STATS 770 Report

Jinghong Zeng

Department of Statistics, University of Auckland

E-mail: jzen696@aucklanduni.ac.nz

Abstract

An analysis was conducted on a subset data of a government sponsored cohort study of adults aged 65 years and older, to explore the association between the cognitive function measurements, DSST and MMMSE, demographic, behavioral and biological factors and mortality, and also to estimate the causal effect of the cognitive function measurements on mortality. Exploratory quasi-poisson models showed that DSST and MMMSE measurements are both negatively associated with age. Male or nonwhite people tend to have lower DSST and MMMSE measurements. A small alcohol intake is positively associated with the two measurements. People with diabetes tend to have lower DSST and MMMSE measurements and among people without diabetes, the two measurements are negatively associated with glucose. Hypertension , peripheral artery and heart diseases are also negatively associated with the two measurements, and systolic and diastolic blood pressure have clear interaction on their association with the two measurements. Lung function indicated by FEV and FVC is positively associated with the two measurements. However, no clear association is found between the cognitive function measurements and smoking, kidney or liver function.

Ordinary Cox models and cross validation LASSO Cox models were used to explore the association between the cognitive functions and mortality as well as compare the predictability of the cognitive function and other factors on mortality. The cognitive function measurements are negatively associated with mortality. They and the other factors including age, sex, smoking, creatinine and FEV are both predictive of mortality and can improve the prediction in the presence of each other.

Ordinary Cox models and the instrumental variable approach were used to discover the causal direct effect of the cognitive function measurements on mortality and examine potential residual confounding. The estimated causal direct effect of the cognitive function measurements from the instrumental variable approach is smaller than that from the Cox models, and there is weak evidence to support that some unmeasured confounder such as unmeasured risk factor or disease exists on the relationship between the cognitive function measurements and mortality. Clearer evidence comes from MMMSE. The two approaches both support that good cognitive function decreases mortality causally in elderly people.

Introduction

A government sponsored cohort study of adults aged 65 years and older was conducted to observe the incidence of cardiovascular and cerebrovascular diseases in the elderly over an 11 year period, and to relate the incidence of those diseases to various risk factors measured in the population on a regular basis. In this study, elderly, generally healthy, adults were randomly selected from Medicare rolls. Agreement to participate was high, and thus the sample can be regarded as a fairly accurate representation of healthy older Americans. At the time of study enrollment, and on annual visits over the length of the study, the participants' data regarding various behavioral (e.g., smoking, alcohol consumption), functional (e.g., ability to perform routine tasks), and clinical (e.g., blood pressure, laboratory tests) measures are recorded.

This analysis was based on a representative subset of 3,660 participants and 26 variables from the study and aimed to explore the relationship between cognitive function and demographics, behavior, body function, mortality. Results from this subset can be reasonably generalized to the older American population.

Statistical Analysis

Statistical analyses were completed using R statistical software. In all analyses, the significance level was chosen as 0.05 and a $p < 0.05$ was considered statistically significant. Confidence intervals were all 95% intervals. Missing data from the variables `fev` and `fvc` was imputed by Bayesian linear regression (R `mice` package), because missingness in the two variables was about 10%, and much greater than about 4% in the other variables. `fev` and `fvc` was only significantly correlated with height, so non-missing height data was used for their imputation. After imputation, the data still had a small amount of missingness and then a complete case analysis was conducted. In any step of analysis, two cognitive function measurements were analyzed separately.

To explore the association between two cognitive function measurements, DSST and MMMSE, and demographics, behavior, body function, various graphs and exploratory quasi-poisson generalized linear models with an identity link were used. Piecewise linear functions and categorization on numeric variables were adopted as a main tool to adjust the functional form of the variables. Sandwich estimator was used for variance estimation in the quasi-poisson models and Wald tests with a sandwich estimate were used. Variables with the plausible functional form were significantly associated with any cognitive function measurement if they were statistically significant in the quasi-poisson models.

Kaplan-Meier plots and ordinary Cox models were used to explore the the association between two cognitive function measurements, DSST and MMMSE, and mortality. All the Cox models afterwards assumed ignorable censoring. Robust variance estimation was enabled in Cox models and built-in tests with robust variance estimates from R function `coxph` were used. The proportional hazard assumption was diagnozed for Cox models. Any cognitive function measurement was significantly associated with mortality if either one was statistically significant in Cox models.

Cox models with LASSO cross validation were used to compare the predictability from two cognitive function measurements and the other variables. The model predictability was measured by minimal cross validation error which was computed from deviance. Smaller cross validation error indicated better prediction in the models. The strategy was to compare the reduction in the cross validation error after adding any cognitive function measurement in the presence of some other variables to the reduction in the cross validation error after adding some other variables in the presence of any cognitive function measurement. The best predictor set of the other variables was selected by LASSO cross validation Cox models, so this set of variables and the penalty if possible were fixed when adding a new cognitive function measurement in the model.

Ordinary Cox models were used to estimate the causal direct effect of any cognitive function measurement on mortality, based on a causal graph. Confounders were adjusted with the plausible functional form. Further interest was to learn if there was any other underlying and unmeasured disease that can act as a confounder between cognitive function and mortality. An instrumental variable approach was also carried out to estimate the direct effect of any cognitive function measurement on mortality. Education level was re-categorized into two categories, (1) did not graduate high school and (2) graduated high school, from the original three categories, and was used as the instrument, because (1) it affects cognitive function and tends to affect cognitive function in a consistent direction (e.g. higher education level leads to better cognitive function due to knowledge training), (2) it cannot affect mortality directly (e.g. this is no clear evidence that people with higher education level tend to have a greater risk of death), and (3) it affects mortality only through cognitive function in this study (e.g. education level does not affect demographics, smoking, alcohol intake and other diseases due to lack of clear evidence). Then the direct effect of any cognitive function from the Cox models was compared to that from the instrumental variable approach, and any significant difference in the direct effect estimate indicated possible residual confounding.

Results

Exploratory analysis of cognitive function measurements and other variables

Estimates from a quasi-poisson GLM with an identity link for DSST are shown in Table 1. The functional form of alcohol intake, glucose and diastolic blood pressure was piecewise-linear: alcohol intake had a knot at 50 grams, glucose had a knot at 126 mg/dl, and diastolic blood pressure had a knot at 60 mmHg. LDL, albumin, creatinine, systolic blood pressure were considered both in a piecewise linear form and a categorical form. Their categorical forms were included in the model because they were significant or easy for comparison. Systolic blood pressure levels were defined as low, normal and high, because systolic blood pressure and diastolic blood pressure had interaction which can be easily seen when systolic blood pressure was categorized. From the residual diagnostics, the model had no influential data; no clear heteroscedasticity was found. From all the estimates,

age, sex, race, education level, alcohol intake, glucose, ankle:arm index, FEV, interaction between diastolic and systolic blood pressure are significant.

Table 1: Estimates of main variables in the association with DSST

Variable	Estimate	SE	P	CI
Intercept	94.151	7.478	0.000	[79.495, 108.808]
Age (years)	-0.753	0.043	0.000	[-0.838, -0.668]
Sex (male)	-5.147	0.653	0.000	[-6.428, -3.867]
Height (centimeters)	-0.012	0.037	0.739	[-0.084, 0.060]
Weight (pounds)	0.013	0.008	0.127	[-0.004, 0.029]
Race (nonwhite)	-9.805	0.593	0.000	[-10.968, -8.642]
Education (graduate and college)	8.629	0.475	0.000	[7.699, 9.560]
Education (post graduate)	12.512	0.710	0.000	[11.121, 13.904]
Alcohol intake when <= 50 grams	0.076	0.014	0.000	[0.049, 0.103]
Alcohol intake when > 50 grams	-0.004	0.004	0.330	[-0.011, 0.004]
High level LDL (mg/dl)	-0.512	0.459	0.265	[-1.413, 0.388]
Very high level LDL (mg/dl)	-1.547	1.298	0.233	[-4.091, 0.996]
Glucose when <= 126 mg/dl	-0.042	0.018	0.018	[-0.077, -0.007]
Glucose when > 126 mg/dl	-0.002	0.010	0.837	[-0.023, 0.018]
Low level albumin (g/l)	2.581	3.463	0.456	[-4.205, 9.368]
High level creatinine (mg/dl)	-0.520	0.517	0.315	[-1.533, 0.494]
Systolic BP (mmHg)	-0.018	0.021	0.397	[-0.058, 0.023]
Ankle:arm index	3.539	1.266	0.005	[1.058, 6.020]
FEV (liters per second)	2.313	0.394	0.000	[1.542, 3.085]
Smoking pack years for non-current smokers	0.017	0.009	0.066	[-0.001, 0.035]
Smoking pack years for current smokers	-0.021	0.017	0.210	[-0.054, 0.012]
Diastolic BP when <= 60 mmHg, normal systolic BP	0.243	0.098	0.013	[0.051, 0.434]
Diastolic BP when <= 60 mmHg, high systolic BP	0.443	0.171	0.009	[0.108, 0.777]
Diastolic BP when <= 60 mmHg, low systolic BP	-0.014	0.134	0.918	[-0.277, 0.250]
Diastolic BP when > 60 mmHg, normal systolic BP	-0.068	0.028	0.017	[-0.123, -0.012]
Diastolic BP when > 60 mmHg, high systolic BP	-0.063	0.029	0.029	[-0.119, -0.006]
Diastolic BP when > 60 mmHg, low systolic BP	-0.081	0.032	0.010	[-0.143, -0.019]

DSST decreases by 0.753 on average per one year increase in age. In addition, a piecewise linear form of age with a knot at 70 years old (Figure 1) was also tried and turned out to be insignificant. Then interaction between age and sex was also insignificant.

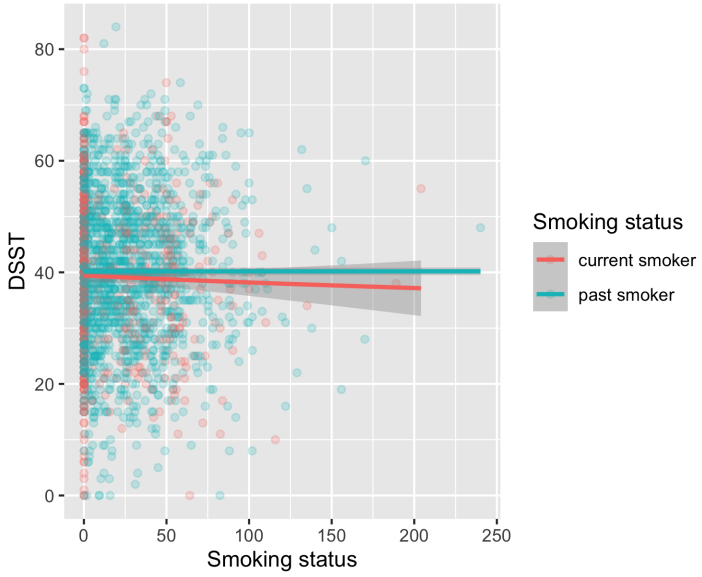
In the elderly people, DSST is 5.147 lower on average in males than in females, and 9.805 higher on average in white people than in nonwhite people. People who had graduate and college education have a DSST 8.629 higher on average than people who did not graduate high school, and people who had post graduate education have a DSST 12.512 higher on average than people who did not graduate high school.

The model means that smoking is not associated with DSST. There were 3 types of smoking status: a person was a never smoker if he had no smoking history and years since quitting smoking, was a current smoker if he had smoking history but no years since quitting, was a past smokers if he had smoking history and years since quitting smoking. There was little difference in average DSST among the three smoking groups. Figure 2 and 3 showed smoking history and years since quitting smoking had no clear interaction.

When average weekly alcohol intake is no greater than 50 grams, DSST increases on average 0.076 per one unit increase in alcohol intake, but when average weekly alcohol intake is greater than 50 grams, DSST is not significantly associated with alcohol intake. This indicates a small amount of weekly alcohol intake is helpful to

Figure 1: Scatter plot of DSST and age

Figure 2: Scatter plot of DSST and smoking history



maintain good cognitive function.

When glucose is no greater than 126 mg/dl, which means no diabetes, DSST decreases on average 0.042 per one unit increase in glucose, but when glucose is greater than 126 mg/dl, which means diabetes, DSST has no significant association with glucose. This trend change can be seen in Figure 4. It turns out that people with diabetes have a lower DSST score on average than people without diabetes.

DSST increases by 3.539 on average per one unit increase in the ankle:arm index. Hence, people with peripheral artery disease tend to have a lower DSST score on average. However, very high ankle:arm index can also indicate other kinds of artery diseases, then this conclusion may be unreliable, because we may expect to see cognitive function decreases in the presence of another disease. Hence, this conclusion should only be interpreted as peripheral artery disease is negatively associated with cognitive function.

DSST increases by 2.313 on average per one unit increase in FEV. FVC can also have similar association with DSST since FEV and FVC are highly correlated.

Systolic and diastolic blood pressure themselves are not significantly associated with DSST, but their interaction effects are. Figure 5 and 6 showed their interaction. When systolic blood pressure is at the normal level and diastolic blood pressure is no greater than 60 mmHg, DSST increases by 0.243 on average per one unit increase

Figure 3: DSST of past smokers in different pack years

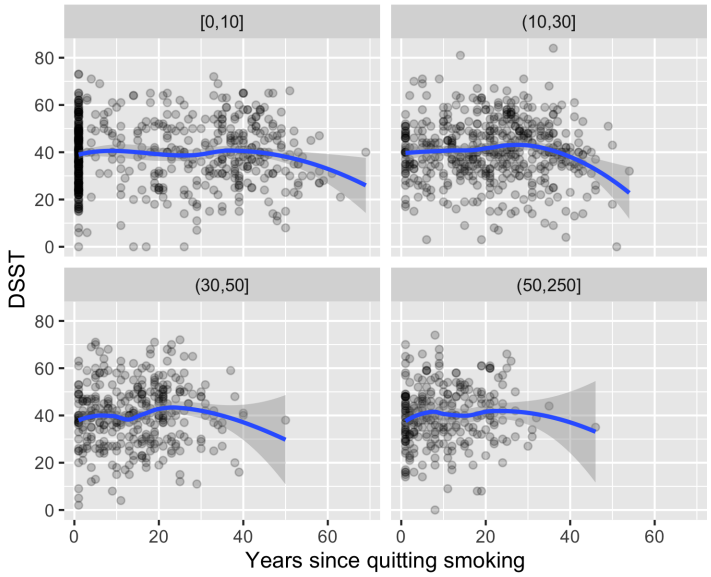
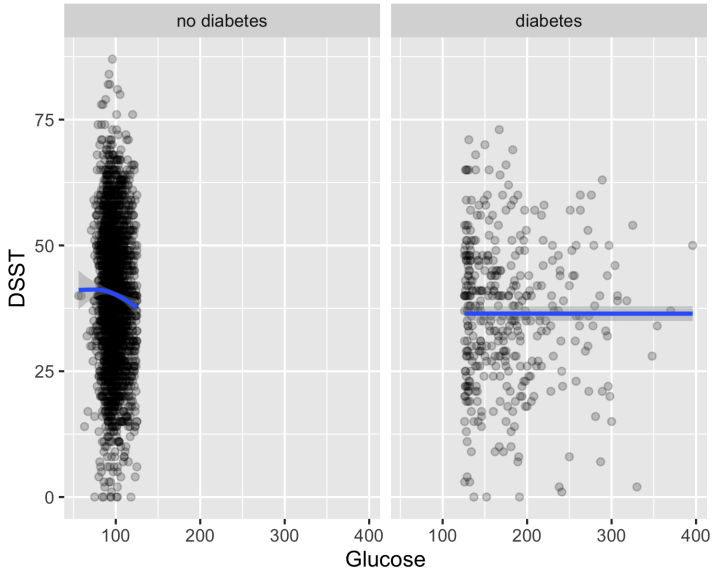


Figure 4: DSST without and with diabetes



in diastolic blood pressure. When systolic blood pressure is at the high level and diastolic blood pressure is no greater than 60 mmHg, DSST increases by 0.443 on average per one unit increase in diastolic blood pressure. When systolic blood pressure is at the normal level and diastolic blood pressure is greater than 60 mmHg, DSST decreases by 0.068 on average per one unit increase in diastolic blood pressure. When systolic blood pressure is at the high level and diastolic blood pressure is greater than 60 mmHg, DSST decreases by 0.063 on average per one unit increase in diastolic blood pressure. When systolic blood pressure is at the low level and diastolic blood pressure is greater than 60 mmHg, DSST decreases by 0.081 on average per one unit increase in diastolic blood pressure. This shows that compared with people with normal systolic and diastolic blood pressure levels, people with hypertension have a lower DSST score on average. Hence, hypertension is negatively associated with cognitive function.

Estimates from a quasi-poisson GLM with an identity link for MMMSE are shown in Table 2. All the variables had the same functional form and definitions as in the model for DSST, after different functional forms were tried. From the residual diagnostics, the model had no influential data, but clear heteroscedasticity was found. Since sandwich estimator was used, the variance estimation would become valid. The exploratory analysis approach for MMMSE was similar as for DSST. From all the estimates, age, sex, weight, race, education level,

Figure 5: DSST in different diastolic blood pressure

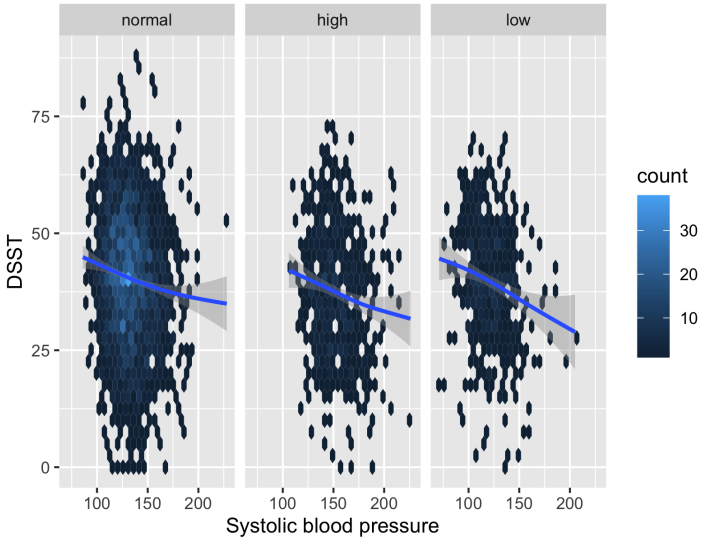
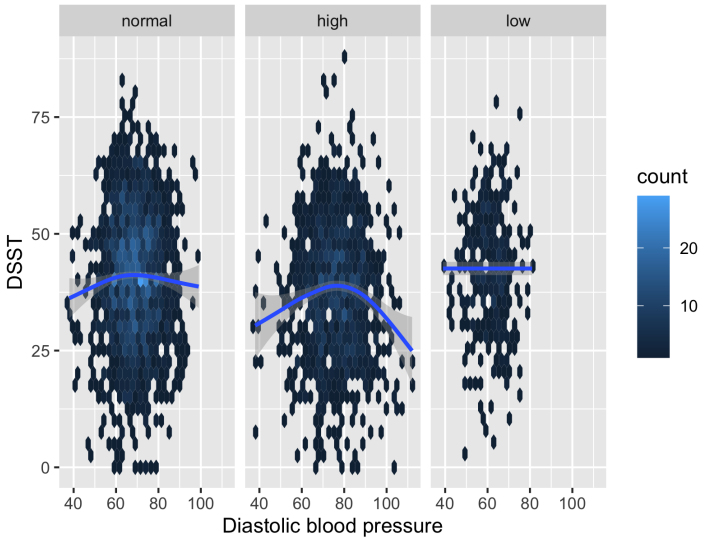


Figure 6: DSST in different systolic blood pressure



LDL, alcohol intake, ankle:arm index, FEV are significant.

MMMSE decreases by 0.32 on average per one year increase in age. In the elderly people, MMMSE is 1.565 lower on average in males than in females, and 5.6 higher on average in white people than in nonwhite people. People who had graduate and college education have a MMMSE 4.282 higher on average than people who did not graduate high school, and people who had post graduate education have a MMMSE 6.449 higher on average than people who did not graduate high school.

When average weekly alcohol intake is no greater than 50 grams, MMMSE increases on average 0.030 per one unit increase in alcohol intake, but when average weekly alcohol intake is greater than 50 grams, MMMSE is not significantly associated with alcohol intake.

MMMSE increases by 2.511 on average per one unit increase in the ankle:arm index. This should only be interpreted as peripheral artery disease is negatively associated with cognitive function.

MMMSE increases by 0.663 on average per one unit increase in FEV. FVC can also have similar association with MMMSE.

Differences from DSST are: glucose, systolic and diastolic blood pressure and their interaction are not significantly associated with MMMSE; one unit increase in weight increases MMMSE by 0.011 on average, and very high level LDL decreases MMMSE by 1.757 on average.

DSST and MMMSE are both associated with people demographics, behavior and body function. Age, sex,

Table 2: Estimates of main variables in the association with MMMSE

Variable	Estimate	SE	P	CI
Intercept	110.675	4.428	0.000	[101.996, 119.353]
Age (years)	-0.320	0.029	0.000	[-0.377, -0.263]
Sex (male)	-1.565	0.359	0.000	[-2.268, -0.862]
Height (centimeters)	0.002	0.022	0.925	[-0.040, 0.044]
Weight (pounds)	0.011	0.005	0.016	[0.002, 0.021]
Race (nonwhite)	-5.600	0.393	0.000	[-6.371, -4.829]
Education (graduate and college)	4.282	0.306	0.000	[3.683, 4.882]
Education (post graduate)	6.449	0.413	0.000	[5.640, 7.258]
Alcohol intake when <= 50 grams	0.030	0.008	0.000	[0.015, 0.046]
Alcohol intake when > 50 grams	-0.002	0.002	0.438	[-0.007, 0.003]
High level LDL (mg/dl)	-0.150	0.263	0.570	[-0.665, 0.366]
Very high level LDL (mg/dl)	-1.757	0.878	0.045	[-3.477, -0.037]
Glucose when <= 126 mg/dl	-0.007	0.010	0.498	[-0.027, 0.013]
Glucose when > 126 mg/dl	-0.006	0.007	0.354	[-0.019, 0.007]
Low level albumin (g/l)	0.333	1.635	0.839	[-2.872, 3.538]
High level creatinine (mg/dl)	-0.089	0.285	0.755	[-0.647, 0.470]
Systolic BP (mmHg)	0.002	0.011	0.857	[-0.020, 0.024]
Ankle:arm index	2.511	0.706	0.000	[1.128, 3.894]
FEV (liters per second)	0.663	0.227	0.004	[0.217, 1.109]
Smoking pack years for non-current smokers	0.001	0.006	0.862	[-0.011, 0.013]
Smoking pack years for current smokers	-0.001	0.008	0.920	[-0.017, 0.015]
Diastolic BP when <= 60 mmHg, normal systolic BP	0.042	0.054	0.435	[-0.064, 0.148]
Diastolic BP when <= 60 mmHg, high systolic BP	-0.044	0.075	0.557	[-0.190, 0.103]
Diastolic BP when <= 60 mmHg, low systolic BP	-0.043	0.078	0.584	[-0.197, 0.111]
Diastolic BP when > 60 mmHg, normal systolic BP	-0.020	0.019	0.277	[-0.057, 0.016]
Diastolic BP when > 60 mmHg, high systolic BP	-0.021	0.019	0.265	[-0.059, 0.016]
Diastolic BP when > 60 mmHg, low systolic BP	-0.025	0.020	0.216	[-0.065, 0.015]

race and education level are significant demographic factors. Alcohol intake is a significant behavior factor. Lung function is a significant biological factor. As age increases, cognitive function tends to decrease. Elderly males and nonwhite people tend to have lower cognitive function. A small amount of alcohol intake tend to increase cognitive function. Good lung function tends to increase cognitive function, but liver and kidney function are not associated with cognitive function. Perhaps peripheral artery disease, heart disease, hypertension, diabetes can lower cognitive function, so cognitive function also depends on the health.

DSST and MMMSE have significant correlation ($\rho = 0.589$). But some participants have very low DSST score but relatively high MMMSE score. This may indicate that DSST score and MMMSE score are not perfectly transferable. 11 people had relatively high MMMSE score but 0 DSST score, who possibly did not take the DSST test. They were not excluded from the data.

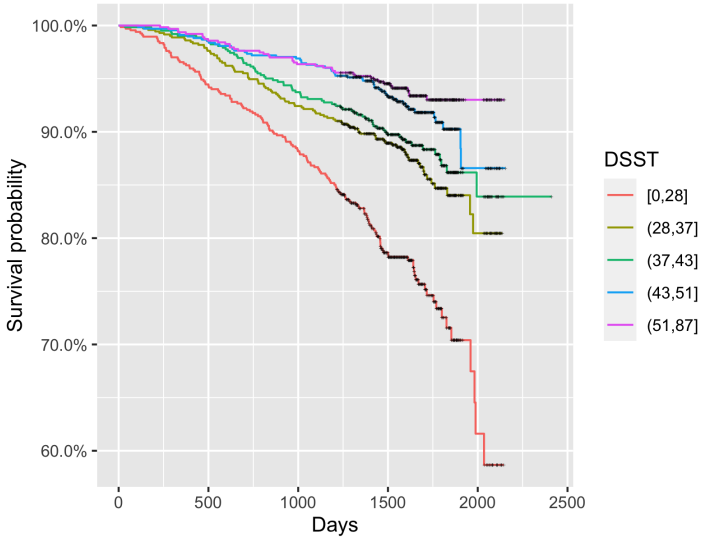
Exploratory analysis of cognitive function measurements and mortality

Mortality was interpreted as survival probability (Kaplan-Meier estimate) and hazard (ordinary Cox model).

Figure 7 was the Kaplan-Meier plot of the survival probability against DSST. DSST was categorized into 5 groups with nearly equal size. Figure 7 showed that the estimated survival probability is higher whenever DSST is higher, so this means that DSST is positively associated with the survival of people. Higher DSST means better cognitive function and means a higher survival probability and a higher probability to survive a longer

time.

Figure 7: Kaplan-Meier plot of survival against DSST



Cox model was constructed to explore the association between DSST and hazard as a measure of mortality. The Schoenfeld residuals from the Cox model did not show time trends, so the proportional hazard assumption was met and the Cox model was valid. Likelihood ratio, Wald and score tests all showed that DSST is significantly negatively associated with hazard. Hazard decreases by 4% on average per one unit increase in DSST (Table 3). People with good cognitive function tend to have lower hazard and better survival. Figure 8 showed again, cognitive function is positively associated with survival, and negatively associated with mortality.

Table 3: Cox model with DSST

Log hazard ratio	SE	P	CI	Hazard ratio	CI
-0.039	0.004	0.000	[-0.046, -0.032]	0.962	[0.955, 0.969]

Figure 8: Fitted survival probabilities

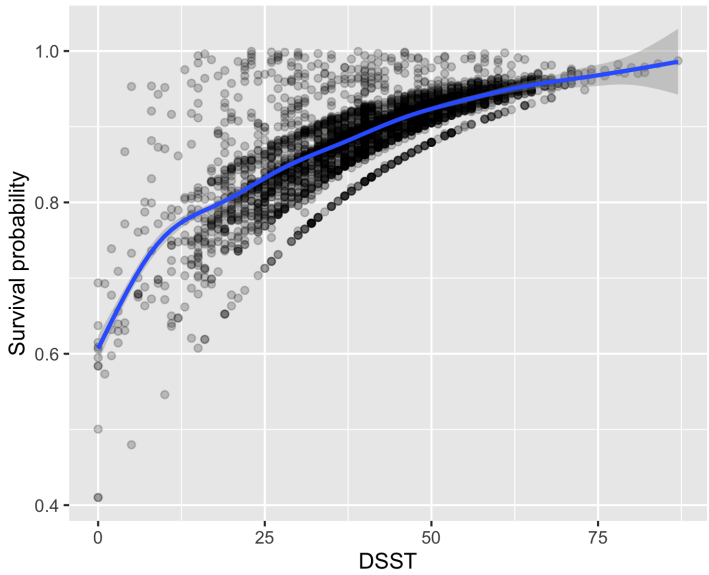
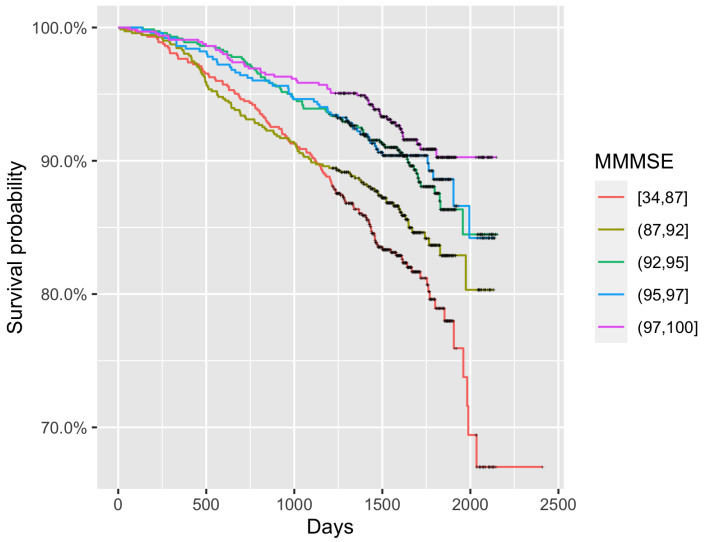


Figure 9 was the Kaplan-Meier plot of the survival probability against MMMSE. MMMSE was categorized

into 5 groups with nearly equal size. Figure 9 showed that some survival curves have crosses, such as the curve with MMMSE in 92-95 becomes higher first and then lower than the curve with MMMSE in 95-97. The overall tendency after a relatively long survival period is that higher MMMSE leads to a higher survival probability. From MMMSE, cognitive function is positively associated with the survival of people to some extent.

Figure 9: Kaplan-Meier plot of survival against MMMSE



Cox model was constructed to explore the association between MMMSE score and hazard. The Schoenfeld residuals from the Cox model did not show time trends, so the proportional hazard assumption was met and the Cox model was valid. Likelihood ratio, Wald and score tests all showed that MMMSE score is significantly negatively associated with hazard. Hazard decreases by 4% on average per one unit increase in MMMSE (Table 4). Figure 10 showed again, cognitive function is positively associated with survival, and negatively associated with mortality. The association with mortality is quite similar for DSST and MMMSE.

Table 4: Cox model with MMMSE

Log hazard ratio	SE	P	CI	Hazard ratio	CI
-0.042	0.005	0.000	[-0.052, -0.032]	0.959	[0.949, 0.968]

Predictive analysis of cognitive function measurements and other variables

To examine the predictability of cognitive function measurements and other variables, ordinary Cox models were used. Kaplan-Meier plots of survival probability against some variables were explored. Then cross validation LASSO Cox models were constructed for DSST and MMMSE separately, to compare each cognitive function measurement with the other variables including those created from existing variables.

Table 5 shows cross validation results from LASSO Cox models with combinations of cognitive function measurements and the other variables. From model 1 and 2, adding some other variables decreases the cross validation error significantly in the presence of DSST in the model, because the confidence interval of the cross validation error from model 2 is on the left-hand side of that from model 1 with no overlap. From model 5 and 6, adding some other variables decreases the cross validation error significantly in the presence of MMMSE in the model. Hence, the other variables can improve the prediction in the presence of any cognitive function measurement in the model.

From model 3 and 4, adding DSST decreases the cross validation error in the presence of the other variables in the model, but not significantly. From model 3 and 7, adding MMMSE increases the cross validation error in the presence of the other variables in the model, but not significantly. The two comparisons may not be accurate, because the penalty of the other variables in the presence of any cognitive function measurement was not equal to that in the absence of any cognitive function measurement. Then Cox models were used to do

Figure 10: Fitted survival probabilities

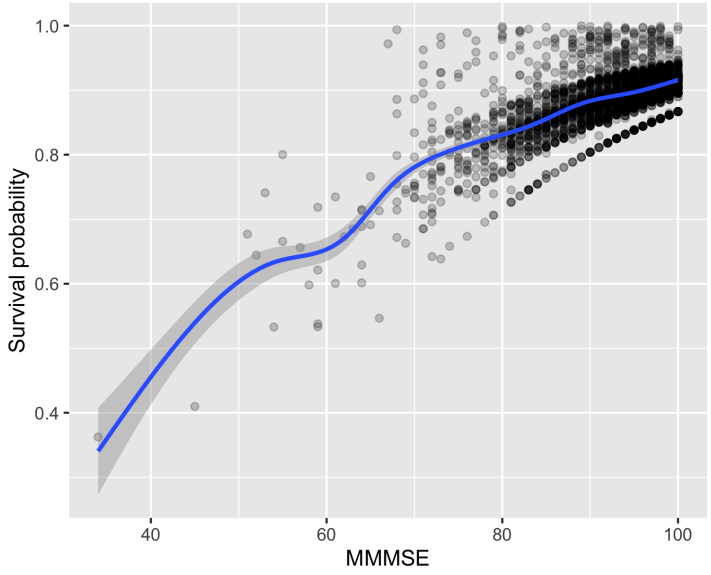


Table 5: Cross validation LASSO Cox models

Model	Variable	Penalty	CV error	CI
1	DSST only	0	16.750	[16.671, 16.829]
2	DSST + best other variables (1)	0 + 0.004	16.383	[16.264, 16.502]
3	Best other variables (2)	0.004	16.431	[16.283, 16.579]
4	Best other variables (2) + DSST	0 + 0	16.386	[16.271, 16.501]
5	MMMSE only	0	16.910	[16.777, 17.044]
6	MMMSE + best other variables (3)	0 + 0.004	16.428	[16.255, 16.600]
7	Best other variables (2) + MMMSE	0 + 0	16.469	[16.360, 16.579]

Note 1: Variables followed by + were added later.

Note 2: Best other variables had 3 sets because of the order they were in the model.

further comparisons. Table 6 shows that when the fitted values from the LASSO model with only the best other variables are used as a predictor, and DSST or MMMSE is also added in the model, then DSST or MMMSE is significant. Hence, any cognitive function measurement can improve the predictions in the presence of the other variables in the model.

Table 6: Cox models for model 4 and 7

Model	Variable	Estimate	SE	P
4	DSST	-0.019	0.004	0.000
	Predictions from model 3	0.991	0.064	0.000
7	MMMSE	-0.015	0.006	0.010
	Predictions from model 3	1.054	0.062	0.000

In conclusion, any cognitive function measurement and the other variables do not outperform each other in prediction. They are both predictive to mortality.

Causal analysis of cognitive function measurements on mortality

A causal graph (Figure 11) was drawn to show the causal relationship between the cognitive function measurements and mortality. DSST and MMMSE represented cognitive function, and the effect of each cognitive function measurement was analyzed separately. They were assumed to have the same confounders, which were confounders for cognitive function and mortality. Even though some confounders did not have significant association with one cognitive function measurement, it may still be appropriate to have them in the model, so as to avoid potential residual confounding. The functional forms of the confounders were the same as they were in the exploratory models. The confounders were known risk factors or diseases in this study including diabetes, heart and peripheral artery diseases, hypertension, lung and kidney and liver malfunction. Some other factors including albumin, creatinine and white blood cells count formed confounding paths with age. Smoking can also affect mortality and thus was included to improve the model estimation.

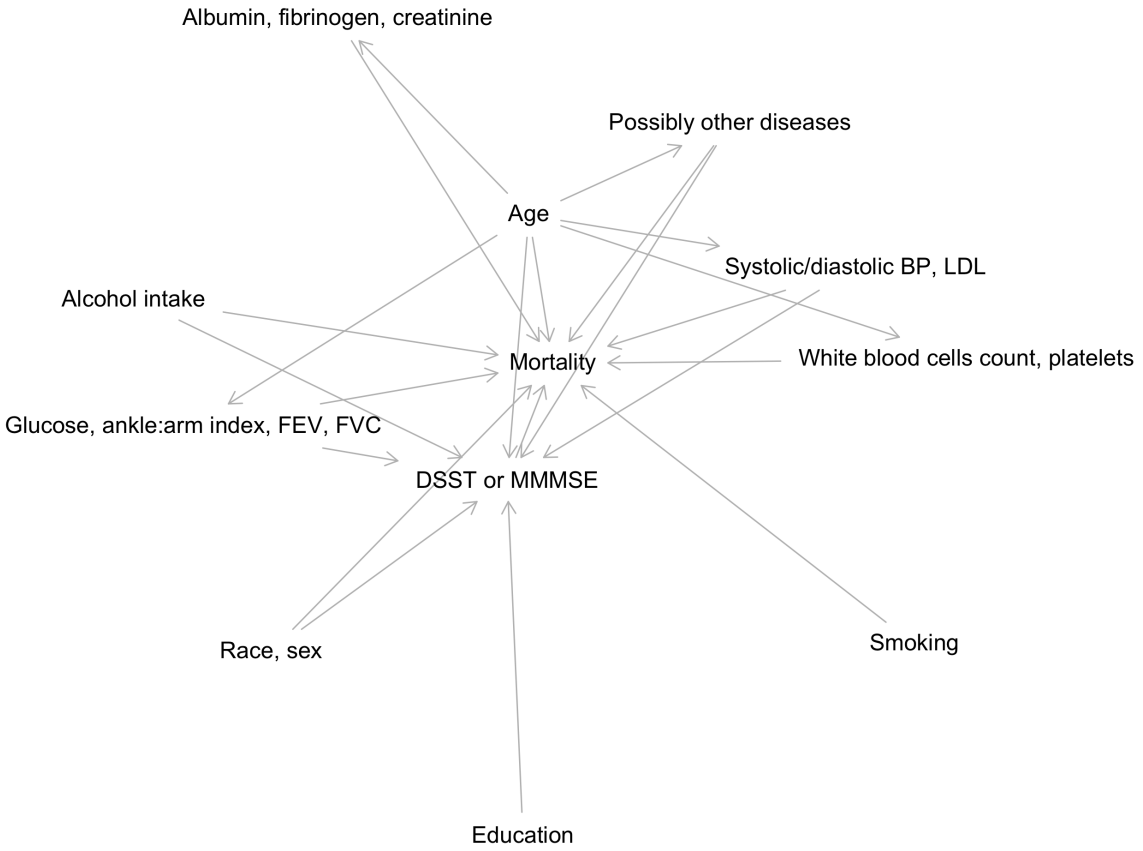


Figure 11: Causal graph

Ordinary Cox models were constructed to estimate the direct effect of DSST on mortality and the direct effect of MMMSE on mortality, adjusting for all the identified confounders. Then education level was used as an instrument to estimate the direct effect of DSST or MMMSE. People who graduated high school and people who had post-graduate schooling were combined in a single group who graduated high school. This was plausible because the hazard ratio between these two groups was insignificantly away from 1.

Table 7 shows the causal direct effect estimates of the two cognitive function measurements from Cox models and the instrumental variable approach. DSST is very significant in the Cox model. MMMSE is moderately significant in the Cox model.

For DSST, the log hazard ratio and hazard ratio estimates from the instrumental variable approach are slightly smaller than those from the Cox model. But the confidence intervals of the instrumental variable estimates entirely cover the confidence intervals of the Cox model estimates. This indicates that there may be some residual confounding, though with no sound evidence.

For MMMSE, the log hazard ratio and hazard ratio estimates from the instrumental variable approach are much smaller than those from the Cox model. The confidence intervals of the instrumental variable have a little overlap on those of the Cox model estimates, but the confidence intervals from the two approaches have clearer

Table 7: Direct causal effect estimates

Measure	Approach	Log hazard ratio	CI	Hazard ratio	CI
DSST	Cox model	-0.023	[-0.031, -0.014]	0.978	[0.969, 0.986]
	Instrument	-0.028	[-0.045, -0.010]	0.973	[0.956, 0.990]
MMMSE	Cox model	-0.018	[-0.031, -0.005]	0.982	[0.970, 0.995]
	Instrument	-0.057	[-0.096, -0.018]	0.945	[0.909, 0.982]

separation than for DSST. This gives more evidence to support that there is some other unmeasured confounding path between MMMSE and mortality.

The estimated log hazard ratio and hazard ratio from the instrumental variable approach are all smaller than those from the Cox model, so the true direct effect of DSST or MMMSE should be greater than the Cox model estimates. This implies the direct effect of cognitive function on mortality should be more significant, and this effect is likely to be confounded by some other diseases which are not identified in this study.

Discussion

This study is a cohort study and the cohort is a representative sample of the true elderly population, so the study design is reasonably good. However, it is an observational study, so when the causal effect needs to be estimated, they may suffer bias and error.

The analysis did not adjust for measurement error. However, it is very likely that DSST, MMMSE and some biological measures involve measurement errors.

Instrumental variable is an efficient approach to estimate the causal effect in observational studies. In this analysis, it helped estimate the direct effect of DSST and MMMSE measurements. When education level was used as an instrument, exclusion restriction and randomization assumptions may be still under question. This subset data can be regarded as a random sample, so education level has some randomization characteristics, but itself is not a perfect substitute of randomization. There is no evidence to support that education level affects mortality, but it is likely that education level can affect something else to affect mortality, such as income. This may violate exclusion restriction. Hence, education level is a useful but non-optimal instrument.

Conclusion

Among elderly people, cognitive function measurements are significantly associated with demographic factors (age, sex, race, education level), behavioral factors (alcohol intake), and biological factors (glucose, blood pressure, FEV). Age has most significant negative association with cognitive function measurements. Diabetes, hypertension and lung malfunction can also lower cognitive function measurements. In addition, cognitive function measurements have significant negative association with mortality. High cognitive function measurements indicate good survival status of people. When it comes to prediction of mortality, cognitive function measurements cannot replace the other demographic, behavioral and biological factors, and vice versa. Cognitive function measurements and the other factors are both predictive of mortality, possibly because they measure different underlying things which can account for mortality.

Cognitive function measurements are a good indication for cognitive function. Through these measurements, one more goal is to estimate the direct causal effect of cognitive function on mortality. Many diseases in this study can affect cognitive function and thus affect the cognitive function measurements, such as heart disease and diabetes, and they confound the causal relationship between cognitive function measurements and mortality. After these factors are adjusted, there is some weak evidence that the causal relationship between cognitive function measurements and mortality is still confounded by some other unmeasured factors or diseases. This evidence is plausible in the real world, because cognitive function can also be affected by some other diseases such as dementia and depression.

On the other side, the causal analysis of the cognitive function measurements also suggests that cognitive function is negatively associated with mortality. Higher cognitive function lead to a lower risk of mortality.