{aligned}$$

The two **models** are the same.
The **parametrization** is different.

The age-curve refers either
- to a period (cross-sectional rates) or
- to a cohort (longitudinal rates).



Which age-curve is period and which is cohort?

Age-Period-Cohort model

Bendix Carstensen

An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time
CDC, Atlanta,June 2019

`http://BendixCarstensen/APC`

The age-period-cohort model

$$\log[\lambda(a,p)] = \alpha_a + \beta_p + \gamma_c$$

- Three effects:

The age-period-cohort model

$$\log[\lambda(a, p)] = \alpha_a + \beta_p + \gamma_c$$

- Three effects:
  - a — `Age` (at diagnosis)

The age-period-cohort model

$$\log[\lambda(a,p)] = \alpha_a + \beta_p + \gamma_c$$

- Three effects:
  - a — Age (at diagnosis)
  - p — Period (of diagnosis)

The age-period-cohort model

$$\log[\lambda(a,p)] = \alpha_a + \beta_p + \gamma_c$$

- Three effects:
  - a — Age (at diagnosis)
  - p — Period (of diagnosis)
  - c — Cohort (of birth)

The age-period-cohort model

$$\log[\lambda(a,p)] = \alpha_a + \beta_p + \gamma_c$$

- Three effects:
  - a — Age (at diagnosis)
  - p — Period (of diagnosis)
  - c — Cohort (of birth)
- No assumptions about the **shape** of effects.

The age-period-cohort model

$$\log[\lambda(a,p)] = \alpha_a + \beta_p + \gamma_c$$

- Three effects:
  - a — `Age` (at diagnosis)
  - p — `Period` (of diagnosis)
  - c — `Cohort` (of birth)
- No assumptions about the **shape** of effects.
- Levels of `A`, `P` and `C` are assumed **exchangeable**

The age-period-cohort model

$$\log[\lambda(a,p)] = \alpha_a + \beta_p + \gamma_c$$

- Three effects:
  - a — `Age` (at diagnosis)
  - p — `Period` (of diagnosis)
  - c — `Cohort` (of birth)
- No assumptions about the **shape** of effects.
- Levels of `A`, `P` and `C` are assumed **exchangeable**
- *i.e.* no assumptions about the relationship between parameter estimates and the **scaled values** of `A`, `P` and `C`

Fitting the model in R I

```
> m.apc <- glm( D ~ 0 + factor(A) + factor(P) + factor(C),
+               offset = log(Y), family = poisson, data = tc )
> round( ci.lin( m.apc ), 4 )
```

```
                Estimate StdErr        z      P     2.5%    97.5%
factor(A)17.5   -11.3989 0.2332 -48.8886 0.0000 -11.8559 -10.9419
factor(A)22.5   -10.2022 0.2552 -39.9849 0.0000 -10.7023  -9.7021
factor(A)27.5    -9.7634 0.2755 -35.4328 0.0000 -10.3035  -9.2233
factor(A)32.5    -9.6795 0.2974 -32.5482 0.0000 -10.2624  -9.0966
factor(A)37.5    -9.8283 0.3201 -30.7015 0.0000 -10.4557  -9.2009
factor(A)42.5   -10.1047 0.3435 -29.4182 0.0000 -10.7779  -9.4315
factor(A)47.5   -10.5268 0.3676 -28.6390 0.0000 -11.2472  -9.8064
factor(A)52.5   -10.8863 0.3921 -27.7650 0.0000 -11.6548 -10.1179
factor(A)57.5   -11.2709 0.4082 -27.6079 0.0000 -12.0710 -10.4707
factor(P)1950.5   0.2029 0.0825   2.4598 0.0139   0.0412   0.3645
factor(P)1955.5   0.4204 0.0908   4.6297 0.0000   0.2424   0.5984
factor(P)1960.5   0.6410 0.1055   6.0769 0.0000   0.4343   0.8477
```

Fitting the model in R II

```
factor(P)1965.5   0.8214 0.1241   6.6199 0.0000   0.5782   1.0645
factor(P)1970.5   1.0644 0.1444   7.3689 0.0000   0.7813   1.3474
factor(P)1975.5   1.2780 0.1665   7.6738 0.0000   0.9516   1.6044
factor(P)1980.5   1.4344 0.1896   7.5651 0.0000   1.0628   1.8060
factor(P)1985.5   1.5058 0.2134   7.0565 0.0000   1.0875   1.9240
factor(P)1990.5   1.5880 0.2356   6.7396 0.0000   1.1262   2.0498
factor(C)1893     0.5056 0.4289   1.1786 0.2385  -0.3351   1.3463
factor(C)1898     0.5644 0.3840   1.4699 0.1416  -0.1882   1.3170
factor(C)1903     0.2843 0.3556   0.7995 0.4240  -0.4126   0.9812
factor(C)1908     0.2068 0.3284   0.6299 0.5288  -0.4367   0.8504
factor(C)1913     0.2230 0.3034   0.7350 0.4624  -0.3717   0.8177
factor(C)1918     0.0271 0.2815   0.0964 0.9232  -0.5246   0.5789
factor(C)1923     0.0328 0.2597   0.1263 0.8995  -0.4762   0.5418
factor(C)1928     0.0215 0.2394   0.0900 0.9283  -0.4478   0.4909
factor(C)1933     0.0252 0.2199   0.1145 0.9088  -0.4058   0.4561
factor(C)1938    -0.0724 0.2027  -0.3572 0.7209  -0.4696   0.3248
factor(C)1943    -0.3528 0.1871  -1.8862 0.0593  -0.7195   0.0138
factor(C)1948    -0.3047 0.1731  -1.7606 0.0783  -0.6440   0.0345
factor(C)1953    -0.1792 0.1626  -1.1020 0.2705  -0.4978   0.1395
```

Fitting the model in R III

```
factor(C)1958    -0.1174 0.1558  -0.7532 0.4513  -0.4228   0.1881
factor(C)1963    -0.1088 0.1541  -0.7062 0.4801  -0.4108   0.1932
factor(C)1968    -0.1681 0.1623  -1.0353 0.3005  -0.4863   0.1501
factor(C)1973     0.0000 0.0000      NaN    NaN   0.0000   0.0000
```

No. of parameters

`A` has $9(A)$ levels
`P` has $10(P)$ levels
`C=P-A` has $18(C = A + P - 1)$ levels
Age-drift model has $A + 1 = 10$ parameters
Age-period model has $A + P - 1 = 18$ parameters
Age-cohort model has $A + C - 1 = 26$ parameters
Age-period-cohort model has $A + P + C - 3 = 34$ parameters:

```
> length( coef(m.apc) ) ; sum( !is.na(coef(m.apc)) )
```

```
[1] 35
```

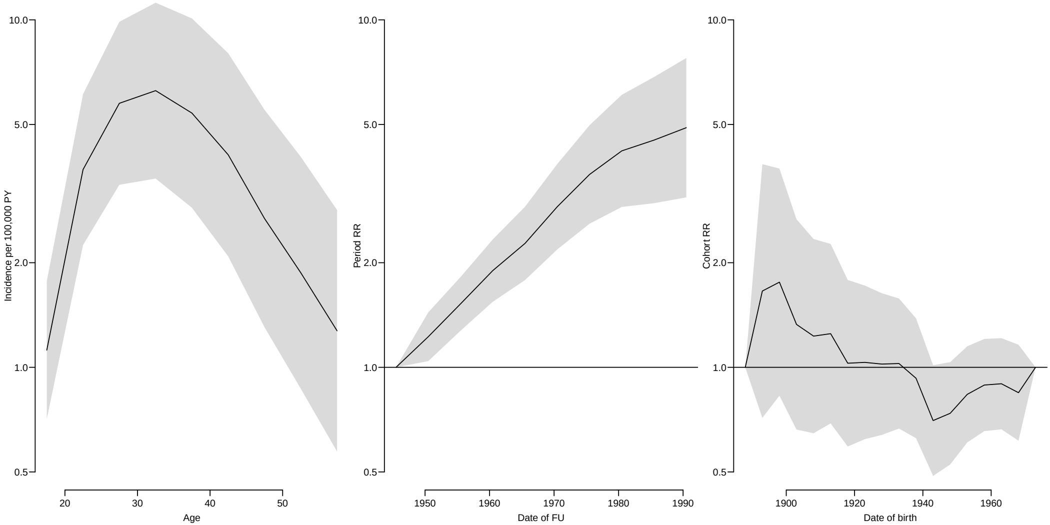
```
[1] 34
```

The missing parameter is because of the **identifiability problem**.

A, P, C effects

```
> par( mfrow=c(1,3), mar=c(3,3,0.1,0.1), mgp=c(3,1,0)/1.6 )
> m.apc <- glm( D ~ 0 + factor(A) + factor(P) + factor(C),
+               offset = log(Y), family = poisson, data = tc )
> #
> matshade( seq(17.5,57.5,5), ci.exp(m.apc,subset="A")*10^5, plot=TRUE,
+           log="y", ylab="Incidence per 100,000 PY", xlab="Age", ylim=c(0.5,10)
> #
> matshade( seq(1945.5,1990.5,5), rbind(1,ci.exp(m.apc,subset="P")), plot=TRUE,
+           log="y", ylab="Period RR", xlab="Date of FU", ylim=c(0.5,10) )
> abline( h=1 )
> #
> matshade( seq(1888,1973,5), rbind(1,ci.exp(m.apc,subset="C")), plot=TRUE,
+           log="y", ylab="Cohort RR", xlab="Date of birth", ylim=c(0.5,10) )
> abline( h=1 )
```

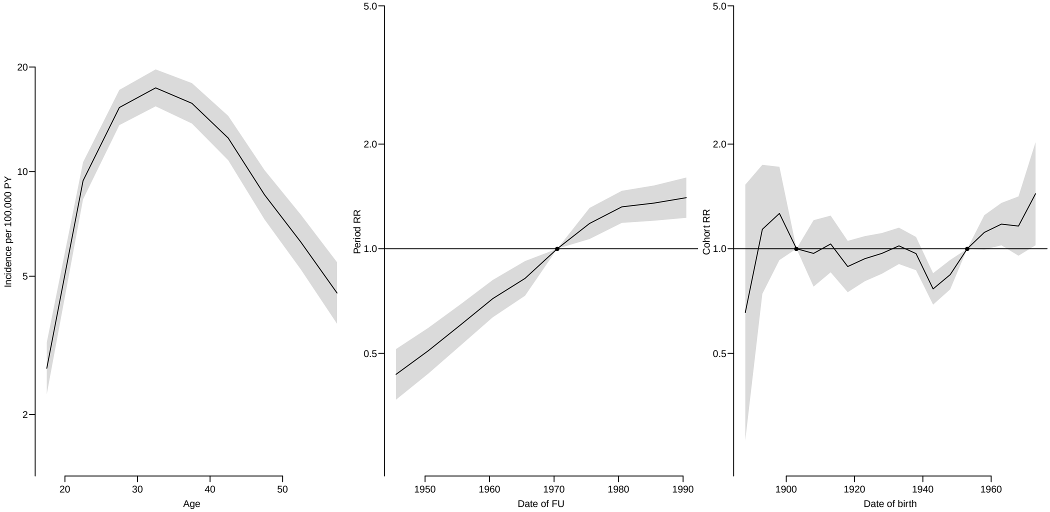
A, P, C effects


10.0
5.0
2.0
1.0
0.5
Incidence per 100,000 PY
20
30
40
50
Age
10.0
5.0
2.0
1.0
0.5
Period RR
1950
1960
1970
1980
1990
Date of FU
10.0
5.0
2.0
1.0
0.5
Cohort RR
1900
1920
1940
1960
Date of birth

A, P, C effects, different reference

```
> m.apc <- glm( D ~ 0 + factor(A) + relevel(factor(P),6) +
+                                   Relevel(factor(C),c(4,1:3,5:13,15:18,14)),
+                 offset = log(Y), family = poisson, data = tc )
> #
> par( mfrow=c(1,3), mar=c(3,3,0.1,0.1), mgp=c(3,1,0)/1.6 )
> matshade( seq(17.5,57.5,5), ci.exp(m.apc,subset="A")*10^5, plot=TRUE,
+           log="y", ylab="Incidence per 100,000 PY", xlab="Age", ylim=c(0.5,10)*
> #
> matshade( seq(1945.5,1990.5,5), rbind(1,ci.exp(m.apc,subset="P"))[c(2:6,1,7:10)
+           log="y", ylab="Period RR", xlab="Date of FU", ylim=c(0.5,10)/2 )
> abline( h=1 ) ; points( 1970.5, 1, pch=16 )
> #
> matshade( seq(1888,1973,5), rbind(1,ci.exp(m.apc,subset="C"))[c(2:4,1,5:13,18,1
+           log="y", ylab="Cohort RR", xlab="Date of birth", ylim=c(0.5,10)/2 )
> abline( h=1 ); points( c(1903,1953), c(1,1), pch=16 )
```

A, P, C effects


Incidence per 100,000 PY
20
10
5
2
20
30
40
50
Age
Period RR
5.0
2.0
1.0
0.5
1950
1960
1970
1980
1990
Date of FU
Cohort RR
5.0
2.0
1.0
0.5
1900
1920
1940
1960
Date of birth

Test for effects

```
> tc.acp <- apc.fit( tc, model="factor", ref.c=1943, print.AOV=FALSE )
```

```
> print( tc.acp$Anova, digits=4 )
```

```
              Model Mod.df. Mod.dev. df.    dev.   Pr(>Chi)  dev/df          H0
1               Age      81  1114.65  NA      NA         NA      NA
2         Age-drift      80   131.77   1 982.879 9.458e-216 982.879 zero drift
3        Age-Cohort      64    70.20  16  61.570  2.840e-07   3.848 Coh eff|dr.
4 Age-Period-Cohort      56    38.78   8  31.418  1.183e-04   3.927 Per eff|Coh
5        Age-Period      72   122.23  16  83.451  3.950e-11   5.216 Coh eff|Per
6         Age-drift      80   131.77   8   9.538  2.990e-01   1.192 Per eff|dr.
```

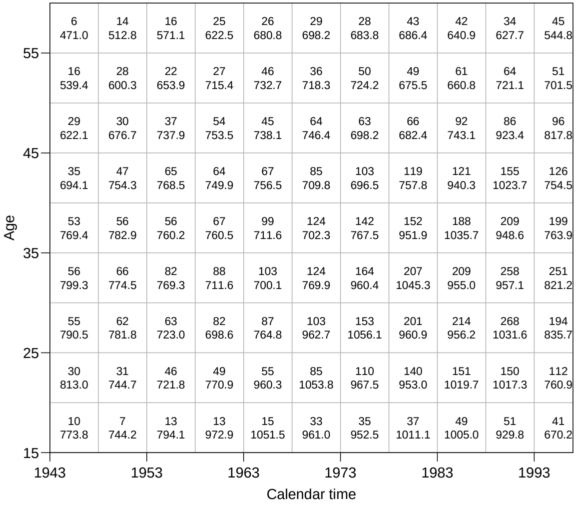
Tabulation in the Lexis diagram

Bendix Carstensen

An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time
CDC, Atlanta,June 2019

`http://BendixCarstensen/APC`

Tabulation of register data

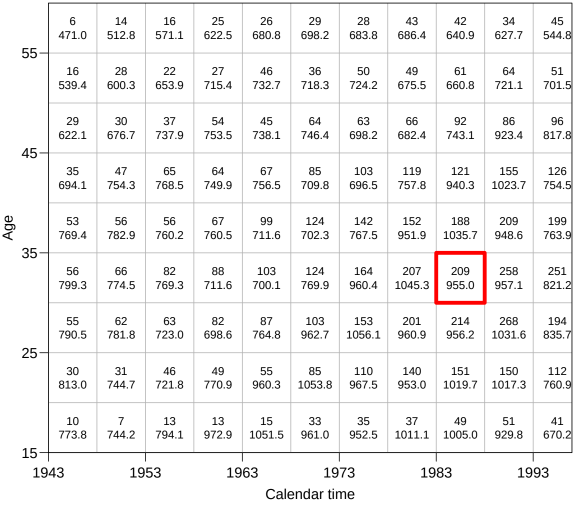


6 471.0	14 512.8	16 571.1	25 622.5	26 680.8	29 698.2	28 683.8	43 686.4	42 640.9	34 627.7	45 544.8
16 539.4	28 600.3	22 653.9	27 715.4	46 732.7	36 718.3	50 724.2	49 675.5	61 660.8	64 721.1	51 701.5
29 622.1	30 676.7	37 737.9	54 753.5	45 738.1	64 746.4	63 698.2	66 682.4	92 743.1	86 923.4	96 817.8
35 694.1	47 754.3	65 768.5	64 749.9	67 756.5	85 709.8	103 696.5	119 757.8	121 940.3	155 1023.7	126 754.5
53 769.4	56 782.9	56 760.2	67 760.5	99 711.6	124 702.3	142 767.5	152 951.9	188 1035.7	209 948.6	199 763.9
56 799.3	66 774.5	82 769.3	88 711.6	103 700.1	124 769.9	164 960.4	207 1045.3	209 955.0	258 957.1	251 821.2
55 790.5	62 781.8	63 723.0	82 698.6	87 764.8	103 962.7	153 1056.1	201 960.9	214 956.2	268 1031.6	194 835.7
30 813.0	31 744.7	46 721.8	49 770.9	55 960.3	85 1053.8	110 967.5	140 953.0	151 1019.7	150 1017.3	112 760.9
10 773.8	7 744.2	13 794.1	13 972.9	15 1051.5	33 961.0	35 952.5	37 1011.1	49 1005.0	51 929.8	41 670.2

Testis cancer cases in Denmark.

Male person-years in Denmark.

Tabulation of register data

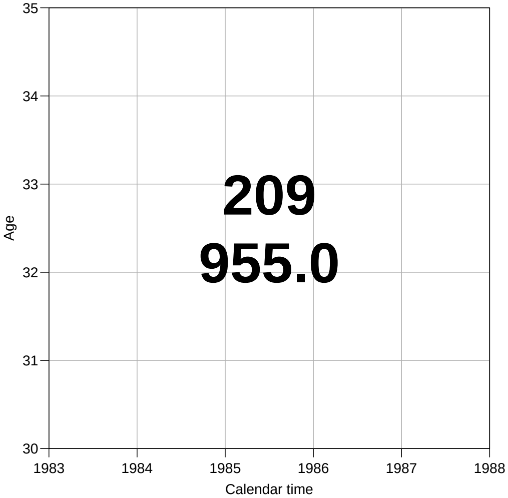


6 471.0	14 512.8	16 571.1	25 622.5	26 680.8	29 698.2	28 683.8	43 686.4	42 640.9	34 627.7	45 544.8
16 539.4	28 600.3	22 653.9	27 715.4	46 732.7	36 718.3	50 724.2	49 675.5	61 660.8	64 721.1	51 701.5
29 622.1	30 676.7	37 737.9	54 753.5	45 738.1	64 746.4	63 698.2	66 682.4	92 743.1	86 923.4	96 817.8
35 694.1	47 754.3	65 768.5	64 749.9	67 756.5	85 709.8	103 696.5	119 757.8	121 940.3	155 1023.7	126 754.5
53 769.4	56 782.9	56 760.2	67 760.5	99 711.6	124 702.3	142 767.5	152 951.9	188 1035.7	209 948.6	199 763.9
56 799.3	66 774.5	82 769.3	88 711.6	103 700.1	124 769.9	164 960.4	207 1045.3	209 955.0	258 957.1	251 821.2
55 790.5	62 781.8	63 723.0	82 698.6	87 764.8	103 962.7	153 1056.1	201 960.9	214 956.2	268 1031.6	194 835.7
30 813.0	31 744.7	46 721.8	49 770.9	55 960.3	85 1053.8	110 967.5	140 953.0	151 1019.7	150 1017.3	112 760.9
10 773.8	7 744.2	13 794.1	13 972.9	15 1051.5	33 961.0	35 952.5	37 1011.1	49 1005.0	51 929.8	41 670.2

Testis cancer cases in Denmark.

Male person-years in Denmark.

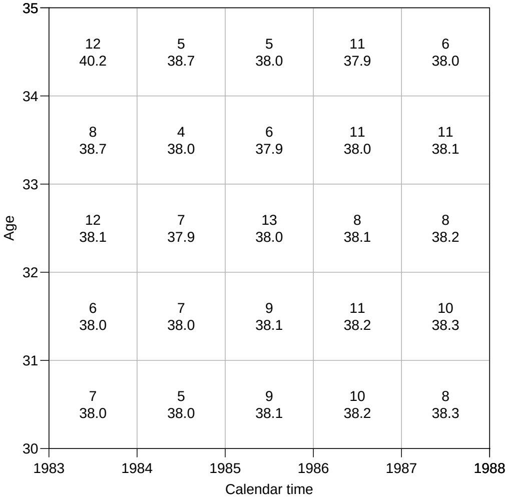
Tabulation of register data



Testis cancer cases in Denmark.

Male person-years in Denmark.

Tabulation of register data

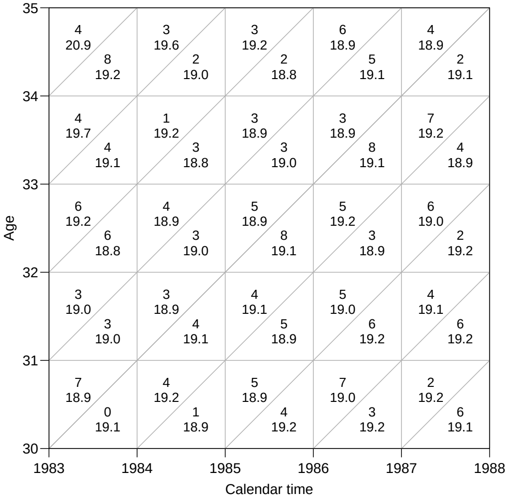


Age	1983	1984	1985	1986	1987
34–35	12 40.2	5 38.7	5 38.0	11 37.9	6 38.0
33–34	8 38.7	4 38.0	6 37.9	11 38.0	11 38.1
32–33	12 38.1	7 37.9	13 38.0	8 38.1	8 38.2
31–32	6 38.0	7 38.0	9 38.1	11 38.2	10 38.3
30–31	7 38.0	5 38.0	9 38.1	10 38.2	8 38.3

Testis cancer cases in Denmark.

Male person-years in Denmark.

Tabulation of register data



Testis cancer cases in Denmark.

Male person-years in Denmark.

Subdivision by year of birth (cohort).

Major sets in the Lexis diagram

A-sets: Classification by age and period. (□)

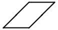
Major sets in the Lexis diagram

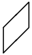
A-sets: Classification by age and period. (□)

B-sets: Classification by age and cohort. (▱)

Major sets in the Lexis diagram

A-sets: Classification by age and period. ()

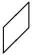
B-sets: Classification by age and cohort. ()

C-sets: Classification by cohort and period. ()

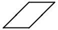
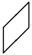
Major sets in the Lexis diagram

A-sets: Classification by age and period. (□)

B-sets: Classification by age and cohort. (▱)

C-sets: Classification by cohort and period. ()

Major sets in the Lexis diagram

- A-sets: Classification by age and period. (□)
- B-sets: Classification by age and cohort. ()
- C-sets: Classification by cohort and period. ()

The mean age, period and cohort for these sets is just the mean of the tabulation interval.

The mean of the third variable is found by using $a = p - c$.

Lexis triangles

Analysis of rates from a complete observation in a Lexis diagram need not be restricted to these classical sets classified by two factors.

Lexis triangles

Analysis of rates from a complete observation in a Lexis diagram need not be restricted to these classical sets classified by two factors.

We may classify cases and risk time by all three factors **Lexis triangles**:

Lexis triangles

Analysis of rates from a complete observation in a Lexis diagram need not be restricted to these classical sets classified by two factors.

We may classify cases and risk time by all three factors

Lexis triangles:

Upper triangles: age and period, earliest born cohort. (◸)

Lexis triangles

Analysis of rates from a complete observation in a Lexis diagram need not be restricted to these classical sets classified by two factors.

We may classify cases and risk time by all three factors
Lexis triangles:

Upper triangles: age and period, earliest born cohort. (◸)
Lower triangles: age and period, latest born cohort. (⊿)

Mean a, p and c during FU in triangles

Modeling requires that each set (=observation in the dataset) be assigned a value of age, period and cohort. So for each triangle we need:

- mean age at risk.

Mean a, p and c during FU in triangles

Modeling requires that each set (=observation in the dataset) be assigned a value of age, period and cohort. So for each triangle we need:

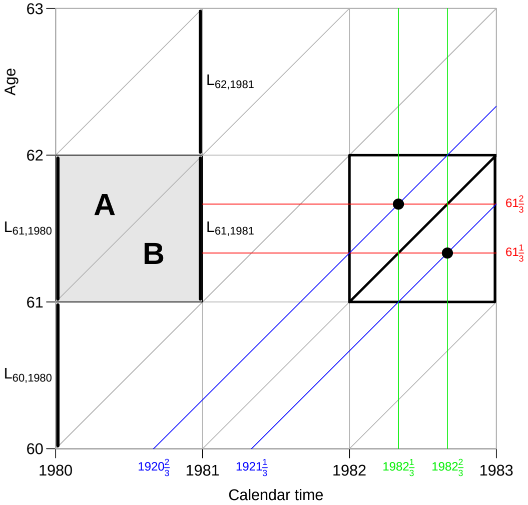
- mean age at risk.
- mean date at risk.

Mean a, p and c during FU in triangles

Modeling requires that each set (=observation in the dataset) be assigned a value of age, period and cohort. So for each triangle we need:

- mean age at risk.
- mean date at risk.
- mean cohort at risk.

Tabulation by age, period and cohort

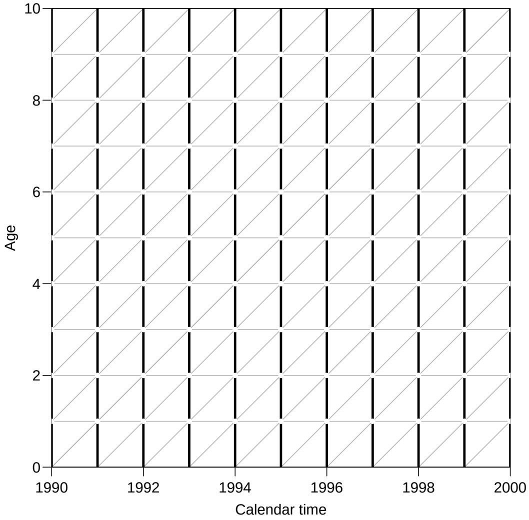


Gives triangular sets with differing mean age, period and cohort:

These are correct midpoints for age, period and cohort must be used in modeling.

From population figures to risk time

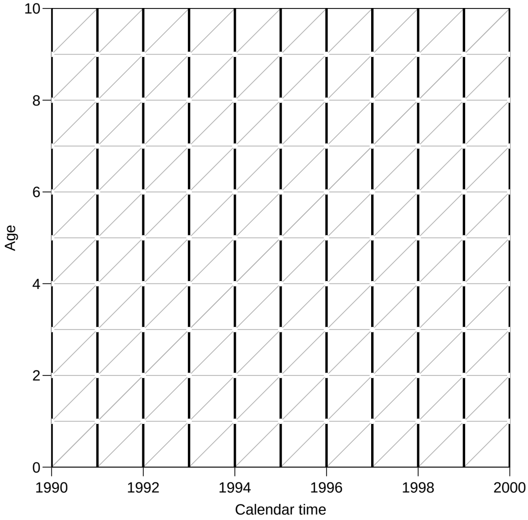
Population figures in the form of size of the population at certain date are available from most statistical bureaus.



From population figures to risk time

Population figures in the form of size of the population at certain date are available from most statistical bureaus.

This corresponds to population sizes along the vertical lines in the diagram.

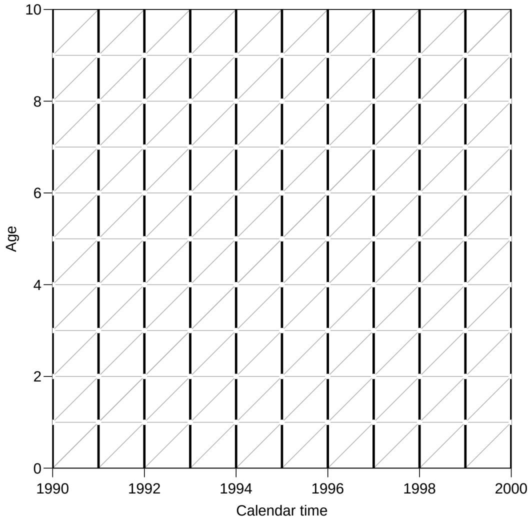


From population figures to risk time

Population figures in the form of size of the population at certain date are available from most statistical bureaus.

This corresponds to population sizes along the vertical lines in the diagram.

We want risk time figures for the population in the squares and triangles in the diagram.



Summary:

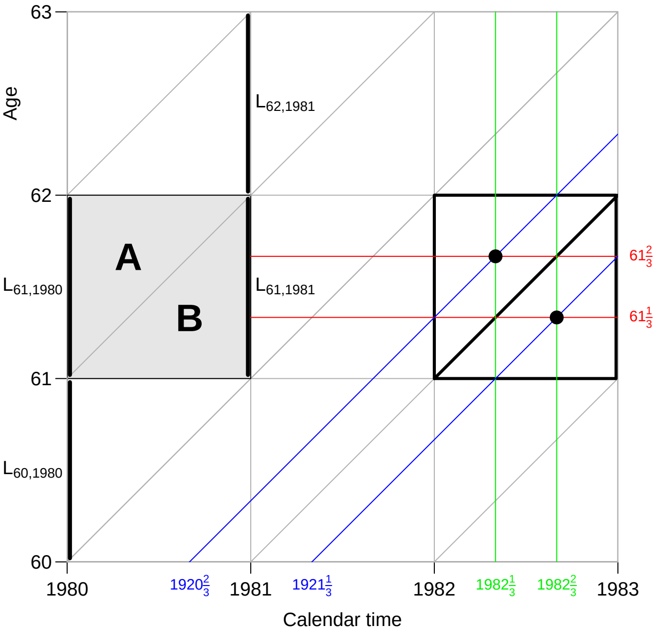
Population risk time (`N2Y`):

A: $\left(\frac{1}{3}\mathrm{L}_{a,p} + \frac{1}{6}\mathrm{L}_{a+1,p+1}\right) \times 1\mathrm{y}$

B: $\left(\frac{1}{6}\mathrm{L}_{a-1,p} + \frac{1}{3}\mathrm{L}_{a,p+1}\right) \times 1\mathrm{y}$

Mean age, period and cohort:

$\frac{1}{3}$ into the interval.



APC-model: Parametrization

Bendix Carstensen

An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time
CDC, Atlanta,June 2019

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

…but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

…but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

$$\log(\lambda_{ap}) \quad = \quad f(a) + g(p) + h(c)$$

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

. . . but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

$$\log(\lambda_{ap}) \quad = \quad f(a) + g(p) + h(c) + \gamma(p - a - c)$$

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

…but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

$$\begin{array}{rcllcl}
\log(\lambda_{ap}) & = & \multicolumn{4}{l}{f(a) + g(p) + h(c) + \gamma(p - a - c)} \\
 & = & f(a) & - & \gamma a & + \\
 & & g(p) & + & \gamma p & + \\
 & & h(c) & - & \gamma c &
\end{array}$$

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

…but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

$$\begin{array}{rcllllll} \log(\lambda_{ap}) & = & \multicolumn{6}{l}{f(a) + g(p) + h(c) + \gamma(p - a - c)} \\ & = & f(a) & - & \mu_p & - & \gamma a & + \\ & & g(p) & + & \mu_p & + & \gamma p & + \\ & & h(c) & & & - & \gamma c & \end{array}$$

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

…but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

$$\begin{aligned}
\log(\lambda_{ap}) &= f(a) + g(p) + h(c) + \gamma(p - a - c) \\
&= f(a) - \mu_p + \mu_c - \gamma a + \\
& g(p) + \mu_p + \gamma p + \\
& h(c) - \mu_c - \gamma c
\end{aligned}$$

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

…but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

$$\begin{array}{rcl}
\log(\lambda_{ap}) & = & f(a) + g(p) + h(c) + \gamma(p - a - c) \\
 & = & \begin{array}{llllllll}
f(a) & - & \mu_p & + & \mu_c & - & \gamma a & + \\
g(p) & + & \mu_p & & & + & \gamma p & + \\
h(c) & & & - & \mu_c & - & \gamma c &
\end{array}
\end{array}$$

A decision on parametrization is needed.

Age-Period-Cohort model

$$\log(\lambda_{ap}) = \alpha_a + \beta_p + \gamma_c = f(a) + g(p) + h(c)$$

…but $\quad c = p - q \quad \Leftrightarrow \quad p - a - c = 0$

$$\begin{aligned}
\log(\lambda_{ap}) \quad &= \quad f(a) + g(p) + h(c) + \gamma(p - a - c) \\
&= \quad f(a) \; - \; \mu_p \; + \; \mu_c \; - \; \gamma a \; + \\
& g(p) \; + \; \mu_p \; \; + \; \gamma p \; + \\
& h(c) \; \; - \; \mu_c \; - \; \gamma c
\end{aligned}$$

A decision on parametrization is needed.

…it must be **external** to the **model**.

Parametrization principle

The problem is to choose μ_a, μ_c and γ according to some (**external!**) criterion for the functions.

Parametrization principle

The problem is to choose μ_a, μ_c and γ according to some (**external!**) criterion for the functions.

1. The age-function should be interpretable as log age-specific rates in a cohort c_0 after adjustment for the period effect.

Parametrization principle

The problem is to choose μ_a, μ_c and γ according to some (**external!**) criterion for the functions.

1. The age-function should be interpretable as log age-specific rates in a cohort c_0 after adjustment for the period effect.
2. The cohort function is 0 at a reference cohort c_0, interpretable as log-RR relative to cohort c_0.

Parametrization principle

The problem is to choose μ_a, μ_c and γ according to some (**external!**) criterion for the functions.

1. The age-function should be interpretable as log age-specific rates in a cohort c_0 after adjustment for the period effect.
2. The cohort function is 0 at a reference cohort c_0, interpretable as log-RR relative to cohort c_0.
3. The period function is 0 on average with 0 slope, interpretable as log-RR relative to the age-cohort prediction. (residual log-RR).

Parametrization principle

The problem is to choose μ_a, μ_c and γ according to some (**external!**) criterion for the functions.

1. The age-function should be interpretable as log age-specific rates in a cohort c_0 after adjustment for the period effect.
2. The cohort function is 0 at a reference cohort c_0, interpretable as log-RR relative to cohort c_0.
3. The period function is 0 on average with 0 slope, interpretable as log-RR relative to the age-cohort prediction. (residual log-RR).

Parametrization principle

The problem is to choose μ_a, μ_c and γ according to some (**external!**) criterion for the functions.

1. The age-function should be interpretable as log age-specific rates in a cohort c_0 after adjustment for the period effect.
2. The cohort function is 0 at a reference cohort c_0, interpretable as log-RR relative to cohort c_0.
3. The period function is 0 on average with 0 slope, interpretable as log-RR relative to the age-cohort prediction. (residual log-RR).

This will yield cohort age-effects a.k.a. **longitudinal** age effects.

Parametrization principle

The problem is to choose μ_a, μ_c and γ according to some (**external!**) criterion for the functions.

1. The age-function should be interpretable as log age-specific rates in a cohort c_0 after adjustment for the period effect.
2. The cohort function is 0 at a reference cohort c_0, interpretable as log-RR relative to cohort c_0.
3. The period function is 0 on average with 0 slope, interpretable as log-RR relative to the age-cohort prediction. (residual log-RR).

This will yield cohort age-effects a.k.a. **longitudinal** age effects.

Biologically interpretable: what happens in the lifespan of a cohort?

Period-major parametrization

- Alternatively, the period function could be constrained to be 0 at a reference date, p_0.

Period-major parametrization

- Alternatively, the period function could be constrained to be 0 at a reference date, p_0.
- Age-effects would refer to age apecific rates at p_0

Period-major parametrization

- Alternatively, the period function could be constrained to be 0 at a reference date, p_0.
- Age-effects would refer to age apecific rates at p_0
- Cohort effects constrained to be 0 on average with 0 slope.

Period-major parametrization

- Alternatively, the period function could be constrained to be 0 at a reference date, p_0.
- Age-effects would refer to age apecific rates at p_0
- Cohort effects constrained to be 0 on average with 0 slope.
- Gives period or **cross-sectional** age-effects

Period-major parametrization

- Alternatively, the period function could be constrained to be 0 at a reference date, p_0.
- Age-effects would refer to age apecific rates at p_0
- Cohort effects constrained to be 0 on average with 0 slope.
- Gives period or **cross-sectional** age-effects

Period-major parametrization

- Alternatively, the period function could be constrained to be 0 at a reference date, p_0.
- Age-effects would refer to age apecific rates at p_0
- Cohort effects constrained to be 0 on average with 0 slope.
- Gives period or **cross-sectional** age-effects

Bureaucratically interpretable: what was seen at a given date?

Implementation:

1. Obtain any set of parameters $f(a)$, $g(p)$, $h(c)$.

Implementation:

1. Obtain any set of parameters $f(a)$, $g(p)$, $h(c)$.
2. Extract the trend from the period effect (find μ and β):

   $$\tilde{g}(p) = \hat{g}(p) - (\mu + \beta p)$$

   (regression of $\hat{g}(p)$ on p)

Implementation:

1. Obtain any set of parameters $f(a)$, $g(p)$, $h(c)$.
2. Extract the trend from the period effect (find μ and β):

   $$\tilde{g}(p) = \hat{g}(p) - (\mu + \beta p)$$

   (regression of $\hat{g}(p)$ on p)
3. Decide on a reference cohort c_0.

Implementation:

1. Obtain any set of parameters $f(a)$, $g(p)$, $h(c)$.
2. Extract the trend from the period effect (find μ and β):

$$\tilde{g}(p) = \hat{g}(p) - (\mu + \beta p)$$

   (regression of $\hat{g}(p)$ on p)
3. Decide on a reference cohort c_0.
4. Use the functions:

$$\begin{aligned}\tilde{f}(a) &= \hat{f}(a) + \mu + \beta a + \hat{h}(c_0) + \beta c_0 \\ \tilde{g}(p) &= \hat{g}(p) - \mu - \beta p \\ \tilde{h}(c) &= \hat{h}(c) \qquad\quad + \beta c - \hat{h}(c_0) - \beta c_0\end{aligned}$$

“Extract the trend”

- **Not** a well-defined concept:

“Extract the trend”

- **Not** a well-defined concept:
    - Regress $\hat{g}(p)$ on p for all units in the dataset.

"Extract the trend"

- **Not** a well-defined concept:
  - Regress $\hat{g}(p)$ on p for all units in the dataset.
  - Regress $\hat{g}(p)$ on p for all different values of p.

“Extract the trend”

- **Not** a well-defined concept:
    - Regress $\hat{g}(p)$ on p for all units in the dataset.
    - Regress $\hat{g}(p)$ on p for all different values of p.
    - Weighted regression — what weights?

“Extract the trend”

- **Not** a well-defined concept:
  - Regress $\hat{g}(p)$ on p for all units in the dataset.
  - Regress $\hat{g}(p)$ on p for all different values of p.
  - Weighted regression — what weights?
- A better founded solution is needed…

“Extract the trend”

- A solution from linear algebra:

“Extract the trend”

- A solution from linear algebra:
    - Take the columns from the parametric period effect,

“Extract the trend”

- A solution from linear algebra:
  - Take the columns from the parametric period effect,
  - projection on the orthogonal to $(1, p)$

“Extract the trend”

- A solution from linear algebra:
  - Take the columns from the parametric period effect,
  - projection on the orthogonal to $(1, p)$
  - requires an innner product in the observation space

“Extract the trend”

- A solution from linear algebra:
  - Take the columns from the parametric period effect,
  - projection on the orthogonal to $(1, p)$
  - requires an innner product in the observation space
  - should be an inner product using person-years as weights

“Extract the trend”

- A solution from linear algebra:
  - Take the columns from the parametric period effect,
  - projection on the orthogonal to $(1, p)$
  - requires an innner product in the observation space
  - should be an inner product using person-years as weights
- Stepwise process:

“Extract the trend”

- A solution from linear algebra:
  - Take the columns from the parametric period effect,
  - projection on the orthogonal to $(1, p)$
  - requires an innner product in the observation space
  - should be an inner product using person-years as weights
- Stepwise process:
  - Fit Age-Cohort model

“Extract the trend”

- A solution from linear algebra:
  - Take the columns from the parametric period effect,
  - projection on the orthogonal to $(1, p)$
  - requires an innner product in the observation space
  - should be an inner product using person-years as weights
- Stepwise process:
  - Fit Age-Cohort model
  - compute the predicted values for the observed data

“Extract the trend”

- A solution from linear algebra:
    - Take the columns from the parametric period effect,
    - projection on the orthogonal to $(1, p)$
    - requires an innner product in the observation space
    - should be an inner product using person-years as weights
- Stepwise process:
    - Fit Age-Cohort model
    - compute the predicted values for the observed data
    - use the log of these as offset in a model with only Period

“Extract the trend”

- A solution from linear algebra:
  - Take the columns from the parametric period effect,
  - projection on the orthogonal to $(1, p)$
  - requires an innner product in the observation space
  - should be an inner product using person-years as weights
- Stepwise process:
  - Fit Age-Cohort model
  - compute the predicted values for the observed data
  - use the log of these as offset in a model with only Period
  - longitudinal age-effects, cohort with a reference and period as residuals

“Extract the trend”

- A solution from linear algebra:
    - Take the columns from the parametric period effect,
    - projection on the orthogonal to $(1, p)$
    - requires an innner product in the observation space
    - should be an inner product using person-years as weights
- Stepwise process:
    - Fit Age-Cohort model
    - compute the predicted values for the observed data
    - use the log of these as offset in a model with only Period
    - longitudinal age-effects, cohort with a reference and period as residuals
- Both implemented in `apc.fit`

ML and residual modeling

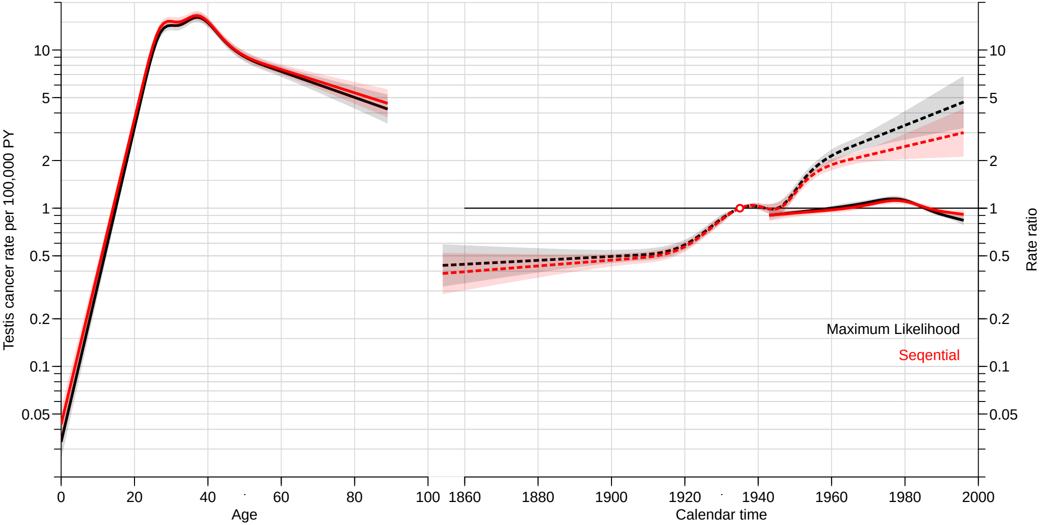
```
> library( Epi )
> data( testisDK )
> head( testisDK )
```

```
  A    P D        Y
1 0 1943 1 39649.50
2 1 1943 1 36942.83
3 2 1943 0 34588.33
4 3 1943 1 33267.00
5 4 1943 0 32614.00
6 5 1943 0 32020.33
```

```
> mm <- apc.fit( data=testisDK, ref.c=1935, parm="ACP" , npar=c(6,5,8), scale=10^
```

```
[1] "ML of APC-model Poisson with log(Y) offset : ( ACP ):\n"
                Model Mod. df. Mod. dev. Test df.  Test dev.     Pr(>Chi) Test dev/
1                 Age     4854  6008.406       NA         NA           NA
2           Age-drift     4853  4864.393        1 1144.01295 8.976155e-251 1144.0129
3          Age-Cohort     4847  4758.975        6  105.41779  1.853664e-20   17.5696
4 Age-Period-Cohort     4844  4704.333        3   54.64241  8.184605e-12   18.2141
5          Age-Period     4850  4846.349        6  142.01605  3.762037e-28   23.6693
6           Age-drift     4853  4864.393        3   18.04415  4.307234e-04    6.0147
```

Two ways of fixing parameters



Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
  - Age is the primary effect

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
  - Age is the primary effect
  - Cohort the secondary, reference c_0

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
  - Age is the primary effect
  - Cohort the secondary, reference c_0
  - Period is the residual

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
  - Age is the primary effect
  - Cohort the secondary, reference c_0
  - Period is the residual
  - Inner product for trend extraction

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
  - Age is the primary effect
  - Cohort the secondary, reference c_0
  - Period is the residual
  - Inner product for trend extraction
- . . . or sequential fitting of models (different model)

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
  - Age is the primary effect
  - Cohort the secondary, reference c_0
  - Period is the residual
  - Inner product for trend extraction
- ...or sequential fitting of models (different model)
- There is no magical fix that allows you to escape this, it comes from using variables a, p and $p - a$

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
  - Age is the primary effect
  - Cohort the secondary, reference c_0
  - Period is the residual
  - Inner product for trend extraction
- …or sequential fitting of models (different model)
- There is no magical fix that allows you to escape this, it comes from using variables a, p and $p - a$
- Any fix has some (hidden) assumption(s)

Parametrization of the APC model is arbitrary

- Separation of the three effects relies on arbitrary principles, *e.g.*:
    - Age is the primary effect
    - Cohort the secondary, reference c_0
    - Period is the residual
    - Inner product for trend extraction
- …or sequential fitting of models (different model)
- There is no magical fix that allows you to escape this, it comes from using variables a, p and $p - a$
- Any fix has some (hidden) assumption(s)
- …but the **fitted values** are the same (except for the sequential method).

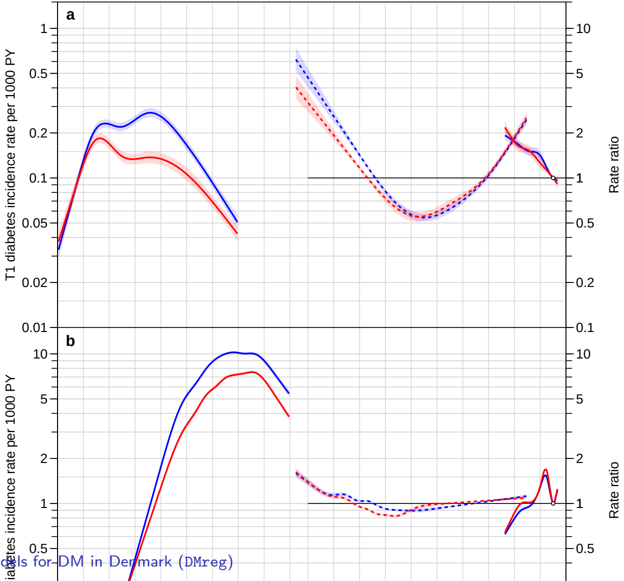
APC-models for DM in Denmark

Bendix Carstensen

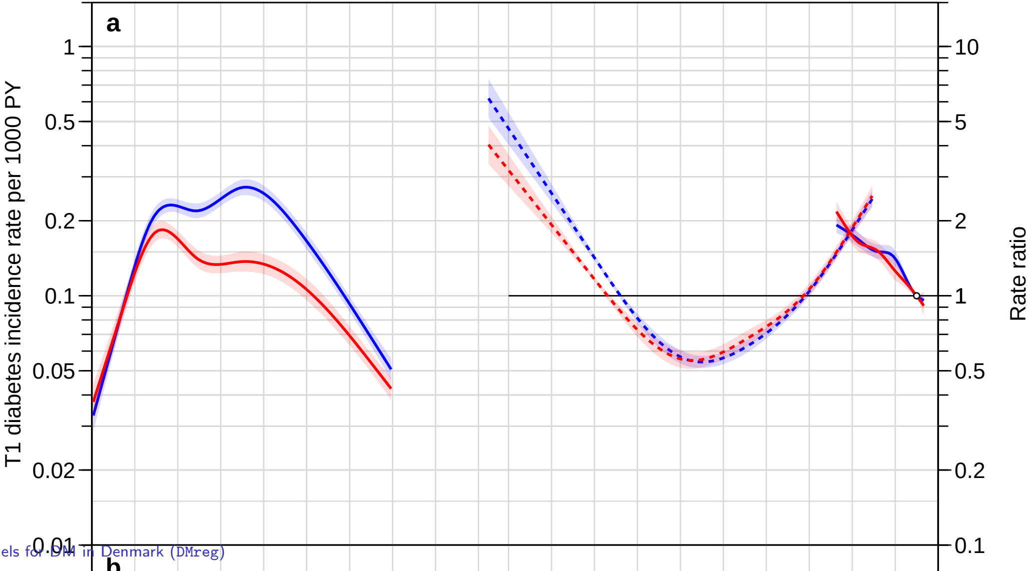
An APC Analytic Approach to Analyzing and Predicting National Trends in Diabetes Incidence over Time
CDC, Atlanta,June 2019

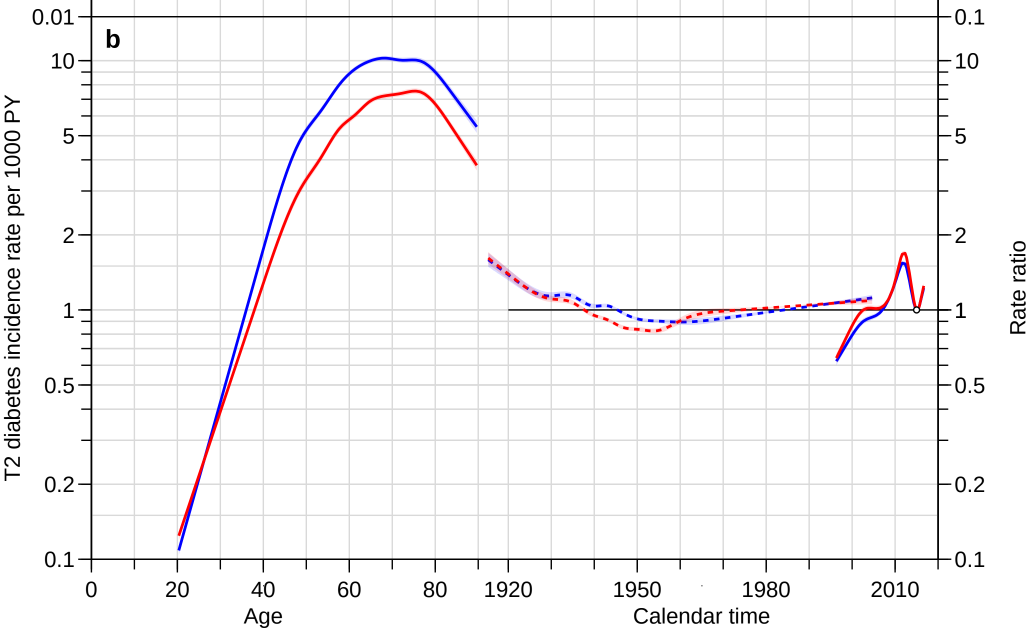
`http://BendixCarstensen/APC`

Age-Period-Cohort analysis of DM in Denmark



Age-Period-Cohort analysis of DM in Denmark




b
T2 diabetes incidence rate per 1000 PY
0.01
10
5
2
1
0.5
0.2
0.1
0
20
40
60
80
Age
1920
1950
1980
2010
Calendar time
0.1
10
5
2
1
0.5
0.2
0.1
Rate ratio

Incidence rates

- T1D:

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)
  - change by calendar time: -3.5% /year

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)
  - change by calendar time: -3.5% /year
- T2D:

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)
  - change by calendar time: -3.5% /year
- T2D:
  - peaks ages 65–80

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)
  - change by calendar time: -3.5% /year
- T2D:
  - peaks ages 65–80
  - decrease after 80

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)
  - change by calendar time: -3.5% /year
- T2D:
  - peaks ages 65–80
  - decrease after 80
  - peak rates at 7–10 cases per 1000 PY (2015)

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)
  - change by calendar time: -3.5% /year
- T2D:
  - peaks ages 65–80
  - decrease after 80
  - peak rates at 7–10 cases per 1000 PY (2015)
  - change by calendar time: 3.3% /year

Incidence rates

- T1D:
  - peaks ages 15–40, weak increase for men, weak decrease for women.
  - decrease after age 40
  - peak rates at 10-20 cases per 100,000 PY (2015)
  - change by calendar time: -3.5% /year
- T2D:
  - peaks ages 65–80
  - decrease after 80
  - peak rates at 7–10 cases per 1000 PY (2015)
  - change by calendar time: 3.3% /year
  - very irregular calendar time pattern

Age-Period-Cohort analysis of DM in Denmark

- Alternative to showing the (arbitrarily fixed) age-, period- and cohort-components, is to show the predicted rates

Age-Period-Cohort analysis of DM in Denmark

- Alternative to showing the (arbitrarily fixed) age-, period- and cohort-components, is to show the predicted rates
- . . . for a fixed age (50 years, say) as a function of calendar time

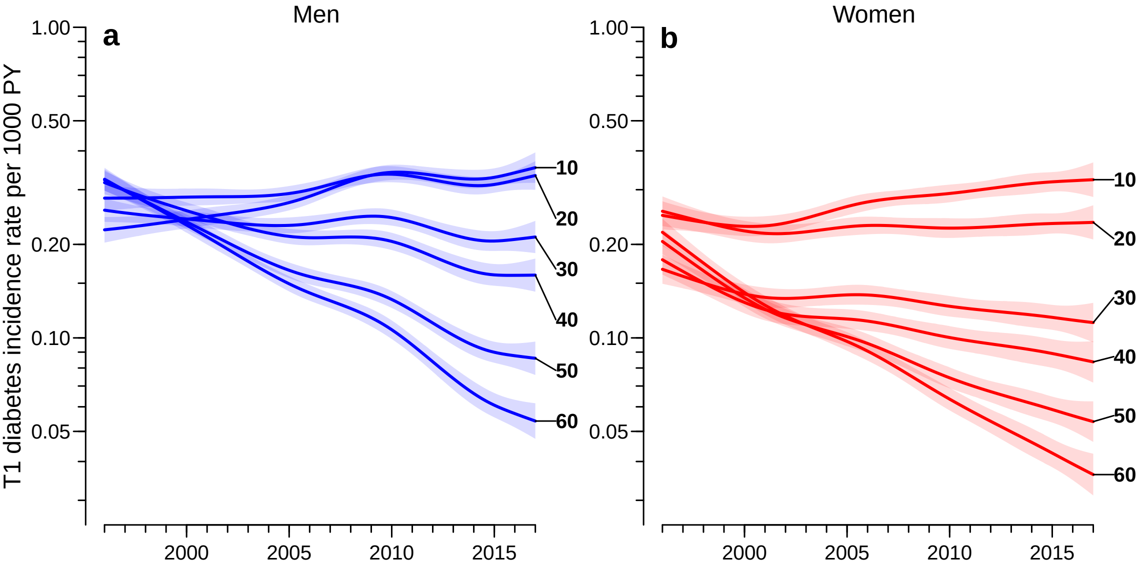
Age-Period-Cohort analysis of DM in Denmark

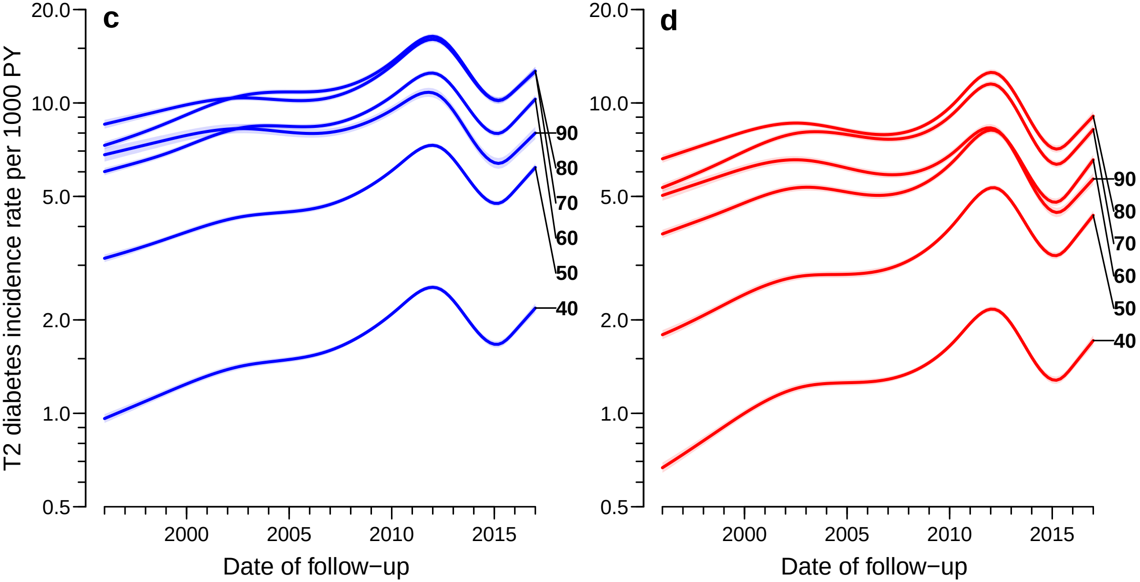
- Alternative to showing the (arbitrarily fixed) age-, period- and cohort-components, is to show the predicted rates
- . . . for a fixed age (50 years, say) as a function of calendar time
- The natural splines constrain P and C components to be linear at the end, so easy to extrapolate rates at any desired age into the future

Age-Period-Cohort analysis of DM in Denmark

- Alternative to showing the (arbitrarily fixed) age-, period- and cohort-components, is to show the predicted rates
- …for a fixed age (50 years, say) as a function of calendar time
- The natural splines constrain P and C components to be linear at the end, so easy to extrapolate rates at any desired age into the future
- …but may overshoot

Age-Period-Cohort analysis of DM in Denmark




c
d
T2 diabetes incidence rate per 1000 PY
20.0
10.0
5.0
2.0
1.0
0.5
90
80
70
60
50
40
2000
2005
2010
2015
Date of follow−up

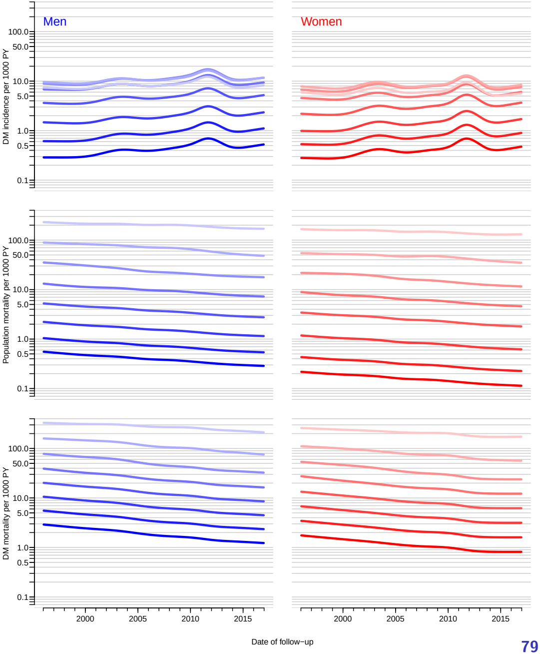
Predictions for total DM

Incidence of total DM

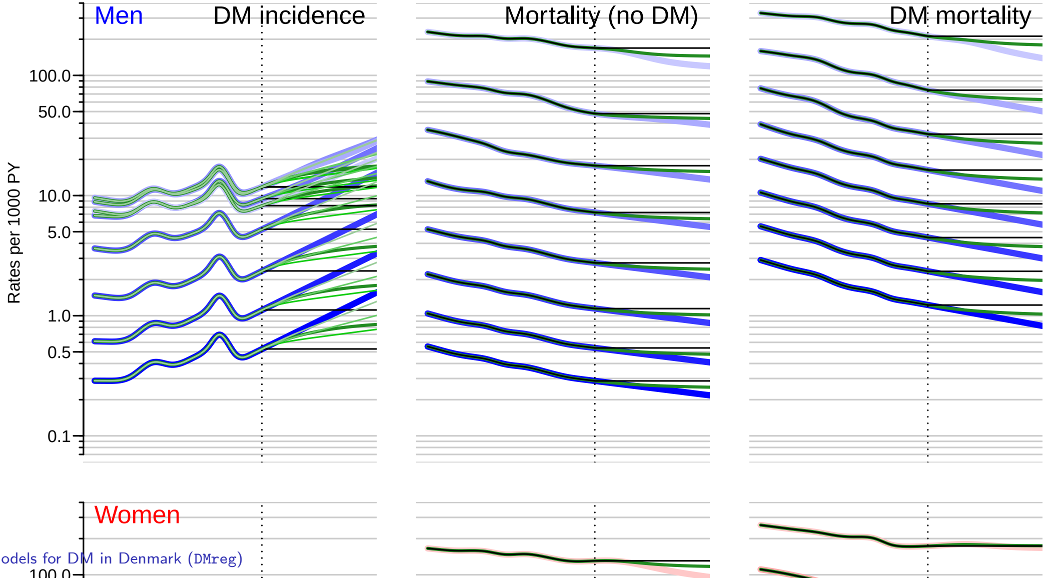
Mortality in total DM

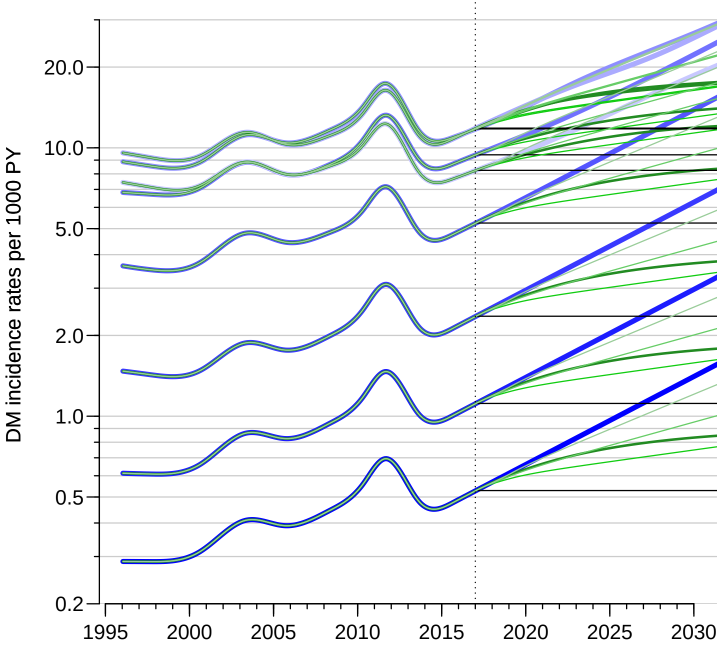
Mortality in no DM

Ages 20, 30,. . . ,90
(strong to weak color)



Future rates for total DM




DM incidence rates per 1000 PY
20.0
10.0
5.0
2.0
1.0
0.5
0.2
1995
2000
2005
2010
2015
2020
2025
2030

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)
  - Simple linear projection of period and cohort effects

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)
    - Simple linear projection of period and cohort effects
    - Attenuation of **slopes** of age-specific rates:
      Every 5 years the slope is halved

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)
    - Simple linear projection of period and cohort effects
    - Attenuation of **slopes** of age-specific rates:
      Every 5 years the slope is halved
    - Simple linear increase in incidence rates 2017–2030:
      0%/year, 2%/year, 4%/year, 6%/year,

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)
  - Simple linear projection of period and cohort effects
  - Attenuation of **slopes** of age-specific rates:
    Every 5 years the slope is halved
  - Simple linear increase in incidence rates 2017–2030:
    $0\%/$year, $2\%/$year, $4\%/$year, $6\%/$year,
- Mortality rates (3 scenarios)

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)
  - Simple linear projection of period and cohort effects
  - Attenuation of **slopes** of age-specific rates:
    Every 5 years the slope is halved
  - Simple linear increase in incidence rates 2017–2030:
    0%/year, 2%/year, 4%/year, 6%/year,
- Mortality rates (3 scenarios)
  - Simple linear projection of period and cohort effects

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)
    - Simple linear projection of period and cohort effects
    - Attenuation of **slopes** of age-specific rates:
      Every 5 years the slope is halved
    - Simple linear increase in incidence rates 2017–2030:
      0%/year, 2%/year, 4%/year, 6%/year,
- Mortality rates (3 scenarios)
    - Simple linear projection of period and cohort effects
    - Attenuation of **slopes** of age-specific rates:
      Every 5 years the slope is halved

Projection scenarios for incidence rates

- Incidence rates (6 scenarios)
  - Simple linear projection of period and cohort effects
  - Attenuation of **slopes** of age-specific rates: Every 5 years the slope is halved
  - Simple linear increase in incidence rates 2017–2030: 0%/year, 2%/year, 4%/year, 6%/year,
- Mortality rates (3 scenarios)
  - Simple linear projection of period and cohort effects
  - Attenuation of **slopes** of age-specific rates: Every 5 years the slope is halved
  - Constant rates as of 2017

Future number of prevalent cases

1. Start with prevalence as of 2017-01-01:
   The predicted prevalences for each month of age (1200 classes)

Future number of prevalent cases

1. Start with prevalence as of 2017-01-01:
   The predicted prevalences for each month of age (1200 classes)
2. Use incidence rates to predict the fraction of non-DM that will be DM one month later (and one month older)

Future number of prevalent cases

1. Start with prevalence as of 2017-01-01:
   The predicted prevalences for each month of age (1200 classes)
2. Use incidence rates to predict the fraction of non-DM that will be DM one month later (and one month older)
3. Use mortality for DM to predict the fraction of the prevalent cases that will survive one month (and be one month older)

Future number of prevalent cases

1. Start with prevalence as of 2017-01-01:
   The predicted prevalences for each month of age (1200 classes)
2. Use incidence rates to predict the fraction of non-DM that will be DM one month later (and one month older)
3. Use mortality for DM to predict the fraction of the prevalent cases that will survive one month (and be one month older)
4. Use mortality for non-DM to predict how many of the non-DM will survive one month (and be one month older)

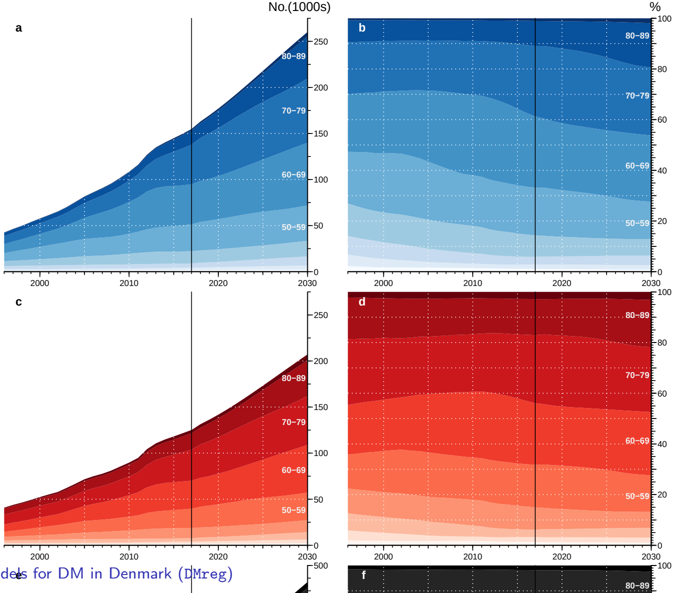
Future number of prevalent cases

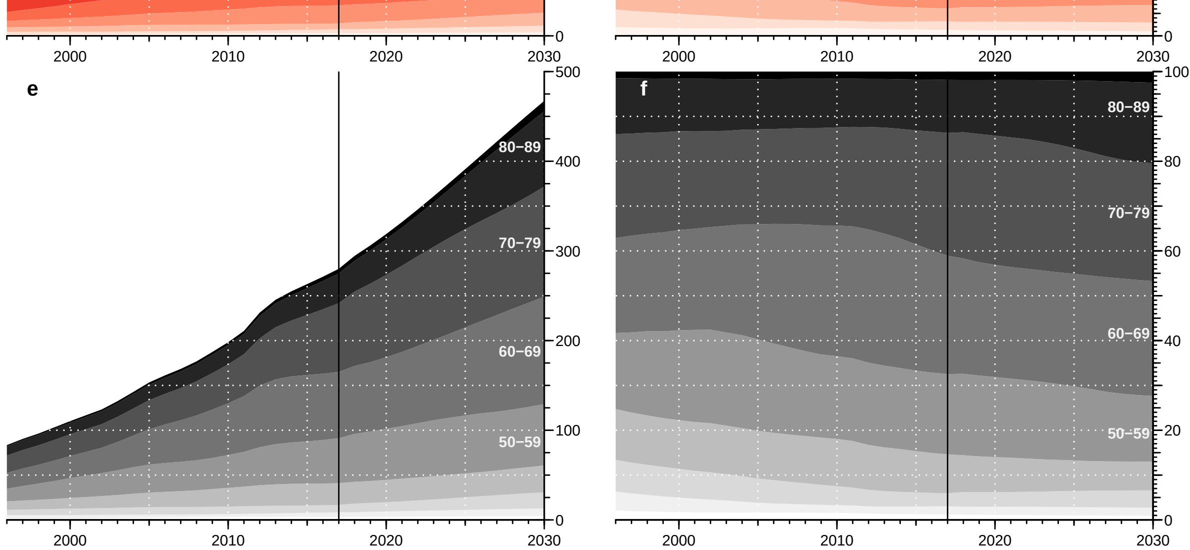
1. Start with prevalence as of 2017-01-01:
   The predicted prevalences for each month of age (1200 classes)
2. Use incidence rates to predict the fraction of non-DM that will be DM one month later (and one month older)
3. Use mortality for DM to predict the fraction of the prevalent cases that will survive one month (and be one month older)
4. Use mortality for non-DM to predict how many of the non-DM will survive one month (and be one month older)
5. From this we know the prevalence of DM as of 2017-**02**-01, in one month older age

Future number of prevalent cases

1. Start with prevalence as of 2017-01-01:
   The predicted prevalences for each month of age (1200 classes)
2. Use incidence rates to predict the fraction of non-DM that will be DM one month later (and one month older)
3. Use mortality for DM to predict the fraction of the prevalent cases that will survive one month (and be one month older)
4. Use mortality for non-DM to predict how many of the non-DM will survive one month (and be one month older)
5. From this we know the prevalence of DM as of 2017-**02**-01, in one month older age
6. Multiply with population forecast from Statistics Denmark to get the **number** of prevalent cases at any future time

Future number of prevalent cases (M/W)




0
2000
2010
2020
2030
0
2000
2010
2020
2030
e
500
400
300
200
100
0
80–89
70–79
60–69
50–59
2000
2010
2020
2030
f
100
80
60
40
20
0
80–89
70–79
60–69
50–59
2000
2010
2020
2030

Future number of prevalent cases (M/W)

- Total no. prevalent cases increase from 287,000 in 2017 to 467,000 in 2030.

Future number of prevalent cases (M/W)

- Total no. prevalent cases increase from 287,000 in 2017 to 467,000 in 2030.
- The population of DM cases will be older — the over-80 will increase from 13 to 20%

Future number of prevalent cases (M/W)

- Total no. prevalent cases increase from 287,000 in 2017 to 467,000 in 2030.
- The population of DM cases will be older — the over-80 will increase from 13 to 20%
- The incidence raes are erratic toward the end of the observation period, so prediction to 2040 is not feasible

Future number of prevalent cases (M/W)

- Total no. prevalent cases increase from 287,000 in 2017 to 467,000 in 2030.
- The population of DM cases will be older — the over-80 will increase from 13 to 20%
- The incidence raes are erratic toward the end of the observation period, so prediction to 2040 is not feasible
- Scenarios with 2%, resp. 4% annual increase from 2017 level of incidence gives predictions of 445,000 and 482,000 prevalent cases.

Mehodological points

- Incidence and mortality in tables by age, period and cohort in 1-year classes (Lexis triangles)

Mehodological points

- Incidence and mortality in tables by age, period and cohort in 1-year classes (Lexis triangles)
- Score the correct mean age, period and cohort in each

Mehodological points

- Incidence and mortality in tables by age, period and cohort in 1-year classes (Lexis triangles)
- Score the correct mean age, period and cohort in each
- Model with smooth functions for age, period and cohort — a kind of parametric smoothing of the rates over the Lexis diagram

Mehodological points

- Incidence and mortality in tables by age, period and cohort in 1-year classes (Lexis triangles)
- Score the correct mean age, period and cohort in each
- Model with smooth functions for age, period and cohort — a kind of parametric smoothing of the rates over the Lexis diagram
- Use the predicted rates in 1-month steps to project future prevalence

Mehodological points

- Incidence and mortality in tables by age, period and cohort in 1-year classes (Lexis triangles)
- Score the correct mean age, period and cohort in each
- Model with smooth functions for age, period and cohort — a kind of parametric smoothing of the rates over the Lexis diagram
- Use the predicted rates in 1-month steps to project future prevalence
- Small steps important — we assume that DM and death cannot occur in the same interval. 1 year intervals rendes this too probable

Mehodological points

- Incidence and mortality in tables by age, period and cohort in 1-year classes (Lexis triangles)
- Score the correct mean age, period and cohort in each
- Model with smooth functions for age, period and cohort — a kind of parametric smoothing of the rates over the Lexis diagram
- Use the predicted rates in 1-month steps to project future prevalence
- Small steps important — we assume that DM and death cannot occur in the same interval. 1 year intervals rendes this too probable
- The parametric compnent of age, period and cohort can only be derived using explicit constraints (3 of them to be precise)

More

A complete account of all analyses is in:
`http://bendixcarstensen.com/DMreg/NewAna.pdf`

A more complete account of APC-modeling can be found in the course material from the European Doctoral School of Demography:
`http://bendixcarstensen.com/APC/EDSD-2019/`