Journal of Statistical and Econometric Methods, vol.2, no.4, 2013, 119-127
ISSN: 2241-0384 (print), 2241-0376 (online)
Scienpress Ltd, 2013

Cohort Coefficients Describing the secular development in protective and detrimental cohort effects associated with apoplexy

Gustav N. Kristensen[1]

Abstract

Cohorts are the aggregate of individuals who experience the same event within the same time interval.
Cohorts can be based on people born in a given year, for example in 1940 or within a span of years, e.g. born in 1940-1944. The year of birth is here the defining event for cohorts.
The health differs between cohorts. This article focuses on the protective and detrimental cohort effect in relation to the risk of death from apoplexy (stroke).
A dummy variable method is recommended to describe the changing cohort effect over a century. Likewise it is shown how information in the data diagonals in an age-period-cohort model can be transferred to a time series model.

Mathematics Subject Classification: Statistics, numerical analysis, special functions
Keywords. Cohorts, apoplexy, ischaemic stroke, century, age-period-cohort.

[1] University of Southern Denmark. Odense.

Article Info: *Received* : September 3, 2013. *Revised* : October 15, 2013
Published online : December 1, 2013

1 Introduction

The purpose of this article is to describe the cohort effect over a century in relation to the death rate from apoplexy (stroke).

Cohorts are according to the words of Ryder (1965) defined as: “The aggregate of individuals (within some population definition) who experience the same event within the same time interval.” According to Ryder: “the defining event [is] birth but this is [of course] only a special case of a more general approach.”
Age, period, and cohorts are crucial elements in describing mortality or death rates.

The death rate from apoplexy for each gender depends on age, time (period), and cohort.

The multicollinearity problem (or identification problem) related to these three elements in case of linearity is well known and described elsewhere (Glenn 1976; Osmond and Gardner 1982; Clayton and Schifflers 1987; Holford 1991; Robertson and Boyle 1998; Rostgaard et al. 2001). A model building based on dummies for cohorts is applied in order to avoid the classical identification or multicollinearity problem in age-period-cohort models.

In statistic publications age and periods are often grouped. In this study “Age” is the lower limit of a 5-year age group. For e.g. the age group 80-84 “Age” is indicated by 80. For the age group 85+ for convenience the oldest persons are set to be 89 years. In total there are 15 age groups from (15-19) to (85+). Periods are annual, and where the calculation starts refer to a birth of the youngest in the age group 85-89 in 1892 while the empirical data cover the span of years 1977-2010, that is 34 calendar years.

2 Data

The Danish data are from: Sundhedsstyrelsen (the Danish Health and Medicines Authority): Dødsårsager (B-061 Karsygdomme i hjerne (Causes of death from haemorrhagia and apoplexia cerebralis) I60-69).

The applied notation is:	Women	Men
Death rate. Number of deaths per 100 000 persons	Dw	Dm

Women and men are distributed into five-year age groups named after the lowest age.

Age group 55-59		Age55
Death rate for apoplexy for age group 55-59	Dw55	Dm55
Historical period		1892 - 1985
Empirical period		1977 - 2010
Cohort born in “year”		CohYear
Time trend		T = 1 for 1977

The overlapping age intervals for cohorts can e.g. be written as 40-44, 41-45, 42-46, etc.

Cohorts are named by the birth year of the youngest in the age group. In Coh1892 the youngest was born in 1892 and the oldest in 1888. That implies that the age group 0-4 is complete in 1892, and that the age group 85-89 in 1977 started as complete in 1892.

The last year in which all five age groups in Coh1892 were alive was in 1977 - the year where the data available for this study starts. Similarly Coh95 indicates that the youngest persons were born in 1995. As the youngest age group is 15-19, it is seen that 1995 + 15 = 2010 is the last included empirical data. Due to the declining observations and due to several "0" death rates in the youngest age groups, the estimation applied stops in 1985.

3 The Age-Period-Cohort Model

The basis for the present model for the death rate from apoplexy is the Gompertz-Makeham law of mortality; see e.g. Gavrilova and Gavrilov (2011).

$$Dw = \alpha e^{\beta_1 Age} \tag{1}$$

$$Log(Dw) = \beta_0 + \beta_1 Age \tag{2}$$

By including a time trend we have

$$Log(Dw) = \beta_0 + \beta_1 Age - \beta_2 T \tag{3}$$

Assumption: The cohort effect is multiplicatively connected to age and period and (proportionally) constant over lifetime for the individual but varies over generations.

The cohorts are included by dummies. Each dummy in principle covers 15 age groups in an age-period diagonal. However, due to the short time span of mpirical data the covered age groups are smaller.

The Age-Period-Cohort Model from the death rate for apoplexy shall now be estimated. The best model of the death rate for women became

$$Log(Dw) = \alpha_0 + \alpha_1 Age + \alpha_2 Age^3 + \alpha_3 Age^3 T^2 + \alpha_4 Age^4 T^2 + \beta_1 Coh1892 + \ldots + \beta_{104} Coh95 + \varepsilon \tag{4}$$

The cohorts go from 1892 to 1995. β_1 is the coefficient for the dummy Coh1892 referring to 1892 in Table 1 below. β_{104} is the coefficient for the dummy Coh95.

The cohorts are related to 34 (calendar) years for 15 age groups. In total there are 34 + 5*14 = 104 cohorts.

Therefore the limits of the coefficients in the estimations of the cohort effects are

β_1	$\rightarrow \beta_i$	$\rightarrow \beta_{104}$	coefficient
Coh1892	- CohYear -	Coh95	dummy variable

The way the dummies are created means that there is no linear connection between age period and cohort. However, not all coefficients β_1 to β_{104} could be included; the death rate for the youngest are in some cases zero and thus could not be described by dummies. The zero death rates here must become a rounding error – that is, in reality an omitted variable problem. Therefore the 10 last cohorts were dropped in the estimation of model (4). After the ten youngest cohorts were dropped due to a too tiny data material, the model includes 94 cohorts.

Because the death rate from apoplexy is increasing with age (and consequently has more observations in high age groups), the random element including the rounding error declines with age. The applied weight in the estimation was therefore Age. An alternative weight could be the number of observations CohYear was based on. For simplicity, this weight was not applied.

Table 1 shows the dummies related to empirical data for the period 1977-2010, while historically they go back to births in 1892. Coh1892 takes the value 1 on the indicated places, otherwise 0. Similar for all other dummies CohYear.

Figure 1 shows the coefficients of the CohYear dummies for women and men. The coefficients indicate the detrimental and protective cohort effects over the years 1892-1985.

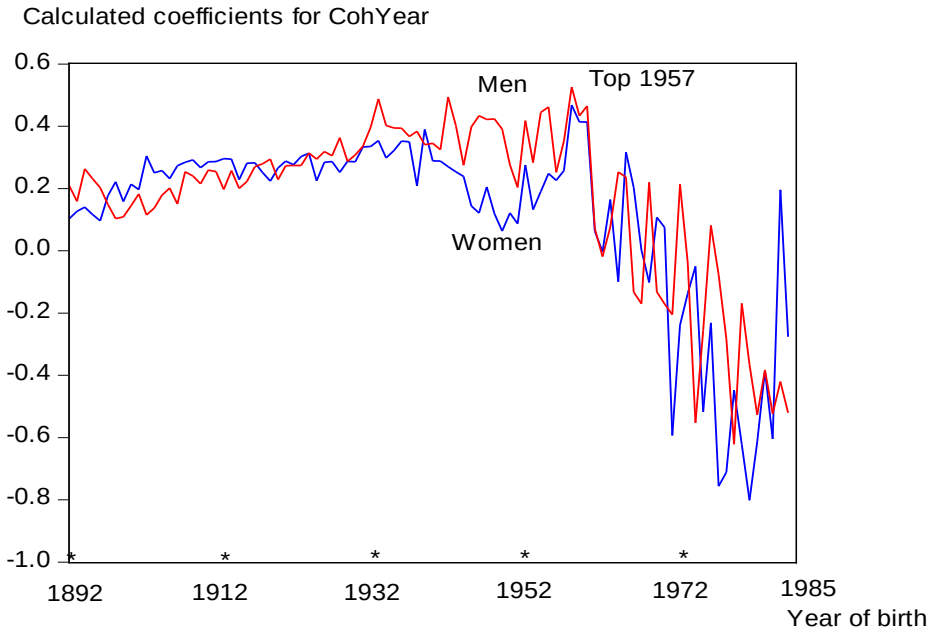

Figure 1: The dummy coefficients for cohorts after (youngest) year of birth 1892-1985.

Table 1: The Cohorts variables (104 years, 1907-2010)

Per\Age	15-19	20-24	25-29…	50-54…	80-84	85+
1907	Coh1892					
…						
1912	Coh1897	**Coh1892**				
…	…	…				
1917	Coh02	Coh1897	**Coh1892**			
…	…	…	…			
1942	Coh27	Coh22	Coh17	**Coh1892**		
…	…	…	…			
1972	Coh**57**	Coh52	Coh47	Coh22	**Coh1892**	
…	…	…	…	…	…	
1977	Coh62	**Coh57**	Coh52	Coh27	Coh1897	**Coh1892**
1978	Coh63	Coh58	Coh53	Coh28	Coh1898	Coh1893
1979	Coh64	Coh59	Coh54	Coh29	Coh1899	Coh1894
1980	Coh65	Coh60	Coh55	Coh30	Coh00	Coh1895
1981	Coh66	Coh61	Coh56	Coh31	Coh01	Coh1896
1982	Coh67	Coh62	Coh**57**	Coh32	Coh02	Coh1897
1983	Coh68	Coh63	Coh58	Coh33	Coh03	Coh1898
1984	Coh69	Coh64	Coh59	Coh34	Coh04	Coh1899
1985	Coh70	Coh65	Coh60	Coh35	Coh05	Coh00
1986	Coh71	Coh66	Coh61	Coh36	Coh06	Coh01
1987	Coh72	Coh67	Coh62	Coh37	Coh07	Coh02
1988	Coh73	Coh68	Coh63	Coh38	Coh08	Coh03
…	…	…	…	…	…	…
1999	Coh84	Coh79	Coh74	Coh49	Coh19	Coh14
2000	Coh**85**	Coh80	Coh75	Coh50	Coh20	Coh15
2001	*Coh86*	Coh81	Coh76	Coh51	Coh21	Coh16
2002	*Coh87*	Coh82	Coh77	Coh52	Coh22	Coh17
2003	*Coh88*	Coh83	Coh78	Coh53	Coh23	Coh18
2004	*Coh89*	Coh84	Coh79	Coh54	Coh24	Coh19
2005	*Coh90*	Coh**85**	Coh80	Coh55	Coh25	Coh20
2006	*Coh91*	*Coh86*	Coh81	Coh56	Coh26	Coh21
2007	*Coh92*	*Coh87*	Coh82	Coh**57**	Coh27	Coh22
2008	*Coh93*	*Coh88*	Coh83	Coh58	Coh28	Coh23
2009	*Coh94*	*Coh89*	Coh84	Coh59	Coh29	Coh24
2010	*Coh**95***	*Coh90*	Coh**85**	Coh60	Coh30	Coh25

Data in italics indicate omitted cohorts due to too few observations of death from apoplexy.

For men and women the cohort effect is detrimental (positive coefficients) until about 1970, and after that basically increasingly protective.
A decline in the coefficients starts from the detrimental top for men and women born in 1957 (Coh57). That is unexplained improvement in survivorship in relation to the death rate from apoplexy, which is beyond the general good development in the trend.

It should be noted that this development is hardly created by (at least direct) medical progress Wardlaw et al. (2009). Modrau et al. (2007, p 1) state that "thrombolysis is the only effective therapy for acute ischaemic stroke. The treatment has been approved in Europe since 2002."

4 Calculations with the model

Inclusion of the cohort effect in a time-series model is useful (Kristensen 2013) and can be made in the following way:

Step 1. In order to use the information in the data diagonals the model must be estimated on the form given in (4).
Step 2. However, calculations can nevertheless be made for the individual age groups in the following way by inserting age in (4) and recalculating the α_j coefficients for each age group:

$$\begin{aligned}\text{Log(Dw)} = {} & [\alpha_0 + \alpha_1 \text{Age} + \alpha_2 \text{Age}^3] + [\alpha_3 \text{Age}^3 + \alpha_4 \text{Age}^4]*\text{T}^2 \\ & + \beta_1 \text{Coh1892} + \ldots\ldots + \beta_{104} \text{Coh95} + \varepsilon \end{aligned} \quad (5)$$

The empirical data covers only 1977-2010. For the specific age groups models for the *columns* in Table 1 over 34 years can now be written:

$$\text{Log(Dw85)} = \gamma_{0,85} + \gamma_{1,85}\text{T}^2 + \beta_1 \text{Coh1892} + \ldots\ldots + \beta_{34} \text{Coh25} \quad (5.85)$$

$$\text{Log(Dw80)} = \gamma_{0,80} + \gamma_{1,80}\text{T}^2 + \beta_6 \text{Coh1897} + \ldots\ldots + \beta_{39} \text{Coh30} \quad (5.80)$$

….

$$\text{Log(Dw45)} = \gamma_{0,45} + \gamma_{1,45}\text{T}^2 + \beta_{41} \text{Coh32} + \ldots\ldots + \beta_{74} \text{Coh65} \quad (5.45)$$

….

$$\text{Log(Dw15)} = \gamma_{0,15} + \gamma_{1,15}\text{T}^2 + \beta_{71} \text{Coh62} + \ldots\ldots + \beta_{104} \text{Coh95} \quad (5.15)$$

In this way the *age specific* death rate for the period 1977-2010 can be calculated with the cohort effects calculated on the *entire* dataset by equation (4).

Step 3. As CohYear is a dummy equal to one, the form of the models in practical estimation for women age group 45-49 becomes (remember T=1 for 1977):

$$
\begin{aligned}
&\text{Log}(\text{Dw45}_{1977}) = \gamma_{0,45} + \gamma_{1,45} 1^2 + \beta_{41} \\
&\text{Log}(\text{Dw45}_{1978}) = \gamma_{0,45} + \gamma_{1,45} 2^2 + \beta_{42} \\
&\text{Log}(\text{Dw45}_{1979}) = \gamma_{0,45} + \gamma_{1,45} 3^2 + \beta_{43} \\
&:::: \\
&\text{Log}(\text{Dw45}_{2010}) = \gamma_{0,45} + \gamma_{1,45} 34^2 + \beta_{74}
\end{aligned}
\qquad (6)
$$

The β_j is thus (in this form) an extra explaining variable B with the expected coefficient one. Having β_j estimated with equation (4) we can estimate the equation:

$$
\begin{aligned}
&\text{Log(Dw45)} = 2.45 - .00055*T^2 + 1.40*B \\
&\text{t} \qquad\qquad (27.27) \quad (-5.59) \qquad (5.01)
\end{aligned}
$$

$R^2 = .714$ adj. $R^2 = .696$ DW = 2.09

5 Discussion

In their article Clayton and Schifflers (1987 II, p 478) criticize Osmond and Gardner (1982), who (in order to come around the identification problem) "introduced a mathematical constraint in the model", and continue: "such a strategy can only be justified if the property which identifies the unique solution has any biological basis and no such justification has been offered."
The present model is not a "mathematical" solution nor a solution based on mathematical constraints but an empirical driven solution that points upon an interesting biological development over a century, which still needs an explanation.

The secular description might thus become an important tool in understanding disease etiology. Holford (1991, p 426) mention that: "children born under the years when it was not uncommon to prescribe diethylstilbestrol (DES) to pregnant women might face lifetime risk for certain types of cancer that differs from that faced by children born at another time".
DES was given to pregnant women from 1938-1971. Sales of DES peaked in 1953. The detrimental cohort effect related to apoplexy peaked in 1957. This proves of course nothing but could inspire to more research on the subject.

The model can look backwards and calculate the trend development back to 1892, but it cannot look forward because there is no explanation as for the future development of the trend.

In Denmark 1957 is known as the start of “the good years” and the development of the welfare state.

6 Conclusion

The dummy model uses data for 1977-2010, but nevertheless can describe the development in cohort effects back to 1892. The dummy method to reveal the cohort effect over a century seems efficient.

Over a span of 100 years the protective and detrimental effects of the birth-defined cohorts have changed in relation to the death rate from apoplexy, with an overwhelming progress for the protective effect.

The cohort effect indicates that people become better at resisting death from apoplexy over years by *birth*.

For men and women a positive development in the protective cohort effect started around 1957.

The period effect “T” indicates a clear *overall* decline in the death rate from apoplexy.

The two effects seem to come from the general improvements in living conditions in Denmark as no medicine against apoplexy was efficient before 2002.

The article likewise show how the cohort effect “diagonal” over age and calendar years can be transformed into a time-series variable.

The estimated results for men and women of equation (4) can be delivered by addressing the author.

Acknowledgements. I am indebted to Klaus Rostgaard for good advice in writing this article.

References

[1] N.B. Ryder, The Cohort as a Concept in the Study of Social-Change, *American Sociological Review*, **30,** (1965), 843-861.

[2] N.D. Glenn, Cohort Analysts’ Futile Quest: Statistical Attempts to Separate Age, Period and Cohort Effects, *American Sociological Review*, **41**(5), (1976), 900-904.

[3] C. Osmond and M.J. Gardner, Age, Period and Cohort Models applied to Cancer Mortality Rates, *Statistics in Medicine*, **1**, (1982), 245-259.

[4] D. Clayton and E. Schifflers, Models for temporal variation in cancer rates. I. Age-period and age-cohort models. *Statistics in Medicine,* **6**, (1987), 449-467.

[5] D. Clayton and E. Schifflers, Models for temporal variation in cancer rates. II. The age-period-cohort model. *Statistics in Medicine,* **6**, (1987), 469-481

[6] T.R. Holford, Understanding the effects of age, period, and cohort on incidence and mortality rates, *Annu. Rev. Publ. Health*, **12**, (1991), 425-457.
[7] C. Robertson and P. Boyle, Age-period-cohort analysis of chronic disease rates. I: Modeling approach. *Statistics in Medicine*, **17**, (1998), 1305-1323.
[8] K. Rostgaard, M. Vath, H. Holst, M. Madsen and E. Lynge. Age-period-cohort modeling of breast cancer incidence in Nordic countries, *Statistics in Medicine*, **20**, (2001), 47-61.
[9] Sundhedsstyrelsen: Dødsårsager (B-61 Karsygdomme i hjerne (haemorrhagia et apoplexia cerebralis) I60-69), Register data.
[10] N.S. Gavrilova and L.A. Gavrilov, Aging and Longevity: Mortality laws and mortality forecasts for aging populations. *Demografie*, **53**(2), (2011), 109-128. (English text).
[11] J.M. Wardlaw, V. Murray, E. Berge and G.J. del Zoppo. Thrombolysis for acute ischaemic stroke (Review). *The Corchrane Library*, **4**, (2009), 1-18.
[12] B. Modrau, K. Vestergaard, H.K. Iversen et al. Thrombolysis for acute ischaemic stroke. (Danish text). *Ugeskr Laeger*, **169**, (2007), 3383-3385.
[13] G.N. Kristensen. The Vintage Waves in the Death Rate Data for Apoplexy, Paper presented on the *Symposium in Applied Statistics (2013)*, pp 209-219, University of Aarhus, 2013.