Delirium and the risk of developing dementia: a cohort study of 12 949 patients

Samuel P Leighton 1, James W Herron1, Eric Jackson2, Matthew Sheridan2, Fani Deligianni3, Jonathan Cavanagh4

ABSTRACT

Background Delirium is an important risk factor for subsequent dementia. However, the field lacks large studies with long-term follow-up of delirium in subjects initially free of dementia to clearly establish clinical trajectories.

Methods We undertook a retrospective cohort study of all patients over the age of 65 diagnosed with an episode of delirium who were initially dementia free at onset of delirium within National Health Service Greater Glasgow & Clyde between 1996 and 2020 using the Safe Haven database. We estimated the cumulative incidence of dementia accounting for the competing risk of death without a dementia diagnosis. We modelled the effects of age at delirium diagnosis, sex and socioeconomic deprivation on the cause-specific hazard of dementia via cox regression.

Results 12 949 patients with an incident episode of delirium were included and followed up for an average of 741 days. The estimated cumulative incidence of dementia was 31% by 5 years. The estimated cumulative incidence of the competing risk of death without dementia was 49.2% by 5 years. The cause-specific hazard of dementia was increased with higher levels of deprivation and also with advancing age from 65, plateauing and decreasing from age 90. There did not appear to be a relationship with sex.

Conclusions Our study reinforces the link between delirium and future dementia in a large cohort of patients. It highlights the importance of early recognition of delirium and prevention where possible.

INTRODUCTION

Delirium and dementia are two of the most common causes of cognitive impairment in the elderly population, but their interrelationship is poorly understood.1 Dementia is characterised by an irreversible progressive global cognitive decline. It is associated with huge financial and wider societal costs. In the UK, the annual cost of dementia is £35 billion, two-thirds of which is borne by people with dementia and their families.2 Delirium is characterised by an acute and fluctuating disturbance in attention and awareness with associated disturbance in cognition (eg, memory deficit, disorientation, language, visuospatial ability or perception), which cannot be explained by another neurocognitive disorder and does not occur in the context of a severely reduced level of arousal, such as coma. It is a serious and life-threatening neuropsychiatric syndrome, which is a direct physiological consequence of another medical condition, substance intoxication or withdrawal, toxins or multiple aetiologies.3 Delirium is very common in the elderly and present in up to 50% of patients over the age of 65 admitted to hospital.4 Delirium is a clinical diagnosis, which is often under-recognised and frequently overlooked. This has led to a number of high-profile campaigns to increase the awareness and recognition of delirium across the UK and the wider world.5

Dementia is the primary risk factor for delirium and delirium is a major risk factor for subsequent dementia.1 6 It is not yet clear if delirium is a simply a marker of brain vulnerability, whether the impact of delirium on dementia is derived from its precipitating cause or whether delirium itself leads to permanent neuronal damage. Delirium is preventable in 30%-40% of cases and is, therefore, an important modifiable risk factor for dementia.4 Several clinical studies provide evidence to support the relationship between delirium and dementia. A 2010 meta-analysis of two studies (n=241) found that delirium was associated with an increased risk of dementia (RR 5.7, 95% CI 1.3 to 24.0).7 A 2021 meta-analysis of six studies (n=901) showed that delirium was associated with increased odds of developing new dementia compared with patients without delirium (OR 11.9, 95% CI 7.29 to 19.6).8 The relationship has also been explored in a small population-based cohort study of 533 individuals aged 85+, which found an increased risk of incident dementia following episode of delirium (OR 8.7, 95% CI 2.1 to 35).9

However, to date, the field lacks large studies with long-term follow-up of delirium in subjects initially free of dementia to clearly establish outcomes.10 11 Our study has two objectives:

1. To estimate the cumulative incidence of dementia among those who experience an episode of delirium but who have not yet been diagnosed with dementia prior to that episode.
2. To model the effect of age at delirium diagnosis, sex and socioeconomic deprivation on the rate of occurrence of dementia among those still at risk (ie, the cause-specific hazard of dementia).

METHODS

We adhere to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) and the Reporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statements.10 11

We undertook a retrospective cohort study of patients over the age of 65 who had been diagnosed with an index episode of delirium but who had...
not been diagnosed with dementia prior to their index episode of delirium. Patients from the National Health Service (NHS) Greater Glasgow & Clyde (GG&C) health board were included. Patients with a diagnosis of delirium made before 1 May 2020 were included back as far as the records allowed. The earliest delirium diagnosis was 21 April 1996. Patients were followed from their first episode of delirium up until 1 October 2020 when the data were collected. The primary outcome event of interest was diagnosis of dementia. A competing event, death before dementia diagnosis, was observed. Patients who had not experienced either event before the end of the follow-up period were censored as censored. Patients who experienced their outcome event on the same day as their index delirium diagnosis were considered to have survived 0.5 days.

West of Scotland Safe Haven at NHS GG&C created the study population from the database population. The diagnoses of delirium and dementia were clinical diagnoses based on the International Classification of Diseases 10th Revision made by the treating clinician (see online supplemental materials). Diagnoses could have been made in accident & emergency (A&E), as an inpatient or outpatient or on death. Age at delirium diagnosis, sex and Scottish Index of Multiple Deprivation (SIMD) 2009 quintile (lowest equals most deprived) were included as covariates. SIMD 2009 was based on their most recent postal address. All subjects had information about covariates—there were no missing data. The total number of relevant delirium patients in the NHS GG&C Safe Haven database determined the sample size.

As outlined above, competing risks are present as a participant is at risk of two mutually exclusive events. Using the Kaplan-Meier estimate of the survival function to estimate the incidence function in the presence of competing risks generally results in upward biases in the estimation of the incidence function. Instead, we used the cumulative incidence function (CIF), which allows for the estimation of the incidence of the occurrence of an event (dementia) while taking competing risk (death without a dementia diagnosis) into account. The CIF for the kth cause is defined as: \( \text{CIF}(t) = \Pr(T \leq t, D = k) \), where \( D \) denotes the type of event that occurred, and \( T \) denotes the time from baseline time until the occurrence of the event. The function \( \text{CIF}(t) \) denotes the probability of experiencing the \( k \)th event before time \( t \) and before the occurrence of a different type of event.\(^{12}\)

We also modelled the effect of covariates (age at incident delirium, sex and deprivation quintile) on the cause-specific hazard function. The cause-specific hazard function is the instantaneous rate of occurrence of the primary event (dementia) in subjects who have not yet experienced either event (dementia or death without dementia). The exponentiated regression coefficient from the cause-specific hazard model represents the amount of relative change in the cause-specific hazard function associated with a 1-unit change in the covariate. The cause-specific hazard model is well suited to studying the aetiology of a disease.\(^3\) We fit the cause-specific hazard model by estimating a Cox proportional hazards model and treating subjects who experience a competing event as being censored at the time of occurrence of the competing event. Postmodel assumption testing included testing the proportional hazard’s assumption via Schoenfeld residuals, using the difference in beta values (DFBETAS) to check for influential observations and assessing the functional form of covariates via Martingale residuals. Age had a non-linear functional form, so the final Cox model was refitted using a penalised cubic spline term for age. The results of our postmodel assumption testing are available in online supplemental materials.

All analyses were performed using R, CRAN V.4.0.0\(^{14}\) (with the ‘survival’,\(^15\) \( \text{cmprsk} \)^\(17\) and ‘survminer’\(^18\) packages) and code is available in online supplemental materials.

RESULTS

12 949 patients with a relevant index episode of delirium followed up for an average of 741 days (minimum=0.5 days, maximum=8855 days) were included in the study. 3530 (27%) of these patients had a subsequent diagnosis of dementia and 5788 (45%) died without a diagnosis of dementia, leaving 3631 (28%) who were coded as censored by the study end date. The diagnosis of dementia was made on death in 643 (18%) of patient who were diagnosed with dementia. This information is summarised in figure 1.

The diagnosis of new index episodes of delirium increased in frequency over time with some seasonal variation as per figure 2.

Descriptive statistics for the patients in the study are reported in table 1.

The estimated cumulative incidences of dementia and for the competing risk of death without a dementia diagnosis are presented in figure 3. The estimated cumulative incidence of dementia, accounting for the competing risk of death without a dementia diagnosis, was 9.0% (95% CI 8.5% to 9.5%) by 6 months, 13.6% (95% CI 13.0% to 14.2%) by a year, 31.0% (95% CI 30.1% to 31.9%) by 5 years, 35.5% (95% CI 34.5% to 36.5%) by 10 years, and 36.3% (95% CI 35.2% to 37.3%) by 20 years. The estimated cumulative incidence of the competing risk of death without a dementia diagnosis was 20.0% (95% CI 19.3% to 20.7%) by 6 months, 27.1% (95% CI 26.3% to 27.9%) by a year, 49.2% (95% CI 48.2% to 50.2%) by 5 years, 55.3% (95% CI 54.3% to 56.4%) by 10 years and 57.4% (95% CI 56.2% to 58.5%) by 20 years.

The multivariable adjusted cause-specific HRs for sex and SIMD 2009 deprivation quintile are illustrated in figure 4A. The multivariable-adjusted cause-specific HR for age at delirium diagnosis is illustrated in figure 4B.

DISCUSSION

To the best of our knowledge, this study represents the largest cohort (\( n = 12\ 949 \)) followed up for the longest period of time (up to 8855 days; mean 741 days) within the published...
literature examining the new diagnosis of dementia following an episode of delirium. The results show that a first episode of delirium after the age of 65 is associated with a substantial risk of subsequently developing dementia (31% by 5 years). This is in line with data from smaller previously published studies.\textsuperscript{7,9-19} Our data also show that delirium is associated with substantial mortality, in addition to the risk of dementia. This underlines the seriousness of delirium and the importance of prompt diagnosis and treatment of underlying cause. Our research supports the concept of delirium as both an indicator of physiological frailty as well as a possible precipitating and accelerating factor in cognitive and physical decline. Within NHS GG&C, there has been a trend of increases in diagnosis of delirium over time. This may indicate that recent high-profile delirium recognition campaigns are having the desired impact including the Think Delirium campaign, which was introduced in NHS GG&C in 2015.\textsuperscript{5,20} Findings from the Cox-regression analysis show that the multivariable-adjusted cause-specific hazard of dementia among those diagnosed with delirium increases with higher levels of deprivation and also with advancing age, plateauing and decreasing in extremes of age. However, there does not appear to be a relationship with sex.

The most frequent causes of delirium involve significant systemic inflammation. Inflammation is well recognised as a major precipitant of delirium.\textsuperscript{21} There exists an extensive network of mechanisms that allow neuroimmune communication.\textsuperscript{22} In recent years, the effects of inflammatory insult on central nervous structure and function have become increasingly well characterised.\textsuperscript{21} Dementia is a disorder which, except in rare single-gene inherited syndromes, has a complex aetiology involving multiple contributory interacting factors. These include ageing, obesity, diabetes, hypertension and smoking—the common strand to these risk factors is the systemic preponderance of inflammatory molecules.\textsuperscript{23} Inflammation is thought to have a central mechanistic role in the pathogenesis of both Alzheimer’s dementia\textsuperscript{24} and vascular dementia,\textsuperscript{25} the two most common subtypes. While acute inflammation is protective to the brain under most circumstances, prolonged or excess release of proinflammatory molecules within the vulnerable or aged brain may activate various downstream cellular cascades relevant to the emergence of dementia.\textsuperscript{26}

These phenomena may be relevant in the context of our findings that support the link between hospitalisation with delirium and subsequent dementia diagnosis. It remains a matter of discussion whether delirium is purely a marker of susceptibility to developing dementia, or unmasks/accelerates unrecognised dementia, or indeed, whether delirium may have direct neurotoxic effects that can be causal in the pathogenesis of dementia.\textsuperscript{1} Evidence from the Vantaa 85+ population-based study may provide evidence to support the latter hypothesis. Neuropathological correlates of dementia such as neurofibrillary tau, β-amyloid plaque burden, vascular lesions, Lewy-body pathology and ApoE4 allele status were not found to be positively associated with subjects who developed dementia following delirium, while in contrast, a strong association existed in those that developed dementia without a delirium history.\textsuperscript{9} Although the Vantaa study

![Figure 2](http://jnnp.bmj.com/) The monthly frequency of new index delirium diagnoses in patients who had not been diagnosed with dementia prior to this episode of delirium.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total sample (n=12,949)</th>
<th>Dementia diagnosis (n=3,530)</th>
<th>Death without a dementia diagnosis (n=5,788)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at index episode of delirium Mean (SD)</td>
<td>82.3 (7.8)</td>
<td>83.2 (7.0)</td>
<td>82.7 (8.1)</td>
</tr>
<tr>
<td>Male sex number (%)</td>
<td>5,036 (39%)</td>
<td>1,262 (36%)</td>
<td>2,458 (42%)</td>
</tr>
<tr>
<td>SIMD 2009 quintile number (%)</td>
<td>First—4,976 (38%)</td>
<td>First—1,359 (38%)</td>
<td>First—2,247 (39%)</td>
</tr>
<tr>
<td></td>
<td>Second—2,341 (18%)</td>
<td>Second—595 (17%)</td>
<td>Second—1,050 (18%)</td>
</tr>
<tr>
<td></td>
<td>Third—1,986 (15%)</td>
<td>Third—620 (18%)</td>
<td>Third—885 (15%)</td>
</tr>
<tr>
<td></td>
<td>Fourth—1,673 (13%)</td>
<td>Fourth—515 (15%)</td>
<td>Fourth—672 (12%)</td>
</tr>
<tr>
<td></td>
<td>Fifth—1,973 (15%)</td>
<td>Fifth—441 (12%)</td>
<td>Fifth—934 (16%)</td>
</tr>
</tbody>
</table>

SIMD, Scottish Index of Multiple Deprivation.
Cognition was not powered to be conclusive, it may suggest that, in some cases, dementia following delirium represents a different aetiologic pathway to the development of dementia, rather than being purely a vulnerability marker/accelerator of pre-existing disease.

In our multivariable analysis, the cause-specific hazard of dementia increases with age of delirium diagnosis from age 65 until around age 90, when it plateaus then decreases. This is consistent with existing evidence in the general population demonstrating a doubling of both the prevalence and incidence of dementia every 5–6 years until the age of 90.27 Evidence for trends in dementia diagnosis among the oldest old is limited by sample size. However, two large population-based cohort studies found the increases in the incidence of dementia plateau or even decline beyond age 90. It is suggested that among the oldest old, risk factors for dementia may not be related to the ageing process itself but with age-associated risk factors such as hypertension, hyperlipidaemia and heart disease.28–30

We found that living in an area of deprivation is associated with an increased cause-specific hazard of developing dementia following an incident episode of delirium after adjusting for age at delirium diagnosis and sex. This supports an earlier finding that the hazard of dementia is increased among those living in areas with higher levels of deprivation in an English population cohort study of 6220 adults over the age of 65.31 Unfortunately, we did not have information available to adjust for personal indicators of socioeconomic status like personal wealth, educational attainment or occupation, so we are not able to clearly determine whether individual factors were driving this area deprivation effect. However, previous research has shown that living in an area of higher deprivation is associated with poorer health outcomes even after adjusting for personal wealth, education and employment.32

Our study has a number of strengths including the large sample size and long length of follow-up. Furthermore, by virtue of being registry based, our study is pragmatic and the setting is entirely naturalistic. We have properly accounted for the impact of competing risks by using the CIF rather than Kaplan-Meier estimator and we have modelled the effect of covariates on the cause-specific hazard of dementia in those who experience an episode of delirium. We adhere to gold-standard STROBE reporting guidelines.

There were several limitations. The cohort largely consisted of patients diagnosed within secondary care. Only diagnoses made at death were included from primary care. This introduced a selection bias for more severe cases of delirium requiring assessment at A&E, on admission to hospital or on death. Equally, it is possible that our cases could have had earlier incidences of delirium, perhaps within primary care, which were milder and not coded and indexed to our data set. Similarly, the majority of dementia diagnoses were made within secondary care. If patients moved out of NHS GG&C after their index delirium diagnosis but before their outcome occurred, their outcome would not be known except if it was made at death. As such, it is possible that the proportion of patients who developed dementia was underestimated due to attrition bias (patients were censored when they should not have been). Furthermore, in those whose dementia diagnosis was made on death, it is possible that this dementia diagnosis was made in primary care at an earlier time point patient and, thus, dementia survival was overestimated. In addition, our cohort is drawn from all medical records over a specific timeframe rather than being set up as a prospective cohort study. We rely on clinicians accurately and reliably coding the diagnosis of delirium at the point of clinical care being administered rather than trained research assistants. While we believe the system of diagnostic coding to be robust within NHS GG&C, it is likely that some cases of dementia or delirium may be missed or inaccurately diagnosed or coded. For example, there is a clear trend of increasing diagnosis of delirium over time within NHS GG&C. This is unlikely to represent a true increase in the underlying

Figure 3 Cumulative incidence function for dementia (blue) and for death without dementia (red) in patients with an index episode of delirium by time in years with 95% CIs.
Cognition

rates of delirium but rather represent an increase in the recognition and coding of delirium, perhaps driven by a number of high-profile delirium recognition campaigns, leading to a general increase in awareness of the condition. Moreover, our multivariable model lacks several important covariates like medical comorbidities, lifestyle factors like diet and smoking or genetics which have been clearly identified as important risk factors for dementia. Finally, when we designed our study, we set it up as a cohort study of patients with an incident episode of delirium to determine the risk of dementia, not as a case–control study, with patient with delirium and matched controls without delirium.

As such, we were unable to determine the net effect of delirium itself on dementia diagnosis. Future work should consider a case–control design to answer this important question.

In conclusion, our study reinforces the link between delirium and future dementia within a unique and well-powered data set. It has key clinical implications. We have shown that delirium in over 65s carries a 31% risk of developing dementia and an even greater risk of death in the 5-year postdiagnosis. This highlights the importance of recognising delirium and preventing it where possible. Future research is required to determine whether the recognition and early treatment of delirium could
reduce the risk of subsequent dementia or death. Moreover, at present, there is no consensus about follow-up and monitoring of cognitive function after an episode of delirium in the elderly. Our findings seem to support closer follow-up of delirium and proactive screening for dementia, but this has implications for service provision, particularly as the population ages. Indeed, it may be that those who experience an episode of delirium represent an ‘at risk’ group who could be candidates for future novel targeted therapies for dementia prevention and early-stage treatment. Finally, important questions about the pathophysiology of delirium remain to be answered. It is unclear whether delirium is a marker or an accelerant of irreversible cognitive decline. The field lacks strong data on the mechanistic relationship between delirium and dementia and indeed the cellular/molecular landscape in delirium and dementia. This is best generated through a combination of neuroimaging approaches, quality animal research and human biomarker studies.6

Twitter Samuel P Leighton @samleighton87 and Matthew Sheridan @mattdoc1982

Acknowledgements The authors would like to acknowledge that this work uses data provided by patients and collected by the NHS as part of their routine care and support. The authors would like to acknowledge the work of the West of Scotland Safe Haven team in supporting the extractions and linkage to de-identified NHS patient datasets.

Contributors SLP, JWH, EJ, MS and JC formulated the research question and designed the study. SLP analysed the data. SLP and JWH drafted the manuscript. EJ, MS, FD and JC critically evaluated and revised the manuscript. SLP serves as guarantor for the study.

Funding SLP is funded by a Clinical Academic Fellowship from the Chief Scientist Office, Scotland (CAF/19/04). JC is supported by funding from the Medical Research Council (MR/R035733/1), the Welcome Trust (104025/2/14/Z), Versus Arthritis (224535) and the Inger and George M Simpson Donation.

Competing interests EJ has received honoraria from Biogen and General Electric Healthcare. MS has received an honorarium from General Electric Healthcare. JC attended a Glasgow masterclass in dermatology funded by Janssen.

Patient consent for publication Not applicable.

Ethics approval The West of Scotland Safe Haven has ethical approval (17/WS/0237) to create a research database using routinely collected, un-consented patient data. Delegated research ethics approval was granted for linkage to NHS patient data by the Local Privacy and Advisory Committee at NHS GG&C under approval (GSH/18/AM/004).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. The study data is available by application to West of Scotland Safe Haven.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

ORCID iD
Samuel P Leighton http://orcid.org/0000-0002-3999-4204

REFERENCES
ICD10 Codes

**Delirium Coding**

F05    Delirium not induced by alcohol and other psychoactive substances

**Dementia Coding**

F00    Dementia in Alzheimer’s disease
F01    Vascular Dementia
F02    Dementia in other diseases classified elsewhere
F03    Unspecified dementia
F1073  Residual & Late-onset Psychotic Dementia Due to Use of Alcohol
F1173  Residual & Late-onset Psychotic Dementia Due to Use of Opioids
F1273  Residual & Late-onset Psychotic Dementia Due to Use of Cannabinoids
F1373  Residual & Late-onset Psychotic Dementia Due to Use of Sedatives/Hypnotics
F1473  Residual & Late-onset Psychotic Dementia Due to Use of Cocaine
F1573  Residual & Late-onset Psychotic Dementia Due to Use of Other Stimulants Including Caffeine
F1673  Residual & Late-onset Psychotic Dementia Due to Use of Hallucinogens
F1773  Residual & Late-onset Psychotic Dementia Due to Use of Tobacco
F1873  Residual & Late-onset Psychotic Dementia Due to Use of Volatile Solvents
F1973  Residual & Late-onset Psychotic Dementia Due to Use of Multiple Drugs/Psychoactive Substances
POST-MODEL ASSUMPTION TESTING

Figure 1 - The proportional hazard assumption is supported by a non-significant relationship between residuals and time for each of the covariates and the global test. The plots of the Schoenfeld residuals are independent of time.

Schoenfeld Individual Test p: 0.3728

Schoenfeld Individual Test p: 0.6658

Schoenfeld Individual Test p: 0.0781
Figure 2 - The above index plots show that comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations are influential individually according to the cut off proposed by Belsley, Kuh and Welch (values larger than 2/sqrt(n) are considered highly influential, in our case values larger than 0.018). [1]
Figure 3 – We assessed for nonlinearity by plotting Martingale residuals of the null Cox proportional hazards model against the continuous covariate, age (nonlinearity is not a concern for categorical variables). The fitted line suggests that age has a non-linear functional form, so the final Cox model was refitted using a penalised cubic spline term for age.

References

library(cmprsk)
library(survminer)
library(readr)
library(ggplot2)
library(scales)
library(dotwhisker)
library(survival)

# working directory
setwd("S:/Sam/Analysis_3/")

options(scipen=999)

data = read_csv("final_ggc_65.csv")
data = data[which(data$TIME_DEMENTIA_DEATH>0),]
data$DELIRIUM_DATE = as.Date(data$DELIRIUM_DATE, format = "%d/%m/%Y")
data$DELIRIUM_YEAR = format(data$DELIRIUM_DATE, "%Y")
data$DEMENTIA_ON_DEATH = as.factor(data$DEMENTIA_ON_DEATH)
data$DEC = data$AGE_DELIRIUM/10

data$SIMD_2009_QUINTILE = as.factor(data$SIMD_2009_QUINTILE)
data$SIMD_2009_QUINTILE = relevel(data$SIMD_2009_QUINTILE, ref = 5)

# TIME_DEMENTIA_DEATH denotes the survival time to the occurrence of the first event
# STATUS_DEMENTIA_DEATH is the event type indicator:
# 1: Dementia diagnosis
# 2: Death before dementia diagnosis
# 0: Censored observation: alive at end of follow-up

# Demographics
summary(as.factor(data$STATUS_DEMENTIA_DEATH))
summary(as.factor(data$DELIRIUM_YEAR))
summary(as.factor(data$SIMD_2009_QUINTILE))
summary(as.factor(data$SEX_MALE))
summary(data$AGE_DELIRIUM)
sd(data$AGE_DELIRIUM)

by(data$SIMD_2009_QUINTILE, as.factor(data$STATUS_DEMENTIA_DEATH), summary)
by(as.factor(data$SEX_MALE), as.factor(data$STATUS_DEMENTIA_DEATH), summary)
by(data$AGE_DELIRIUM, as.factor(data$STATUS_DEMENTIA_DEATH), summary)
by(data$AGE_DELIRIUM, as.factor(data$STATUS_DEMENTIA_DEATH), sd)
by(data$DEMENTIA_ON_DEATH, as.factor(data$STATUS_DEMENTIA_DEATH), summary)

# Stacked bar chart of outcomes
Outcome = c(rep("Censored",2), rep("Dementia diagnosis",2), rep("Death without dementia",2))
dem_on_death = rep(c("No","Dementia diagnosed on death"),3)
Patients = c(3631,0,2887,643,5788,0)

graph_data = data.frame(Outcome, dem_on_death, Patients)
pdf("outcomes.pdf", 10, 7)
ggplot(graph_data, aes(fill=dem_on_death, y=Patients, x=Outcome)) +
  geom_bar(position="stack", stat = "identity", alpha = 0.5) +
annotate("text", x=c(3,3), y=c(3250, 1000), label = c("On death","Before death"), color = "white", size = 6, fontface="bold")+
theme_bw()+theme(legend.position = "none", text = element_text(size=20), axis.title.x = element_blank())+
scale_y_continuous(minor_breaks = seq(0,6000,500), breaks = seq(0,6000,1000))+
scale_fill_manual(values = c("red","blue"))+
ggtitle("Outcomes for patients with an index episode of delirium")
dev.off()

#hist(data$DELIRIUM_DATE, breaks = "months", freq = TRUE)
#grid()

#Count by year and month
new = data.frame(table(format(data$DELIRIUM_DATE, "%Y-%m")))
#Append a day
new$Var1 = paste0(new$Var1, ",-1")
#Turn back into a date
new$Var1 = as.Date(new$Var1, format = "%Y-%m-%d")
pdf("delirium.pdf",10,5)
#plot using scale_x_date with 6 month breaks
ggplot(data = new, aes(x=Var1, y=Freq)) + geom_bar(stat="identity", colour ="grey20", fill = "white", size = 0.25)+
  scale_x_date(labels = date_format("%b %Y"), breaks = date_breaks("12 months"))+theme_bw()+theme(axis.text.x = element_text(angle = 90, vjust =0.5, hjust=1))+xlab("Date")+ylab("Monthly
Frequency of Delirium")+
scale_y_continuous(minor_breaks = seq(0,225,5), breaks = seq(0,225,50))
dev.off()

summary(data$TIME_DEMENTIA_DEATH)

# Figure 1: Plot cumulative incidence functions for dementia and death without dementia

#cumulative incidence 6 months

#cumulative incidence 1 year

ci_fit$`1 1`$est[which(ci_fit$`1 1`$time >1)][1]+1.96*sqrt(ci_fit$`1 1`$var[which(ci_fit$`1 1`$time >1)][1])
ci_fit$`1 2`$est[which(ci_fit$`1 2`$time >1)][1]  #Death without dementia
ci_fit$`1 2`$est[which(ci_fit$`1 2`$time >1)][1]-1.96*sqrt(ci_fit$`1 2`$var[which(ci_fit$`1 2`$time >1)][1])
ci_fit$`1 2`$est[which(ci_fit$`1 2`$time >1)][1]+1.96*sqrt(ci_fit$`1 2`$var[which(ci_fit$`1 2`$time >1)][1])
#cumulative incidence 5 years

#Dementia

ci_fit$`1 1`$est[which(ci_fit$`1 1`$time >5)][1]  #Death without dementia

#cumulative incidence 10 years

#Dementia

#cumulative incidence 15 years

#cumulative incidence 20 years

ggcompetingrisks(ci_fit, multiple_panels = F, conf.int = TRUE) + theme_bw() +
guides(linetype = FALSE) + scale_y_continuous(minor_breaks = seq(0,1,0.05), breaks = seq(0,1,0.1)) +
scale_x_continuous(minor_breaks = seq(0,25,1), breaks = seq(0,25,5)) + xlab("Time (years)") +
scale_color_manual(labels=c("Dementia","Death without dementia"), values = c("blue","red")) +
scale_fill_manual(labels=c("Dementia","Death without dementia"), values = c("blue","red")) +
theme(text = element_text(size = 20)) + labs(fill = "Outcome", colour = "Outcome")

dev.off()

# confidence intervals are 1.96+/- sqrt variance from cuminc function of cmpsk package as per
Frank Harrell
# discourse.datamethods.org/t/95-ci-around-cumulative-incidence-estimate/3948

#Cause specific hazard model for dementia treats death before dementia as censored
#
#time in years
res.cox.dem = coxph(Surv(TIME_DEMENTIA_DEATH/365.25, STATUS_DEMENTIA) ~ AGE_DELIRIUM +
SEX_MALE + SIMD_2009_QUINTILE, data = data)
summary(res.cox.dem)

#sthda.com/english/wiki/cox-model-assumptions
#test proportional-hazards assumption
test.ph = cox.zph(res.cox.dem)
test.ph
ggcovxph(test.ph)

#influential observations
ggcovdiagnostics(res.cox.dem, type = "dfbeta", linear.predictions = FALSE, ggtheme = theme_bw())
#values larger than 2/sqrt(n) are considered highly influential = 0.018

ggcovdiagnostics(res.cox.dem, type = "deviance", linear.predictions = FALSE, ggtheme =
theme_bw())
#deviance residuals should be roughly symmetry distributed around zero with a standard
deviation of 1

ggcovfunctional(Surv(TIME_DEMENTIA_DEATH/365.25, STATUS_DEMENTIA) ~ AGE_DELIRIUM, data = data, ylim = c(-1,1))
#Is the functional form of age linear. Assess using martingale's residuals from null model

#95%CI for loess fit
smoothSECurve = function(yy, xx){
  # use after a call to "plot"
  # fit a lowess curve and 95% confidence interval curve
  # make list of x values
  xx.list = min(xx) + ((0:100)/100)*(max(xx) - min(xx))
  # Then fit loess function through the points (xx, yy)
  # at the listed values
  yy.xx = predict(loess(yy ~ xx), se=T, newdata=data.frame(xx=xx.list))
  lines(yy.xx$fit ~ xx.list, lwd=2)
#SAME AS ABOVE ggcoxfunctional
res.cox.dem.0 = coxph(Surv(TIME_DEMENTIA_DEATH/365.25, STATUS_DEMENTIA) ~ 1, data = data)
rr.0 = residuals(res.cox.dem.0, type = "martingale")
plot(rr.0 ~ AGE_DELIRIUM, data=data)
smoothSEcurve(rr.0, data$AGE_DELIRIUM)
#Is the functional form of age linear. Assess using martingale's residuals from null model

rr.final = residuals(res.cox.dem, type = "martingale")
plot(rr.final~ AGE_DELIRIUM, data = data)
smoothSEcurve(rr.final, data$AGE_DELIRIUM)
#Is the functional form of age linear. Assess using martingale's residuals from fully adjusted model including age as covariate

#Age does not have a linear functional form so remodel using a penalised spline term for age
#See https://stats.stackexchange.com/questions/362510 for interpretation of above martingale's residuals plots

#add spline setting df = 0 then the AIC = (loglik -df) is used to choose an "optimal" degrees of freedom
res.cox.dem.spline = coxph(Surv(TIME_DEMENTIA_DEATH/365.25, STATUS_DEMENTIA) ~ pspline(AGE_DELIRIUM, df=0) + SEX_MALE + SIMD_2009_QUINTILE, data = data)
summary(res.cox.dem.spline)

#without age to do likelihood ratio test
res.cox.dem.noage = coxph(Surv(TIME_DEMENTIA_DEATH/365.25, STATUS_DEMENTIA) ~ SEX_MALE + SIMD_2009_QUINTILE, data = data)
#p value for age term
#according to Robertson
anova(res.cox.dem.spline, res.cox.dem.noage)
#according to stack exchange https://stats.stackexchange.com/questions/197179
anova(res.cox.dem.spline, res.cox.dem)

#Plot the age term on a log-hazard scale
termplot(res.cox.dem.spline, se=T, terms = 1, ylabs = "Log hazard")

pdf("age_spline.pdf",10,7)
#Plot the age term on a hazard scale
predicted = predict(res.cox.dem.spline, type = "terms", se.fit = TRUE, terms = 1)
plot(data$AGE_DELIRIUM, exp(predicted$fit), type="n", ylim = c(0,2), ylab = "Hazard ratio (df = 5.3)", xlab = "Age at delirium diagnosis")
lines(smooth.spline(data$AGE_DELIRIUM, exp(predicted$fit)), col = "red", lty = 1)
lines(smooth.spline(data$AGE_DELIRIUM, exp(predicted$fit + 1.96 * predicted$se.fit)), col = "orange", lty = 2)
lines(smooth.spline(data$AGE_DELIRIUM, exp(predicted$fit - 1.96 * predicted$se.fit)), col = "orange", lty = 2)
abline(h=1, lty=2)
boxplot(add=T, data$AGE_DELIRIUM, horizontal = T, at = 0, boxwex = 0.1)
dev.off()

# forest plot
# set up data
term = c("Male", "Female", "SIMD2009 Quintile 1", "SIMD2009 Quintile 2", "SIMD2009 Quintile 3", "SIMD2009 Quintile 4", "SIMD2009 Quintile 5")
estimate = c(1.057, 1, 1.386, 1.190, 1.454, 1.465, 1)
conf.low = c(0.9858, 1, 1.2439, 1.0518, 1.2870, 1.2903, 1)
conf.high = c(1.134, 1, 1.544, 1.346, 1.643, 1.664, 1)
plot_data = data.frame(term, estimate, conf.low, conf.high)

pdf("forest.pdf", 10, 5)
# forest plot
dwplot(plot_data, vline = geom_vline(xintercept = 1, colour = "grey60", linetype = 2),
    dot_args = list(color = "#F8766D"), # color for the dot
    whisker_args = list(color = "Grey47") # color for the whisker
) +
ggtitle("Multivariable Adjusted Hazard Ratios for Dementia") +
theme_bw() +
xlim(0.75, 1.75) +
theme(plot.margin = unit(c(0.5,1,1), "cm")) +
xlabel("Hazard Ratio (95% CI)") +

gtext(x = 1.82,
y = 7, label = "1.06 (0.99, 1.13); p = 0.12",
hjust = 0,
size = 3) +
gtext(x = 1.82,
y = 6, label = "Reference",
hjust = 0,
size = 3) +
gtext(x = 1.82,
y = 5, label = "1.39 (1.24, 1.54); p = <0.001",
hjust = 0,
size = 3) +
gtext(x = 1.82,
y = 4, label = "1.19 (1.05, 1.35); p = 0.006",
hjust = 0,
size = 3) +
gtext(x = 1.82,
y = 3, label = "1.45 (1.29, 1.64); p = <0.001",
hjust = 0,
size = 3) +
gtext(x = 1.82,
y = 2, label = "1.47 (1.29, 1.66); p = <0.001",
hjust = 0,
size = 3) +
gtext(x = 1.82,
y = 1, label = "Reference",
hjust = 0,