Analysis of the effects of several risk factors on cognitive outcomes and the interactions between age and the other risk factors

Haoran Hu, Department of Biostatistics, Columbia University

Zhezhen Jin, Department of Biostatistics, Columbia University

Gladys E Maestre, Department of Biomedical Sciences, Division of Neurosciences, University of Texas at Rio Grande Valley

Joseph H. Lee, G.H. Sergievsky Center, Columbia University

Ⅰ.Overview and Student Role

I worked as a research assistant with Dr. Zhezhen Jin. We were primarily interested in investigating the effects of the risk factors on the cognitive outcome levels as well as the interactions between aging and the other risk factors using data from Maracaibo Aging Study (MAS).[1] My role in the research included following tasks:

- Processed data for modeling, along with exploratory analysis
- Fitted linear or generalized linear models with Generalized Estimating Equations (GEE) and included risk factor-by-time interaction terms to evaluate associations stated above
- Checked linearity assumption by plotting marginal residuals vs. explanatory variables
- Assessed the effects of missing data by multiple imputation and pooled analysis

Ⅱ.Backgrounds

More than half of all people over 55 years lived in the less developed countries[1]. In 2000, approximately 11% of the population in this age group in the developing countries resides in Latin America. However, there were no registries for common disease for elderly people such as dementia in Latin America. Hence, there were not enough data to study dementia at population level. Under this situation, researchers conducted the Maracaibo Aging Study (MAS) to fill this gap.[1] The MAS is a longitudinal study with a focus on memory-related diseases, among people over 55 years living in the city of Maracaibo, Venezuela. The data collection process can be divided into three phases. In the first phase, the researchers conducted a door-to-door survey to obtain sociodemographic characteristics; In the second phase, they collected data regarding changes in the subjects' abilities. The third phase consists of genetic analysis, neuropsychiatric examination, etc. In our analysis, we

used the data from MAS to explore the associations between six repeatedly measured cognitive outcomes and age, education level, gender, and APOE Ɛ4 genotype. We are primarily interested in investigating the effects of the risk factors on the cognitive outcome levels as well as the interactions between aging and the other risk factors. The results of our analysis should be representative for elderly people in the city of Maracaibo. Then, we compared our results to the associations between dementia and these risk factors established in previous studies. [2,3,4]

III. Methods

3.1 Data description

The study sample consists of 2452 subjects who took one or multiple cognitive assessments. The number of cognitive assessments that a subject went through ranged from 1 to 5 (median, 2).Time intervals between the cognitive assessments were irregular, ranging from 10 days to 11 years (mean[SD], 3.51[2.21] years). Each assessment included tests on six cognitive outcomes, which were EMEMs, EMEMf, total memory, long term memory(LTM), short term memory (STM), and recognition memory. The first assessment of the first subject was in 1991, and the last assessment of the last subject was in 2010. Risk factors of interest include age, gender, education, and APOE ε4 genotype. The subjects' gender, years of education, APOE ε4 genotype were recorded at baseline, and their age was recorded at each assessment. Age and years of education were considered as continuous variables, and gender and APOE ε4 genotype were considered as binary variables.

3.2 Cognitive outcomes

EMEMs, EMEMf, total memory, and LTM are continuous outcomes. STM (range, 0 - 11) and recognition memory (range, 0 - 12) were recorded in ordinal scales, but I converted them into binary outcomes. STM levels and recognition memory levels that were higher than or equal to their corresponding medians (5 for STM; 11 for recognition memory) were recoded as "high-performance level," while STM levels and recognition memory levels lower than their corresponding medians were recoded as "low-performance level."

3.3 Analytic methods

We are primarily interested in investigating the effects of the risk factors on cognitive outcome levels as well as the interactions between aging and the other risk factors. Since there were repeated measurements on the subjects, I fitted linear models or generalized linear models with generalized estimating equations (GEE) method to analyze the effects of the risk factors on each of the outcomes separately, assuming working independence correlation structures. Models included covariates for age at assessment, gender, years of education, APOE ε4 genotype, and interaction terms between age at assessment and the other risk factors. For the convenience of interpretation, I centered years of education and age at assessment at their means. For EMEMs, EMEMf, total memory, and LTM, which were continuous outcomes, I fitted the models with identity link functions. For STM and recognition memory, which were binary outcomes, I fitted the models with logit link functions. Then, I used backward elimination to exclude the insignificant interaction terms under significance level 0.05. The main effects reflect the average effects of the risk factors on the cognitive outcomes at mean age. The interaction effects reflect if and how the risk factors other than age affect the effects of aging on the cognitive outcomes. To deal with missing data, I first applied listwise deletion and did complete case analysis. I also used Little's missing completely at random (MCAR) test to test the missing data mechanism, and then used multiple imputation and pooled analysis to evaluate the impact of missing data. The data in this study are multilevel data, with missingness in one level-2 predictor, i.e., APOE ε4 genotype, and in all the six level-1 outcomes, including EMEMs, EMEMf, total memory, LTM, STM, and recognition memory. According to some literature, missingness in level-2 predictors are typically fixed by deleting all records in the cluster[5]. Therefore, I applied listwise deletion to observations with missing APOE ε4 genotype, and used lineal mixed model with probability mean matching to impute missing level-1 outcomes.

Ⅳ. Results

Data on age, gender, and education are complete for all the 2452 participants, while APOE ε4 genotypes of 306 participants are missing. Baseline demographic statistics are given in Table 1.

Table 1

Baseline demographic statistics

Number of subjects: 2452

Baseline age		**Gender**		**Education**		**APOE ε4**	
Mean (SD)	67.41 (9.02)	Men	807 (32.9%)	Mean (SD)	5.88 (4.19)	N-Miss	306
Range	54 - 101	Women	1645 (67.1%)	Range	0 - 22	0	1691 (78.8%)
						1	455 (21.2%)

Before generating descriptive statistics for the outcomes, I applied listwise deletion to observations with missing APOE ε4 genotype. The descriptive statistics for the outcomes at each measurement are displayed in Table 2.

Table 2

Descriptive statistics of the outcomes

Measurement	1st (N =2452)	2nd (N = 1830)	3rd (N = 621)	4th (N = 100)	5th (N = 11)	Total(N=5014)
Time from baseline(years)						
Mean(SD)	0.00(0.00)	3.38(2.11)	6.34(2.77)	8.59(2.31)	9.85(1.24)	N/A
Range	0.00-0.00	0.003-11.27	1.20-12.49	2.68-11.80	7.39-11.06	N/A
EMEMs						
N-Miss	17	71	15	5	1	109
Mean(SD)	41.78(10.00)	42.04(10.24)	40.10(10.26)	36.34(10.85)	30.00(6.65)	41.54(10.18)
Range	0.00-57.00	0.00-58.00	0.00-57.00	0.00-55.00	17.00-39.00	0.00-58.00
EMEMf						
N-Miss	36	62	22	10	1	131
Mean(SD)	22.87(5.47)	22.93(5.48)	22.21(6.27)	20.59(7.78)	18.00(5.25)	22.76(5.64)
Range	0.00-72.00	0.00-70.00	0.00-81.00	0.00-71.00	5.00-23.00	0.00-81.00
Total memory						
N-Miss	180	136	87	22	3	428
Mean(SD)	35.74(10.52)	35.55(10.11)	34.11(9.79)	30.90(12.71)	30.63(8.16)	35.39(10.35)
Range	0.00-69.00	0.00-61.00	0.00-61.00	0.00-62.00	0.00-43.00	0.00-69.00
LTM						
N-Miss	181	136	89	22	3	431
Mean(SD)	22.19(12.11)	22.12(11.35)	20.51(11.16)	17.83(12.88)	16.25(11.79)	21.88(11.76)
Range	0.00-67.00	0.00-60.00	0.00-56.00	0.00-56.00	0.00-35.00	0.00-67.00
STM						
N-Miss	188	153	91	25	3	460
low	1355(59.8%)	1026(61.2%)	343(64.7%)	59(78.7%)	6(75.0%)	2789(61.2%)
high	909(40.2)	651(38.8%)	187(35.3%)	16(21.3%)	2(25.0%)	1765(38.8%)
Recognition Memory						
N-Miss	192	159	94	26	3	474
low	1410(62.4%)	1007(60.3%)	397(75.3%)	62(83.8%)	8(100%)	2884(63.5%)
high	850(37.6%)	664(39.7%)	130(24.7%)	12(16.2%)	0(0.00%)	1056(36.5%)

For each of the outcomes, I first excluded observations with the outcome missing, and then fitted a linear model or a generalized linear model with GEE method. For continuous outcomes, including EMEMs, EMEMf, total memory, and long term memory, I fitted linear models with identity links. The models I built are shown in the appendix, and the results are displayed in Table 3. Coefficients for the main effects shows the average effects of the risk factors on the cognitive outcomes at mean age, and all of them are significantly different from zero for all the outcomes. Coefficients for age and APOE $\varepsilon 4$ genotype are negative for all the cognitive outcomes, suggesting the cognitive outcome levels decrease with age, and the presence of APOE $\varepsilon 4$ genotype is associated with lower cognitive outcome levels. Coefficients for education are positive for all the cognitive outcomes, suggesting higher education levels are associated with higher cognitive outcome levels. Females have lower EMEMs and EMEMf levels compared to males, while they have higher total memory levels and long term memory levels than males. Coefficients for the interaction effects reflect if and how the risk factors other than age affect the effects of aging on the cognitive outcomes. The results suggest that higher education level is associated with lower rate of decline with age in EMEMs and EMEMf. However, higher education level is associated with higher rate of decline with age in long term memory.

For binary outcomes, including short term memory and recognition memory, I fitted generalized linear models with logit links. The results are displayed in Table 4. All the main effects are significant for both the outcomes. Specifically, old age and the presence of APOE $\varepsilon 4$ genotype are associated with lower odds of high performance levels in the cognitive tests; female and people with higher education levels tend to have higher odds of high performance levels in the cognitive tests. Gender and education have significant interaction with age in the model for recognition memory. The results suggest that higher education level and female gender are associated with larger harmful effects of age on recognition memory.

Table 3

Differences in average cognitive test scores and average annual rate of change in the scoresassociated with the given difference in risk-factor level*

	EMEMs		EMEMf		Total Memory		LTM	
Risk factor	Difference in average test scores for people with average age (se)	P-value	Difference in average test scores for people with average age (se)	P-value	Difference in average test scores for people with average age (se)	P-value	Difference in average test scores for people with average age (se)	P-value
Age	-0.42 (0.023)	<0.001	-0.22(0.013)	<0.001	-0.45 (0.022)	<0.001	-0.46 (0.025)	<0.001
Sex(female)	-2.45 (0.347)	<0.001	-0.97 (0.207)	<0.001	2.23 (0.394)	<0.001	2.87 (0.468)	<0.001
Education	0.97 (0.046)	<0.001	0.45 (0.026)	<0.001	0.66 (0.047)	<0.001	0.56 (0.057)	<0.001
APOE ε4	-1.87 (0.417)	<0.001	-0.94 (0.225)	<0.001	-1.97 (0.440)	<0.001	-2.00 (0.519)	<0.001
Risk factor	Difference in average annual rate of change in the test scores (se)	P-value	Difference in average annual rate of change in the test scores (se)	P-value	Difference in average annual rate of change in the test scores (se)	P-value	Difference in average annual rate of change in the test scores (se)	P-value
Sex(female)		ns		ns		ns		ns
Education	0.02 (0.005)	<0.001	0.01 (0.003)	<0.001		ns	-0.01 (0.006)	0.027
APOE ε4		ns		ns		ns		ns

*For continuous risk factors, the given difference is one-unit increase in their values; for sex, the given difference represents females vs. males; for APOE ε4, the given difference represents presence vs. non-presence of the APOE ε4 genotype

Table 4

Associations between risk factors and cognitive test performance and interactions between age and the other risk factors

	STM		Recognition Memory	
Risk factor	Odds ratio of high performance in cognitive tests associated with the given difference in risk factors for people with mean age (se)*	P-value	Odds ratio of high performance in cognitive tests associated with the given difference in risk factors for people with mean age (se)	P-value
Age	0.93 (0.005)	<0.001	0.95 (0.008)	<0.001
Sex(female)	1.42 (0.130)	<0.001	1.18 (0.101)	0.055
Education	1.07 (0.011)	<0.001	1.12 (0.013)	<0.001
APOE ε4	0.76 (0.073)	0.004	0.81 (0.077)	0.025
Risk factor	With the given difference in risk factors, ratios of the odds ratios associated with one year increase in age (se)	P-value	With the given difference in risk factors, ratios of the odds ratios associated with one year increase in age (se)	P-value
Sex(female)		ns	0.98 (0.010)	0.016
Education		ns	1.00 (0.001)	0.012
APOE ε4		ns		ns

*For continuous risk factors, the given difference is one-unit increase in their values; for sex, the given difference represents females vs. males; for APOE ε4, the given difference represents presence vs. non-presence of the APOE ε4 genotype

I used Little's MCAR test to test if the missing data in the outcomes are missing completely at random. The p-value given by the test was close to 0, suggesting that the data are not missing completely at random. Therefore, I did multiple imputation, assuming the data are missing at random. I used "mice" package in R to impute missing data in all the outcomes. For short term memory and recognition memory, I did imputation based on their original scale, and then transformed them into binary variables using the same cut off values as used previously. In the multiple imputation process, I generated 5 imputed datasets, using 10 iterations for each imputed dataset. To impute each outcome, I used all the risk factors and all the other outcomes in the dataset as predictors, and I also included age × education and age × APOE ε4 genotype interaction terms in the imputation models. Then, I did pooled analysis on the 5 imputed data set based on Rubin's rules. In the pooled analysis, I fitted the same models as I did previously in the complete case analysis, and the results are given in Table 5 and Table 6.

These results show that the multiple imputation with pooled analysis gives similar results to the complete case analysis. Though there are slight differences in values and standard errors of the coefficients in the models, the significant predictors remained almost the same. The similarity in the analysis results is possibly due to insufficiency of predictive covariates in the imputation models[6]. The multiple imputation is based on the assumption of missing at random. If this assumption does not hold, the analysis based on multiple imputation will be problematic.

Table 5

Results from pooled analysis on the imputed datasets: Differences in average cognitive test scores and average annual rate of change in the scores associated with the given difference in risk-factor level*

	EMEMs		EMEMf		Total Memory		LTM	
Risk factor	Difference in average test scores for people with average age (se)	P-value	Difference in average test scores for people with average age (se)	P-value	Difference in average test scores for people with average age (se)	P-value	Difference in average test scores for people with average age (se)	P-value
Age	-0.42 (0.023)	<0.001	-0.22 (0.013)	<0.001	-0.49 (0.022)	<0.001	-0.49 (0.024)	<0.001
Sex(female)	-2.27 (0.363)	<0.001	-0.96 (0.210)	<0.001	2.17 (0.416)	<0.001	2.77 (0.470)	<0.001
Education	0.98 (0.049)	<0.001	0.46 (0.026)	<0.001	0.64 (0.050)	<0.001	0.54 (0.057)	<0.001
APOE ε4	-1.73 (0.415)	<0.001	-0.93 (0.227)	<0.001	-2.09 (0.472)	<0.001	-2.03 (0.515)	<0.001
Risk factor	Difference in average annual rate of change in the test scores (se)	P-value	Difference in average annual rate of change in the test scores (se)	P-value	Difference in average annual rate of change in the test scores (se)	P-value	Difference in average annual rate of change in the test scores (se)	P-value
Sex(female)		ns		ns		ns		ns
Education	0.02 (0.005)	<0.001	0.01 (0.003)	<0.001		ns	-0.01 (0.005)	0.008
APOE ε4		ns		ns		ns		ns

*For continuous risk factors, the given difference is one-unit increase in their values; for sex, the given difference represents females vs. males; for APOE ε4, the given difference represents presence vs. non-presence of the APOE ε4 genotype

Table 6

Results from pooled analysis on the imputed datasets: Associations between risk factors and cognitive test performance and interactions between age and the other risk factors

	STM		Recognition Memory	
	Odds ratio of high performance in cognitive tests associated with the given difference in risk factors for people with mean age (se)*	P-value	Odds ratio of high performance in cognitive tests associated with the given difference in risk factors for people with mean age (se)	P-value
Risk factor				
Age	0.93 (0.004)	<0.001	0.94 (0.008)	<0.001
Sex(female)	1.43 (0.128)	<0.001	1.20 (0.104)	0.035
Education	1.07 (0.012)	<0.001	1.12 (0.014)	<0.001
APOE ε4	0.73 (0.069)	0.001	0.79 (0.073)	0.010
Risk factor	With the given difference in risk factors, ratios of the odds ratios associated with one year increase in age (se)	P-value	With the given difference in risk factors, ratios of the odds ratios associated with one year increase in age (se)	P-value
Sex(female)		ns	0.98 (0.010)	0.027
Education		ns	1.00 (0.001)	0.034
APOE ε4		ns		ns

*For continuous risk factors, the given difference is one-unit increase in their values; for sex, the given difference represents females vs. males; for APOE ε4, the given difference represents presence vs. non-presence of the APOE ε4 genotype

Ⅴ. Conclusions/Discussion

Our results are consistent with previous work in general[4,7]. All the cognitive outcomes decline with age, and the presence of APOE ε4 genotype is associated with lower levels in all the cognitive outcomes. Education is a protective factor for all the outcomes. Female gender is associated with lower performance in EMEMs and EMEMf, but it is associated with higher performance in all the other outcomes.

Education is the only covariate that appear to significantly impact the effect of age on the outcomes. Higher education level can reduce the harmful effects of age on EMEMs and EMEMf, but can increase the harmful effects of age on long term memory and recognition memory. Though higher education level is associated with larger harmful effects of age on some of the memory outcomes, the magnitude of that association is much smaller compared to the direct associations between education and the outcomes. Therefore, education is still considered as a protective factor.

Though some previous work shows the presence of APOE ε4 genotype is associated with higher rates of decline in cognitive outcomes[4,7], that association is not seen in my result. One important fact is, there is some evidence showing that data on APOE ε4 status are missing not at random. I compared the variable values between the subgroup with APOE ε4 status recorded and the subgroup with APOE ε4 status missing. Kruskal Wallis tests were used to test the differences in continuous variables and Chi-square tests were used to test the differences in categorical variables between the two subgroups. The result shows that subjects in the subgroup with APOE ε4 status missing have significant better performance in almost all the cognitive tests compared to the subgroup with APOE ε4 status recorded. That's possibly because subjects that had already shown cognitive impairment at baseline were more likely to be tested on APOE ε4 status, as they are more likely to possess the APOE ε4 genotype. If that's the case, the proportion of the presence of APOE ε4 genotype would be lower in the subgroup with APOE ε4 status missing, compared to the subgroup with APOE ε4 status recorded. As a result of this, we might have underestimated the effects of APOE ε4 genotype on the cognitive outcomes.

Ⅵ. References

[1] Maestre GE, Pino-Ramírez G, Molero AE, Silva ER, Zambrano R, Falque L, Gamero MP, Sulbarán TA. The Maracaibo Aging Study: population and methodological issues. Neuroepidemiology. 2002; 21:194–201. doi: 10.1159/000059524

[2] Ngandu T., von Strauss E., Helkala EL., et al. Education and dementia: what lies behind the association?Neurology.2007;69:1442-1450.2007;69:1442–1450.

[3] Mielke M.M., Vemuri P., Rocca W.A. Clinical epidemiology of Alzheimer's disease: Assessing sex and gender differences.Clin. Epidemiol.2014;6:37–48. doi: 10.2147/CLEP.S37929.

[4] Knopman DS, Mosley TH, Catellier DJ, Coker LH; Atherosclerosis Risk in Communities Study Brain MRI Study (2009) Fourteen-year longitudinal study of vascular risk factors, APOE genotype, and cognition: The ARICMRI Study. Alzheimers Dement 5, 207–214.

[5] Stef van Buuren. Flexible Imputation of Missing Data, Second Edition: CRC Press; 2018

[6] Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. ClinEpidemiol. 2017;9:157–166. doi: 10.2147/CLEP.S129785.

[7] Blair CK, Folsom AR, Knopman DS, Bray MS, Mosley TH, Boerwinkle E, et al. Apolipoprotein E genotype and cognitive decline in a middle-aged cohort. Neurology 2005;64:268–76.