The association of pathology patterns and AD risk: with and without imputing pathology data

Hiroko H. Dodge, PhD

Associate Professor of Neurology
Biostatistics and Data Core Leader
Layton Aging and Alzheimer's Disease Center,
Oregon Health and Science University, Portland, OR, USA

Kevreson Research Professor of Neurology
Associate Professor of Neurology
Biostatistics and Data Core Leader
Michigan Alzheimer's Disease Center
University of Michigan, Ann Arbor, MI, USA

OREGON HEALTH & SCIENCE UNIVERSITY

ORCATECH
SUCCESSFUL AGING WITH TECHNOLOGY

University of Michigan Health System

Outline

- SMART (Statistical Modeling of Aging and Risk of Transition) study data sets
- Research question
- Brief explanation of Imputation methods used in this project
- Results

Statistical Modeling of Aging and Risk of Transition study (SMART)

- **PIs: Fred Schmitt & Dick Kriscio (Kentucky)**

- NIA R01AG386561 (2011-2016)

- A consortium of **11 different** high-quality longitudinal studies of aging and cognition (N=11,541 participants)

MAPWU/Memory and Aging Project (Washington University)

OBAS/ Oregon Brain Aging Study I & II + AADAPt/African American Dementia and Aging Project + KEAP/Klamath Exceptional Aging Project (Oregon Health & Science University)

BRAiNS/Biologically Resilient Adults in Neurological Studies (Kentucky)

HAAS/Honolulu Asia Aging Study (Kuwakini Medical Center)

ROS/Religious Orders Study + MAPRU /Memory and Aging Project (Rush University)

Nun study (U. of Minnesota)

EAS/Einstein Aging Study (Yeshiva University)

Data: Statistical Modeling of Aging and Risk of Transition study (SMART)

- **Community samples**, not clinical samples (i.e., more representative of **mixed pathologies**, not limited to pure AD pathology). Ideal for characterizing risk and protective factors associated with subtypes of age-associated **mixed neuropathologies**
- **Over 6000 death**
- **Over 3000 autopsies (but not all pathology variables available)**

Abner et al., "The Statistical Modeling of Aging and Risk of Transition Project: Data Collection and Harmonization Across 11 Longitudinal Cohort Studies of Aging, Cognition, and Dementia" *Observational Studies* 2015:1:56-73

Research Question

Contribution of vascular factors on incidence of clinically diagnosed AD

How much additional risks are contributed by pathology-confirmed vascular factors, beyond definitive AD pathology, in developing clinically diagnosed AD

Background

- For a given level of AD pathology in the brain, the greater the number of cerebrovascular lesions, the greater the likelihood of clinically significant cognitive impairment / dementia

ELSEVIER

Alzheimer's & Dementia 7 (2011) 356–360

Alzheimer's & Dementia

Short Report

Risk of Alzheimer's disease incidence attributable to vascular disease in the population

Hiroko H. Dodge[a,b,c,*], Chung-Chou H. Chang[d], Ilyas M. Kamboh[e], Mary Ganguli[b,f]

[a]Department of Neurology, Oregon Health & Science University, Portland, OR, USA
[b]Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA, USA

Population Attributable Risk (PAR):

Risk of dementia could be reduced by 10.8% by eliminating overt cerebrovascular disease (stroke/TIA), and the risk of AD by 9.1%.

Abstract

carriers.

Keywords: Population-Attributable Risk % (PAR%); Epidemiology; Vascular disease; AD; *APOE*ε4

Objective

To examine the risk of incident AD associated with **pathology-confirmed presence** of vascular factors: lacunes (small artery infarcts) and one or more large artery cerebral infarct(s).

(Could also add, e.g., atherosclerotic vascular pathology (npavas), subcortical arteriosclerotic leukoencephalopathy (npart), hemorrhages (nphem))

Inclusion Criteria in the Current Study

Due to the limitation in harmonization of some autopsy variables, the following 6 cohorts were used in the current analyses.

- *MAPWU/Memory and Aging Project (Washington University)*
- *OBAS/ Oregon Brain Aging Study I & II + AADAPt/African American Dementia and Aging Project + KEAP/Klamath Exceptional Aging Project (Oregon Health & Science University)*
- *BRAiNS/Biologically Resilient Adults in Neurological Studies (Kentucky)*
- *ROS/Religious Orders Study + MAPRU /Memory and Aging Project (Rush University)*

Inclusion criteria: normal cognition at baseline with Apoe4 information

Pathology Patterns Created

AD definitive pathology (ADDP) was defined as having frequent or moderate neuritic plaques scores, or Braak& Braak neurofibrillary stage ≥ 5

6 pathology patterns:

1) ADDP only,

2) lacunes without ADDP,

3) ADDP and lacunes,

4) large infarcts without ADDP,

5) ADDP and large infarct(s) with or without lacunes,

6) no ADDP or vascular pathologies (control group). (LB pathology—controlled)

Approach

MAR assumption?
Among those with normal at baseline, apoe 4 information

Mean (std)	With autopsy variables of our interests N=1054	Missing autopsy N=512	p-value*
age at death	88.5 (7.26)	89.49 (7.31)	0.119
Women (%)	63.1	61.3	0.015
Years of Education	15.9 (3.43)	14.38 (3.46)	<0.001
Apoe*4 (at least one e4 allele, %)	22.1	22.8	0.523
Duration of follow-up	7.84 (4.72)	6.92 (5.14)	0.017
Duration from the last assessment to death	0.95 (1.29)	1.82 (2.22)	<0.001

*: p-values are based on a logistic regression model (outcome=having autopsy data or not), including all variables in this table in the model

Why impute?

ADVANTAGE

- Increase power
- Reduce estimation bias

e.g. 1: Autopsy cases—younger than general age of death –> over-represent the pathology-cog association among younger group

e.g. 2: Those who die with mixed pathologies might live longer -> their death = less likely to be included in the autopsy sample –> under-represent the pathology- cog association among those died with mixed pathologies

Why impute?

ALTERNATIVE APPROACHES

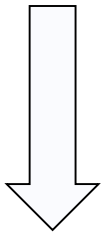
Use weights

- ✓ good for marginal representation, but not good for “associations”
- ✓ Wasting a lot of information available

MAR Assumption
Limitation

X	Y
observed	observed
observed	?

MAR: X is always observed and
Pr(Y is observed|Xobs,Yobs,Ymis)
=Pr(Y is observed|Xobs,Yobs)



Pr(Ymis|Xobs)=Pr(Yobs|Xobs)

Multiple Imputation (MI)

- MI analysis has many desirable advantages over other missing data analysis methods (e.g. weighting) for our data
- The assumption for missingness in our data is missing at random (MAR, Rubin 1976). We have evidence showing that the missing completely at random (MCAR) assumption is not satisfied.

Data with Missing Values

	Pathology Variables from Autopsy				Clinical Diagnosis from Survey			
Fully Observed Variables*	Large Infarct (Y/N?)	Lacunar Infarct (Y/N?)	ADDP (Y/N?)	Lewy Body (Y/N?)	Age at MCI Onset	Age at General Dementia Onset	"Dementia =AD?" at Dementia Onset	Age at AD onset
observed	observed	Observed	Observed	observed	observed	observed	observed	observed
observed	observed	observed	observed	observed	observed	observed	observed	?
observed	**?**	observed	observed	**?**	observed	observed	observed	observed
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
observed	observed	**?**	observed	observed	observed	**?**	**?**	**?**
observed	**?**	observed	**?**	observed	**?**	**?**	**?**	**?**

* Fully observed variables: Gender, Education, Age at death, Clinical diagnosis at last available survey, Participants survey duration, Participants duration between end of survey and death, etc.

Challenges for MI Modeling

- Different types of variables: binary, continuous, or even survival variables
- Boundary restrictions: e.g., imputed age at dementia onsets must be later than last normal clinical diagnosis date, or normal dx does not come after observed or imputed AD onset age
- Logical restrictions: e.g., only those with dementia onset need additional imputed AD onset date (if participants were AD demented at that time, then AD onset date is equal to general dementia onset date)

Sequential Regression Multiple Imputation (SRMI)

- We adopt a sequential regression multiple imputation (SRMI, Raghunathan et al 2001) approach, also known as multivariate imputation by chained equations (MICE, van Buuren 2011), to impute the missing values.
- The SRMI approach uses an iterative algorithm with a sequence of fully conditionally specified models (similar to Gibbs samplers but with key differences).
- It's particularly useful in our study as it can easily handle the challenging features.
- Algorithm convergence need to be closely monitored due to SRMI's theoretical weakness (no joint distribution proposed)

The First Iteration

Fill missing values with initially imputed values

	Pathology Variables from Autopsy				Clinical Diagnosis from Survey			
Fully Observed Variables*	Large Infarct	Lacunar Infarct	ADDP	Lewy Body	Age at MCI Onset	Age at General Dementia Onset	"Dementia =AD?" at Dementia Onset	Age at AD onset
observed	observed	observed	observed	observed	observed	observed	observed	observed
observed	observed	observed	observed	observed	observed	observed	observed	**imputed**
observed	**imputed**	observed	observed	**imputed**	observed	observed	observed	observed
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
observed	observed	**imputed**	observed	observed	observed	**imputed**	**imputed**	**imputed**
observed	**imputed**	observed	**imputed**	observed	**imputed**	**imputed**	**imputed**	**imputed**

* Fully observed variables: Gender, Education, Age at death, Last clinical diagnosis, Participants survey duration, Participants duration between end of survey and death, etc.

SRMI Model Specification for iteration t ≥ 2

- Pathology variables (4 binary variables)
  - **m1(Large Infarct|Predictors, θ1)** -- logistic regression with covariates: Lacunar Infarct, ADDP, Lewy body, Apoe4, last available clinical diagnosis, gender, education, age at death, survey duration, survey end to death duration, age at MCI onset, age at dementia onset, age at AD onset
  - **m2(Lacunar Infarct|Predictors, θ2)** -- logistic regression with covariates: Large Infarct, ADDP, Lewy body, Apoe4, last available clinical diagnosis, gender, education, age at death, survey duration, survey end to death duration, age at MCI onset, age at dementia onset, age at AD onset
  - **m3(ADDP|Predictors, θ3)** -- logistic regression with covariates: Large Infarct, Lacunar Infarct, Lewy body, Apoe4, last available clinical diagnosis, gender, education, age at death, survey duration, survey end to death duration, age at MCI onset, age at dementia onset, age at AD onset
  - **m4(Lewy Body|Predictors, θ4)** -- logistic regression with covariates: Large Infarct, Lacunar Infarct, ADDP, Apoe4, gender, education, age at death, survey duration, survey end to death duration, age at MCI onset, age at dementia onset, age at AD onset

SRMI Model Specification for iteration t ≥ 2

- Clinical variables (3 continuous and 1 binary)
  - **m5(MCI age of onset|Predictors, θ5)** – linear regression with covariates: Large Infarct, Lacunar Infarct, ADDP, Lewy body, Apoe4, gender, education, age at death, survey duration, survey end to death duration
  - **m6(General Dementia Onset|Predictors, θ6)** – linear regression with covariates: Large Infarct, Lacunar Infarct, ADDP, Lewy body, Apoe4, gender, education, age at death, survey duration, survey end to death duration
  - **m7(Dementia type=AD?|Predictors, θ7)** – logistic regression with covariates: Large Infarct, Lacunar Infarct, ADDP, Lewy body, Apoe4, gender, education, age at death, survey duration, survey end to death duration, age at MCI onset, age at dementia onset
  - **m8(AD age of onset |Predictors, θ8)** – linear regression with covariates: Large Infarct, Lacunar Infarct, ADDP, Lewy body, Apoe4, gender, education, age at death, survey duration, survey end to death duration

Imputing missing values for Large Infarct

at the *t*th iteration

Large Infarct	Other Variables (updated at (t-1)th iteration)
observed	observed+Imputed(t-1)
?	observed+Imputed(t-1)

m1(Large Infarct|updated predictors, θ1) is fit, and θ1(t) is drawn from its approximate posterior distribution

Missing values are drawn from m1(Large Infarct|updated predictors, θ1(t))

Imputing missing values for Lacunar Infarct

at the *t*th iteration

Lacunar Infarct	Large Infarct (updated)	Other Variables (updated)
observed	observed+ Imputed(t)	observed+ Imputed(t-1)
?	observed+ Imputed(t)	observed+ Imputed(t-1)

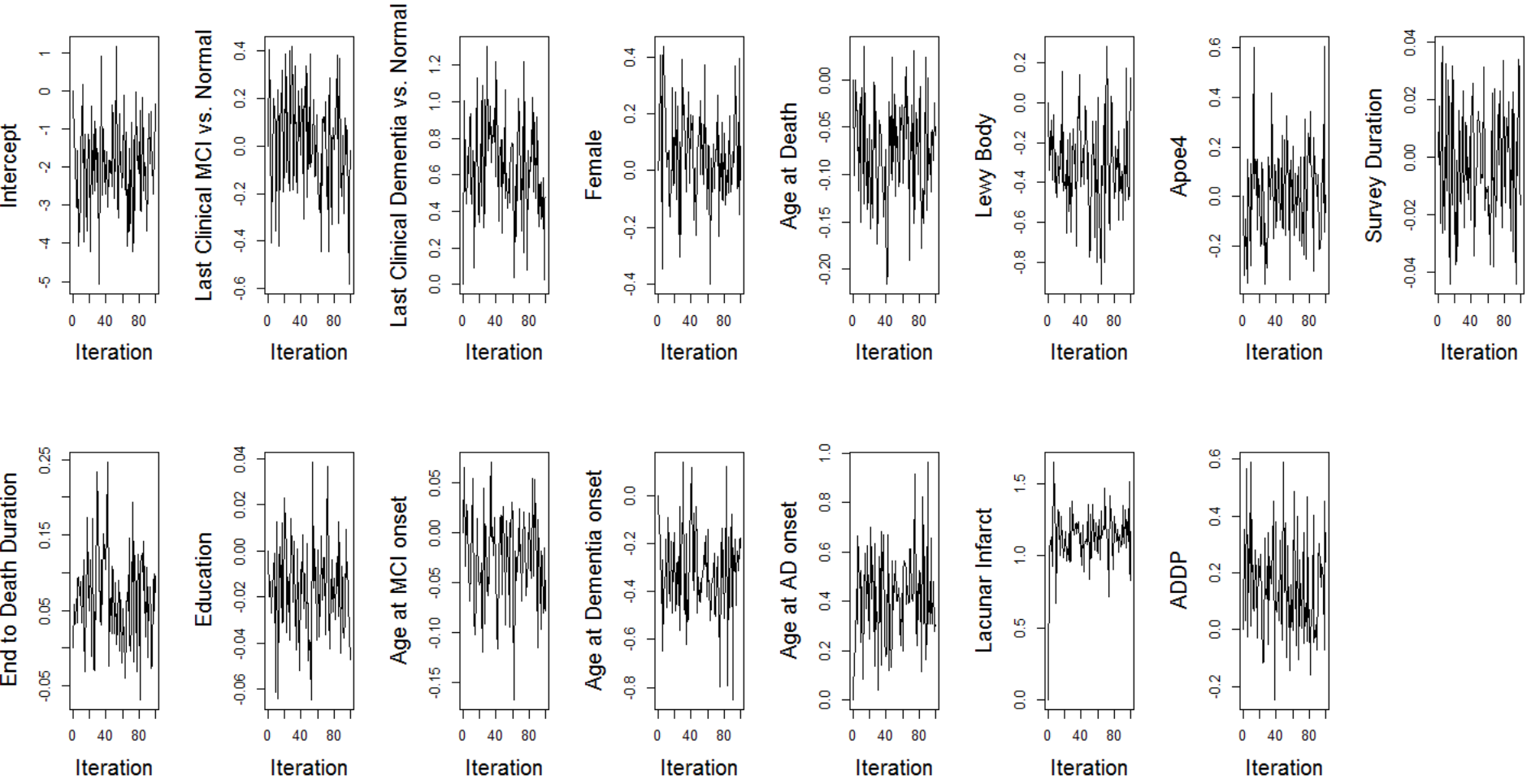
m2(Lacunar Infarct|updated predictors, θ2) is fit, and θ2(t) is drawn from its approximate posterior distribution

Missing values are drawn from m2(Lacunar Infarct|updated predictors, θ2(t))

Monitoring Algorithm Convergence

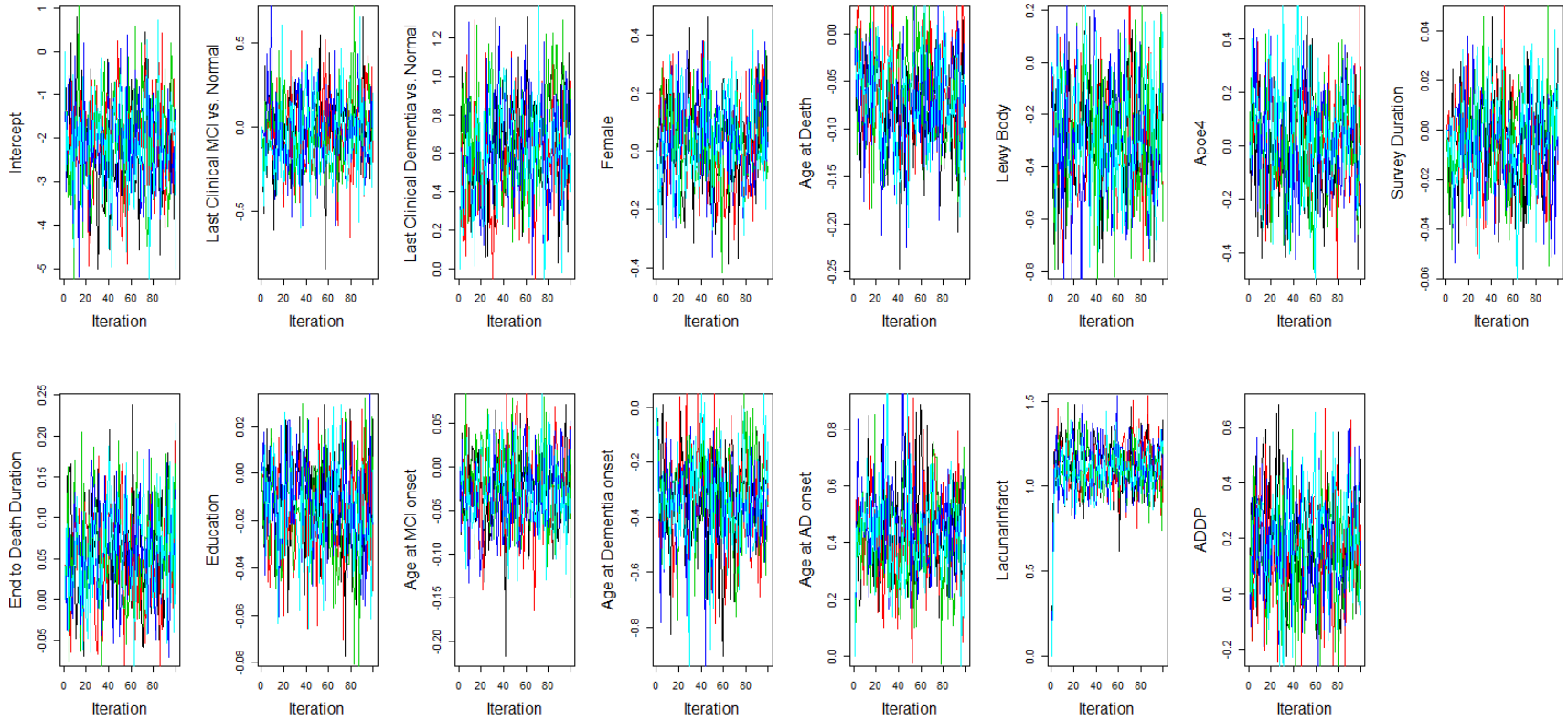
- Parameters from M=5 replicates should all reach the same convergence point

NPLINF_reg=glm(high_NPLINF~last_group+female+Age_at_Death+LEWY+apoe4+survey_duration+end_death_duration+EDUC+MCI_onset2+dementia_onset+high_NPLAC+new_ad_path,data=imp1[m_high_NPLINF==0,],family=binomial(),x=TRUE)

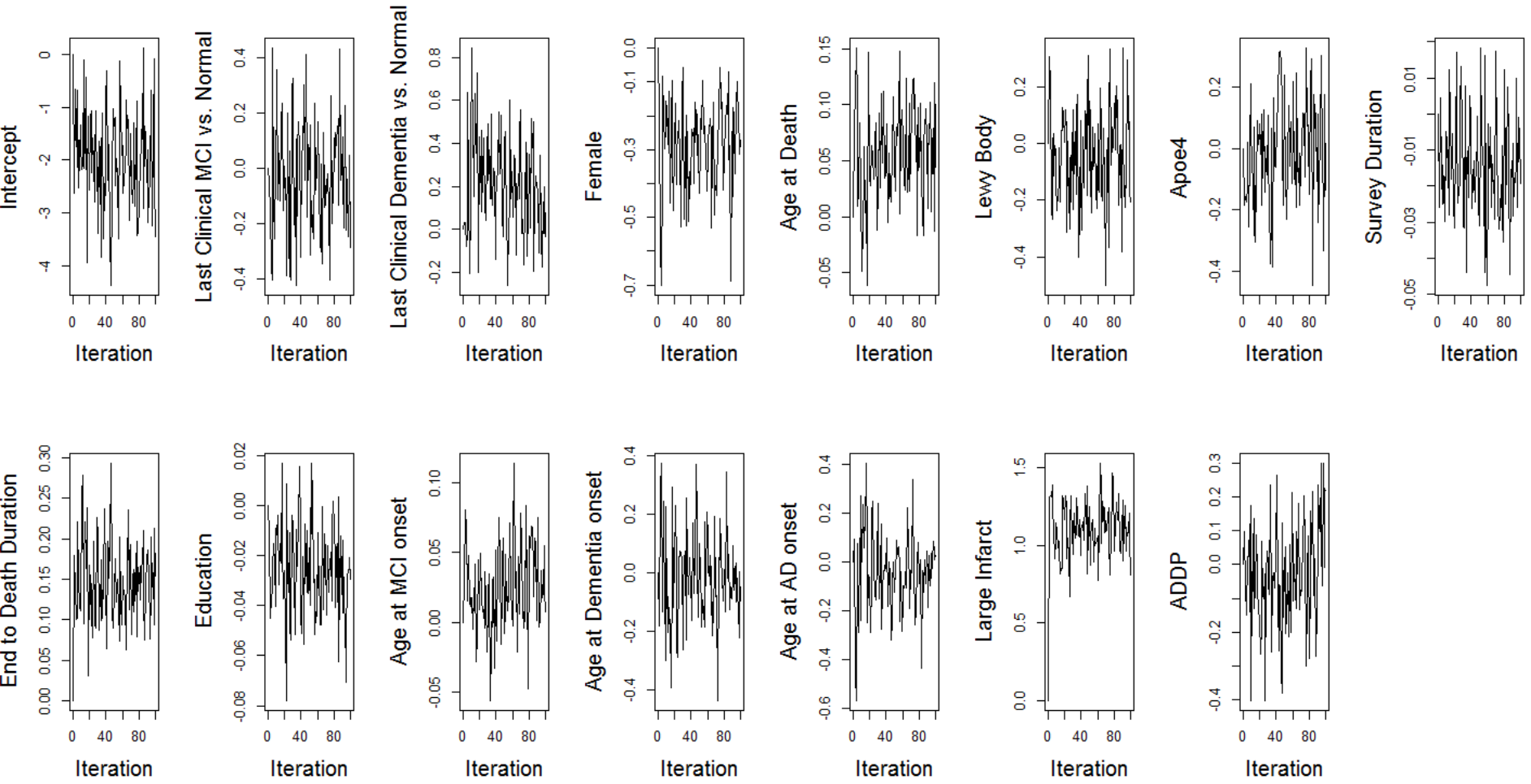


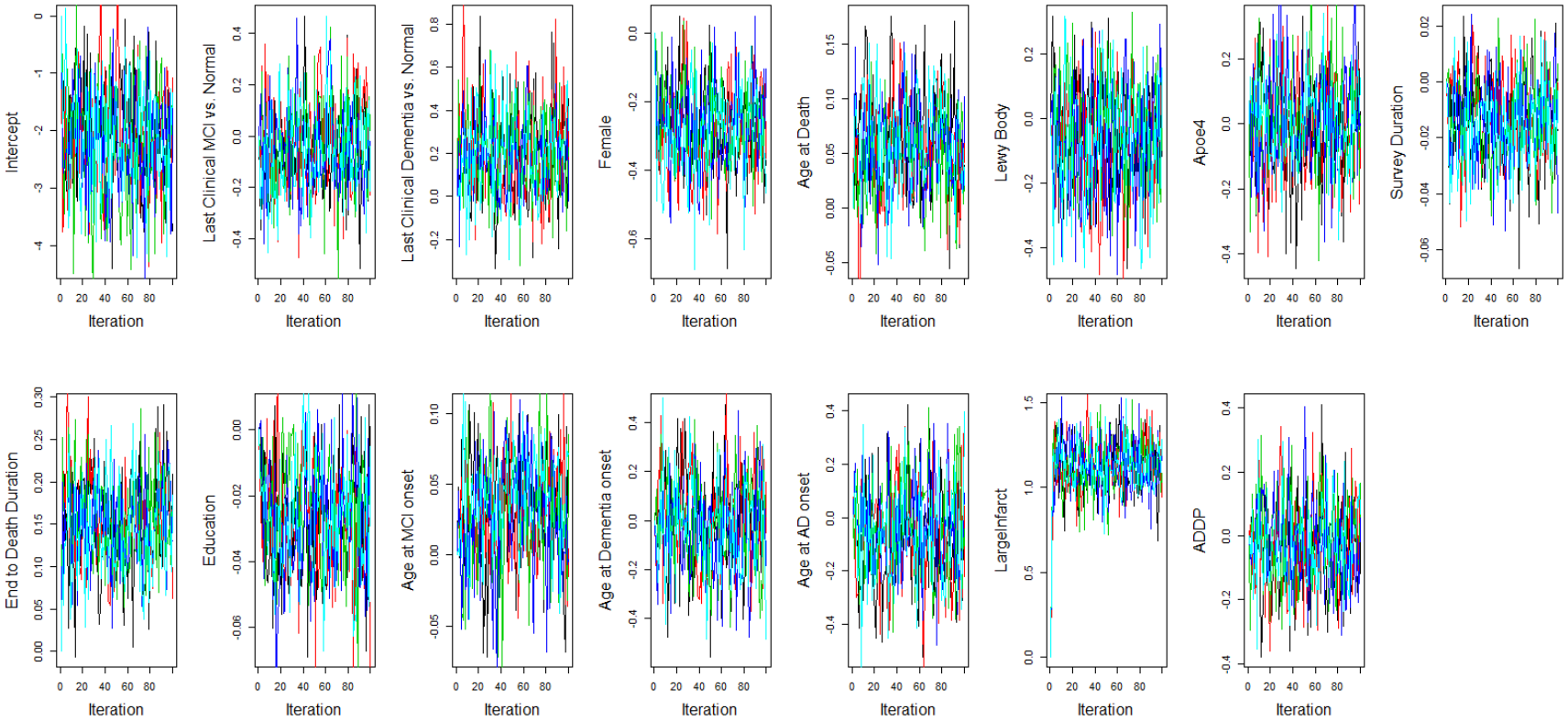
last_group: 1-normal, 2-MCI and 3-dementia; so two parameters 2vs1 and 3vs1 were estimated

100 iterations X 5 times Overlaid

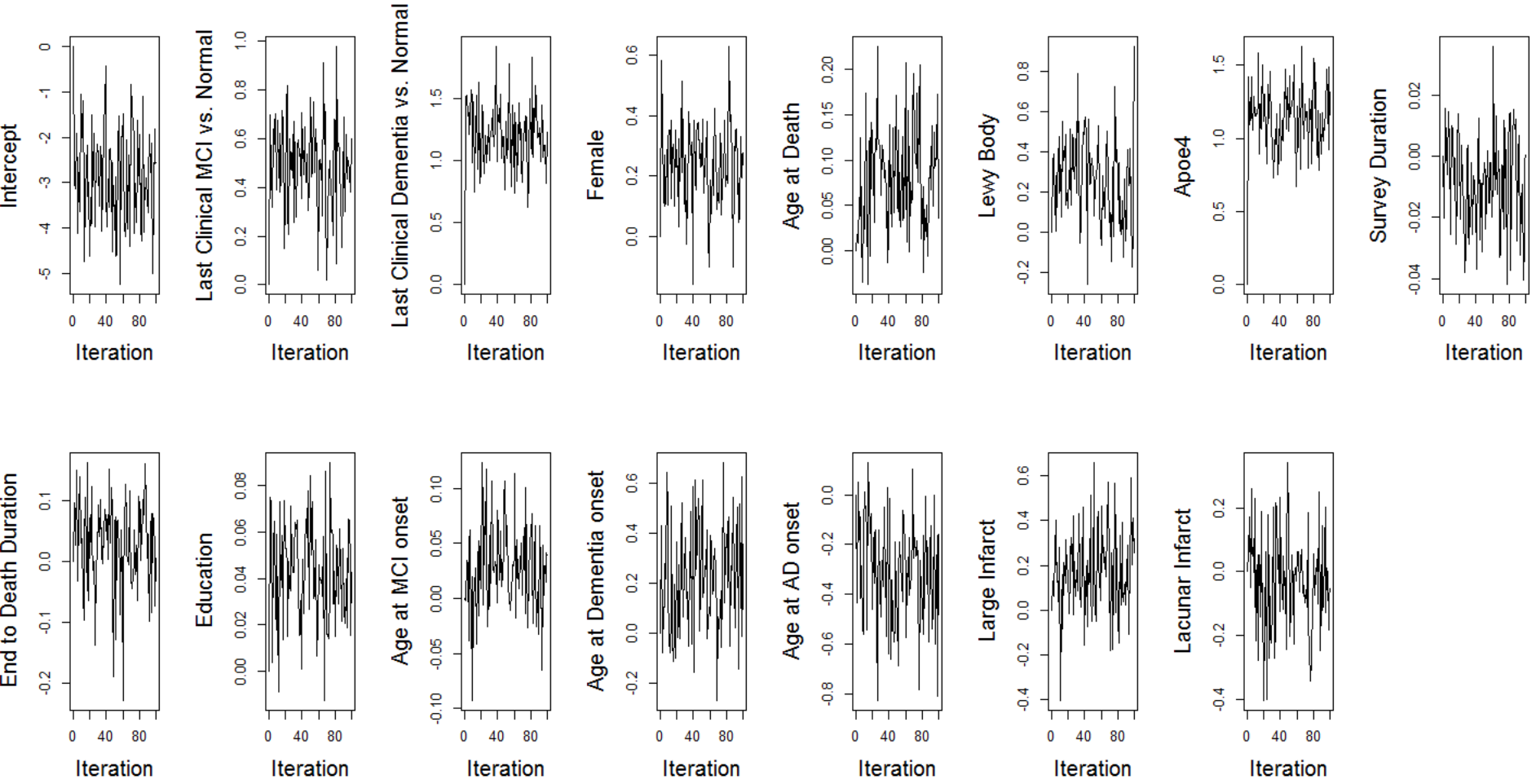


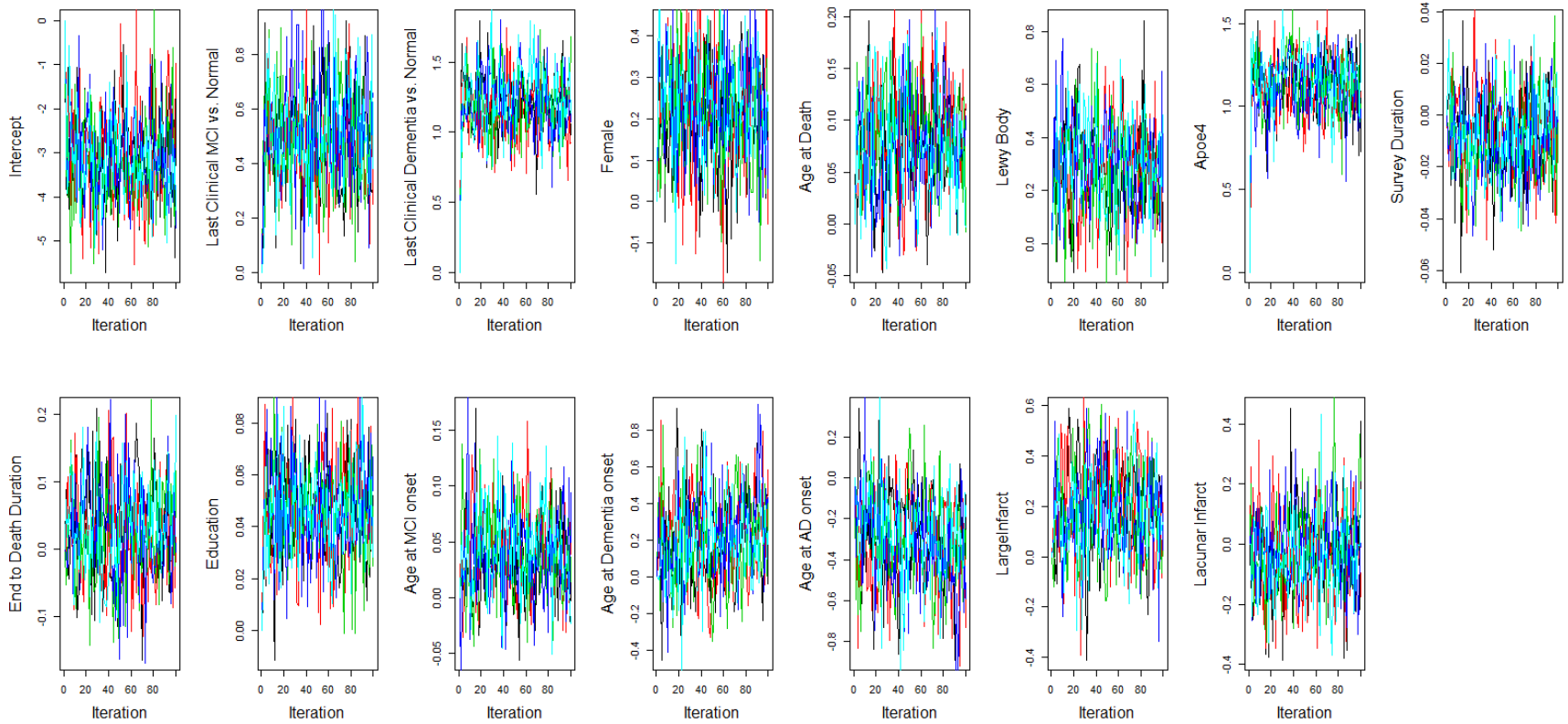
NPLAC_reg=glm(high_NPLAC~last_group+female+Age_at_Death+LEWY+apoe4+survey_duration+end_death_duration+EDUC+MCI_onset2+dementia_onset+high_NPLINF+new_ad_path,data=imp1[m_high_NPLAC==0,],family=binomial(),x=TRUE)



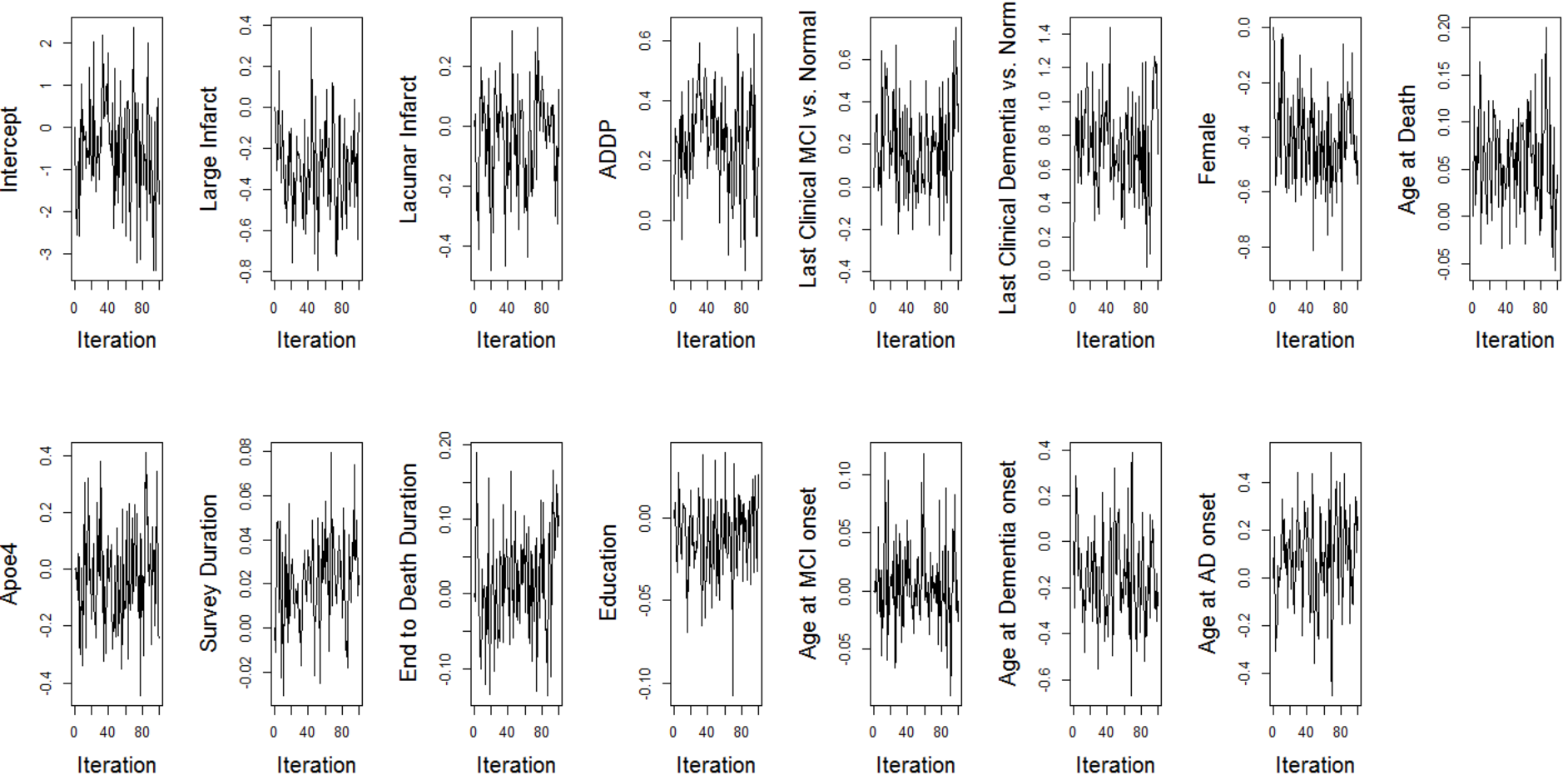

Intercept
Last Clinical MCI vs. Normal
Last Clinical Dementia vs. Normal
Female
Age at Death
Lewy Body
Apoe4
Survey Duration
End to Death Duration
Education
Age at MCI onset
Age at Dementia onset
Age at AD onset
LargeInfarct
ADDP
Iteration

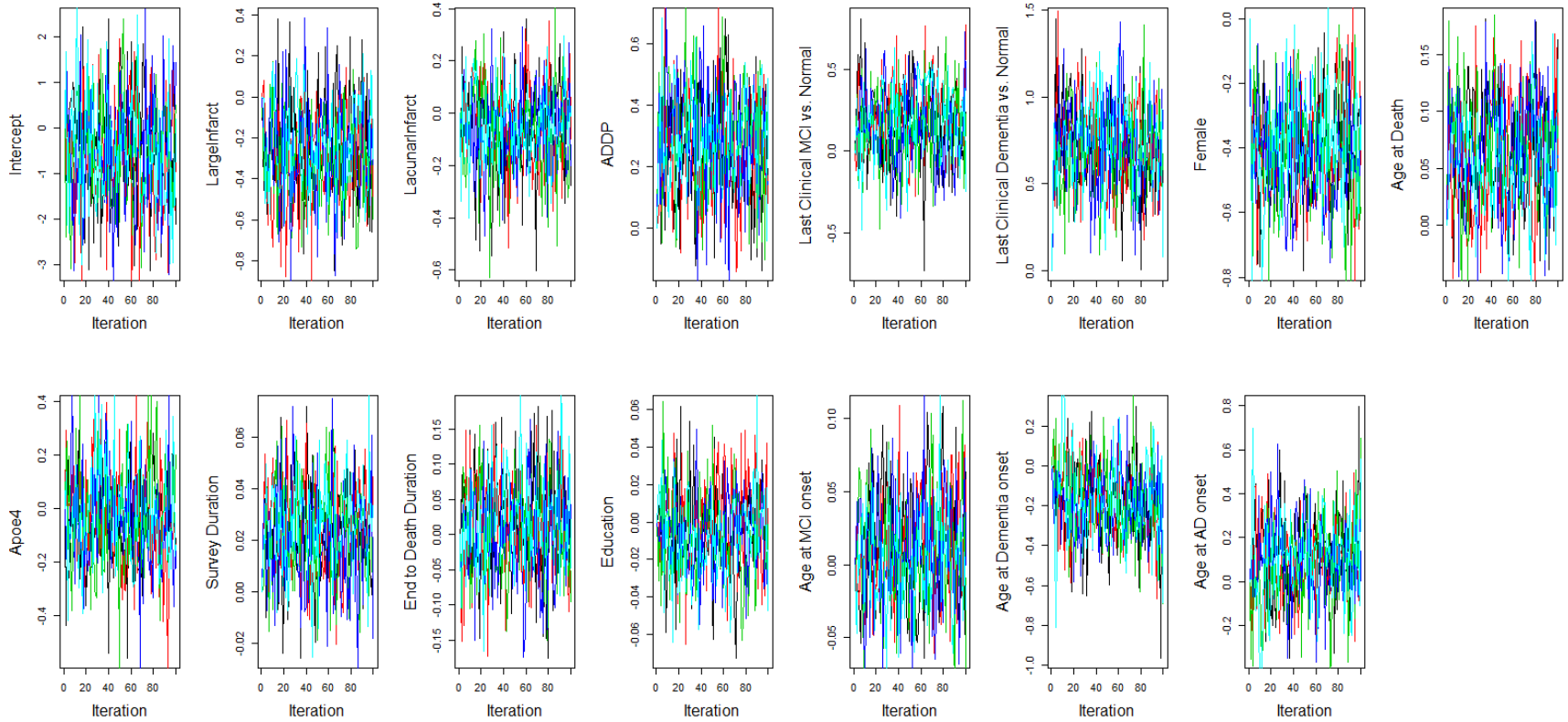
ad_reg=glm(new_ad_path~last_group+female+Age_at_Death+LEWY+apoe4+survey_duration+end_death_duration+EDUC+MCI_onset2+dementia_onset+high_NPLINF+high_NPLAC,data=imp1[m_new_ad_path==0,],family=binomial(),x=TRUE)



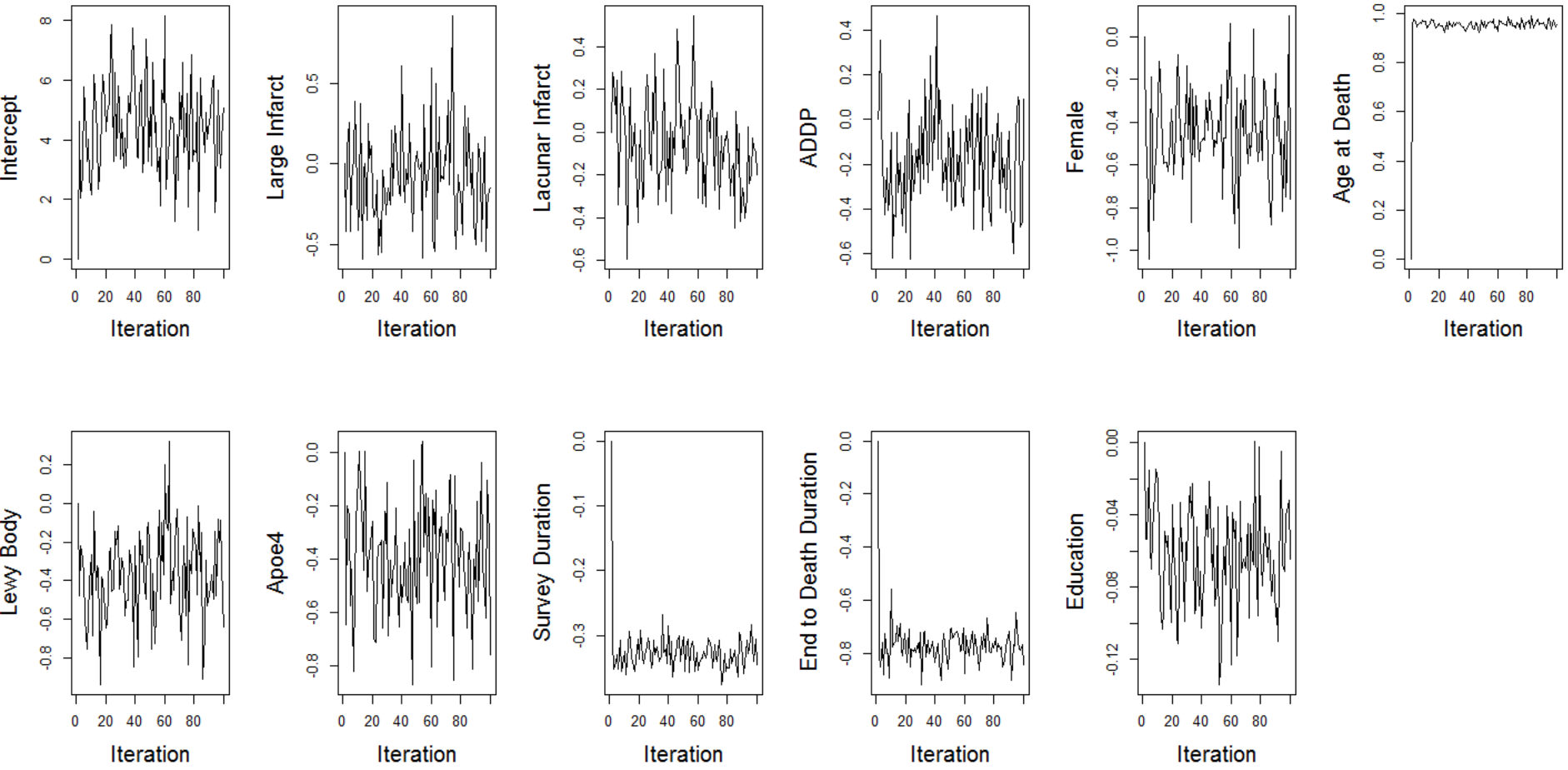

Intercept
Last Clinical MCI vs. Normal
Last Clinical Dementia vs. Normal
Female
Age at Death
Lewy Body
Apoe4
Survey Duration
End to Death Duration
Education
Age at MCI onset
Age at Dementia onset
Age at AD onset
LargeInfarct
Lacunar Infarct
Iteration

LEWY_reg=glm(LEWY~high_NPLINF+high_NPLAC+new_ad_path+
last_group+female+Age_at_Death+apoe4+survey_duration+end_death_duration+EDUC+
MCI_onset2+dementia_onset,data=imp1[m_LEWY==0,],family=binomial(),x=TRUE)

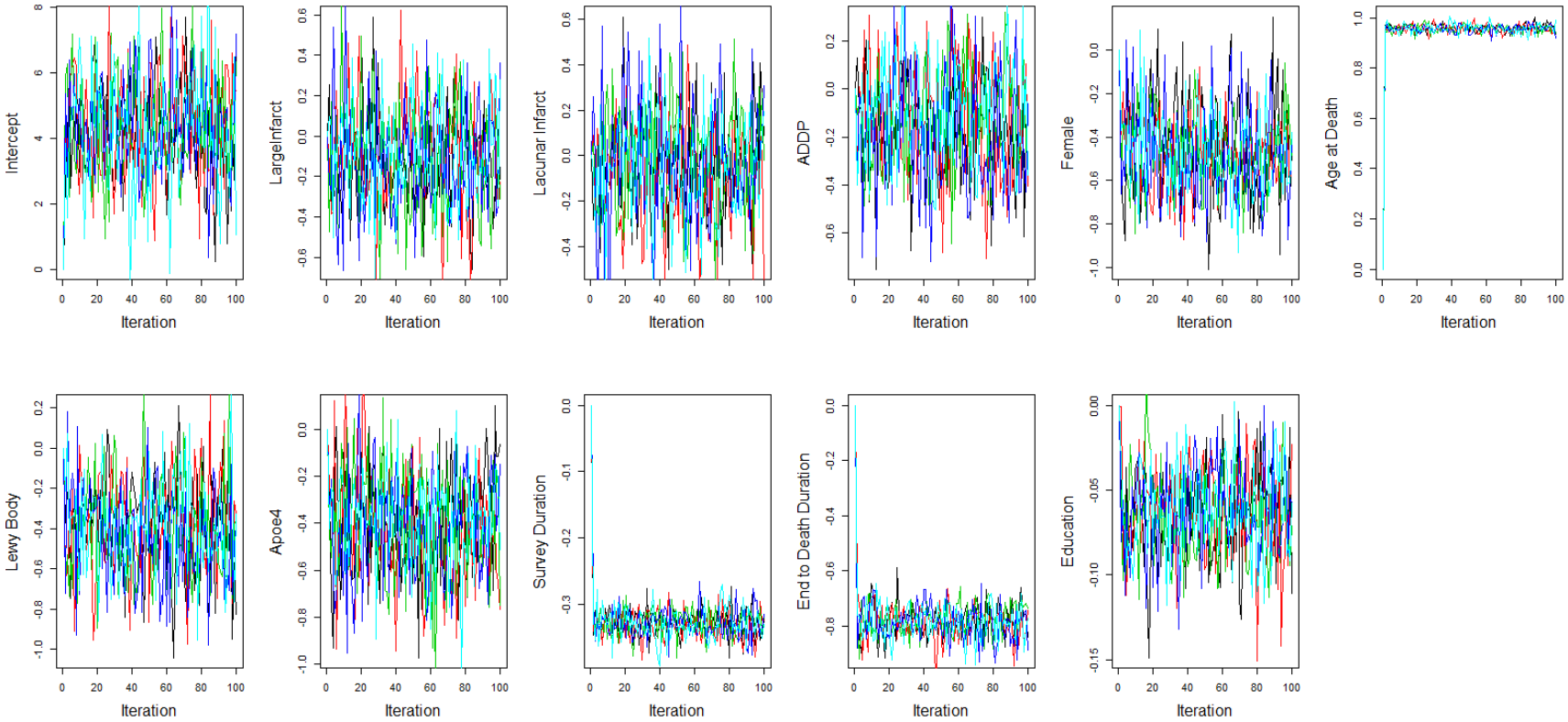



Intercept
LargeInfarct
LacunarInfarct
ADDP
Last Clinical MCI vs. Normal
Last Clinical Dementia vs. Normal
Female
Age at Death
Apoe4
Survey Duration
End to Death Duration
Education
Age at MCI onset
Age at Dementia onset
Age at AD onset
Iteration

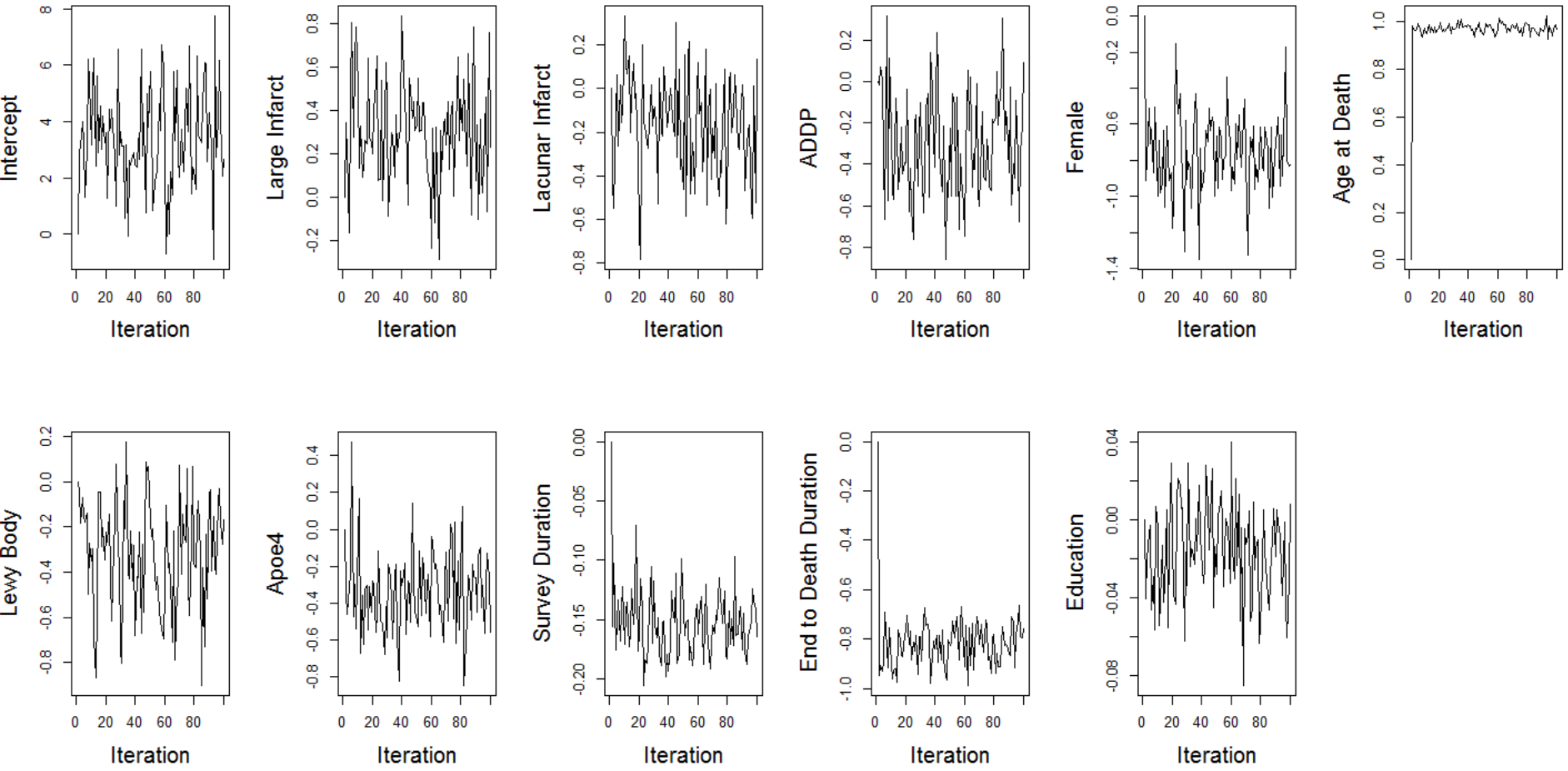
MCI_onset2_reg=lm(MCI_onset2~
high_NPLINF+high_NPLAC+new_ad_path+female+Age_at_Death+LEWY+apoe4+survey_duration+end_death_duration+EDUC,data=imp1[m_MCI_onset2==0,],x=TRUE)



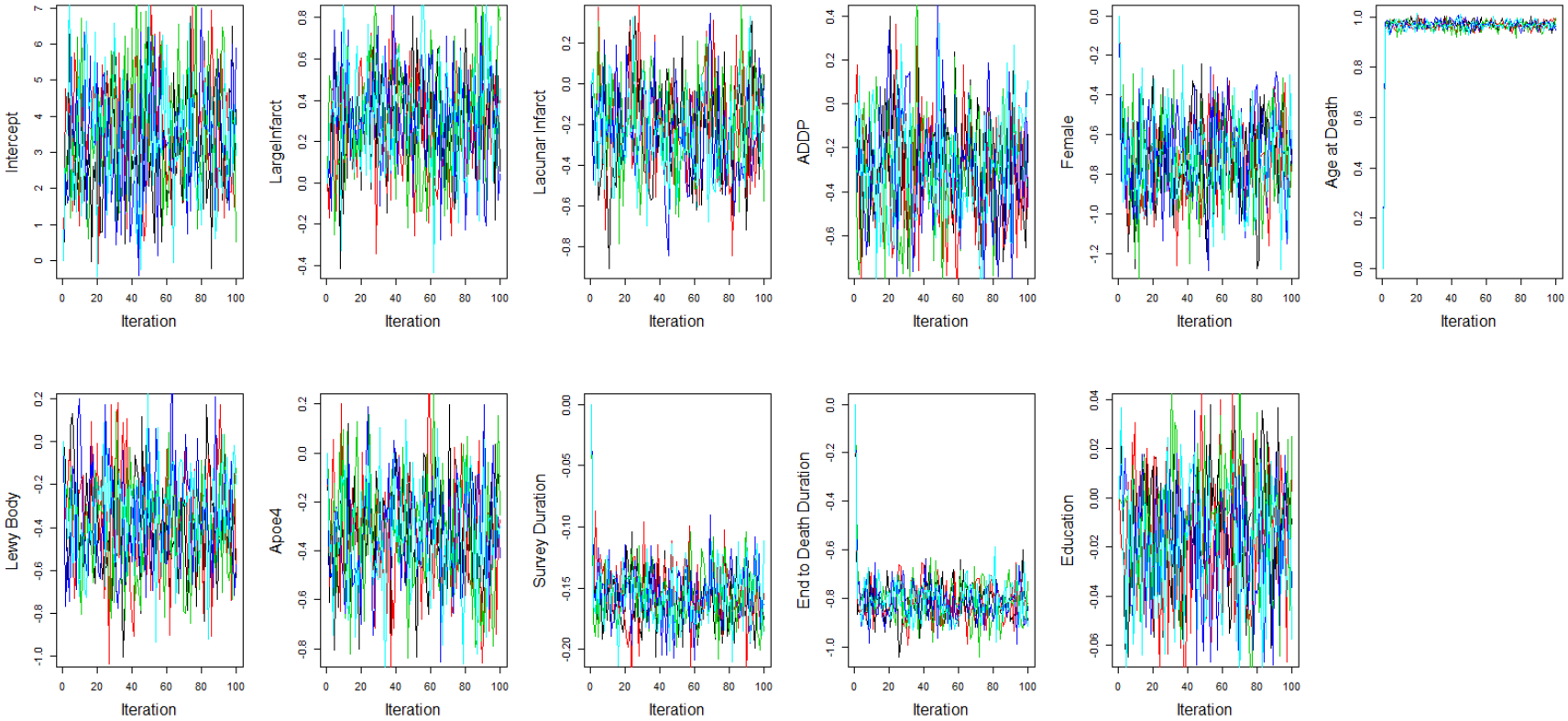
Imputed MCI_onset2 is drawn from the approximate posterior distribution truncated above last_normal so that the imputed value is meaningful


Intercept
LargeInfarct
Lacunar Infarct
ADDP
Female
Age at Death
Lewy Body
Apoe4
Survey Duration
End to Death Duration
Education
Iteration

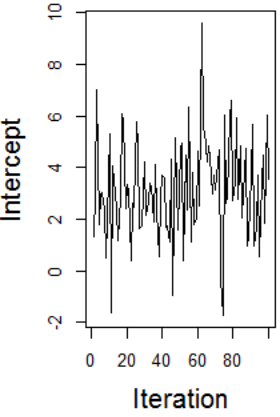
dementia_onset_reg=lm(dementia_onset~
high_NPLINF+high_NPLAC+new_ad_path+female+Age_at_Death+LEWY+apoe4+survey_durat
ion+end_death_duration+EDUC,data=imp1[m_dementia_onset==0,],x=TRUE)

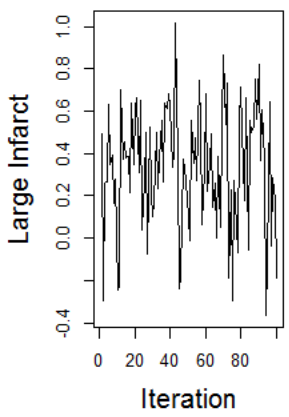


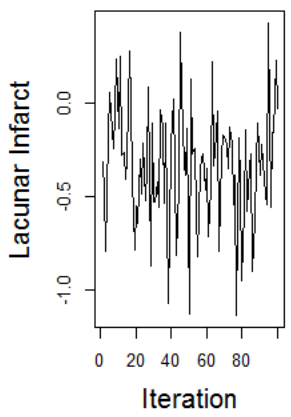
Imputed dementia_onset is drawn from the approximate posterior distribution truncated above MCI_onset2 so that the imputed value is meaningful

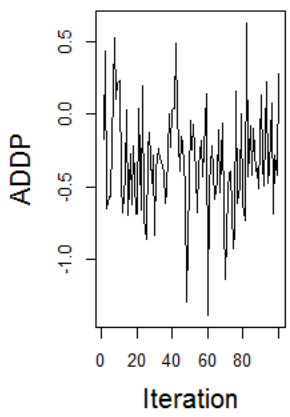

Intercept
LargeInfarct
Lacunar Infarct
ADDP
Female
Age at Death
Lewy Body
Apoe4
Survey Duration
End to Death Duration
Education
Iteration

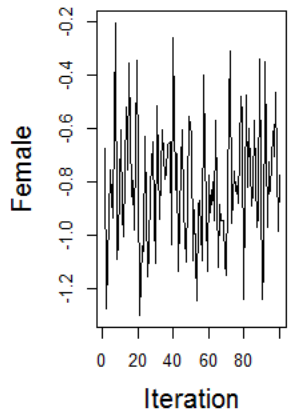
ad_onset_reg=lm(ad_onset~
high_NPLINF+high_NPLAC+new_ad_path+female+Age_at_Death+LEWY+apoe4+surve
y_duration+end_death_duration+EDUC,data=imp1[m_ad_onset==0 ,],x=TRUE)

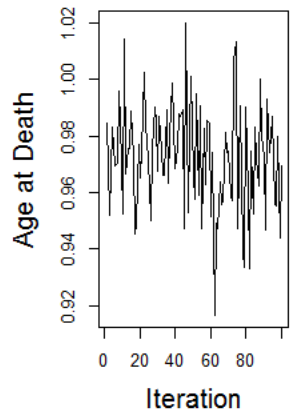


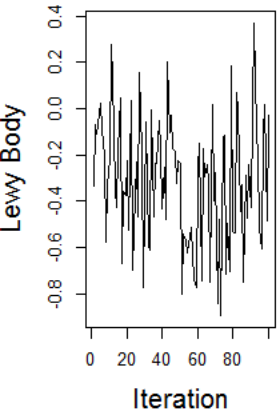


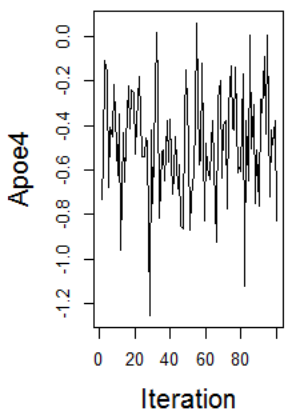


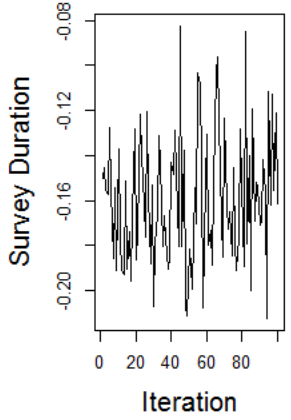


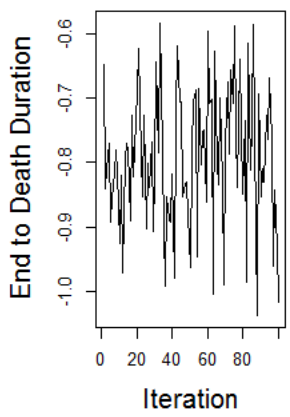


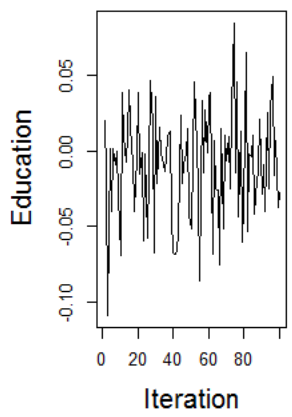


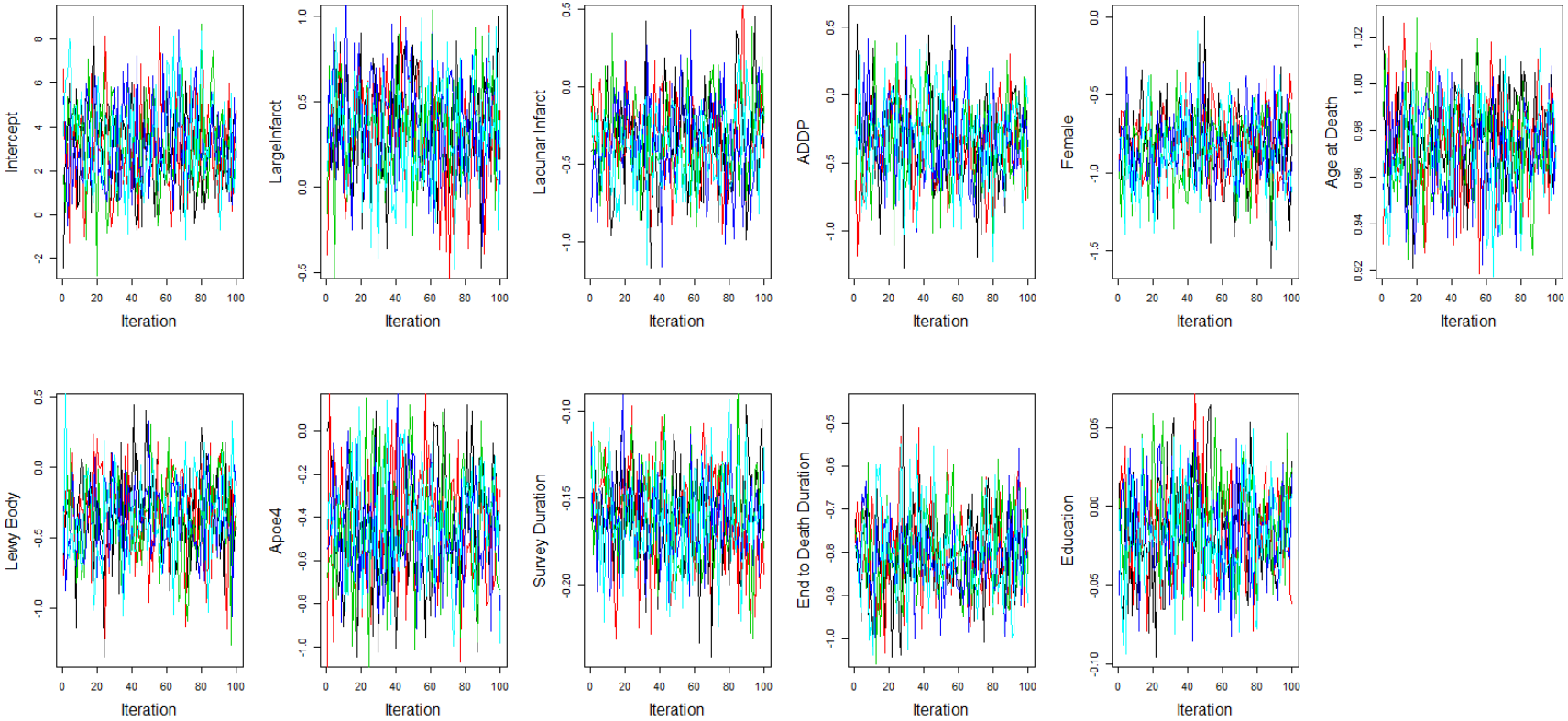




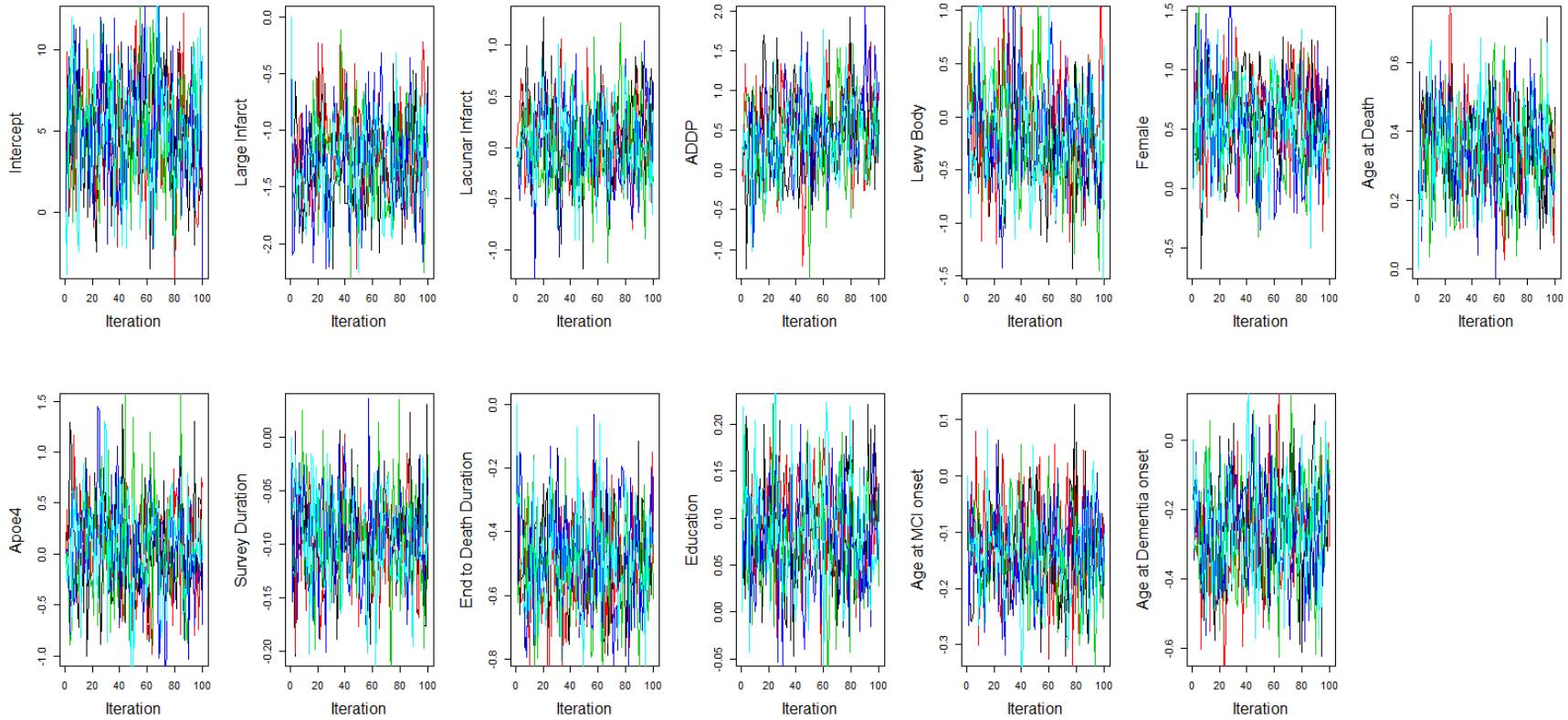







Intercept
LargeInfarct
Lacunar Infarct
ADDP
Female
Age at Death
Lewy Body
Apoe4
Survey Duration
End to Death Duration
Education
Iteration

POSSPROB_AD_reg=glm(POSSPROB_AD~high_NPLINF+high_NPLAC+new_ad_path+LEWY+female+Age_at_Death+apoe4+survey_duration+end_death_duration+EDUC+MCI_onset2+dementia_onset,data=imp1[m_POSSPROB_AD==0,],family=binomial(),x=TRUE)



Results of cox proportional hazard models (outcome=prob/poss AD)

Parameter	Hazard Ratio (1) N=1054	Hazard ratio (2) N=1566
No ADDP, no Lacunes, no Large Infarcts	Reference Group	
AD Definitive Pathology (ADDP)	4.01**	3.60**
Lacunes without ADDP	2.24*	1.93*
ADDP + Lacunes	3.50**	3.01**
Large infarct(s) without ADDP	1.83*	1.81*
ADDP with large infarct(s)	4.93**	3.96**

(1) using only observed autopsy data (2) using observed and imputed data. Models controls for sex, education, Apoe 4 and Lewy body pathology. **: $p < 0.0001$, *: $p<0.01$

Conclusions

- Strong association between missing pathology data and observed variables → MAR assumption (or informal dropout—this cannot be proved statistically)
- By using imputed pathology data, the association between pathology types and incidence of AD gets weaker for all pathology types (especially for ADDP)
- Possibly because those who go through autopsy are more likely to be those with diseases (AD or AD with other diseases)---using observed autopsy cases might be overestimating the risk of AD in relation with pathology types. (Some individuals who remained cognitively intact might be less likely to do autopsy, given the same pathology types)

Special Thanks to:

Michigan Alzheimer's Disease Center , University of Michigan, Ann Arbor, MI

Hank Paulson, MD Jian Zhu

NIA-Layton Aging and Alzheimer's Disease Center, Oregon Health & Science University, Portland, OR

Robin Guariglia, Jeff Kaye, MD

Sanders-Brown Center on Aging, University of Kentucky, Lewisville, KY.

Richard Kryscio, PhD

Frederick Schmitt, PhD

Erin Abner, PhD , Wenjie Lou, PhD

NIH funding: SMART project: R01AG386561, Oregon ADC: P30AG008017 , Kentucky ADC: P30AG028383