

## Supplementary material

### Distribution of age of presentation

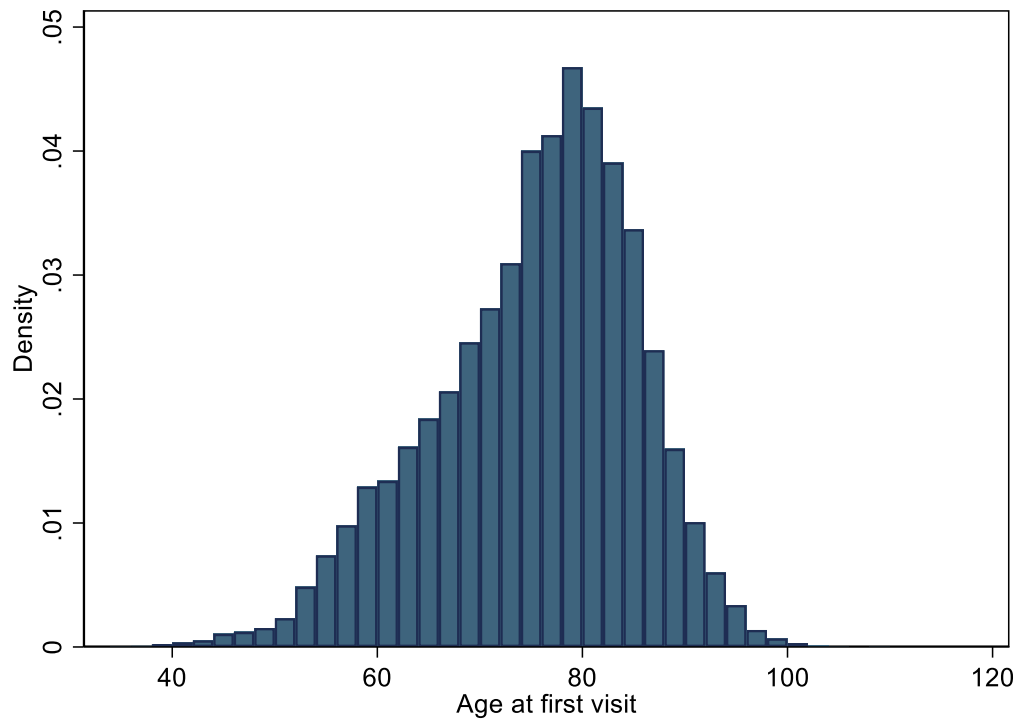


Figure S1: Distribution of age of presentation.

### Missing data

Variable	N Missing
Age at presentation (visit 1)	0
MMSE (visit 1)	1,041
Sex	0
Symptom duration	85
APOE e4 alleles	2640

Table S1: Summary of missing data

## Additional models

Predictors	Second visit symptom			
	Judgement	Language	Visuospatial function	Attention
<b>Patients contributing to model</b>	449	2646	2720	2752
<b>Age at presentation (10 yrs)</b>	0.79 (0.63,0.99)	0.91 (0.82,0.99)	0.89 (0.82,0.98)	0.84 (0.77,0.92)
<b>Symptom length (years)</b>	0.93 (0.87,1.00)	1.03 (1.00,1.06)	1.04 (1.01,1.07)	1.02 (0.99,1.05)
<b>Time since visit 1 (years)</b>	3.59 (1.73,7.45)	1.46 (1.22,1.75)	1.37 (1.14,1.65)	1.43 (1.19,1.73)
<b>Female vs male</b>	0.65 (0.44,0.98)	1.04 (0.87,1.24)	1.14 (0.96,1.35)	1.01 (0.85,1.19)
<b>First predominant cognitive symptom</b>	Memory (baseline)	1.00	1.00	1.00
	Judgement	-	1.41 (0.88,2.25)	1.13 (0.74,1.73)
	Language	1.26 (0.48,3.27)	-	0.85 (0.55,1.32)
	VSF	1.64 (0.43,6.28)	1.58 (0.94,2.67)	-
	Attention	0.86 (0.08,8.87)	2.73 (0.88,8.48)	1.79 (0.73,4.37)
	Fluctuating cognition	*	4.09 (0.37,45.68)	1.00 (1.00,1.00)
	Other	1.00(1.00,1.00)	1.40 (0.24,8.03)	1.35 (0.36,5.10)
	Unknown	*	*	0.81 (0.07,9.49)
<b>Presence of symptom at visit 1</b>	Memory	5.09 (1.67,15.59)	2.85 (0.76,10.74)	2.17 (0.78,6.01)
	Judgement	-	1.63 (1.18,2.24)	2.00 (1.46,2.73)
	Language	1.25 (0.77,2.01)	-	1.24 (1.03,1.48)
	VSF	1.43 (0.86,2.38)	1.30 (1.09,1.56)	-
	Attention	0.78 (0.47,1.29)	1.38 (1.15,1.65)	1.52 (1.27,1.81)
<b>Wald test p-value for First Predominant Symptom</b>	0.89	0.12	0.73	0.78

\* All patients with this symptom as their first predominant developed the outcome symptom at visit 1 (perfect predictor)  
 - These people were omitted as they had already developed the symptom in question

Table S2: Comparative odds ratios (95% CI) for the development of each of four cognitive symptoms at visit 2: logistic regression models as in Figure 3 with additional adjustment for first predominant cognitive symptom.

Visit	APOE-ε4 alleles	Symptoms			
		Judgement	Language	Visuospatial	Attention
1*	1 vs 0	1.08 (0.89, 1.31)	0.99 (0.89, 1.10)	1.07 (0.96,1.19)	1.07 (0.97,1.19)
	2 vs 0	1.19 (0.88, 1.61)	0.83 (0.71, 0.98)	0.98 (0.84,1.15)	1.01 (0.86,1.18)
2†	1 vs 0	1.01 (0.64, 1.60)	0.90 (0.73, 1.10)	1.06 (0.87,1.29)	1.02 (0.83,1.25)
	2 vs 0	1.05 (0.53, 2.09)	0.87 (0.64, 1.17)	1.09 (0.81,1.46)	1.03 (0.77,1.38)

\*Coefficients adjusted for age at presentation, gender, symptom length and first predominant symptom

†Coefficients are adjusted for age at presentation, gender, symptom length, inter-visit interval and presence of symptoms at visit 1

Table S3: Comparative odds ratios for the development of each of four cognitive symptoms according to the number of APOE-ε4 alleles: logistic regression models as in Figures 2 & 3 with number of APOE-ε4 alleles as additional covariates (only APOE coefficients shown).

## Code for analyses

```
cap prog drop dataprep
program define dataprep
```

```
* Create a numeric ID for each patient to replace string *
```

```
cap drop id
gen id = naccid
destring id, replace ignore("NACC")
```

```
* Count the # of unique IDs *
```

```
egen tag = tag(id)
egen distinct = total(tag), by(id)
```

```
* Recode Sex *
```

```
recode sex (1=0) (2=1)
label define sexl 0"Male" 1"Female"
label values sex sexl
```

```
* Re-label first cog/beh symptoms *
```

```
/* Codebook defs
0 = No impairment in cognition
1 = Memory
2 = Orientation
```

```

3 = Executive function - judgment,
planning, problem-solving
4 = Language
5 = Visuospatial function
6 = Attention/concentration
7 = Fluctuating cognition
8 = Other (specify)
99 = Unknown
*/

* MMSE score at visit 1
bysort id: gen mmsev1 = naccmmse[1]

* Fix First predominant symptom to be the same for all visits *
bysort id: gen fps = naccocgf[1]
drop if fps==0 // Drop patients with no FPS

gen cogfrst = naccocgf // Cogfirst reflects the first predominant
symptom recorded at each visit (i.e. can change)
*gen cogfrst = fps // Cogfirst now based on whatever was
recorded as FPS at visit 1 (i.e. cannot change)
label define cogfrst1 1"Memory" 2"Orientation" 3"Judgement" 4"Language"
5"Visuospatial Function" ///
6"Attention/Concentration" 7"Fluctuating
cognition" 8"Other" ///
99"Unknown"

label values cogfrst cogfrst1
label values fps cogfrst1

gen byte cogfrstmem = cogfrst==1

* How many missing values *
cap drop sumcog
egen sumcog=rownonmiss(cogmem cogori cogjudg coglang cogvis cogattn cogothr
cogfluc)

* Max # of visits *
gen maxv=1
replace maxv=2 if visit==2 & sumcog!=0
replace maxv=3 if visit==3 & sumcog!=0
replace maxv=4 if visit==4 & sumcog!=0
replace maxv=5 if visit==5 & sumcog!=0
replace maxv=6 if visit==6 & sumcog!=0
replace maxv=7 if visit==7 & sumcog!=0
replace maxv=8 if visit==8 & sumcog!=0
replace maxv=9 if visit==9 & sumcog!=0
replace maxv=10 if visit==10 & sumcog!=0
replace maxv=11 if visit==11 & sumcog!=0
replace maxv=12 if visit==12 & sumcog!=0

sort id maxv
by id: egen maxvis=max(maxv)

```

```

* Did they attend visit *
gen attend=1 if sumcog>0
replace attend=0 if sumcog==0

* Number of visits attended *
egen visitsatt=total(attend), by(id)

sort naccadc id visit

* Year at first visit *
gen yrvisit1 = birthyr + naccageb

* Convert age of presentation to decades *
gen naccagedec = naccageb/10

* Recode APOE4
gen apoe = naccapoe
recode apoe 1 3 6 = 0
recode apoe 2 5 = 1
recode apoe 4 = 2
recode apoe 9 = .

* Time since last visit (in years)
gen ty = t/365

*----- Relabel all variable names -----*
label variable naccadc "ADC ID"
label variable naccid "NACC ID"
label variable sex "Sex"
label variable visit "Visit"
label variable birthmo "Birth month"
label variable birthyr "Birth year"
label variable naccageb "Age at visit 1"
label variable cogmem "Memory"
label variable cogori "Orientation"
label variable cogjudg "Judgement"
label variable coglang "Language"
label variable cogvis "Visuospatial Function"
label variable cogattn "Attention"
label variable cogfluc "Fluctuation"
label variable cogothr "Other"
label variable cogfrst "First cognitive symptom (as recorded at visit j)"
label variable agediff "Length of first symptom"
label variable t "Time since previous visit (days)"
label variable t1 "Time since visit 1 (days)"
label variable id "ID"
label variable agegp "Age Group"
label variable sumcog "Count of measured cognitive symptoms"
label variable maxvis "Max # of visits"
label variable attend "Visit attended (Y/N)"
label variable visitsatt "# of visits attended"
label variable ty "Time since previous visit (years)"

```

```

label variable naccagedec "Age of onset in decades"
label variable naccmmse "Mini Mental State Exam score (at visit 1)"
label variable apoe "APOE4 group"
label variable fps "First predominant symptom (as recorded at visit 1)"

*-----*
* For missing data *
gen unknowns = 0
qui foreach v of var cogmem-cogothr {
    replace unknowns = unknowns + 1 if inlist(`v', 9)
}

end

*-----*
*drop behavioural variables
cap drop be*

*drop visit 12
drop if visit==12
*----- Add previous visit symptom -----*
cap prog drop createcopy
program define createcopy

    cap drop *prev
    bysort id: gen vprev=visit+1
    *cap drop visit
    rename visit viscurr
    rename vprev visit

    foreach V of varlist cog*{
        rename `V' `V'prev
    }
    keep id visit viscurr cog*
    save copy, replace
end

cap prog drop prevsym
program define prevsym
    use "$dir_1\Data\data", clear
    dataprep

    merge 1:1 id visit using copy
    drop if visit==8

    list id visit cogmem* in 1/20

    label variable cogmemprev "Memory (prev)"
    label variable cogoriprev "Orientation (prev)"
    label variable cogjudgprev "Judgement (prev)"
    label variable coglangprev "Language (prev)"
    label variable cogvisprev "Visuospatial Function (prev)"
    label variable cogattnprev "Attention (prev)"

```

```

        label variable cogflucprev "Fluctuation (prev)"
        label variable cogothrprev "Other (prev)"
end

```

```

global dir_1 "H:\\"

```

```

log using "$dir_1\Log", replace

```

```

* _____
* Data preparation using data prep do file

```

```

cd "$dir_1\Data"

```

```

use data, clear
cd "$dir_1\Do files"
do Dataprep
cd "$dir_1\Data"

```

```

* Create copy of data for previous symptom
dataprep
createcopy

```

```

* Prepare dataset and add previous visit symptoms
prevsym

```

```

save data_clean, replace
use data_clean, clear

```

```

* _____ Descriptives _____

```

```

* Summarise first cognitive symptom

```

```

tab fps
tab fps, nolab
tab cogfrst

```

```

tab cogfrst if visit==2 & cogfrstprev==1
tab cogfrst if visit==2 & cogfrstprev==3
tab cogfrst if visit==2 & cogfrstprev==4
tab cogfrst if visit==2 & cogfrstprev==5
tab cogfrst if visit==2 & cogfrstprev==6
tab cogfrst if visit==2 & cogfrstprev==7
tab cogfrst if visit==2 & cogfrstprev==8

```

```

* Age at visit 1
summ naccageb if tag
return list

```

```

local N = round(r(N),0.1)
local mean = round(r(mean),0.1)
local sd = round(r(sd),0.1)

```

```

* Gender
summ sex if sex==1 & tag

```

```

* Symptom length

```

```

summ agediff if tag

* Max visit
summ maxvis if tag

* Number attending visit 1 and visit 2
tab maxvis if maxvis<=2 & tag
tab attend if visit==1
tab attend if visit==2

* APOE e4
tab apoe if tag, matcell(x)

* MMSE
tab mmsev1 if tag & naccageb!=., m

* Missing Data _____
use data_clean, clear

tab maxvis if tag

tab sumcog visit if visit<=maxvis & visit<3, col

egen sumcog4=rownonmiss(cogmem cogori cogjudg coglang cogvis cogattn
cogothr cogfluc)
tab sumcog4 visit if visit<=maxvis & visit<3, m

* Distn of # missing symptoms for attended visits (exclude visits with all
recorded)
tab sumcog if visit<=maxvis & visit <3 & sumcog!=6

* Which symptoms are missing more often?
tab cogmem if sumcog!=0 & sumcog!=6 & visit<=maxvis & visit<3, m
tab cogori if sumcog!=0 & sumcog!=6 & visit<=maxvis & visit<3, m
tab cogjudg if sumcog!=0 & sumcog!=6 & visit<=maxvis & visit<3, m
tab coglang if sumcog!=0 & sumcog!=6 & visit<=maxvis & visit<3, m
tab cogvis if sumcog!=0 & sumcog!=6 & visit<=maxvis & visit<3, m
tab cogattn if sumcog!=0 & sumcog!=6 & visit<=maxvis & visit<3, m
tab cogothr if sumcog!=0 & sumcog!=6 & visit<=maxvis & visit<3, m

** Other missing variables

mdesc naccageb mmsev1 sex agediff fps apoe if tag

* _____ ANALYSIS QUESTION 1 _____
* VISIT 1
use data_clean, clear

* Obtain # patients with each symptom at visit 1
foreach V of varlist cogjudg-cogothr{
    tab `V' if visit==1
}

* Number of patients developing symptom after FPCS

```



```

tab cogjudg if visit==1 & fps!=3
tab coglang if visit==1 & fps!=4
tab cogvis if visit==1 & fps!=5
tab cogattn if visit==1 & fps!=6

* Judgement
logistic cogjudg ib1.fps ///
      naccagedec agediff i.sex ///           Age at first visit,
symptom length, sex
      if visit==1 & fps != 3, nofvlab      baselevels

* Language
logistic coglang ib1.fps ///
      naccagedec agediff i.sex ///
      if visit==1 & fps != 4, noomit nofvlab baselevels

* Visuospatial
logistic cogvis ib1.fps ///
      naccagedec agediff i.sex ///
      if visit==1 & fps != 5, baselevels

* Attention
logistic cogattn ib1.fps ///
      naccagedec agediff i.sex ///
      if visit==1 & fps != 6, baselevels

* VISIT 2
* Now include presence of other cognitive symptoms at visit 1 (exclude
patients who had the same symptom at visit 1) and include time between
visits
* Obtain # patients with each symptom at visit 2
foreach V of varlist cogjudg-cogothr{
    tab `V' if visit==2
}

* Number of patients developing each symptom
tab cogjudg if visit==2 & fps != 3 & fps!=8 & cogjudgprev != 1
tab coglang if visit==2 & fps != 4 & coglangprev != 1
tab cogvis if visit==2 & fps != 5 & fps!=7 & cogvisprev != 1
tab cogattn if visit==2 & fps != 6 & cogattnprev != 1

* Judgment
logistic cogjudg ib1.fps ///
      cogmemprev coglangprev ///
      cogvisprev cogattnprev ///
      naccagedec ty agediff i.sex ///
      if visit==2 & fps != 3 & cogjudgprev != 1, nofvlabel

* Joint test for First Prom Symptom (FPS)
test 1.fps 4.fps 5.fps 6.fps 8.fps
local pvalfps = r(p)
test cogmemprev coglangprev cogvisprev cogattnprev

* Language

```

```

logistic coglang ib1.fps ///
    cogmemprev cogjudgprev ///
    cogvisprev cogattnprev ///
    naccagedec ty agediff i.sex ///
    if visit==2 & fps != 4 & coglangprev != 1, nofvlab
test 3.fps 5.fps 6.fps 7.fps 8.fps
local pvalfps = r(p)
test cogmemprev cogjudgprev cogvisprev cogattnprev

* VSF
logistic cogvis ib1.fps ///
    cogmemprev cogjudgprev coglangprev ///
    cogattnprev ///
    naccagedec ty agediff i.sex ///
    if visit==2 & fps != 5 & cogvisprev != 1, nofvlab
test 3.fps 4.fps 6.fps 7.fps 8.fps 99.fps
local pvalfps = r(p)
test cogmemprev cogjudgprev coglangprev cogattnprev

* Attention
logistic cogattn ib1.fps ///
    cogmemprev cogjudgprev coglangprev ///
    cogvisprev ///
    naccagedec ty agediff i.sex ///
    if visit==2 & fps != 6 & cogattnprev != 1, nofvlab
test 3.fps 4.fps 5.fps 8.fps 99.fps
local pvalfps = r(p)

test cogmemprev cogjudgprev coglangprev cogvisprev

* VISIT 2
* Excluding FPS, tests equality of previous symptoms at visit 1
* Judgement
logistic cogjudg ///
    cogmemprev coglangprev ///
    cogvisprev cogattnprev ///
    naccagedec ty agediff i.sex ///
    if visit==2 & fps != 3 & fps!=8 & cogjudgprev != 1, nofvlabel
    // Omit FPS==8 because predicts perfectly
* Joint test for First Prom Symptom (FPS)
testparm cogmemprev coglangprev cogvisprev cogattnprev, equal

* Language
logistic coglang ///
    cogmemprev cogjudgprev ///
    cogvisprev cogattnprev ///
    naccagedec ty agediff i.sex ///
    if visit==2 & fps != 4 & coglangprev != 1, nofvlab
testparm cogmemprev cogjudgprev cogvisprev cogattnprev, equal
test cogmemprev=cogjudgprev=cogvisprev=cogattnprev

* VSF

```

```

logistic cogvis ///
    cogmemprev cogjudgprev coglangprev ///
    cogattnprev ///
    naccagedec ty agediff i.sex ///
    if visit==2 & fps != 5 & fps!=7 & cogvisprev != 1, noflab
testparm cogmemprev cogjudgprev coglangprev cogattnprev, equal

* Attention
logistic cogattn ///
    cogmemprev cogjudgprev coglangprev ///
    cogvisprev ///
    naccagedec ty agediff i.sex ///
    if visit==2 & fps != 6 & cogattnprev != 1, noflab
testparm cogmemprev cogjudgprev coglangprev cogvisprev, equal

* _____ ANALYSIS QUESTION 2 _____
* Repeat models above including APOE4 variable

* VISIT 1
* Judgment
logistic cogjudg ib1.fps ///
    naccagedec agediff i.sex i.apoe ///           Age at first
visit, symptom length, sex
    if visit==1 & fps != 3

* Language
logistic coglang ib1.fps ///
    naccagedec agediff i.sex i.apoe ///
    if visit==1 & fps != 4

* VSF
logistic cogvis ib1.fps ///
    naccagedec agediff i.sex i.apoe ///
    if visit==1 & fps != 5

* Attention
logistic cogattn ib1.fps ///
    naccagedec agediff i.sex i.apoe ///
    if visit==1 & fps != 6

* Visit 2
* Judgement
set more off
logistic cogjudg ///
    cogmemprev coglangprev ///
    cogvisprev cogattnprev ///
    naccagedec ty agediff i.sex i.apoe ///
    if visit==2 & fps != 3 & fps!=8 & cogjudgprev != 1, noflabel
// Omit FPS==8 because predicts perfectly

* Joint tests for FPS and visit 1 symptoms
test 1.fps 4.fps 5.fps 6.fps 8.fps

```

```

test cogmemprev coglangprev cogvisprev cogattnprev

* Language
logistic coglang ///
      cogmemprev cogjudgprev ///
      cogvisprev cogattnprev ///
      naccagedec ty agediff i.sex i.apoe ///
      if visit==2 & fps != 4 & coglangprev != 1, nofvlab

* Joint tests for FPS and visit 1 symptoms
test 3.fps 5.fps 6.fps
test cogmemprev cogjudgprev cogvisprev cogattnprev

* VSF
logistic cogvis ///
      cogmemprev cogjudgprev coglangprev ///
      cogattnprev ///
      naccagedec ty agediff i.sex i.apoe ///
      if visit==2 & fps != 5 & fps!=7 & cogvisprev != 1, nofvlab

* Joint tests for FPS and visit 1 symptoms
test 3.fps 4.fps 6.fps 8.fps 99.fps
test cogmemprev cogjudgprev coglangprev cogattnprev

* Attention
logistic cogattn ib1.fps ///
      cogmemprev cogjudgprev coglangprev ///
      cogvisprev ///
      naccagedec ty agediff i.sex i.apoe ///
      if visit==2 & fps != 6 & cogattnprev != 1, nofvlab

* Joint tests for FPS and visit 1 symptoms
test 3.fps 4.fps 5.fps 8.fps 99.fps
test cogmemprev cogjudgprev coglangprev cogvisprev

log close

```