UNIVERSITY OF LEEDS

This is a repository copy of *Age-period-cohort analysis of trends in amyotrophic lateral sclerosis incidence*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/102190/

Version: Accepted Version

Article:
Tobin, K, Gilthorpe, MS orcid.org/0000-0001-8783-7695, Rooney, J et al. (4 more authors) (2016) Age-period-cohort analysis of trends in amyotrophic lateral sclerosis incidence. Journal of Neurology, 263 (10). pp. 1919-1926. ISSN 0340-5354

https://doi.org/10.1007/s00415-016-8215-z

© 2016, Springer-Verlag Berlin Heidelberg. This is an author produced version of a paper published in Journal of Neurology. Uploaded in accordance with the publisher's self-archiving policy.The final publication is available at Springer via http://dx.doi.org/10.1007/s00415-016-8215-z

Reuse
Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Title: Age-Period-Cohort analysis of trends in Amyotrophic Lateral Sclerosis incidence

Authors:
Katy Tobin PhD[1,2], Mark S. Gilthorpe PhD[3], James Rooney MSc[1], Mark Heverin MSc[1], Alice Vajda PhD[1], Anthony Staines PhD[4], Orla Hardiman MD[1]

[1]Academic Unit of Neurology, School of Medicine, Trinity College Dublin
[2]Centre for Health Policy and Management, Trinity Colege Dublin
[3]Division of Epidemiology and Biostatistics, School of Medicine, University of Leeds
[4]School of Nursing and Health Sciences, Dublin City University

Corresponding Author: Katy Tobin
Address: Academic Unit of Neurology, TBSI, 152-160 Pearse Street, Dublin 2, Ireland

Phone: 00353857180805
Email: katy.tobin@tcd.ie

Total word count: 3938 (including title page, references, legends and abstract)

Running title: APC analysis of ALS incidence

Abstract

Background: Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease with an unknown cause. Studies have reported that the incidence rate of ALS might be changing. As ALS is an age related disease, crude incidence could increase as population structure changes, and as overall life expectancy improves.

Methods: Age Period Cohort (APC) models are frequently used to investigate trends in demographic rates such as incidence. Age-specific incidence rate for ALS from 1996-2014 were taken from a population based ALS register in Ireland. In order to circumvent the well-known identifiability issue in APC models, we apply the method of Partial Least Squares Regression to separate the effects of Age, Period and Cohort on ALS incidence over time.

Results: This APC analysis shows no cohort effect and the initial signs of a period effect; increasing incidence of ALS in the most recently diagnosed group.

Conclusions: As further years of data accrue to the Irish register it will become clear if this effect emerges as a strong trend in the incidence of ALS in Ireland and replication of these analyses in other populations will show if our findings on temporal patterns in ALS incidence are shared elsewhere.

Keywords: Age-Period-Cohort analysis, Amyotrophic Lateral Sclerosis, Incidence patterns, Motor Neuron Disease, Partial Least Squares Regression

Amyotrophic Lateral Sclerosis (ALS) is a rare progressive neurodegenerative disease with an unknown cause. The crude incidence of ALS for those aged 18 or over is estimated to be approximately 2.7 per 100,000 person-years in Europe [1]. The annual incidence in Ireland is about 2.6 per 100,000 [2], with approximately 110 new cases diagnosed annually.

A number of studies have been published exploring patterns in ALS incidence over time [3-5]. A recent systematic review of the literature [6] showed that of 21 studies reporting on changes in ALS incidence over similar time periods to this study, results were variable, with more than half reporting stable incidence, five with increased incidence rates and the remainder showing highly variable rates. Trends in overall (crude) incidence rate may be driven by changing population structure; population ageing is a continuing trend in all European Union (EU) member states [7]. As ALS is an age related disease, crude incidence could increase as population demographics change, and as overall life-expectancy improves. The apparent incidence might also rise, solely because of a reduction in competing causes of death earlier in life, for example, the rapid fall in coronary heart disease mortality in many developed countries over the last thirty years [8,9]. Better case ascertainment over time, and subtle changes in diagnostic criteria, may also influence trends in reported incidence,

as the disease becomes more recognisable, and variations in presentation are included (e.g the recognition that frontotemporal dementia is closely associated with the condition). Thus, as diagnostic accuracy improves, there may be a perceived increase in incidence over time. Another potential factor influencing incidence rates arises from variation in early-life conditioning that affects later-life disease risk, thereby invoking variations across birth cohorts. As the causes of ALS are unknown, and possibly associated with exposure to environmental factors, experiences throughout the lifecourse are potentially important but the extent to which there may be critical periods remains unclear. Examining ALS incidence for various birth cohorts might reveal subgroups of the population at higher risk of disease due to exposure to some unknown risk factors during their lifetime.

Age-Period-Cohort (APC) models are frequently used in epidemiological studies to investigate temporal trends in demographic rates, such as incidence and mortality. To our knowledge, only three other APC analyses have been published on ALS [10-12]. Seals et al. [12] applied a method described by Cartensen [13] to develop an APC model for ALS incidence from 1982 to 2009 and mortality from 1970 to 2009 in Denmark. In these instances data were taken from hospitalisation and death records rather than from population-based registers. Their model attributes an increase in ALS incidence and mortality to a birth cohort effect, while period effects are less important in explaining these temporal changes. A 2011 study of APC effects on Motor Neurone Disease (MND) mortality in France between 1968 – 2007 applied a Poisson model to data taken from national records [11]. The model showed that the increasing mortality rate was better explained by a cohort effect, than by time of death (period effect). A third analysis of mortality data from Switzerland between 1942 and 2008 applied a logit model to investigate APC effects in ALS mortality and concluded that

there was no strong evidence of a birth cohort effect on ALS [10].

A longstanding problem in conducting APC analysis is the issue of perfect collinearity between these variables. An identifiability issue arises due to the intrinsic mathematical relation, age + cohort = period. This presents a methodological challenge, as standard regression techniques cannot be applied. A number of methods have been used previously for modelling APC effects in population data [14]. In recent years, the method of partial least squares regression (PLSR) has been used to partition the three components effects of age, period and cohort and applied to health outcomes. PLSR overcomes the identifiability issue by employing an algorithm that optimizes the estimation separately for all three components. Previous PLSR models have explored effects in blood pressure [15], obesity [16] and overall mortality [17]. In this study we seek to describe the APC trends in ALS incidence in Ireland between 1996-2013 in those aged 40-89 using PLSR, thereby partitioning the effects separately for age, period and cohort. This is the first APC study to use population-based ALS Register data.

METHODS/DATA:

Patient Cohorts: Age specific incidence rates were calculated using data from a national based register for ALS established in 1995. Ethical approval was granted by Beaumont Hospital Ethics (Medical Research) Committee and all subjects gave informed consent to their data being used for research purposes. All diagnosed cases of ALS in Ireland captured on the ALS Register between 1996 and 2014 were analysed. The total number of cases included in the analysis was N=1734, 55.1% (n=956) were males. Spinal onset of ALS accounted for 57.6% of the sample (n=999), and 34.1% (n=592) were classified as having bulbar onset. This national register has

almost complete case ascertainment for the Republic of Ireland [18]. Age specific population estimates were taken from national records [19]. The analysis excluded patients younger than 40 and older than 89 years of age due to sparse data. For the purpose of this APC analysis, age was defined as the age of the patient at the time of diagnosis, period was the year of diagnosis, and cohort was the year of birth. Data were split into ten 5-year age groups (ranging 40-89 years) and four time periods (ranging from 1996-2014). Age-specific incidence rates were calculated for each time period; sample size, group definitions and their associated incidence rates are presented in an Age by Period table (Table 1). In aggregated data tables such as this, each diagonal represents a cohort of data. The total number of cohorts is equal to the sum of the number of Age and Period groups, minus one (Age + Period – 1 = 13) [20].

(Table 1 here)

Traditional regression techniques require a covariate matrix that is full rank and thus invertible in order to produce unique coefficient estimates. Due to their perfectly collinear relationship, introducing age, period and cohort as covariates into a single generalised linear regression model results in a covariate matrix that is not full rank and consequently models fail to estimate unique coefficients for these three variables. However, the PLSR method does not require a covariate matrix to be full rank and invertible, and therefore avoids the issue of identifiability. Properties and assumptions of the PLSR method and its application to age-period-cohort problems are discussed in detail elsewhere [15, 20]. Briefly, partial least squares extracts weighted components of the explanatory variables, maximising the covariance between these components and the outcome variable under the assumed constraint that age + period = cohort. The method then sorts the extracted components in order of decreasing

covariance with respect to the outcome variable. The R^2 value produced by the analysis can be used as a guide in choosing the number of components to be extracted, since the first few components will explain the most covariance with the outcome, and a small change in the R^2 value will yield little further explanatory power. In this instance, there are a maximum of two components and we have to decide if one or both are required to explain optimally the relationship between covariates and the outcome.

Here, we apply the methodology to ALS in Ireland over a period of 19 years to explore APC effects on incidence over time. For our analyses, changes in the percentage variance explained in the output variable by the model was used as a criterion for choosing the number of model components to be extracted.

Statistical Methods:

All analyses were carried out using R software version 3.1.2 [21]. Partial least squares analysis used the pls package for R software [22]. Our first analysis was a linear analysis of incidence with three covariates: age, period and cohort. To explore potential non-linear trends, a second analysis was carried out with 27 dummy variables, one for each age, period and cohort grouping. The design matrix for this non-linear analysis is shown in the appendix.

Since PLSR makes no assumption about the distribution for coefficients [23], confidence intervals were calculated using the jack-knife method [20].

RESULTS:

Descriptive analysis:

A basic descriptive analysis was carried out to visualise the data prior to applying the PLSR model. Figures 1A, 1B and 1C show ALS incidence according to age, period of diagnosis and birth cohort fitted with either non-parametic or parametric smoothed curves and a 95% confidence interval. In most population based analyses, ALS incidence increases with age, and peaks in late mid-life, with a slight decrease in incidence for those aged over 70 years. The rate of newly diagnosed cases appears relatively stable over time, as shown in Figure 1B, with a potential increase for the most recent years. Incidence rates vary for different birth cohorts. Figure 1C highlights some important features of the data: first, data are incomplete for the later cohorts, as individuals can develop ALS later in life beyond the current data collection, meaning incidence rates increase with earlier birth cohorts; and second, data are sparse for the earlier cohorts (1907-1916) , in the first years of the register. These data features and their implications are discussed below.

(Figure 1A here)

(Figure 1B here)

(Figure 1C here)

Figure 2 shows the incidence per 100,000 for each age group by diagnosis year. A loess curve ($\alpha = 0.07$) was added to the raw data to show the patterns of change over time. The solid line represents the mean incidence for all ages per diagnosis year. Incidence for those aged 40-50 has remained stable over time, while those in older age groups are subject to greater variation. The 80-89 age group has seen an increase in incidence since approximately 2003.

(Figure 2 here)

PLSR Models:

Our first PLSR model explores the linear effects with one component of the covariates age, period and cohort on ALS incidence. Two components is the maximum number that can be extracted from this linear model. To decide how many components should be used we observe the change in explanatory power of the model, given by R^2, as additional components are added. The model with one component had an R^2 value of 66%, which increased only slightly to 69% with the two-component model, hence the one-component model was preferred due to parsimony. Table 2 shows the regression coefficients and confidence intervals for the one-component model. Both age and cohort showed similar positive associations with incidence.

(Table 2 here)

No effect on incidence rate was found for period; a notable effect was found for age (0.13, 95% CI: 0.09 – 0.16) with incidence rates increasing with each increasing 5 years of age; and a similar effect found for cohort (0.13, 95% CI: 0.09 - 0.17) with incidence rates of ALS increasing amongst older cohorts. This model imposes the assumption of linear effects and does not consider any non-linear trends that may exist in the data. In order to investigate these effects in more detail, a second PLSR model was constructed with 27 dummy variables, one for each subgroup of the three covariates age, period and cohort. Up to 23 components could have been extracted in the model. Table 3 shows the increase in percentage variance explained as the number

of components extracted increases. Very little additional explanatory power is achieved by extracting more than four components; hence we ran our non-linear analysis with only four components. This resulted in an R^2 of 99% for the outcome variable. PLSR coefficients and confidence intervals are given in Table 4.

(Table 3 here)

(Table 4 here)

Figures 3A, 3B and 3C show the PLS coefficients for Age, Period and Cohort with 95% confidence intervals estimated by the jacknife. A loess curve was fitted over the PLS coefficient estimates to explore potential non linear effects.

(Figure 3A here)

(Figure 3B here)

(Figure 3C here)

This non-linear analysis suggests a mostly increasing effect on incidence with age, as shown in Figure 3A and Table 4. This reaches a peak in those aged 75-79 (0.60, 95% CI 0.32 - 0.87), followed by a decrease in incidence for the oldest age groups. The shape of this graph is similar to the crude age effects on incidence shown in Fig 1A, suggesting that period and cohort effects have little impact on the crude age effect.

Figure 3B shows no difference between the first three time periods, and an increase in incidence for the most recently diagnosed group, though confidence intervals are wide and have considerable overlap with previous periods. Visual comparison of Fig 1B

and 3B suggests that there is little conflation with age and cohort effects within the crude period effects.

The non-linear effects of birth cohort on incidence are shown in Figure 3C. Data for the most recent birth cohorts are incomplete as this portion of the population has not yet reached the peak age of onset for ALS, thus incidence in these groups will rise in the coming years. The youngest cohort 1967-1971 was removed from the plot as the confidence interval was very wide due to sparse data. The model results in Fig 3C show less variability across cohorts than the crude birth cohort effects in Fig 1C suggesting that the crude effects in Fig 1C were age and period effects rather than cohort effects.

DISCUSSION:

A cohort effect is usually attributed to variations in early life environment. Exposure to certain unfavourable environmental factors in early life can have adverse effects in later life. Our model shows that there is no substantial cohort effect after we removed the interacting effects of age and period. Incidence measures for the youngest cohorts must be observed with caution, since this rate will increase in the coming years as these individuals age and their risk of developing ALS increases. This is likely to leave a flat graph of cohort effects over time. However, due to the smaller numbers of cases, cohort effects for the earliest and latest birth cohorts are subject to greater uncertainty, and hence it may be more difficult to detect a true cohort effect.

No substansive association was found between period of diagnosis and incidence in either the linear or non-linear models. ALS incidence is stable from 1996-2010 and appears to rise in the final diagnosis period 2010-2014, though the error bounds are

wide and the estimates are not greatly different from the other diagnosis periods. This apparent increase could signify an increase in the overall incidence of ALS in the population. This seems plausible when we consider the data from Figure 2, which shows that the incidence since 2010 is slightly higher overall when compared to previous periods. One possible explanation for this phenomenon is a change in competing risks of death. Improvements in treatment and changes in lifestyle has improved life expectancy for many life limiting conditions. As a result we may see an increased incidence of ALS in older people who, in the past, might have died from another cause. The data from Figure 2 would support this hypothesis, as we see an overall rise in incidence for those aged 80-89. Another potential explanation for this effect could be subtle changes in diagnostic criteria, with better ascertainment of cases not hitherto recognized (e.g. those presenting with frontotemporal dementia. These points serve as explanations for a potential period effect on incidence of ALS. Over time it will become clear if these effects emerge as strong trends in the incidence of ALS.

Two previously published APC models for ALS incidence and mortality reported increased risk of death from ALS for those born between 1880 -1920 [11, 12], while a third study concluded that there was no evidence of a birth cohort effect in ALS in Switzerland [10]. Our model shows no reportable period or cohort effects on incidence of ALS for those diagnosed in Ireland between 1996-2013. Our model is the first to use a population based register for ALS. In addition, we are the first group to successfully separate age, period and cohort effects on ALS incidence without imposing additional limiting constraints. The partial least squares regression method preserves the mathematical relationship between the age, period and cohort variables while circumventing the identification problem. Replication of similar analyses in

other populations would be beneficial to assess whether these patterns of incidence are observable elsewhere.

CONCLUSION:

Although the major cause(s) of ALS remain unidentified, epidemiological analyses may provide further insight and guide future investigations.

Results of this Age-Period-Cohort analysis suggest the initial signs of increasing incidence of ALS in recent years. As further years of data accrue to the Irish register it will become clear if this effect emerges as a strong trend in the incidence of ALS in Ireland and replication of these analyses in other populations will show if our findings on temporal patterns in ALS incidence are shared elsewhere.

Ethical approval for this study was granted by Beaumont Hospital Ethics (Medical Research) Committee

Funding: This work was funded by a Health Research Board Interdisciplinary Capacity Enhancement award, grant number ICE/2012/6.

Disclosures:

Prof. Orla Hardiman has received speaking honoraria from Janssen Cilag, Biogen Idec, Sanofi Aventis, Novartis and Merck-Serono. She has been a member of advisory panels for Biogen Idec, Allergen, Ono Pharmaceuticals, Novartis, Cytokinetics and Sanofi Aventis. She serves as Editor-in-Chief of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia .

Funding is from Health Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 259867, ALSA (the ALS Association), HRB (the Health Research Board, grant H01300), Joint Programme in Neurodegeneration (JPND) , and Research Motor Neuron (previously named Motor Neuron Disease Research Foundation).

Both Dr. Katy Tobin and Dr. James Rooney were funded under separate grants from the Health Research Board during the conduct of the study.

All other authors have no conflicts of interest to declare.

References:

[1] Logroscino G, Traynor BJ, Hardiman O et al. Incidence of Amyotrophic Lateral Sclerosis in Europe. *Journal of Neurology, Neurosurgery & Psychiatry*. 2010;81(4):385–390.

[2] O'Toole O, Traynor BJ, Brennan P et al. Epidemiology and clinical features of amyotrophic lateral sclerosis in Ireland between 1995 and 2004. *Journal of Neurology, Neurosurgery & Psychiatry* 2008;79:30–32.

[3] Murphy M, Quinn S, Young J, Parkin P, Taylor B. Increasing incidence of ALS in Canterbury, New Zealand: a 22-year study. *Neurology*. 2008;71(23),1889–95.

[4] Georgoulopoulou E, Vinceti M, Bonvicini F et al. Changing incidence and subtypes of ALS in Modena, Italy: A 10-years prospective study. *Amyotrophic Lateral Sclerosis*. 2011;12(6),451–7.

[5] Chiò A, Cucatto A, Calvo A, Terreni AA, Magnani C, Schiffer D. Amyotrophic lateral sclerosis among the migrant population to Piemonte, northwestern Italy. *Journal of Neurology*. 1999;246(3),175–80.

[6] Chiò A, Logroscino G, Traynor BJ et al. Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature. *Neuroepidemiology*. 2013;41(2),118–30.

[7] European Commission. Eurostat - Statistics Explained. Population Structure and Ageing [Internet]. 2015 [cited 2015 Nov 20] Available from: http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing

[8] Ford ES, Roger VL, Dunlay SM, Go AS, Rosamond WD. Challenges of ascertaining national trends in the incidence of coronary heart disease in the United States. *Journal of the American Heart Association*. 2014 Dec;3(6):e001097.

[9] Bennett K, Hughes J, Jennings S, Kee F, Shelley E. Comparing the decline in coronary heart disease and stroke mortality in neighbouring countries with different healthcare systems. *Heart.* 2013 Aug 15;99(16):1179–84.

[10] Ajdacic-Gross V, Schmid M, Tschopp A, Gutzwiller F. Birth Cohort Effects in Neurological Diseases: Amyotrophic Lateral Sclerosis, Parkinson's Disease and Multiple Sclerosis. *Neuroepidemiology*. 2012;38:56–63.

[11] Gordon PH, Artaud F, Aouba A, Laurent F, Meininger V, Elbaz A. Changing mortality for motor neuron disease in France (1968-2007): an age-period-cohort analysis. *European Journal of Epidemiology*. 2011;*26*(9):729–37.

[12] Seals RM, Hansen J, Gredal O, Weisskopf MG. Age-period-cohort analysis of trends in amyotrophic lateral sclerosis in Denmark, 1970-2009. *American Journal of Epidemiology*. 2013;178(8):1265–71.

[13] Carstensen B. Age-period-cohort models for the Lexis diagram. *Statistics in Medicine*. 2007;26(15):3018–3045.

[14] McNally RJ, Alexander FE, Staines A, Cartwright RA. A comparison of three methods of analysis for age-period-cohort models with application to incidence data on non-Hodgkin's lymphoma. *International Journal of Epidemiology*. 1997;26(1):32–46.

[15] Tu, YK, Davey Smith G, Gilthorpe MS. A new approach to age-period-cohort analysis using partial least squares regression: the trend in blood pressure in the Glasgow Alumni cohort. *PloS One*. 2011;6(4), e19401.

[16] Jiang T, Gilthorpe MS, Shiely F et al. Age-period-cohort analysis for trends in body mass index in Ireland. *BMC Public Health*, 2013;13(1): 889.

[17] Tu YK, Keyes K, Davey Smith G. Mortality cohort effects from mid 19th to mid 20th century Britain: did they exist? *Annals of Epidemiology* 2014;24(8):570-4

[18] Rooney J, Vajda A, Heverin M et al. Spatial cluster analysis of population amyotrophic lateral sclerosis risk in Ireland. *Neurology*. 2015; 84(15):1537-44

[19] Central Statistics Office. CSO Statistical Databases [Internet]. Unknown [cited 2015 May 15] Available from:http://www.cso.ie/en/databases/

[20] Tu YK, Kramer N, Lee WC. Addressing the Identification Problem in Age-period-cohort Analysis. A Tutorial on the Use of Partial Least Squares and Principal Components Analysis. *Epidemiology*. 2012;23:583–593

[21] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

[22] Mevik BH, Wehrens R. The pls package: principal component and partial least squares regression in R. *Journal of Statistical Software*. 2007;18:1-24.

[23] Haenlein M, Kaplan AM. A Beginner's Guide to Partial Least Squares Analysis. *Understanding Statistics*. 2004;3(4):283–297.

Period	**1996-2000 Incidence (n)**	**2001-2005 Incidence (n)**	**2006-2010 Incidence (n)**	**2011-2014 Incidence (n)**

Age				
40-44	1.19 (15)	0.86 (12)	0.90 (14)	0.66 (9)
45-49	2.50 (29)	1.42 (18)	0.97 (14)	1.61 (20)
50-54	3.19 (33)	3.57 (42)	2.64 (34)	2.58 (29)
55-59	5.52 (45)	4.98 (51)	5.26 (61)	5.11 (51)
60-64	7.96 (57)	7.09 (57)	7.76 (77)	9.47 (84)
65-69	8.91 (57)	11.66 (79)	9.93 (75)	12.17 (90)
70-74	12.69 (71)	12.96 (74)	11.62 (71)	12.49 (68)
75-79	8.04 (35)	11.79 (53)	14.76 (70)	14.79 (62)
80-84	9.40 (26)	8.19 (25)	10.87 (36)	13.88 (40)
85-89	8.05 (11)	7.36 (11)	6.10 (11)	10.41 (17)

Table 1. ALS incidence rates per 100,000 in Ireland between 1996 and 2014 for those aged 40-89 at diagnosis.

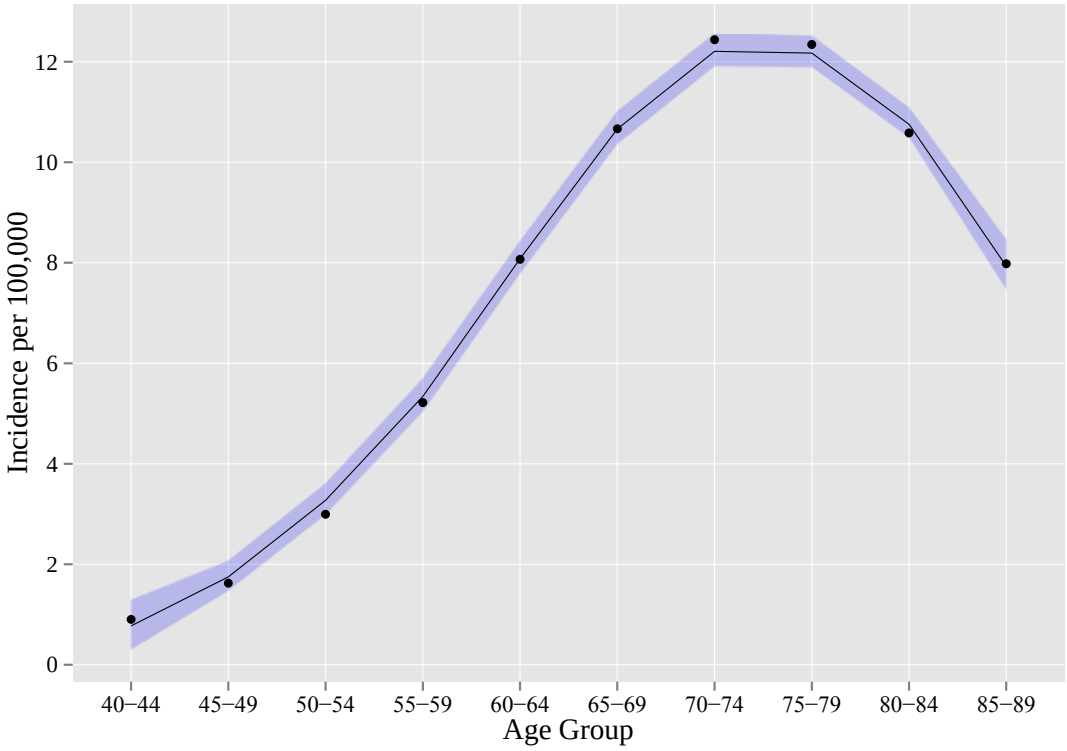

Fig 1A: Mean incidence rate by age of diagnosis (with loess smoother of α = 0.8 applied) and corresponding 95% confidence interval.

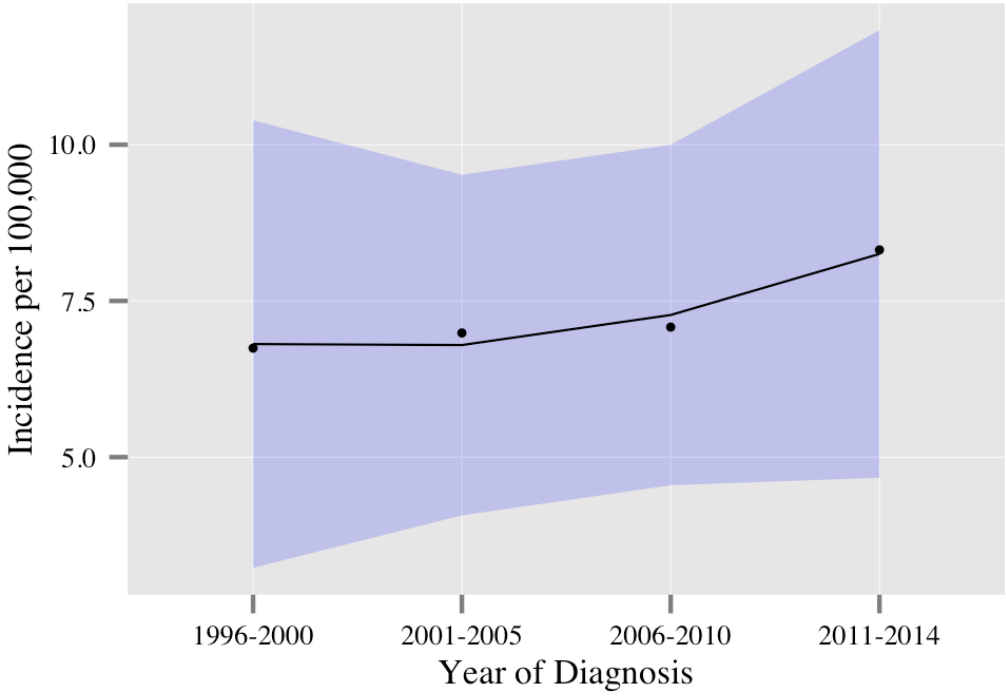

Fig 1B: Mean incidence rate by year of diagnosis (with second degree polynomial fitted) and 95% confidence interval.

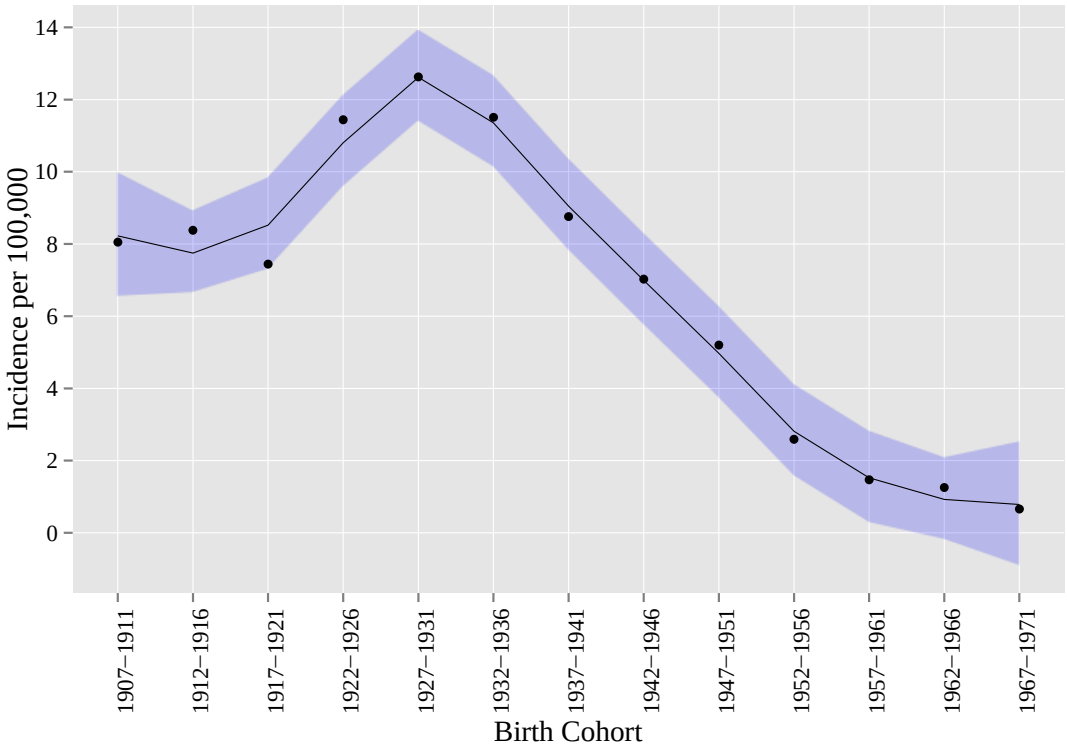

Fig 1C: Mean incidence rate by year of birth (with loess smoother of $\alpha = 0.5$ applied) and 95% confidence interval.

Figure 1: Average incidence per 100,000 for the age at diagnosis of ALS (Fig 1A), the year of diagnosis (Fig 1B) and the birth cohort (Fig 1C).

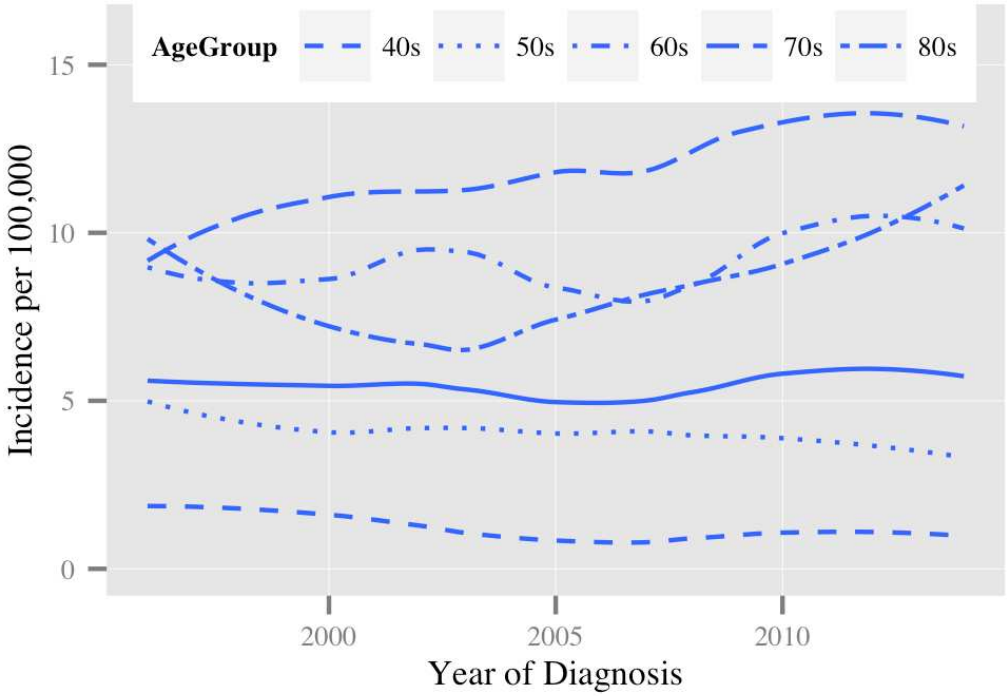

Figure 2: Smoothed loess curve (α = 0.7) showing trends in incidence rate by year of diagnosis for each age group. The solid line represents the overall mean incidence rate for all ages for each diagnosis year.

	Regression Coefficient	**Std. Error**	**95% Confidence Interval**
Age (per 5 years)	0.13	0.02	0.09, 0.16
Period (per 5 years)[a]	0.00	0.01	-0.02, 0.02
Cohort (per 5 years)	0.13	0.02	0.09, 0.17

Table 2: Results of the PLSR linear one-component model of the effects of age, period and cohort on ALS incidence. Positive regression coefficients indicate an increased effect on ALS incidence for the covariate. ([a]The final time period has 4 years. 95% confidence intervals were obtained using the jack-knifing method).

Number of components extracted	R^2 value
1	94.354
2	97.23
3	98.27
4	99.17
5	99.40
6	99.43
23	99.43

Table 3: Percentage variance explained (R^2) according to number of components extracted by the PLSR model.

	Regression Coefficient	Std. Error	95% Confidence Interval
Age Group:			
40-44	-1.24	0.22	(-1.69, -0.80)
45-49	-0.94	0.19	(-1.32, -0.55)
50-54	-0.42	0.17	(-0.76, -0.08)
55-59	-0.06	0.13	(-0.32, 0.20)
60-64	0.26	0.12	(0.01, 0.51)
65-69	0.50	0.17	(0.15, 0.85)
70-74	0.59	0.10	(0.38, 0.80)

75-79	0.60	0.14	(0.32, 0.87)
80-84	0.47	0.12	(0.23, 0.71)
85-89	0.24	0.15	(-0.06, 0.54)
Year Diagnosed:			
1996-2000	-0.10	0.10	(-0.29, 0.10)
2001-2005	-0.07	0.10	(-0.27, 0.13)
2006-2010	-0.04	0.10	(-0.25, 0.17)
2011-2013	0.20	0.11	(-0.02, 0.42)
Birth Cohort:			
1907-1911	0.26	0.27	(-0.29, 0.82)
1912-1916	0.27	0.14	(-0.03, 0.56)
1917-1921	-0.00	0.14	(-0.27, 0.27)
1922-1926	0.36	0.12	(0.12, 0.59)
1927-1931	0.39	0.16	(0.06, 0.72)
1932-1936	0.34	0.12	(0.10, 0.57)
1937-1941	0.15	0.11	(-0.08, 0.38)
1942-1946	0.07	0.17	(-0.28, 0.42)
1947-1951	0.16	0.16	(-0.17, 0.49)
1952-1956	-0.11	0.26	(-0.64, 0.41)
1957-1961	-0.55	0.26	(-1.07, -0.03)
1962-1966	-0.46	0.28	(-1.02, 0.09)
1967-1971	-0.87	0.86	(-2.60, 0.86)

Table 4: Results of the four-component PLSR model with 27 dummy variables for each age, period and cohort subgroup. Positive regression coefficients indicate an increased effect on ALS incidence for the subgroup (95% confidence intervals were obtained using the jack-knifing method).

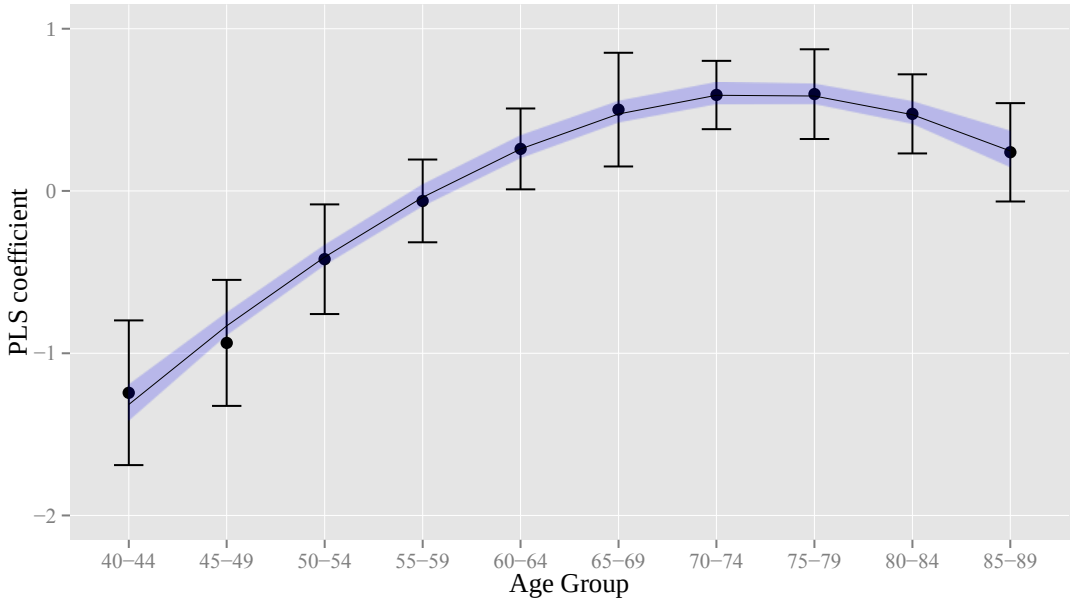

Figure 3A: PLS coefficients by age at diagnosis (with a loess curve of α = 1 applied) and a 95% confidence interval.

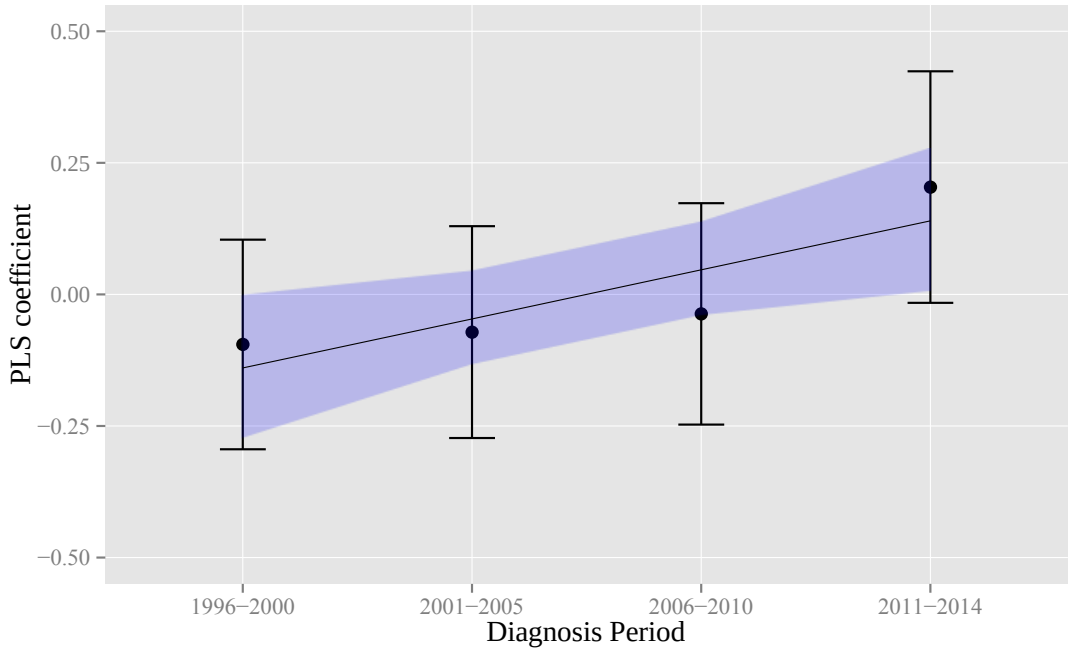

Figure 3B: PLS coefficients by year of diagnosis (with a loess curve of α = 1 applied) and a 95% confidence interval.

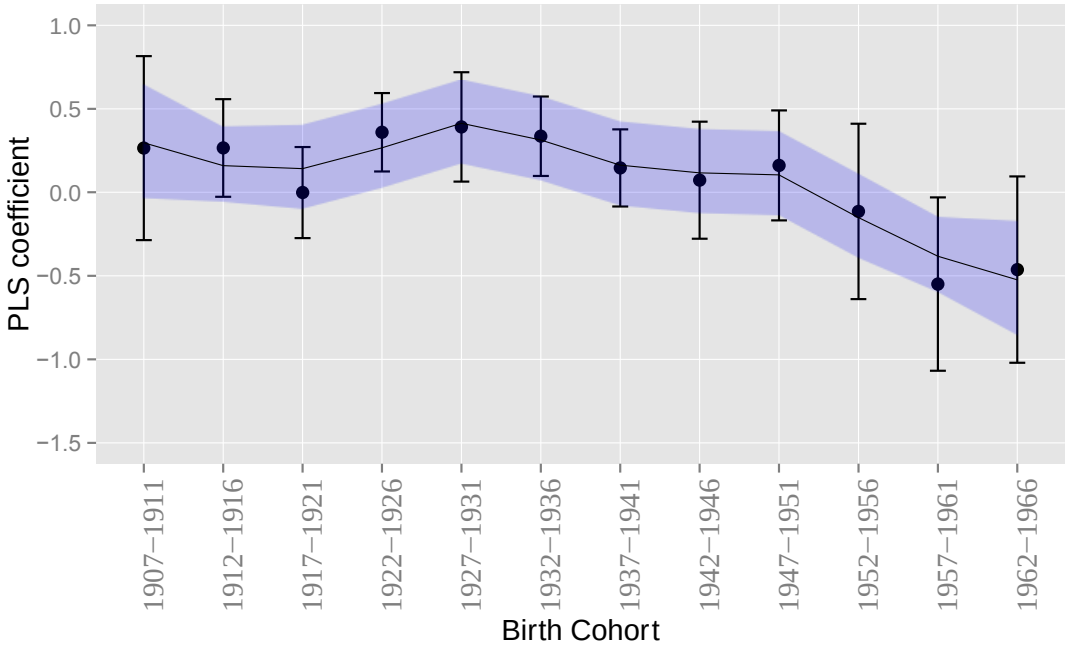

Figure 3C: PLS coefficients by year of birth (with a loess curve of $\alpha = 0.5$ applied) and a 95% confidence interval.

Appendix: Design Matrix for non linear PLSR

age	period	cohort	age1	age2	age3	age4	age5	age6	age7	age8	age9	age10	period1	period2	period3	period4
1	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0
1	2	2	1	0	0	0	0	0	0	0	0	0	0	1	0	0
1	3	3	1	0	0	0	0	0	0	0	0	0	0	0	1	0
1	4	4	1	0	0	0	0	0	0	0	0	0	0	0	0	1
2	1	2	0	1	0	0	0	0	0	0	0	0	1	0	0	0
2	2	3	0	1	0	0	0	0	0	0	0	0	0	1	0	0
2	3	4	0	1	0	0	0	0	0	0	0	0	0	0	1	0
2	4	5	0	1	0	0	0	0	0	0	0	0	0	0	0	1
3	1	3	0	0	1	0	0	0	0	0	0	0	1	0	0	0
3	2	4	0	0	1	0	0	0	0	0	0	0	0	1	0	0
3	3	5	0	0	1	0	0	0	0	0	0	0	0	0	1	0
3	4	6	0	0	1	0	0	0	0	0	0	0	0	0	0	1
4	1	4	0	0	0	1	0	0	0	0	0	0	1	0	0	0
4	2	5	0	0	0	1	0	0	0	0	0	0	0	1	0	0
4	3	6	0	0	0	1	0	0	0	0	0	0	0	0	1	0
4	4	7	0	0	0	1	0	0	0	0	0	0	0	0	0	1
5	1	5	0	0	0	0	1	0	0	0	0	0	1	0	0	0
5	2	6	0	0	0	0	1	0	0	0	0	0	0	1	0	0
5	3	7	0	0	0	0	1	0	0	0	0	0	0	0	1	0
5	4	8	0	0	0	0	1	0	0	0	0	0	0	0	0	1
6	1	6	0	0	0	0	0	1	0	0	0	0	1	0	0	0
6	2	7	0	0	0	0	0	1	0	0	0	0	0	1	0	0
6	3	8	0	0	0	0	0	1	0	0	0	0	0	0	1	0
6	4	9	0	0	0	0	0	1	0	0	0	0	0	0	0	1
7	1	7	0	0	0	0	0	0	1	0	0	0	1	0	0	0

7	2	8	0	0	0	0	0	0	1	0	0	0	0	1	0	0
7	3	9	0	0	0	0	0	0	1	0	0	0	0	0	1	0
7	4	10	0	0	0	0	0	0	1	0	0	0	0	0	0	1
8	1	8	0	0	0	0	0	0	0	1	0	0	1	0	0	0
8	2	9	0	0	0	0	0	0	0	1	0	0	0	1	0	0
8	3	10	0	0	0	0	0	0	0	1	0	0	0	0	1	0
8	4	11	0	0	0	0	0	0	0	1	0	0	0	0	0	1
9	1	9	0	0	0	0	0	0	0	0	1	0	1	0	0	0
9	2	10	0	0	0	0	0	0	0	0	1	0	0	1	0	0
9	3	11	0	0	0	0	0	0	0	0	1	0	0	0	1	0
9	4	12	0	0	0	0	0	0	0	0	1	0	0	0	0	1
10	1	10	0	0	0	0	0	0	0	0	0	1	1	0	0	0
10	2	11	0	0	0	0	0	0	0	0	0	1	0	1	0	0
10	3	12	0	0	0	0	0	0	0	0	0	1	0	0	1	0
10	4	13	0	0	0	0	0	0	0	0	0	1	0	0	0	1

cohort1	cohort2	cohort3	cohort4	cohort5	cohort6	cohort7	cohort8	cohort9	cohort10	cohort11	cohort12	cohort13
1	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0

0	0	0	1	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0

0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	1

Appendix: Design matrix with 10 age groups, 4 time periods and 13 birth cohorts.