Cognitive Impairment Progression Analysis by Comparison of Clinical and Cognitive Scales, including Visualization using Agent-Based Simulation Approach

by

Xue Liang

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Engineering
(Industrial and Systems Engineering)
in the University of Michigan–Dearborn
2016

Master's Thesis Committee:

Associate Professor Yung-wen Liu, Chair
Assistant Professor Armagan Bayram
Assistant Professor Xi Chen

Acknowledgements

I wish to express my deep sense of gratitude to Dr. Yung-wen Liu for his loving inspiration and timely guidance for this research. I am also very much thankful to the National Alzheimer's Coordinating Center for proving access to their dataset.

The NACC database is funded by NIA/NIH Grant U01 AG016976. NACC data are contributed by the NIAfunded ADCs: P30 AG019610 (PI Eric Reiman, MD), P30 AG013846 (PI Neil Kowall, MD), P50 AG008702 (PI Scott Small, MD), P50 AG025688 (PI Allan Levey, MD, PhD), P50 AG047266 (PI Todd Golde, MD, PhD), P30 AG010133 (PI Andrew Saykin, PsyD), P50 AG005146 (PI Marilyn Albert, PhD), P50 AG005134 (PI Bradley Hyman, MD, PhD), P50 AG016574 (PI Ronald Petersen, MD, PhD), P50 AG005138 (PI Mary Sano, PhD), P30 AG008051 (PI Steven Ferris, PhD), P30 AG013854 (PI M. Marsel Mesulam, MD), P30 AG008017 (PI Jeffrey Kaye, MD), P30 AG010161 (PI David Bennett, MD), P50 AG047366 (PI Victor Henderson, MD, MS), P30 AG010129 (PI Charles DeCarli, MD), P50 AG016573 (PI Frank LaFerla, PhD), P50 AG016570 (PI Marie-Francoise Chesselet, MD, PhD), P50 AG005131 (PI Douglas Galasko, MD), P50 AG023501 (PI Bruce Miller, MD), P30 AG035982 (PI Russell Swerdlow, MD) , P30 AG028383 (PI Linda Van Eldik, PhD), P30 AG010124 (PI John Trojanowski, MD, PhD), P50 AG005133 (PI Oscar Lopez, MD), P50 AG005142 (PI Helena Chui, MD), P30 AG012300 (PI Roger Rosenberg, MD), P50 AG005136 (PI Thomas Montine, MD, PhD), P50 AG033514 (PI Sanjay Asthana, MD, FRCP), P50 AG005681 (PI John Morris, MD), and P50 AG047270 (PI Stephen Strittmatter, MD, PhD.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Alzheimer's disease has been identified as the 6th leading cause of death in United States in 2015. One-third seniors die with Alzheimer's and other dementia. Only 45% of AD patients report being told of their diagnosis. Staging the severity of AD is needed of no delay.The "preclinical" phase of AD is known to be a long-term progression process towards severe cognitive impairment. Clinical and Cognitive scaling methods, especially Clinical Dementia Rating (CDR) and Mini-mental State Examination (MMSE), are widely used in mapping the stage of cognitive impairment status. Many studies have also identified important clinical and physiological risk factors, such as age, smoking status, blood pressure, total serum cholesterol, current impairment status, etc. To discuss a better prediction method of AD for mild cognitive impairment patients, a longitudinal study of the impact of these influential factors is essential.

The objective of this research is to firstly analyze the accuracy of both methods and secondly to propose a method that gives higher prediction accuracy. The main contribution of this paper is that we gave a thorough analysis on comparison of CDR and MMSE, commented on which method works better based upon personal demographic performance and brought up a pattern recognition model to predict the probability of patients reaching the unfavorable outcome, i.e. dementia, over time. We gained a method that achieves high accuracy by combining the regression model and MMSE cognitive scales. An agent-based simulation model was brought up to visualize the change of cognitive impairment status of patients over time, in various populations.

Chapter 1: Introduction

1.1 Introduction of Alzheimer's disease

It is very normal for very old people to forget someone's name or misplace things time to time. This kind of forgetfulness is normal in senior's life. However, someone might forget how to get home, get confused about someone's name that he/she is familiar with, or asking simple questions repeatedly. This is a signal for a much more serious problem: Alzheimer's disease.

Alzheimer's disease is an irreversible, progressive brain disorder that slowly destroys memory and thinking skills, and eventually the ability to carry out the simplest tasks[1] National Institute on Aging. "Alzheimer's Disease Fact Sheet." Publication Date: Augus 2016.

[2]. In most people with Alzheimer's, the greatest known risk factor is increasing age; the majority of people with Alzheimer's are 65 and older. As is reported in 2016 Alzheimer's disease Facts and Figures, 5.4 million Americans are suffering Alzheimer's disease, an estimated 5.2 million of which are aged 65 years or older. This means, one in nine people aged 65 or older has Alzheimer's disease. Moreover, the numbers will escalate rapidly in coming years, as the baby boom generation is reaching age 65 and beyond. It is estimated by vary that, by 2050, the number of people with Alzheimer's disease might nearly triple, to projected 13.8 million.

Alzheimer's disease is not only widely affecting a large population but also causing serious problems in cognitive function, which results in more accidentally death of senior people. In 2016, Alzheimer's disease is officially listed as the sixth-leading cause of death in the United

States. As the population of United States ages, Alzheimer's is becoming a more common cause of death. Unlike the other major causes that have decreased significantly in the last decade, deaths from Alzheimer's disease have increased significantly – 71%. Also, Alzheimer's is the only disease among the top 10 causes of death in America that cannot be prevented and cured.

The symptoms of Alzheimer's disease worsen over time and are hard to be early diagnosed. Changes in brain related to Alzheimer's begin years before any signs of the disease. This period of time, known as Preclinical Alzheimer's disease, is a long-term progressive process towards severe cognitive impairment and possibly complications such as cardiovascular problems. People in the preclinical phase of Alzheimer's behave normally, but after that people may suffer from a severe decline in cognitive functions, which might cause inability in living independently, known as progression into a Middle-stage or Late-stage of Alzheimer's disease. However, only 45% of Alzheimer's patients report being told of their diagnosis.

Alzheimer's disease is severely affecting patient's cognitive functions and also influencing a large amount of senior population. Early diagnosis and staging the severity of Alzheimer's disease are needed of no delay. In this case, a thorough literature review was developed and summarized in Chapter 2 about previous researches on the early-diagnosis and rating of Alzheimer's disease. Two methods were investigated and compared with a brief introduction in the next session.

1.2 Introduction of Clinical Dementia Rating and Mini-mental State Examination.

In our research, two popular methods currently been used to help Alzheimer's diagnosis, Clinical Dementia Rating (CDR) and Mini-mental State Examination (MMSE), are compared thoroughly about their accuracy and practicality.

CDR was developed for the evaluation of staging severity of dementia. It's developed primarily for use in persons with dementia of Alzheimer's type. It's a five-point scale in which CDR = 0 connotes no cognitive impairment, CDR = 0.5 indicates very mild dementia, CDR = 1 indicates mild dementia, CDR = 2 and 3 implies a dementia with moderate or severe stage. Six domains were evaluated to construct the overall CDR table: Memory, Orientation, Judgment and Problem-solving, Community Affairs, Home and Hobbies, and Personal Care. Previous researches verified the relative accuracy of CDR. However, to complete a CDR evaluation, considerable amount of time and professional data collection are necessary. It might be problematic if a reliable and professional physician were not available when people perform the CDR evaluation.

MMSE, or Folstein test, is a 30-point questionnaire that is used extensively in clinical and research settings to measure cognitive impairment. A lower score indicates a worse cognitive status. The 30-point score were summarized into 4 stages: Score results in 25 – 30 indicate a questionably significant impairment, 20 – 25 indicates mild cognitive impairment, and 10 – 20, 0 – 10 implies a moderate or severe stage of impairment, respectively. Administration of the test takes between 5 - 10 minutes and examines functions including registration, attention and calculation, recall, language, ability to follow simple commands and orientation. The main advantage of MMSE is that performing this test requires no specialized equipment or training for administration, and the test result is validate and reliable for the diagnosis and longitudinal assessment of Alzheimer's disease, which is proven by the previous researches.

In order to gain an accurate result in staging Alzheimer's disease or one's cognitive status, a test method that not only accurate and reliable but also relative easy and convenient to perform. Thus we brought up the idea of a detailed comparison of both methods. The very first steps of

our research are comparing the results accuracy from both methods and then identify factors that might affect the validation of each method. Detailed analysis method would be discussed in Chapter 3.

1.3 Introduction to logistic regression and its application in current research.

When we were comparing the accuracy of Clinical Dementia Rating (CDR) and Mini-mental State Examination (MMSE), one interesting finding is that demographical factors such as age, education level, living situation and etc. might contribute to the difference of accuracy between the two measurement scales. We then brought up the idea to identify the risk factors that actually contributing to the progression process of cognitive impairment. We were able to identify several demographical variables including gender, age, education level and etc. through statistical analysis. To analyze the combined effects of these risk factors/variables as well as predict the probability of patients' reaching unfavorable outcome over time, we developed a logistic regression model.

In statistics, logistic regression, or logit regression, or logit model is a regression model where the dependent variable is categorical[3]. Especially, the binary logistic model is used to estimate the probability of a binary response based on one or more predictor (or independent) variables (risk factors)[4]. In our research, the dependent variable is the onset of Alzheimer's disease, which is a dichotomous response. Each individual subject has his/her own set of combined values of identified risk factors. In this case, logistic regression model can be applied to represent the influences of demographical input variables.

To accomplish the logistic model, the input information about demographical variables comes from the raw data of the database in National Alzheimer's Coordination Center[5],[6],[7].

R [8] is used to generate the detailed parameters in the logistic regression model. The details of how we achieve the logistic regression model would be discussed in Chapter 4.

1.4 Introduction to Agent-based simulation.

Another objective of our research is to visualize the change of cognitive impairment status of patients, and to predict the probability of patients reaching unfavorable outcome, i.e. Alzheimer's disease, in various populations over time. Simulation approach would be powerful tool to reach this objective. The technique we chose is agent-based simulation.

Agent-based simulation model is applied to simulate the actions and interactions of autonomous agents (either individual or collective ones) based on the clear analysis of the attributes of individual agents, the potential reaction and outcome through interacting with its surrounding agents, and the potential reactions and outcomes through interacting with its local environment[9]. Agent-based simulation is powerful on simulating the heterogeneity of subjects, i.e. the individual differences on progressive process, which is caused by the physical, psychological and social factors.

In our research, agent-based simulation model can easily describe the individual differences in patients' unique set of demographical variables as well as the corresponding different probability of onset of Alzheimer's disease. The detailed process about how we generated the input information, progression rate of each risk factor and the simulation results of various populations would be discussed in Chapter 5.

Chapter 2: Literature Review

A large amount of previous literatures was reviewed in order to gain better understanding of the background of Alzheimer's disease as well as the current research achievements. Some interesting and useful research findings are concluded in the following paragraphs.

The first part of the literatures is focusing on the two main rating methods on Alzheimer's disease: CDR and MMSE. Research papers talking about how both methods were brought up, how to apply both methods in identifying the onset of Alzheimer's, and the advantage and disadvantage for applying each method were broadly reviewed.

CDR, clinical dementia rating, was first introduced by Hughes et al.[10] in 1982 as a clinical scale for staging dementia. Before CDR was brought up, there were no accurate methods that could stage the dementia in old subjects clinically. People were using psychometric testing combined with behavioral rating evaluations to somehow get an estimation of dementia status in Alzheimer's type. The research of Hughes, which states the reliability, accuracy and validity of CDR in staging dementia, especially in Alzheimer's disease, has been of great interest of researchers. Most researchers devoted their efforts in applying such method on various populations to verify and demonstrate the usefulness of CDR in clinically denote the presence of

[11][12][13]. For example, a research from Morris[12] in 1997 explained in detail about how the clinical protocol incorporates semi-structured interviews

with patients and informant to obtain information necessary to rate the subject's cognitive performance. In nutshell, CDR is a widely accepted rating scale in the clinical setting as a reliable and valid global assessment measure for Alzheimer's disease.

Folstein et al. originally introduced MMSE, The Mini-Mental State Examination, in 1975[14]. This method is devised as a simplified, scored form of the cognitive mental status examination. Forlstein in his research demonstrated that, at that time, assessments of cognitive performance typically require much effort to apply. Withers and Hinton's test[15], as an example, includes 33 questions and requires about 30 minutes to administer and score. CDR, as we mentioned before, requires a fully clinical evaluation covering 6 domains of cognition as well as a detailed evaluation form from both patients and informants. Such methods, although they are thorough and reliable in results, is harder to be used in patient's daily life because that elderly patients, particularly those with dementia, cooperate well only for short periods[16]. In this case, MMSE was brought up as a simple cognitive mental status evaluation method, which includes eleven questions and requires only 5 – 10 minutes to administer. Therefore, MMSE can be used serially and routinely while maintaining the validity and reliability of evaluation as well. Nowadays, MMSE is used extensively in clinical and research settings to measure cognitive impairment[17]. Meanwhile, the simplicity of MMSE doesn't affect the validity and reliability of this method as it concentrates on the cognitive aspects of mental functions and examines functions including registration, attention and calculation, recall, language, ability to follow simple commands and orientation[14][18][19][20].

Many researchers brought up comparison between CDR and MMSE. Some of the researchers draw conclusion that CDR is relative more accurate and somehow harder to apply [21][22][23] Forsell et al.[23] found a low correlation of the MMSE with the CDR categories of

community affairs and home/hobbies and they suggested new cut-off points for the MMSE when used as a staging scale. Many other researchers claim that MMSE is less accurate because it's correlated to demographic factors such as age and education level [24]. Changing the cutoff points of MMSE to adjust the influence of age and education were necessary as an improvement of accuracy of MMSE [25].

However, the opposite side stated that MMSE could be used in general-medicine setting as a useful tool for identifying cognitive impairment in individuals and also be used as a simplified method of impairment detection [26][27]. One example would be that Perneczky et al [18]. mapped MMSE scores onto CDR categories to determine how well the MMSE performs as a surrogate of the CDR as a timesaving method of staging dementia and they gave positive conclusion. Their results show that MMSE discriminated well between CDR stages 0.5, 1, 2 and 3, which differentiates the mild, moderate and severe stages of dementia. Another example is that Grut stated that age, sex and education didn't substantially affect the specificity and positive prediction value of MMSE, and even the slight effect on sensitivity is not significant [28].

Through literature review, we brought up the idea of analyzing the effects of demographical factors on the relative accuracy of CDR and MMSE methods. For example, we wanted to bring up a study about in which scenario the relative accurate method might switch between CDR and MMSE for subjects with different education level, age and etc. Based upon this interesting idea, one part of our research is to introduce the demographical effects into the clinical and cognitive scales comparison, which would be discussed in detail in Chapter 3.

After the first part of comparing CDR and MMSE in thorough, we got some interesting findings of the effects of education level, gender and some other factors on the cognitive measurements. We then went deeper is into the analysis of identifying the effect of such factors

on the cognitive progression on MCI subjects. Previous studies have identified clinical and physiological risk factors for Alzheimer's disease. Breteler, et al.(2000) and Newman, et al.(2005) suggested that the presence of cerebrovascular disease and cardiovascular disease intensifies the presence and severity of the clinical symptoms of AD [24][29]. The dependency level of socially and mentally activities, the accessibility of healthcare resources, and family/friend support may influence the speed of progression, which is mentioned in the research of Kotagal, et al [30]. Education level turns out to be a potential risk factor for AD according to Sattler's study[31]. Consumption of up to three servings of wine daily is associated with a lower risk of AD in elderly individuals according to Luchsinger's research in 2004. Besides, Fratiglioni and Gao et al[32][33][34], mentioned in their literatures that the incidence of dementia increases with age, Gender also has demographical influence on Alzheimer's as female tend to have higher risk to develop dementia.

Through literature review, we found that applying similar statistical methods on subjects with different cognitive status or different populations may result in various conclusions. In order to identify the risk factors that might have influence on the subjects we chose, all the demographic factors mentioned above are analyzed statistically in our database from National Alzheimer's Coordination Center, a brief introduction of the database and subjects are to be discussed in Chapter 3. Based upon the statistical results, we identified the potential influential risk factors and analyzed their effects through pattern recognition model. Combining the logistic regression model and the MMSE measurement, we were able to bring up a reliable and validity method to predict the onset of Alzheimer's disease for subjects with MCI.

Moreover, in order to visualize the long-term progressive process of Alzheimer's disease towards severe cognitive impairment including the change of cognitive impairment status of

patients, and the probability of patients reaching the unfavorable outcome, we integrated the evaluation method into agent-based simulation model. Widely literatures were reviewed about how health care problems can gain benefits from modeling and simulation tools[35][36][37] and how safety dynamic model may be used to evaluate various aspects of healthcare[38] [39]. Problems that researchers need to pay attention on when applying simulation models on reality health care topics are also clarified. Such literatures were helpful in our research stage of building up the agent-based simulation model.

Results of our research in first stage were presented on IIE 2015 conference. One of our audiences provided us with valuable information: Patients with Alzheimer's disease can benefit from social activities such as supporting group and mental training activities. The influence of socially connections was of great interest for us. Unfortunately we don't have much information in our database that can be used to analyze the influence of social factors. A meta-analysis was then brought up to identify how social connections related to the progression of Alzheimer's disease in order to consummate our research. Literatures talking about social factors were widely reviewed and some valuable, informative research findings were summarized.

One of the main findings from the longitudinal cohort study of Wilson is that frequent participation in cognitively stimulating activities is associated with reduced risk of AD

[40]

[41]

[42]. In their study, the frequency of participation in such activities was mapped into a 5-scale rating and a one-point increase in cognitive activity score was associated with a 33% reduction in risk of Alzheimer's disease. Another useful finding is from Scarmeas's research

[43]

[44]

[44]. In this research, subjects' leisure activities were grouped into 13 different categories. A 13-scale rating was applied to measure subjects' participation in social activity and their data suggest that engagement in leisure activities may reduce the risk of incident dementia. More literatures that provide supporting information for these two findings were also referred.

Integrating the two findings together would provide us a more complete conclusion about how the social activities, including subjects' frequency in participating and the various categories of activity that subjects attending, might influence the progression of Alzheimer's disease. Applying such conclusion in our agent-based simulation model would also contribute to a more accurate evaluation about how different populations benefit from participating social leisure activities.

Through the thorough literature review, we not only clarified our mind about what results we are looking for and how to proceed on our research to get such result, but also gained support evidence of our research opinions from previous researchers.

Chapter 3: CDR and MMSE Comparison

As is discussed in previous sections, CDR and MMSE are two most popular evaluation methods in staging the severity of cognitive impairment. Some researchers claimed the validity and accuracy of CDR as it rated patients' impairment in six domains and integrated them into one overall category of severity. People also admitted that applying CDR is relative time consuming, as it requires a considerable amount of data collection from both patients and informants. MMSE, as comparison, is a much simpler and briefer assessment tool containing of several short cognitive probes. It takes no more than 30 minutes to accomplish the MMSE questionnaire. However, the main drawback of MMSE that researchers have been discussing is the its relative accuracy and validity. Some hold their opinion that MMSE is not able to provide a fully view of patients' cognitive status while others provided research evidence to show that MMSE results are reliable and trustworthy from statistical point of view.

The aim of this part of research is to identify factors that may affect the validation of each method by applying both methods on the National Alzheimer's Coordinating Center database and give advice about which method to use basing upon the individual's age, education level, living situation and other demographical facts.

3.1 Subjects

In the present study, we sought to describe the cognitive progression of persons with Mild Cognitive Impairment and aged 60 years older. We extracted demographic, cognitive and neuropsychiatric data from National Alzheimer's Coordinate Center (NACC). It is established in

1999 and maintains a cumulative database including clinical evaluations, neuropathology data when available, and now MRI imaging by investigating participants annually. NACC data are freely available to all researchers.

Throughout the whole database, we limited the subjects from the initial database to those who are 60 years or older, having a diagnosis of MCI at the initial data package at the very first visit of data collection and having been followed up for at least seven years in the NACC database.

Totally 505 subjects were selected in our research (241 Female, 264 Male). The mean age of participants was 73.406 years (SD = 7.754). The mean education year was 15.333 (SD = 0.472). Among the participants, 139 were living alone, 288 were living with spouse or partner, 61 were living with family or friends and the rest 17 were having other living situations. Nearly 80% of the participants were able to live independently without help on basic activities. At 7th year of visit, 158 out of 505 subjects progressed to Alzheimer's disease and 347 of them remained MCI.

3.2 Accuracy Comparison Using ROC

According to the global standard of mapping the CDR and MMSE results into the cognitive impairment stages (see Table 3.1), we were able to generate two sets of the evaluation of subjects' cognitive status for both methods. By comparing the predicted cognitive stages, from either MMSE or CDR, with the actual clinical diagnosis of patients' onset of Alzheimer's disease in the NACC database, the accuracy of both methods were analyzed.

Throughout widely literature review, we applied cutoff points 0.5 for CDR and 24 for MMSE in order to compare the accuracy and validity of these two methods. Among the 158 subjects who progressed to Alzheimer's disease after the 7th annual visit, 116 of them are

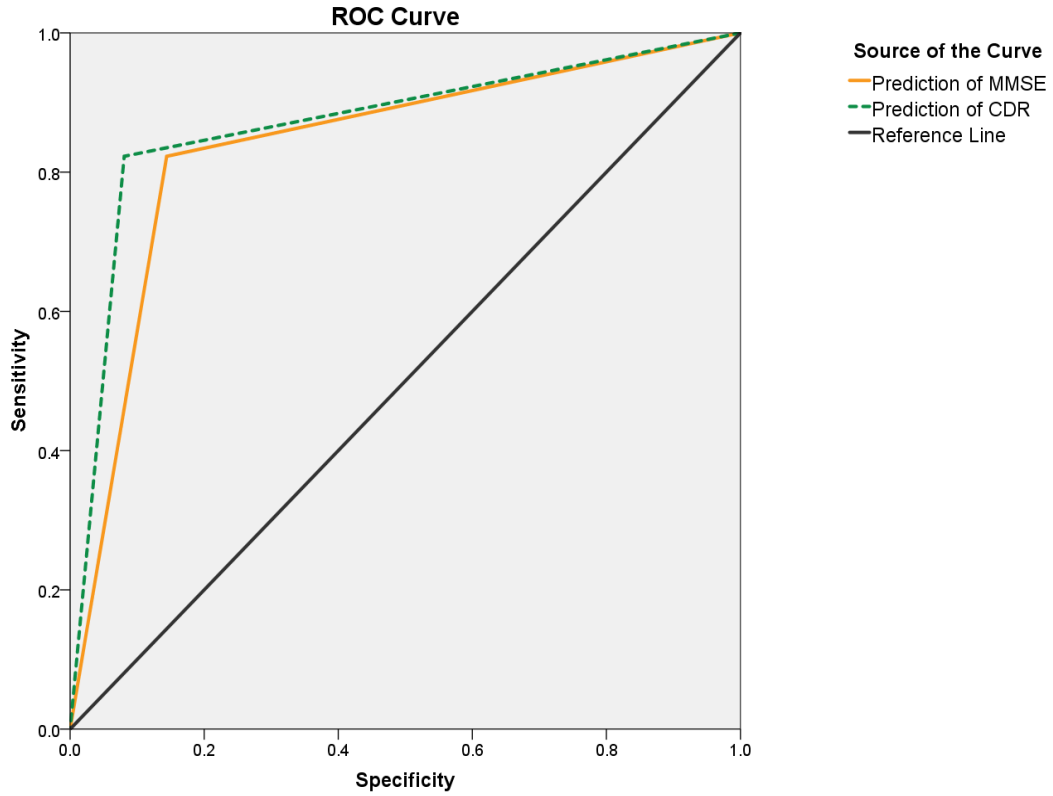
identified by MMSE method, 130 are identified by CDR. Similarly, among the 347 subjects who remained mild cognitive impairment, 310 of them are correctly excluded from Alzheimer's disease by MMSE, 319 are excluded from CDR. Such numbers provide a sense of validity for both methods.

Sensitivity and specificity were first calculated as a general evaluation of the methods' accuracy. As a result, the sensitivity and specificity for CDR are 82.28% and 91.93%, respectively, while 73.42% and 89.34% are for MMSE, respectively. Based upon the sensitivity and specificity results, ROC curve were plotted to visualize the difference for prediction accuracy of both methods, which is shown in Figure 3.1.

Table 3.1: CDR and MMSE Measurement Standard

Cognitive Status Stages / Composite Rating	Normal	Questionable Cognitive Impairment	Mild Cognitive Impairment	Moderate Cognitive Impairment	Severe Cognitive Impairment
CDR	0	0.5	1	2	3
MMSE	≥24		19 - 23	10 - 18	≤ 9

According to the prediction of the actual clinical evaluation of subjects' onset of Alzheimer's disease, CDR has a better performance as it has a larger area under ROC. However, as we stated before, to reach this better prediction of CDR, a much more complicated and professional data collection process is necessary. It might be problematic if a reliable and well-trained caregiver is not available. In comparison, even the area under ROC for MMSE is 3.2% less, the result is still acceptable/reliable in identifying cognitive impairment problems. And sometime MMSE is preferred for most subjects because it's easily accessible for most caregivers without a necessary access to well-train and professional physician.



Area Under the Curve	
Test Result Variable(s)	Area
MMSE_Prediction	0.839
CDR_Prediction	0.871

Figure 3.1: ROC curve for MMSE and CDR Predictions

3.3 Statistical Analysis to identify potential factors influenced the accuracy

As is stated in Chapter 2, some researchers claimed that demographical factors such as education level, age and gender are related with the accuracy of MMSE results in measuring cognitive status. In this case, MMSE cutoff points should be adjusted to reflect these influences. As to verify and investigate whether these demographical factors would be influential on either

one of the two methods and contribute to the difference of accuracy between CDR and MMSE, we applied statistical analysis methods including ANOVA, chi-square test, Fisher's exact test and etc.

According to the evaluation ratings from MMSE and CDR, we grouped the 158 subjects who progressed to Alzheimer's disease into 4 groups, as is shown in Table 3.2. Among the 158 subjects, 107 of them are detected in both methods, 9 are detected only in MMSE, 23 are detected only in CDR and 19 are not detected in either method. Similarly, the 347 subjects who remained MCI status were grouped into 4 groups too (see Table 3.3). For these 347 subjects, 298 of them are predicted correctly as no cognitive status decline in both methods, 12 receive incorrect positive predictions in CDR, 21 receive positive predictions in MMSE and 16 are predicted as Alzheimer's disease in both methods. Similar statistical methods were applied on these subjects with no cognitive decline in order to analyze the False Alarm of both methods.

Based upon widely literature review, the demographical characteristics that we selected for analysis includes Gender, Age, Education Level (in year), Living Situation, Onset of Hypertension/Hypercholesterolemia/Diabetes and Disability on Vision/Hearing. All these factors were somehow concluded related to the progression of cognitive impairment in previous literatures.

Table 3.2: Groups for subjects who progressed to AD

Number of Subjects		**MMSE Prediction**	
		+	-
CDR Prediction	+	107	9
	-	23	19

+ : Positive Predictions: subjects were predicted to AD, correct predictions.
- : Negative Predictions: subjects were predicted to remaining MCI.

Table 3.3: Groups for subjects who remained MCI

Number of Subjects		MMSE Prediction	
		-	+
CDR Prediction	-	298	12
	+	21	16

+: Positive Predictions: subjects were predicted to AD, false alarms.
-: Negative Predictions: subjects were predicted to remaining MCI.

Among these factors, Age and Education Year are continuous variables and we used t-test to compare whether the mean values for these two factors are significantly different in grouped subjects. Gender, Onset of Hypertension/Hypercholesterolemia/Diabetes and Disability on Vision/Hearing are dichotomous categorized variables. Living situation is recorded in four categories: Living alone, Living with spouse/partner, Living with relatives/friends and Living in groups. Such categorized variables were analyzed using chi-square test and fisher's exact test to identify significant differences in grouped subjects. As a result, p-values are summarized in Table 3.4 and Table 3.5.

Table 3.4: Statistical results of significant differences among AD subjects in describe characteristics

Characteristics	P-values		
	Four groups	Two groups Predicted +/- by MMSE	Two groups Predicted +/- by CDR
Gender	0.141	0.371	**0.035***
Age	0.945	0.773	0.747
Education Year	0.824	0.392	0.881
Living Situation	0.329	0.092	0.113
Hypertension	0.199	0.263	0.167
Hypercholesterolemia	0.250	0.843	0.369
Diabetes	0.385	0.418	1
Vision	0.877	0.823	0.810
Hearing	0.178	0.596	0.743

*: Significant factor under α=0.05

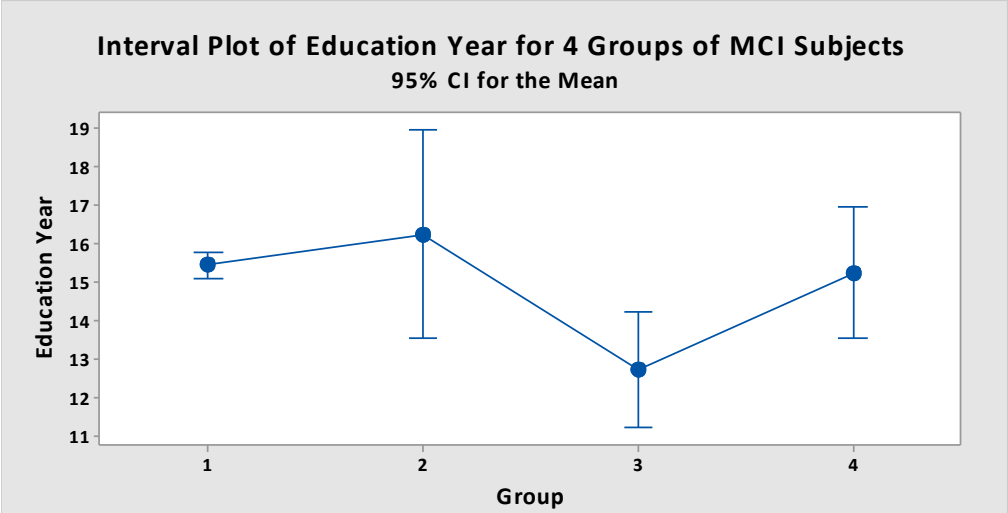
Table 3.5: Statistical results of significant differences among normal subjects in describe characteristics

Characteristics	P-values		
	Four groups	Two groups Predicted +/- by MMSE	Two groups Predicted +/- by CDR
Gender	0.999	1	1
Age	0.157	**0.048 ***	0.312
Education Year	**0.001***	**0.002 ***	0.508
Living Situation	0.590	0.986	0.428
Hypertension	0.604	0.579	0.674
Hypercholesterolemia	0.450	0.155	0.360
Diabetes	0.469	0.651	0.313
Vision	0.597	0.504	0.834
Hearing	0.291	0.423	0.229

*: Significant factor under α=0.05

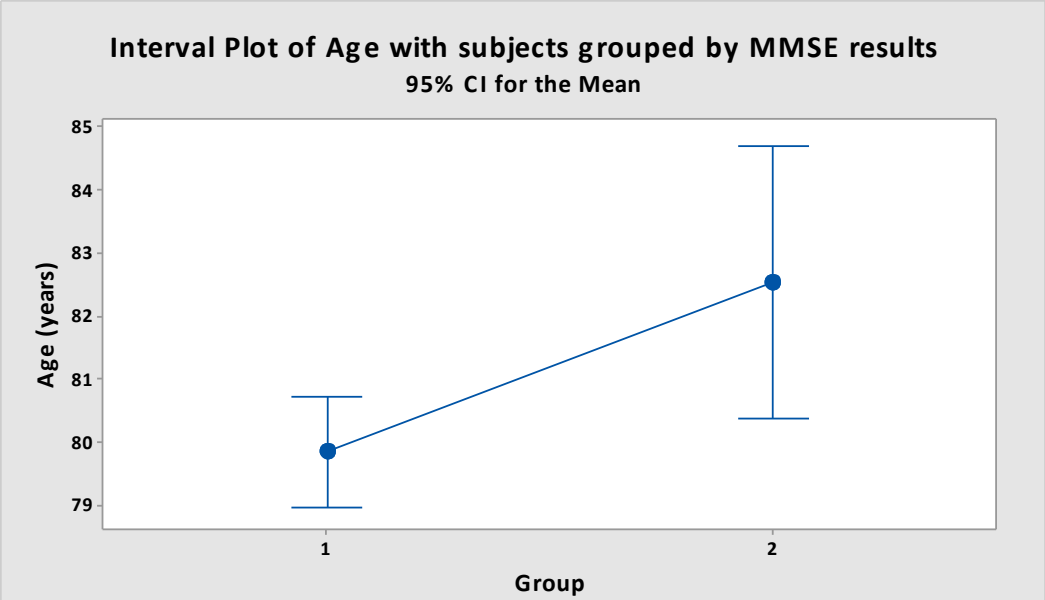
From statistical analysis, we find that, under significant level of 0.05, subject's gender may affect the prediction accuracy of CDR. Age and education year may attribute to more false alarms of MMSE. By looking deeper into the means of these characteristics in each group, we observed that more of the male subjects who actually progressed to Alzheimer's disease after 7 years were not detected correctly when using CDR. CDR prediction for female subjects are relatively more accurate. Similarly, Subjects who received false alarm from MMSE (i.e. normal subjects who had positive prediction using MMSE) tend to have a lower education level and higher age. Interval plots, see Figure 3.2 and 3.3, were generated to help visualize the differences in education level and age among different grouped subjects.

To conclude, in this session, we brought up a thorough analysis to compare the results of prediction from the clinical measurement (CDR) and cognitive scales (MMSE). Simply considering the prediction accuracy, in general view we find that CDR is relative more accurate than MMSE results based upon the specificity and sensitivity results, as well as the ROC curve.



Group 1: MCI Subjcets accurately predicted by MMSE and CDR
Group 2: MCI Subjects false-alarmed only by CDR
Group 3: MCI Subjcets false-alarmed only by MMSE
Group 4: MCI Subjects false-alarmed by CDR and MMSE

Figure 3.2: Interval plot of Education Year among 4 subject subsets.



Group 1: Subjects who remained MCI and got accrate prediction by MMSE
Group 2: Subjects who remained MCI and got false-alarms by MMSE

Figure 3.3: Interval plot of Age among subjects grouped by MMSE.3.4 Discussion

However, to pursue a clinical measurement result, a professional physician or well-trained caregiver is necessary. The difficulty to acquire an accurate result should be taken into consideration when applying such clinical measurement. On contrary to CDR, MMSE could

provide an acceptable result even not as well as CDR. But caregivers get easily apply such method to measure the cognitive status of patients. This easiness for apply is not simply saving time and money. The main advantage is that patients could be evaluated frequently and regularly. In this case caregivers can keep track of the patients' cognitive status and provide treatments correspondingly.

From the detailed analysis by grouping the patients and apply statistical methods in order to investigate whether some demographical variables such as age and gender would affect the prediction accuracy, we were able to conclude clearly that the demographical factors that might influence the accuracy of MMSE include Age and Education Level, which is consistent with the findings from previous researches. Also, gender is influential on the accuracy of CDR as female might gain better results from this method. Once the demographical effects were able to be determined, we can discuss on which method might give a better result on each individual based upon his/her own set of demographic status.

Although adjusting the measurement cut-off points for different education level or age might be a good research idea to discuss, we were not able to determine the rationality behind changing the globally accepted rating standards. Also, a more reasonable explanation is: the demographical variables might results in influencing the cognitive status change instead of simply influencing the prediction results. Along this direction, we brought up the idea to identify the influence of demographical factors on subject's actual cognitive status progress and use logistic regression model to predict whether one might be progressed to Alzheimer's disease or not. This part would be discussed in Chapter 4.

Chapter 4: Logistic Regression Model

We thoroughly compared the accuracy and validity of clinical rating (CDR) and cognitive scaling (MMSE) methods in Chapter 3. Comparing the ROC performance, CDR has better prediction accuracy. However, some interesting findings were brought up by applying statistical analysis on demographic variables to help explain the accuracy differences. From the result, we noticed that age, gender and education level might influence the accuracy MMSE in measuring the cognitive impairment status.

Such findings gave us a hint of identifying the potential demographic risk factors that may influent the onset of Alzheimer's disease. Similar statistical analysis methods in Chapter 3 were applied on the dataset to investigate whether there were significant differences in the demographic variables among subjects who progressed to Alzheimer's disease and those who remained mild cognitive impairment at the 7th annual visit to National Alzheimer's coordination center. Methods and results would be discussed in this section.

4.1 Identify Demographical Risk Factors

According to the research of Fratiglioni (1997) and Gao, et al (1998) [错误! 未定义书签。, 错误! 未定义书签。], in which they claimed that female tend to have higher risks to develop dementia, and the statistical results from Chapter 3, in which we found that gender is an influential factor in the accuracy of measuring, we decided to separate female and male subjects into two groups for further analysis in this session.

Similar to previous discussion, the demographic variables we selected include Age, Education Level (in year), Systolic/Diastolic Blood Pressure, Education Year, Heart Rate, Living Situation, Living Dependence Level, Smoking and Drinking hobbies and Onset of Hypertension/Hypercholesterolemia/Diabetes. T-test and chi-square test were applied to identify significant differences in subjects who remained MCI and who progressed to Alzheimer's disease at 7th visit depending upon whether the variable is continuous or categorized.

Among all these factors, age, education year, Systolic/Diastolic blood pressure and heart rate are continuous variables. Categorized variables are explained as following:

- Handedness is categorized into two groups (left/right). Living situation is recorded in four categories: Living alone, Living with spouse/partner, Living with relatives/friends and Living in groups.
- Living Dependence is grouped into four groups: Able to live independently, Requires some assistance with complex activities, Requires some assistance with basic activities and Completely dependent.
- Marriage status has six groups: Married, Widowed, Divorced, Separated, Never married and Living as married.
- Three measurements were selected for measuring one's smoking history: Whether smoked or not, Year of smoking history and number of packs smoked per day (<0.5, 0.5-1, 1-1.5, 1.5-2, >2).

It can be concluded from the results that for female subjects, the potential risk factors are Age, Education Year, Living Situation and onset of diabetes. Subjects who progressed to Alzheimer's disease tend to be older and having higher education level. Onset of Diabetes might

increase the risk of Alzheimer's disease for female subjects. Similarly, for male subjects, handedness and living situation are influential for those who progressed to Alzheimer's disease. Statistical results were summarized in Table 4.1 and Table 4.2

Table 4.1: Statistical results of risk factors identification for Male

Risk Factor		**Test**	**P-value**	**Description**
Age		t-test	0.622	No significant difference.
Systolic Blood Pressure		t-test	0.572	No significant difference.
Diastolic Blood Pressure		t-test	0.461	No significant difference.
Education Year		t-test	0.100	No significant difference; Larger mean for normal subjects.
Heart Rate		t-test	0.593	No significant difference
Left/Right Handed		chi-square	**0.027***	Less demented subjects for left-handed.
Living Situation		chi-square	**0.005***	More demented subjects for those who don't live alone.
Living Dependence		chi-square	**0.000***	More demented for subjects who can't live independently.
Marriage Status		chi-square	**0.009***	More demented subjects for those married one; More normal subjects for the divorced ones.
Hypertension		chi-square	0.884	No significant difference.
Hypercholesterolemia		chi-square	0.226	No significant difference.
Diabetes		chi-square	0.671	No significant difference.
Alcohol		chi-square	0.369	No significant difference.
Smoke	Smoke (Y/N)	chi-square	0.198	No significant influence on dementia for smoking.
	Yr of Smoke	t-test	0.168	
	Packs/Day	chi-square	0.271	
Vision	Normal	chi-square	0.976	No significant influence on dementia for vision
	Wear lenses	chi-square	0.594	
Hearing	Normal	chi-square	0.320	No significant influence on dementia for hearing.
	Wear aids	chi-square	0.145	

Table 4.2: Statistical results of risk factors identification for Female

Risk Factor		Test	P-value	Description
Age		t-test	**0.002***	Mean age of Demented subjects is greater.
Systolic BP		t-test	0.158	No significant difference; More upper outliers for Normal
Diastolic BP		t-test	0.987	Mean of Demented subjects is lower.
Education Year		t-test	**0.002***	Mean is lower for Normal
Heart Rate		t-test	0.508	No significant difference
Handed		chi-square	0.734	
Living Situation		chi-square	**0.000***	More demented subjects for those who don't live alone.
Living Dependency		chi-square	**0.000***	More demented for subjects who can't live independently.
Marriage Status		chi-square	0.192	More demented subjects for those married one; More normal subjects for the divorced ones.
Hypertension		chi-square	0.11	More demented subjects for those who recently got hypertension or has inactive hypertension; More normal subjects for absent ones.
Hypercholesterolemia		chi-square	0.712	More demented subjects for those who recently had hypercholesterolemia or inactive ones. More normal subjects for absent ones.
Diabetes		chi-square	0.015	
Alcohol		chi-square	0.392	More demented subjects for those abuse alcohol.
Smoke	Smoke?	chi-square	0.061	No significant influence on dementia for smoking.
	Yr of Smoke	t-test	0.009	
	Packs/Day	chi-square	0.150	
Vision	Normal	chi-square	0.722	No significant influence on dementia for vision
	Wear lenses	chi-square	0.664	
Hearing	Normal	chi-square	0.502	More demented subjects for those who have problem in hearing.
	Wear aids	chi-square	0.505	

4.2 Logistic Regression Model

All the identified risk factors identified in the previous session were summarized as the input variables of logistic regression model. The generated logistic model is then used to estimate the likelihood of progressing to Alzheimer's disease (the dichotomous response) for each individual subject who has unique set of combined values of identified risk factors. Calculated results via logistic regression were then mapped into our prediction for whether the subjects progressed to Alzheimer's or not by grouping those whose probability is greater/less than 0.5.

Formula 4.1 shows the generated logistic model that can be used for prediction of Alzheimer's disease. Table 4.3 and Table 4.4 are coefficients for each level of living situating and living dependency in both logistic regression models.

Formula 3.1: Logistic Model for Female subjects.

$$\mu\_male = - 0.99185 - 0.01668 * Age - 0.07453 * EduYr + ls + ld + 0.05446 * Dia + 0.51238 * Hyper - 0.16588 * Hycho$$

$$\mu\_female = -8.11811 + 0.02570 * Age + 0.16443 * EduYr + ls + ld - 0.93669 * Dia - 0.70128 * Hyper + 0.60903 * Hycho$$

$$P(AD) = EXP(\mu) / (1 + EXP(\mu)).$$

Where,

Age is the subject's age,

EduYr is the education year of subjects,

ls is the coefficient indicating subject's living situation[a],

ld is the coefficient indicating subject's living dependence level[b],

Dia,Hyper,Hycho are dichotomous variable showing onset of Diabetes, Hypertension and Hypercholesterolemia.

Table 4.3: Living Situation (ls) coefficient value summary

For Male Subjects:		
LIVSIT	**Coefficient Value**	**Comment**
1	0.00000	Lives alone
2	1.51265	Lives with spouse or partner
3	1.05007	Lives with relative or friend
4	1.09824	Lives with group
5	1.60590	Other

For Female Subjects:		
LIVSIT	**Coefficient Value**	**Comment**
1	0.00000	Lives alone
2	0.92305	Lives with spouse or partner
3	0.93984	Lives with relative or friend
4	1.18757	Lives with group
5	0.80800	Other

Table 4.4: Living Dependenz(ld) coefficient value summary

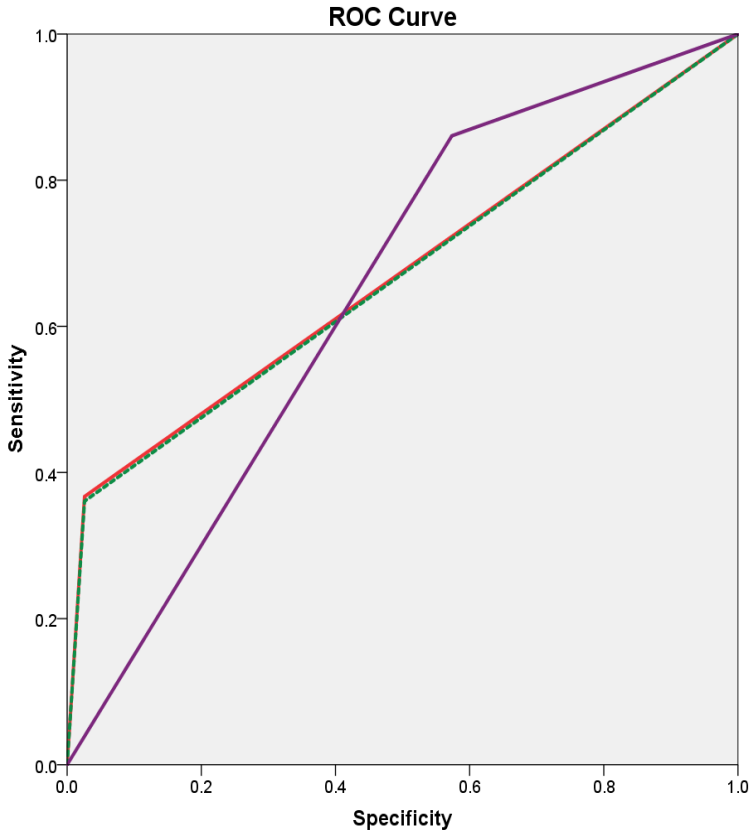
For Male Subjects:		
INDPEND	**Coefficient Value**	**Comment**
1	0.00000	Able to live independently
2	1.71508	Requires complex activity assistance
3	3.10540	Requires basic activity assistance
4	2.82090	Completely dependent

For Female Subjects:		
INDPEND	**Coefficient Value**	**Comment**
1	0.00000	Able to live independently
2	2.94373	Requires complex activity assistance
3	3.92581	Requires basic activity assistance
4	4.24581	Completely dependent

We can get another set of data for the prediction results of logistic model by examining the demographical variables of all the subjects. Sensitivity and Specificity of such prediction outcome from logistic regression model are 88.14% and 41.44% respectively. ROC curve were plotted to compare the prediction results of logistic model, CDR measurement and MMSE scales, see Figure 4.1. The cutoff points we used for CDR and MMSE (1 and 24 respectively) in previous sessions should be adjusted to the criteria of predicting Alzheimer's disease. The criteria were moved closer to the dementia status and resulted in 2 and 18 as new cutoff points for CDR and MMSE.

From Figure 4.1, it's obvious that although the total area under each method was pretty close to each other, the sensitivity of logistic method is much higher than the other two methods. This indicates that logistic regression model is relative "aggressive" of giving more positive predictions. In this case, fewer subjects who might progress to Alzheimer's disease were omitted while more normal subjects may be false alarmed.

This indicates that this model is really "aggressive" of giving more positive predictions. In this case, fewer subjects who might progress to Alzheimer's disease were omitted while more normal subjects may be false alarmed. The clinical measurement (CDR) and cognitive scales (MMSE) might win according to a high specificity performance (less false alarms), but they're still not acceptable because of the extremely low sensitivity, i.e. the ability to give positive prediction for subjects progressed to Alzheimer's disease.



Area Under the Curve	
Test Result Variable(s)	Area
Prediction of Logistic	0.684
Prediction of MMSE	0.671
Prediction of CDR	0.667

Figure 4.1: ROC curve for Logistics, MMSE and CDR Comparison

4.3 Discussion

Along the direction of providing individual suggestions based on people's different demographic status, we brought up a logistic regression model to predict whether one might be progressed to Alzheimer's or not. Such model is generated by first applying statistical methods on the demographic variables from our database. We were able to identify two sets of potential risk factors for both female and male subjects. Putting these factors' information as the input variables of logistic regression model, we were able to develop two logistic regression models for both female and male. By calculating the sensitivity and specificity and plotting the ROC curve, we found that the prediction result has much higher sensitivity than both clinical CDR

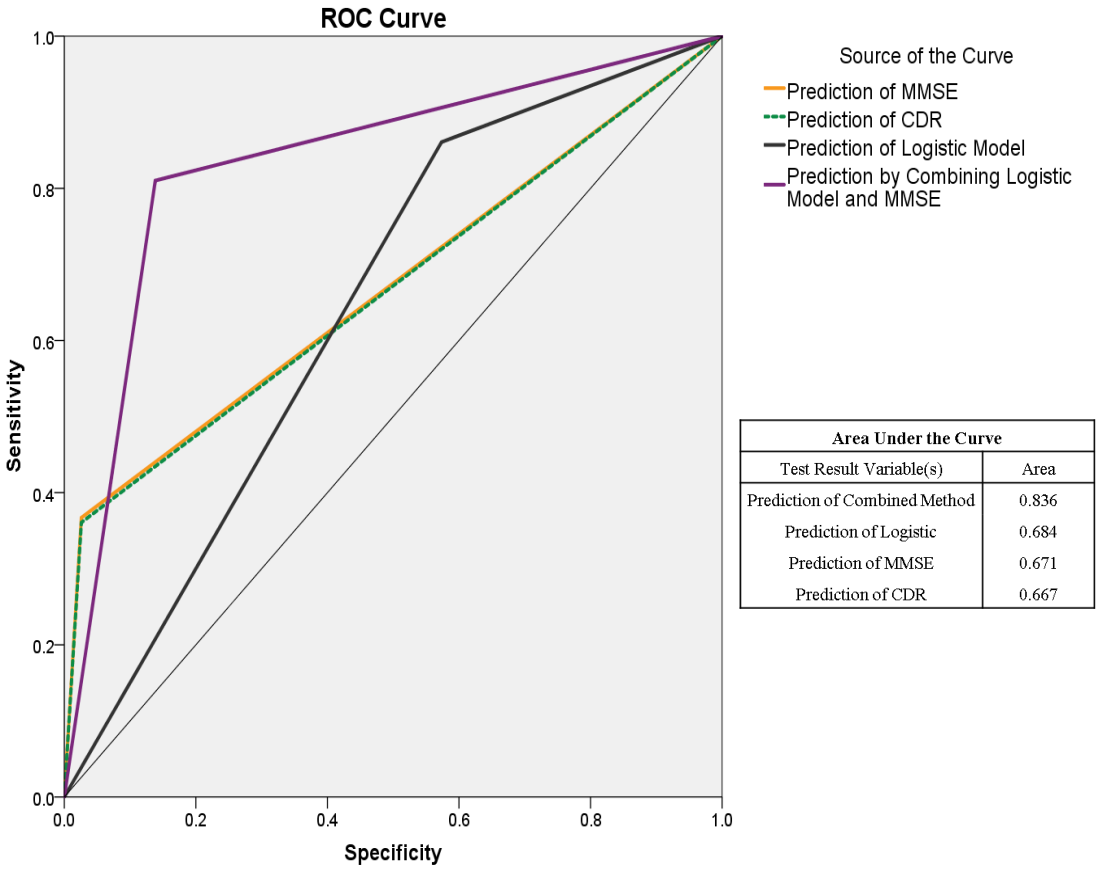
method and cognitive MMSE scale. The low specificity of the regression model is the main drawback of this method, which indicates that people are more likely to get positive results even they're normal, i.e. get false alarms.

The prediction method using logistic regression model is relative aggressive in identifying potential subjects who might progress to Alzheimer's disease. However, combining the regression model and MMSE method might contribute to an easy-to-use, accurate prediction. Better result can be achieved by combining the regression model and easily accessible MMSE measurement. For those who resulted in a positive prediction in the regression model, we will suggest them to do a further test using MMSE measurement. If the MMSE were below the normal cognitive criteria (24 points), one would be confidentially predicted as potentially undergoing a high risk of AD. By so, a prediction measurement of AD, with highly personalization, easily access and relative high accuracy was brought up. In this case, MMSE measurement is used to filter those who are falsely predicted to have high possibility to progress to Alzheimer's disease out of the positive predictions.

The sensitivity and specificity are calculated based upon this method, which integrates the MMSE measurement, and logistic regression method. The results are: 86.71% and 81.01%, respectively. We can conclude that this method achieves a better result than all the other three measurements and its sensitivity and specificity are both acceptable.

We then proposed an idea of visualizing the change of cognitive impairment status of patients by simulating their changes in demographical variables such as age and living situation. Agent-based simulation model is built up based upon all the researches we've done. Through simulation modeling, we can also simulate the performance in Alzheimer's progression in

different populations and investigate the potential effect of mental and physical treatments. This part would be discussed in Chapter 5.



Area Under the Curve	
Test Result Variable(s)	Area
Prediction of Combined Method	0.836
Prediction of Logistic	0.684
Prediction of MMSE	0.671
Prediction of CDR	0.667

Figure 4.2: ROC curve for Logistics, MMSE, CDR and Combined Method Comparison

Chapter 5: Agent-based Simulation Model

From the discussion in Chapter 4, we realized that some demographical factors, including age, education level, gender etc., are influential to the onset and progression of Alzheimer's disease. Their influence about to what degree they are affecting Alzheimer's disease is reflected in the logistic regression model discussed in the previous session. Such influences indicate that different population with various demographical characteristics, such as people's living situation, age and education level, might perform differently when we look into the progression of Alzheimer's disease among the whole population.

To visualize the changes of cognitive impairment status and to predict the probability of patients reaching the unfavorable outcome, i.e. the onset of Alzheimer's disease, among the whole population along time, we integrated the logistic regression model into an agent-based simulation model.

The reason for us to use the agent-based simulation technique is that it's applied to simulate the actions and interactions of autonomous agents based upon the clear analysis of their attributes as well as the interaction between their surrounding agents and local environment. It is also powerful on simulating the heterogeneity of subjects such as their individual differences on progressive process, which is caused by the physical, psychological and social factors.

In this chapter, we would be discussing in detail about how we gathered all the information for the inputs of the simulation model, how we integrated the logistic regression model as well as social factor influences into simulation and what are some findings from the simulation output.

5.1 Simulation Initial Inputs

From the study discussed in Chapter 4, we know that the influential demographical factors that are used as input variables in the logistic regression model are: Age, Education Year, Living situation, Living dependence level, Onset of Diabetes, Onset of Hypertension and Onset of Hypercholesterolemia. Moreover, we brought up two logistic regression models for female and male since they perform significantly different on the progression of Alzheimer's disease.

- Age – Both female and male subjects' age are following Normal distribution. For female subjects, their ages are following Norm (73.79, 8.09). For male subjects, their ages are following Norm (74.41, 7.81). Data frequency plots and normality test results are shown in Figure 5.1 – 5.4

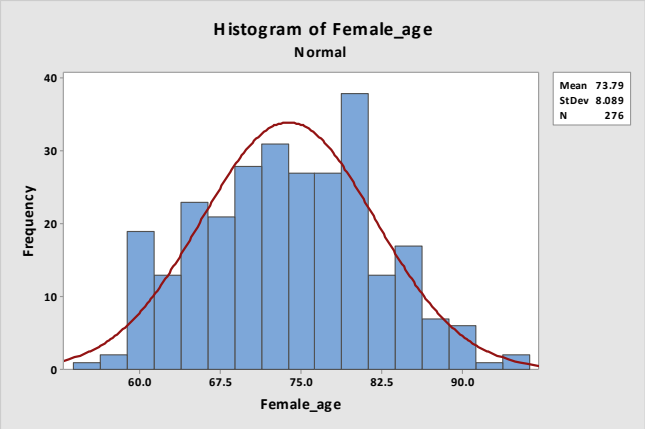


Figure 5.1: Histogram of Age – Female

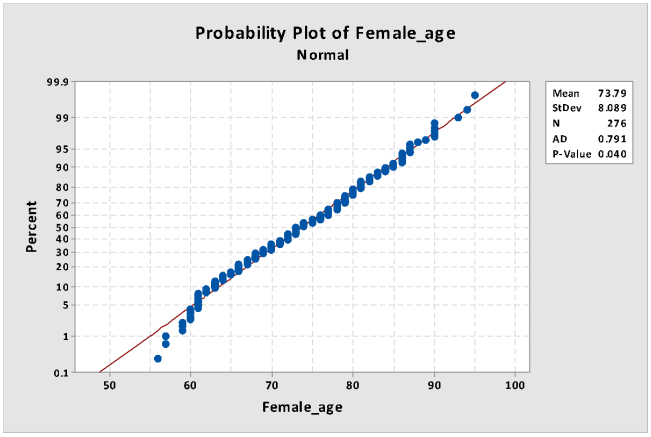


Figure 5.2: Normality Test Plot for Age – Female

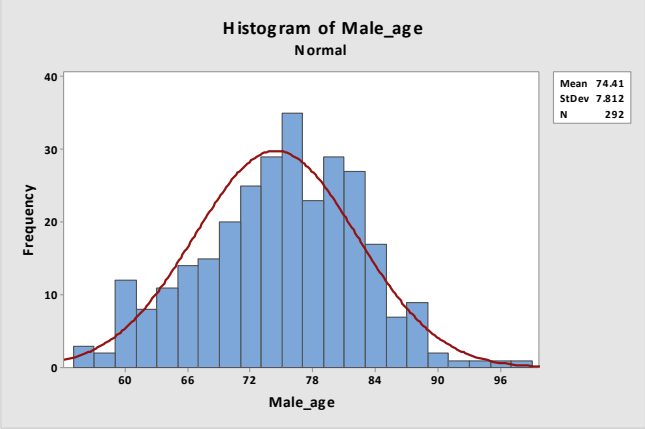


Figure 5.3: Histogram of Age – Male

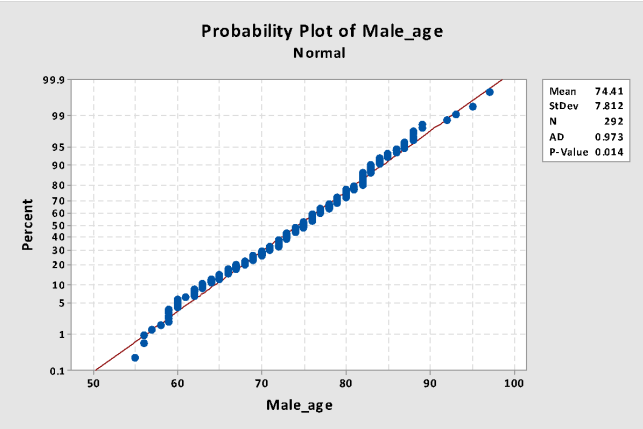


Figure 5.4: Normality Test Plot for Age – Male

• Education Year – In order to get an accuracy input for education year, data points were first grouped into Education levels and then distributed into education year using empirical distribution. Totally 6 levels of education were categorized: Less than elementary school (≤ 6 years of education), Middle school (7–9 years of education), High school (10–12 years of education), Bachelor's degree (13–16 years of education), Master's degree (17–20 years of education) and PhD or higher (≥21 years of education). Both female and male subjects' education levels are following normal distribution. The histograms for education level are shown in Figure 5.5 and 5.6. For female subjects, their education level follows Norm (3.92, 0.94). For male subjects their education follows Norm (4.22, 0.86). There are statistically significant differences among education levels for female and male. Year of education were then generated using empirical distribution based upon the subjects' education level. The histograms of education year are shown in Figure 5.7 and 5.8.

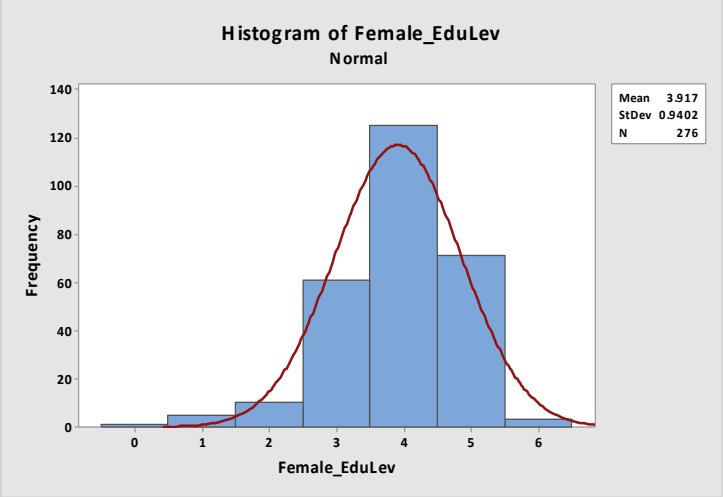


Figure 5.5: Histogram of Categorized Education Level – Female subjects

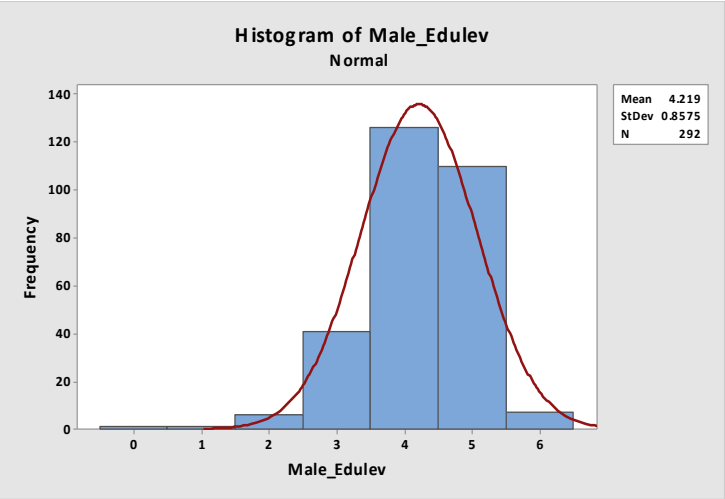


Figure 5.6: Histogram of Categorized Education Level – Male subjects

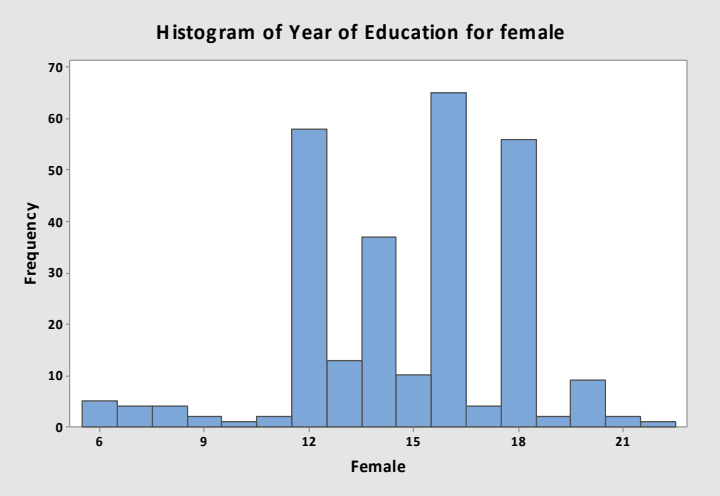


Figure 5.7: Histogram of Education Year – Female subjects

- Living Situation – Subjects are grouped into 5 types of living situation. At the initial year of visit, 41.0% female subjects live alone (level 1), 46.1% female live with spouse/partner (level 2), 11.8% female live with relatives/friends (level 3), the rest 1.1% female live with group such as senior community (level 4), and none of them had other situations (level 5). For male subjects, 11.0% live alone, 84.9% live with spouse/partner, 3.4% live with relative/friends, and the rest 0.7% live in group. Histograms are plotted in Figure 5.9 and 5.10.

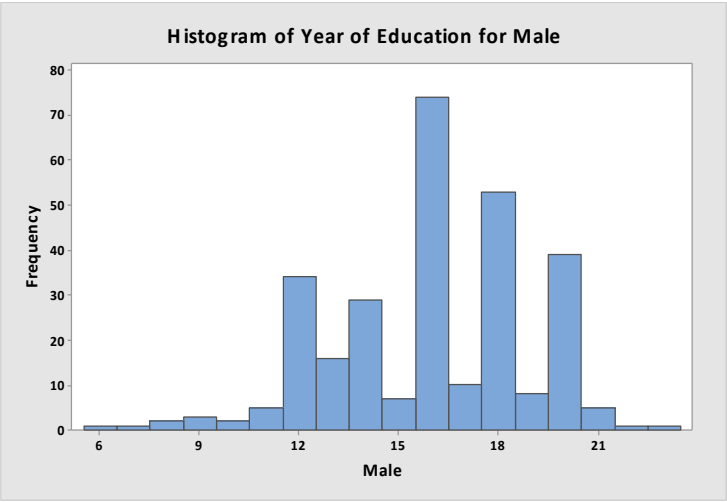


Figure 5.8: Histogram of Education Year – Male subjects

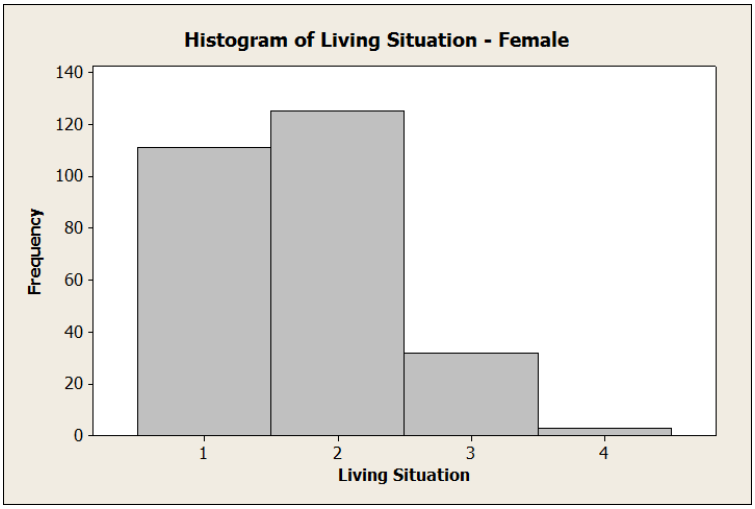


Figure 5.9: Histogram of Living Situation – Female subjects

- Living Dependency Level – Similar as living situation, subjects are grouped into 4 levels of living dependency. Only 3 levels were shown in the dataset for initial visit. Female and male have different distributions among each group. For female subjects, 83.3% are able to live independently, 14.9% require some assistance with complex activities, and the rest 1.8% require some assistance with basic activities. For male subjects, 75.7% are able to live independently, 22.2% require some assistance with complex activities, and the rest 2.1% require some assistance with basic activities. Figure 5.11 and Figure 5.12 are histograms of initial data for living dependency

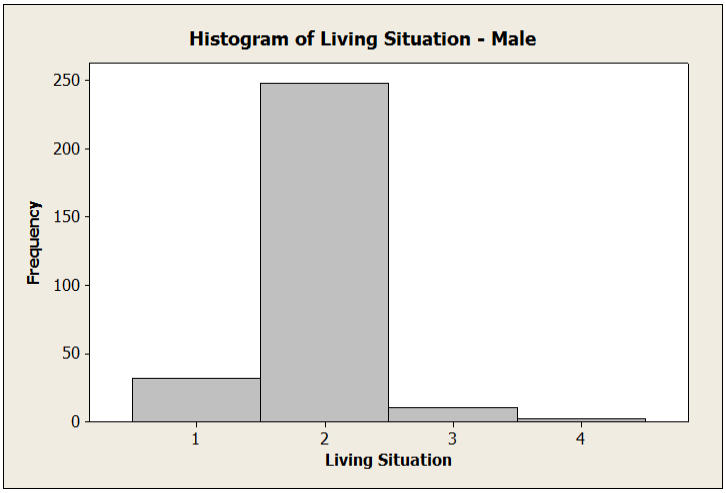


Figure 5.10: Histogram of Living Situation – Male subjects

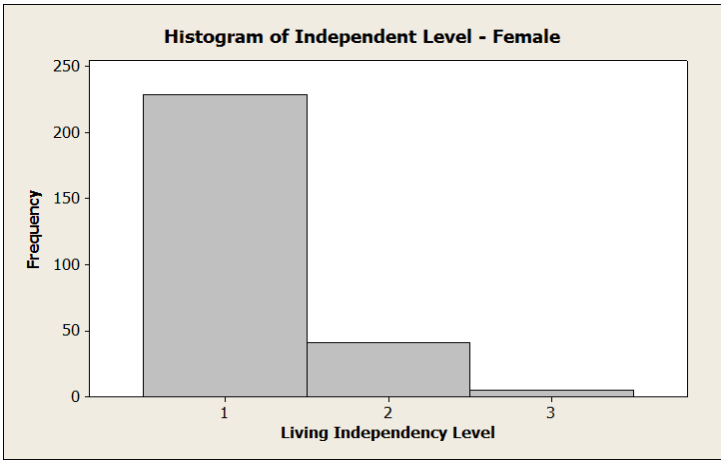


Figure 5.11: Histogram of Living Independence – Female subjects

- Hypertension – Female and Male subjects do not have significant difference among their onset history of hypertension. So only one set of input data were applied for hypertension. Among all subjects, 53.2% of them are reported to have (or used to have) hypertension. 14.4% of these subjects who didn't report absent of hypertension have a Remote/Inactive hypertension status (i.e. used to have hypertension). Histogram is shown in Figure 5.13.

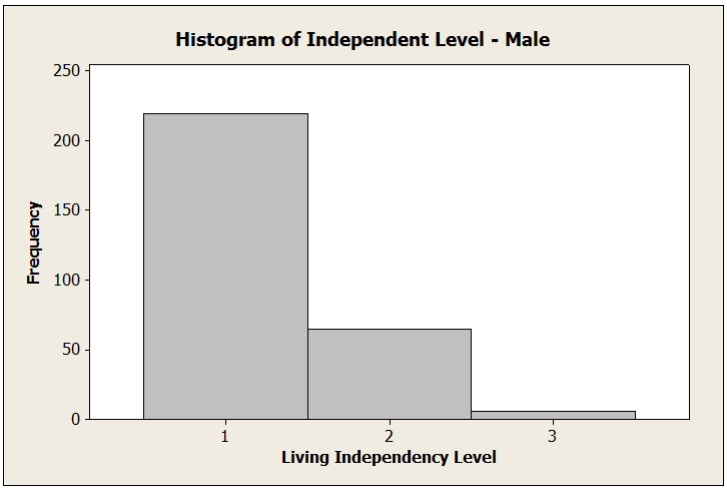


Figure 5.12: Histogram of Living Independence – Male subjects

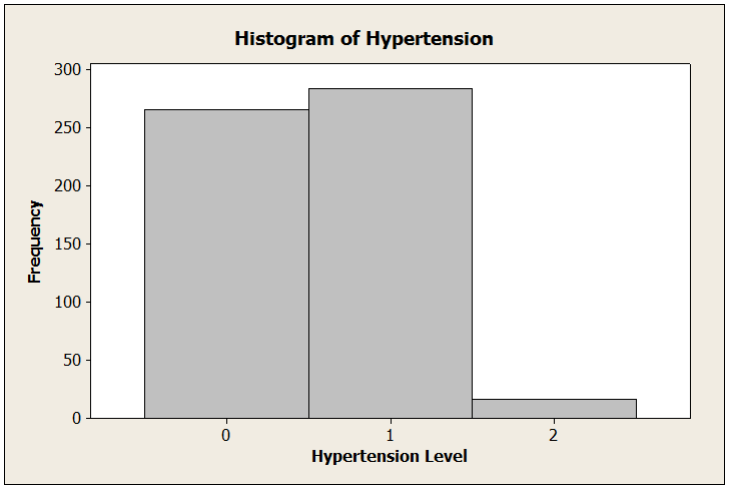


Figure 5.13: Histogram of Hypertension

- Hypercholesterolemia – Two sets of inputs were applied for Hypercholesterolemia because of the significant difference between female and male. Among all female subjects, 43.1% of them reported absent of Hypercholesterolemia. 84.5% of the rest subjects have an active hypercholesterolemia and the other 15.5% of them reported a remote/inactive status. Histogram is shown in Figure 5.14. Similar to Male subjects, 61.3% of them reported to have/used to have Hypercholesterolemia. 84.5% of them are having the active status. See Figure 5.15 for the histogram.

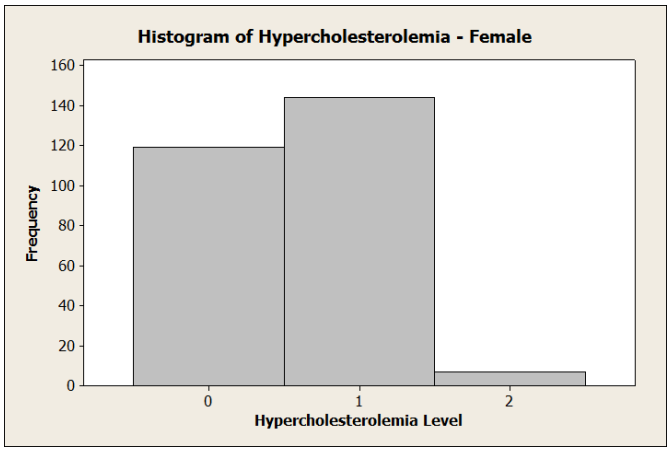


Figure 5.14: Histogram of Hypercholesterolemia – Female

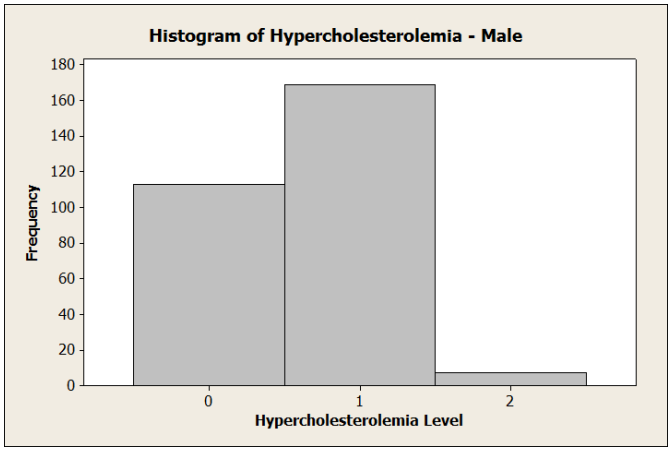


Figure 5.15: Histogram of Hypercholesterolemia – Male

- Diabetes – Similar to the settings for Hypercholesterolemia, only 7.6% Female subjects reported non-absent of Diabetes. 81.4% of them have an active state. For male subjects, 12.3% of them reported non-absent of Diabetes and all of them have an active state.

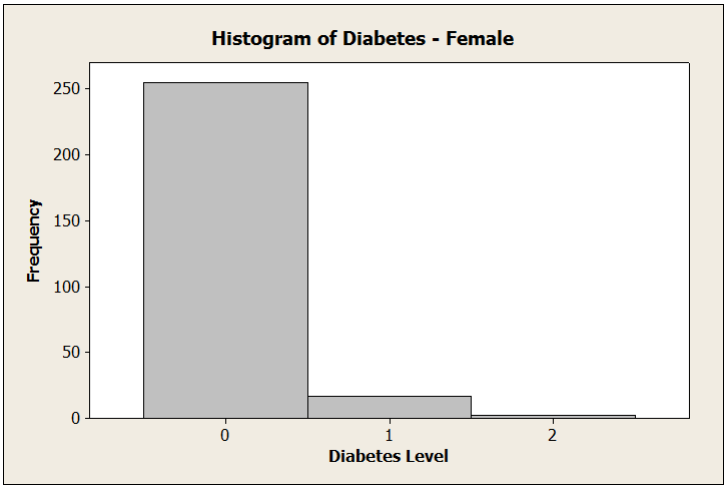


Figure 5.16: Histogram of Diabetes - Female

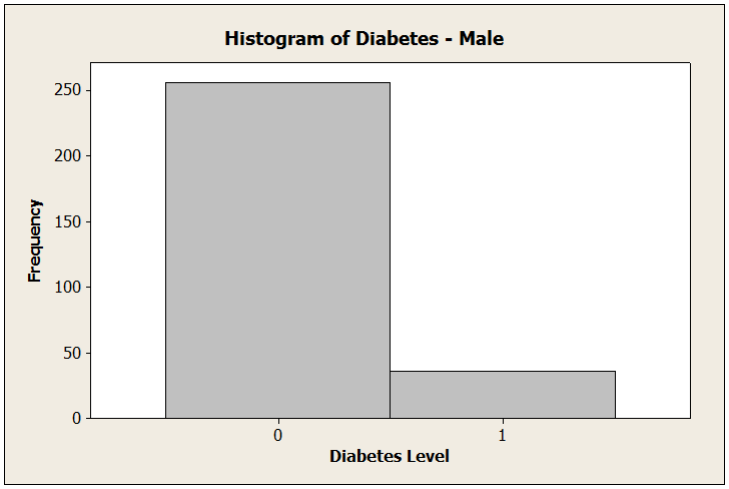


Figure 5.17: Histogram of Diabetes - Male

5.2 Input Variable Progression

Netlogo is used for the simulation as it's a high level platform making it possible to explore the behavior of agents and the dynamic system performances at macro-level. In Netlogo, agents are progressing along ticks. Throughout our simulation model, the tick unit is one month. In this case, the input variables are changing / varying at a monthly pace.

- Age – Subjects' age increase by 0.083 for every tick because the tick unit is one month. Age increases by 1 every 12 ticks

- Education Year – As all the subjects are aged greater than 65 years, we assumed they are not getting education anymore. So the education year wouldn't change.
- Living Situation – Subjects' living situations are grouped into 5 levels: live alone, live with spouse/partner, live with relatives/friends, live with group such as senior community and other situations (level 1 to 5 correspondingly). From the 7-year longitudinal data, number of changes in living situation level within 1 year was calculated and then used as the progression rate for living situation. Since the living situation is not likely to change very frequently (like within 1 month), so we assume that the living situation would only change every 12 ticks (every year). Their progression rates for both female and male were summarized in Table 5.1. The histogram for 7-year data of living situation is also shown in Figure 5.18 and Figure 5.19.

Table 5.1: Progression for Living Situation – Female and Male

Init Level / Delta in Level	Percentage for progression									
	Female					Male				
	1	2	3	4	5	1	2	3	4	5
-4	0.00%	0.00%	0.00%	0.00%	14.89%	0.00%	0.00%	0.00%	0.00%	0.00%
-3	0.00%	0.00%	0.00%	2.50%	0.00%	0.00%	0.00%	0.00%	0.00%	16.67%
-2	0.00%	0.00%	13.40%	0.00%	4.26%	0.00%	0.00%	4.48%	0.00%	0.00%
-1	0.00%	3.58%	1.03%	2.50%	4.26%	0.00%	2.28%	4.48%	7.14%	0.00%
0	90.53%	94.28%	82.47%	90.00%	76.60%	88.53%	96.82%	88.06%	92.86%	83.33%
1	1.18%	0.72%	2.06%	5.00%	0.00%	6.88%	0.28%	2.99%	0.00%	0.00%
2	4.14%	0.43%	1.03%	0.00%	0.00%	3.21%	0.28%	0.00%	0.00%	0.00%
3	1.33%	1.00%	0.00%	0.00%	0.00%	0.92%	0.35%	0.00%	0.00%	0.00%
4	2.81%	0.00%	0.00%	0.00%	0.00%	0.46%	0.00%	0.00%	0.00%	0.00%

- Living Dependency Level – As is mentioned previously, living dependency level are grouped into 4 categories: Able to live independently, Require some assistance with complex activities, Require some assistance with basic activities, and Completely dependent. It's obvious that subject's living independency is getting worse from the first category to the third. Also, the decline of independency ability is a slow progression process. In this case,

we're assuming that, for the progression of Living dependency level, one can only progress by 1 level along 1 tick, which means a subject with ability to live independently cannot progress to the dependent level that needs some assistance with basic activities within one month.

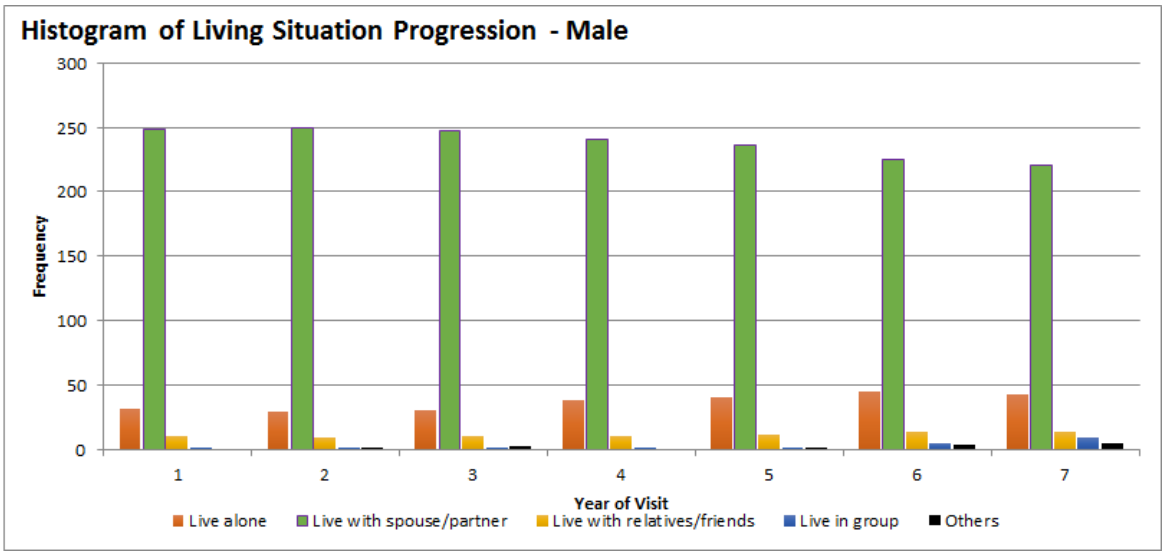


Figure 5.18: Histogram of Living Situation Progression – Male

- As a result, the average progression rates for each level of living dependency, within one tick, were calculated, See Table 5.2. All these progression rates are calculated from the average progression rate among the 7 year of visit data (7-year histogram is shown in Figure 5.20 and Figure 5.21)

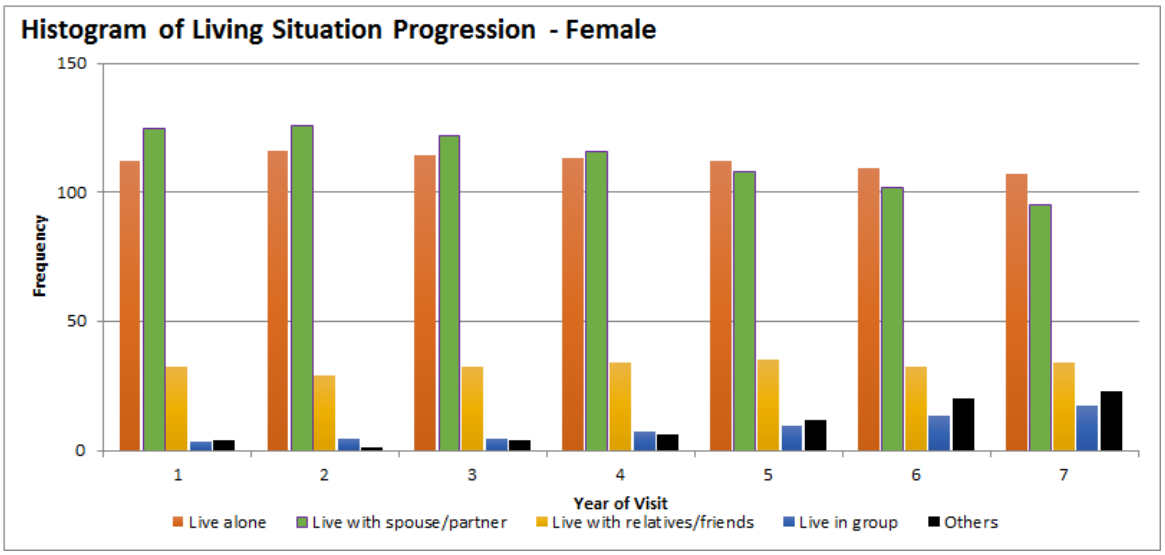


Figure 5.19: Histogram of Living Situation Progression – Female

Table 5.2: Progression Rate for Independency Level – Female and Male

Level	Description	Progression Rate	
		Female	Male
1	Able to live independently	0.68%	0.52%
2	Requires assistance with complex activities	1.64%	0.99%
3	Requires assistance with basic activities	1.99%	0.81%
4	Completely dependent	Final Status	

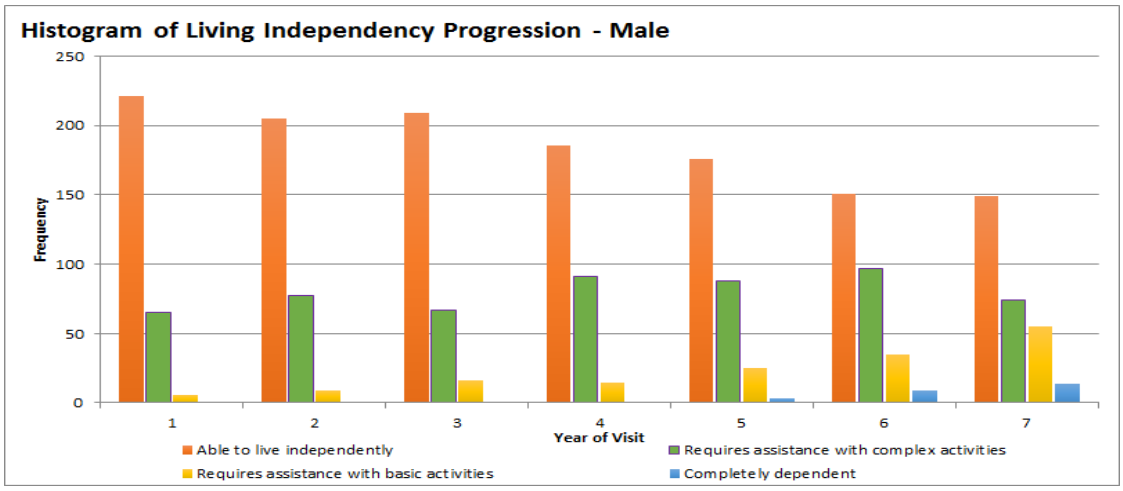


Figure 5.20: Histogram of Living Independency Progression – Male

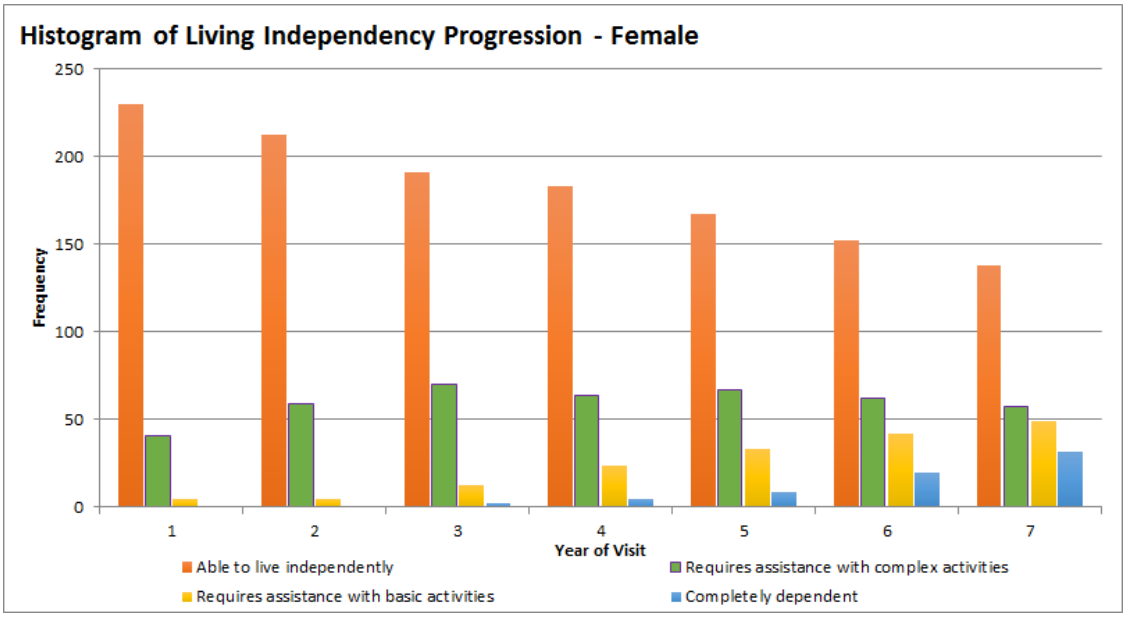


Figure 5.21: Histogram of Living Independency Progression – Female

- Hypertension – Similar to living dependency progression, we have three different categories for Hypertension. These three levels of hypertension are: totally absent, recent/active, and remote/inactive.

  Our assumption on the progression of hypertension is that only normal subjects with absent (level 0) hypertension type can either progress to active (level 1) or progress to inactive (level 2) hypertension. Also, for those who progressed from absent to active/inactive, they cannot progress back to normal.

  The average progression rate for normal subjects, both female and male, calculated from the 7-year data is 0.426%. Among these subjects who recently progressed to hypertension, 85.56% of them got active Hypertension and the rest 14.44% had inactive Hypertension. Histograms are shown in Figure 5.22 and Figure 5.23 for both female and male.

It's obvious that, for the progression of hypertension, there's no significant difference between female and male subjects from the histograms in the following two figures. So only using 1 set of progression rate has been used in our simulation model for hypertension is acceptable.

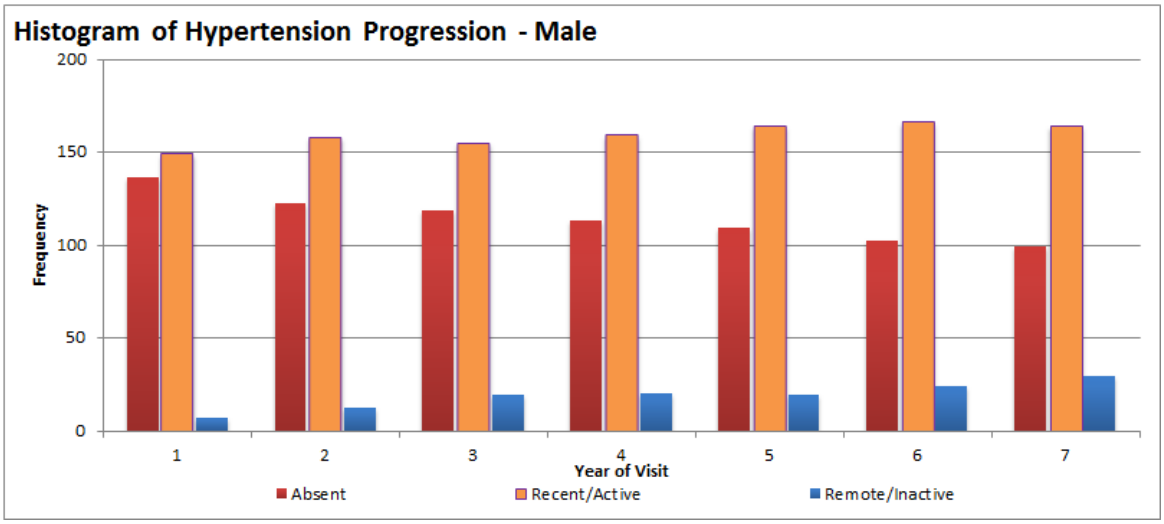


Figure 5.22: Histogram of Hypertension Progression – Male

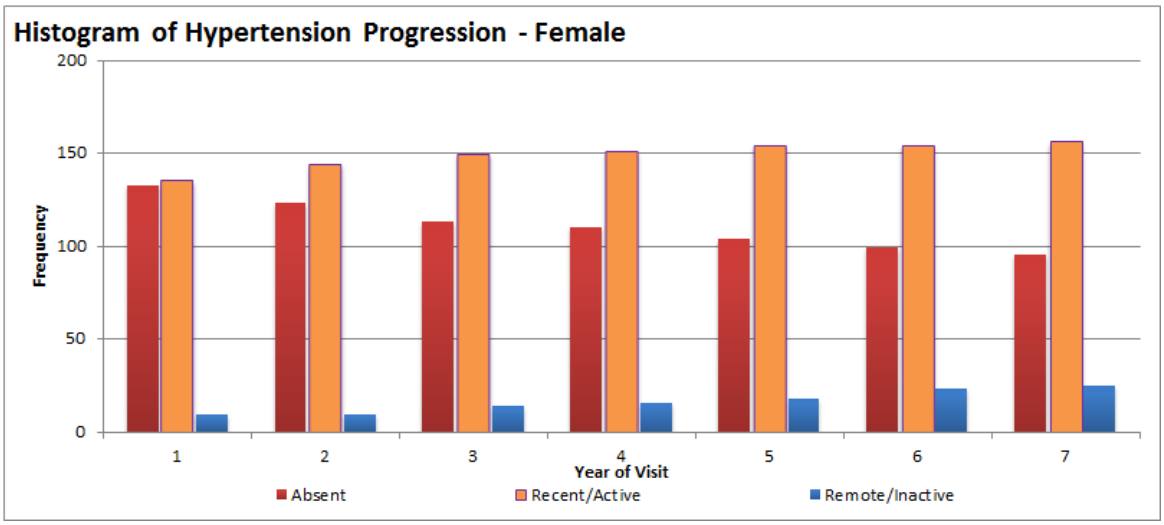


Figure 5.23: Histogram of Hypertension Progression – Female

- Hypercholesterolemia – The assumption for progression of hypercholesterolemia is the same as what we have for hypertension. For male, 0.589% subjects would progress to Hypercholesterolemia Level 1 or Level 2 after each tick. Among these subjects who recently progressed to Hypercholesterolemia, 84.47% of them got active Hypercholesterolemia and

the rest 15.53% had inactive Hypercholesterolemia. For female, the progression rate for normal subjects is 0.458%. 84.47% of them got active Hypercholesterolemia and the rest 15.53% had inactive Hypercholesterolemia. The 7-year histograms are shown in Figure 5.24 and Figure 5.25.

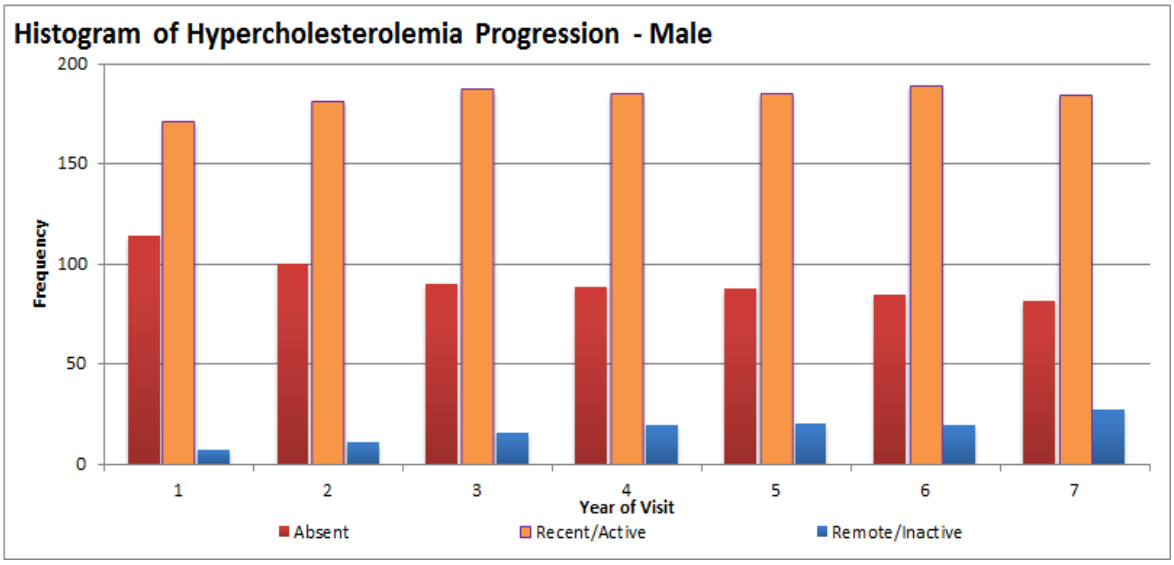


Figure 5.24: Histogram of Hypercholesterolemia Progression – Male

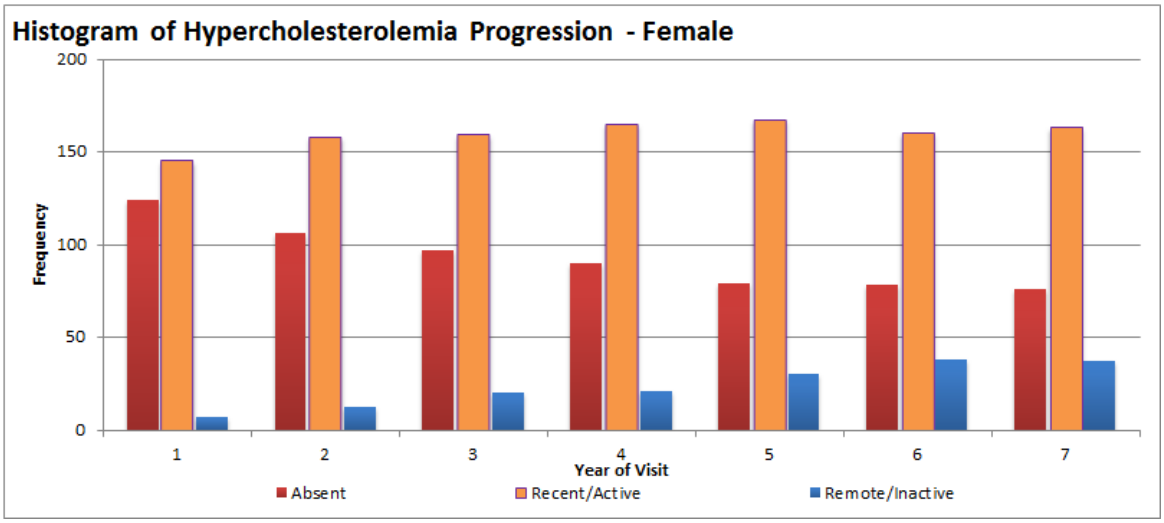


Figure 5.25: Histogram of Hypercholesterolemia Progression – Female

- Diabetes – Similar to assumption for hypertension, for male, 0.0836% subjects would progress to Diabetes Level 1 or Level 2 after each tick. Among these subjects who recently progressed to Diabetes, 81.40% of them got active. For female, the progression rate for

normal subjects is 0.0775%. 81.40% of them got active Diabetes and the rest 15.53% had inactive Diabetes. Histograms are shown in Figure 5.26 and Figure 5.27.

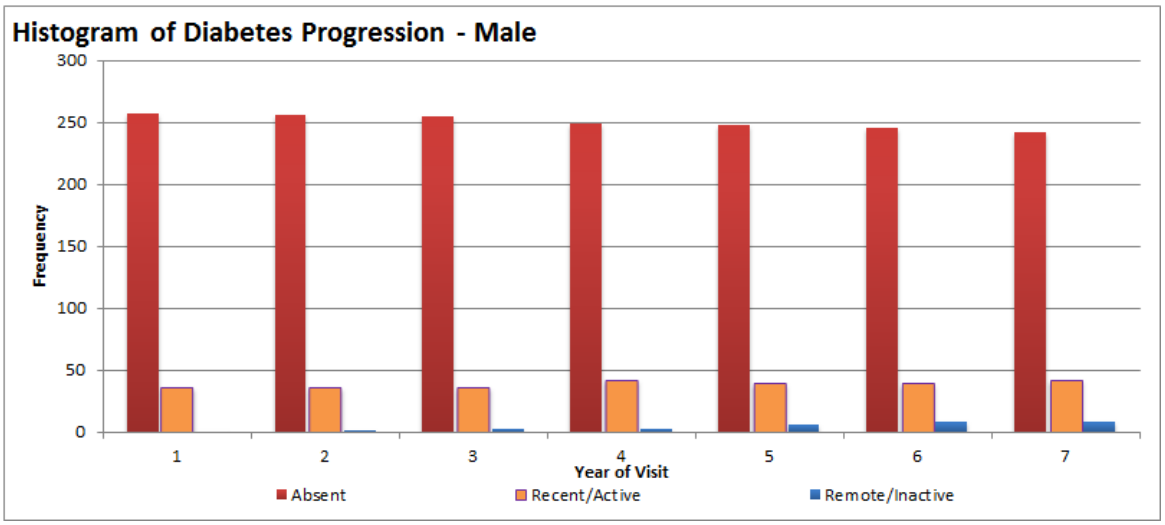


Figure 5.26: Histogram of Diabetes Progression – Male

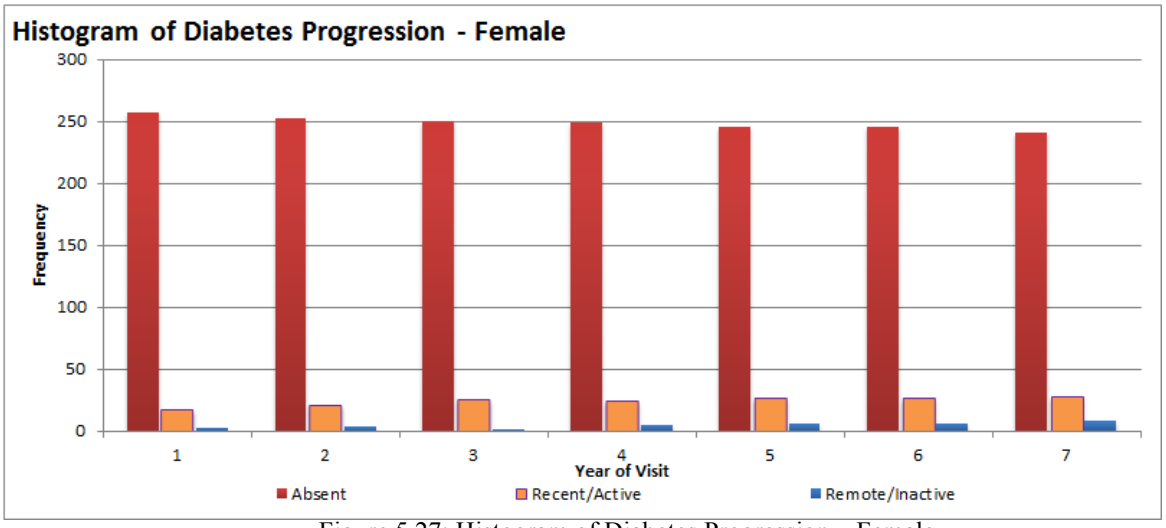


Figure 5.27: Histogram of Diabetes Progression – Female

5.3 Simulation Verification

One way to check the accuracy of the simulation model is to test the statistical significance between the simulation outputs of the demographical variables and the 7th annual data in our database. To verify the output, we run the simulation model for 84 ticks, which represents 7-year progression process, with 570 subjects. After simulation run, we get a set of

outputs for demographical variables. Then significant differences were tested using statistical methods such as t-test for continuous variables and chi-square test for categorized variables. Table 5.3 shows the results of statistical tests.

Table 5.3: Simulation Output Verification

	Simulation Results (mean, s.d.)	**7th annual Data (mean, s.d.)**	**Statistical Test**	**P-value**	**Significant Difference**
Age	80.98, 7.97	80.63, 7.92	t-test	0.453	NO
Education Year	15.55, 3.22	15.40, 3.07	t-test	0.428	NO
Living Situation	2.06, 1.01	2.06, 0.99	chi-squre	0.396	NO
Living Dependency	1.92, 1.03	1.85, 1.04	chi-squre	0.446	NO
Hypertension	0.76, 0.62	0.75, 0.61	chi-squre	0.931	NO
Hypercholesterolemia	0.85, 0.60	0.8, 0.84	chi-squre	0.956	NO
Diabetes	0.19, 0.45	0.18, 0.45	chi-squre	0.574	NO

As is shown in Table 5.3, the p-values for all of the demographical variables are much greater than 0.05, which means none of the outputs are significantly different to the actual data. So the simulation output is verified to be accurate and validity.

We then predict the onset of Alzheimer's disease by adding the logistic regression model into the simulation model. Subjects' demographical variables are the inputs for the logistic model. By so, we should be able to predict the probability of Alzheimer's disease for every subject and then get a general view about what's the percentage about subjects getting unfavorable outcome through the whole population.

In this case, another way to verify the accuracy of the simulation model is to check, by running the simulation model multiple replications, is the mean of the percentages of subjects

getting Alzheimer's, among multiple runs, are statistically significant different from the actual percentage of Alzheimer's subjects (158 out of 505, 31.29%). To test this we run the simulation model for 50 times, which we think is sufficient to perform statistical tests, and we were able to get a set data of the percentage of patients predicted of onset of Alzheimer's disease. Then t-test is performed to see if the mean value for this set of data is significantly different from 31.29%. Below is the table showing that we do not reject the hypothesis that the mean of percentage of patients' getting Alzheimer's disease in simulation outputs is equal to 31.29%.

Table 5.4: T-test for Percentage of Patients in Simulation Output

N	Mean	St Dev.	95% Confidence Interval	T	P-value
50	0.315	0.019	(0.31055,0.31944)	0.94	0.351

Figure 5.28 is the overview of simulation results with initial settings.

5.4 Add in Social Factors

Part of the research results were presented on the IIE 2015 annual conference. During the presentation, one of the good comments brought up by the audience is that the social activities such as supporting groups or mental training classes might slow down the progression process of people with cognitive impairment. However, there's no information recorded to reflect the social activities in the NACC database. We then went through a thorough literature review to implement the influence of social factors in the simulation model.

In another research done by Robert S. Wilson, they figured that every 1-point change in this cognitive activity score is associated with a 33% reduction in risk of Alzheimer's disease. So in our simulation model, subjects' changes in the cognitive activity score were tracked and then used to determine how much reduction of Alzheimer's disease is achieved

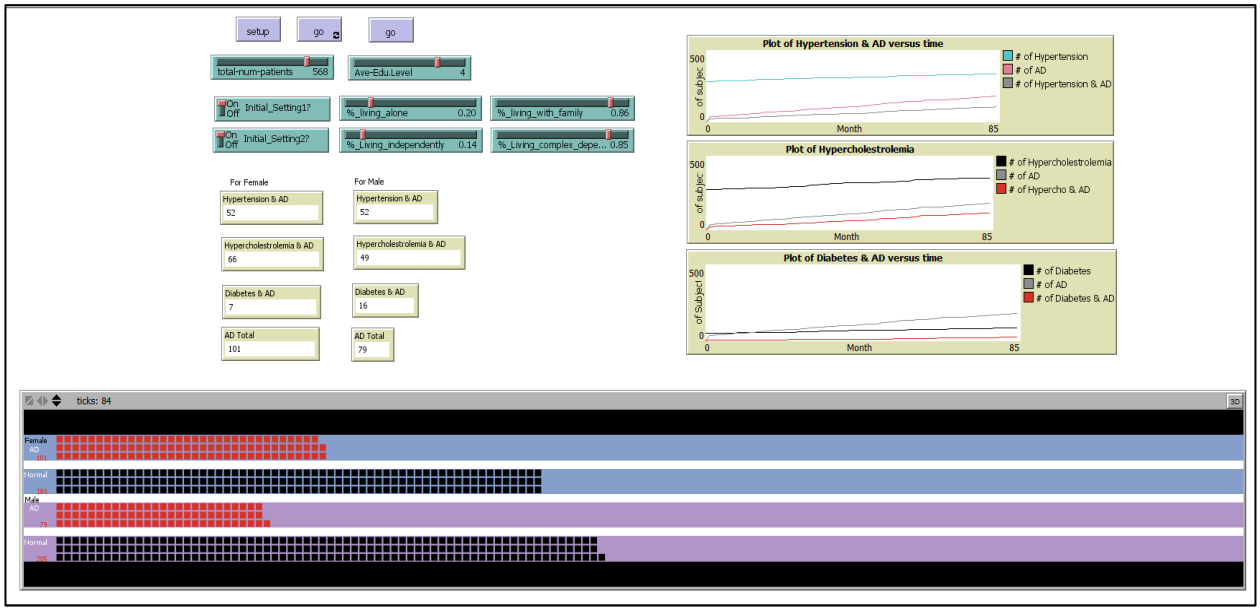


Figure 5.28: Simulation Overview with initial setting

The second measurement is subjects' leisure score. To acquire this score, an interview elicited self-reported participation during the month preceding the interview in the following 13 activities: knitting or music or other hobby, walking for pleasure or excursion, visiting friends or relatives, being visited by relatives or friends, physical conditioning, going to movies or restaurants or sporting events, reading magazines or newspapers or books, watching television or listening to the radio, doing unpaid community volunteer work, playing cards or games or bingo, going to a club or center, going to classes, and going to church or synagogue or temple. One point was given for participation to each of the above activities and an aggregate score was assigned to each subject. Similarly, we're not able to gain this score for subjects from interview. But the research that N.Scarmeas[错误! 未定义书签。] has done indicated that this score is related to the demographical variables such as gender, education level and etc. Also, in this research, it's concluded that regarding this score as a continuous variable, one-point increase would result in an 11% reduction in cognitive impairment progress. The regression model and the influence of the score are both integrated in the simulation model

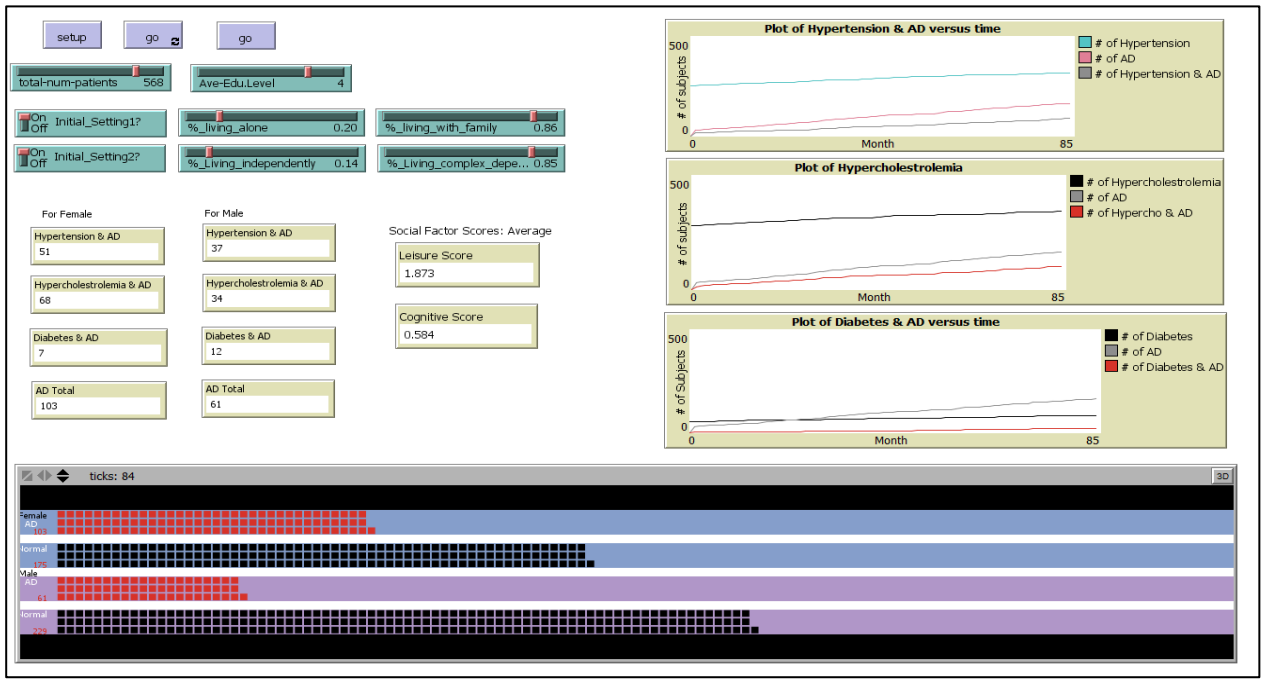


Figure 5.29: Simulation Overview by adding social factors

Table 5.5: Mean Percentage of Patients in Simulation Output Adding Social Factors

N	Mean	St Dev.	95% Confidence Interval
50	0.271	0.016	(0.266,0.276)

5.5 Running Different Scenarios

The following several figures show how the output changes with various population.

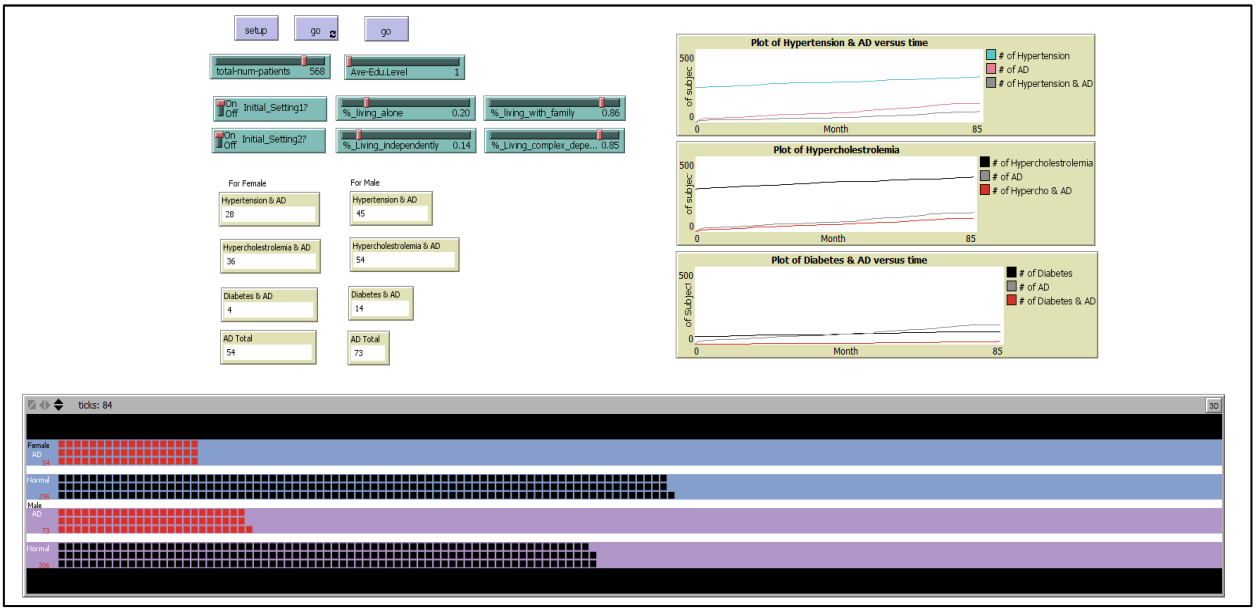


Figure 5.30: Simulation Overview for Population with Lower Education Level.

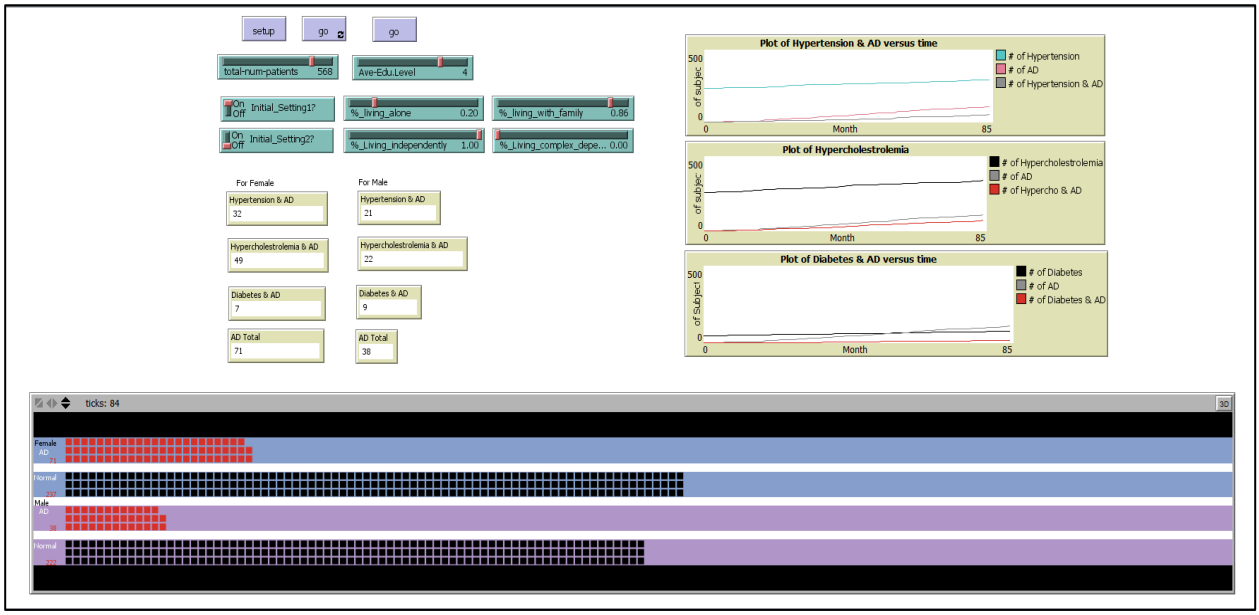


Figure 5.31: Simulation Overview for Population with Better Living Independency.

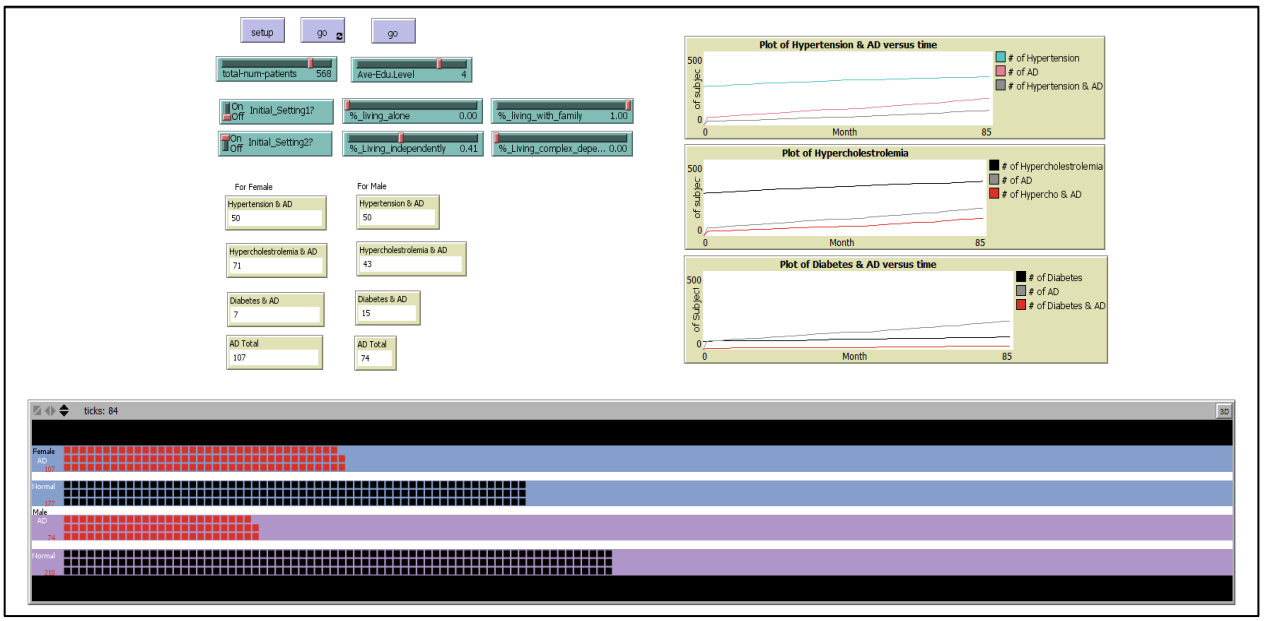


Figure 5.32: Simulation Overview for Population all living with family/friends.

Chapter 6: Conclusion and Discussion

The objective of our research includes:

- Verify and validate the accuracy of the two widely used cognitive impairment staging scales: Clinical Dementia Rating (CDR) and Mini-mental State Examination (MMSE).
- Identify and analyze the combined effects of demographical risk factors on both the accuracy of the two measurement scales and the progression of the preclinical phase of Alzheimer's disease.
- Visualize the change of cognitive impairment status of patients, and to predict the probability of patients reaching unfavorable outcome over time using simulation technique.

To achieve the first objective, we thoroughly compared the accuracy of Clinical Dementia Rating (CDR) and Mini-mental State Examination (MMSE). Comparing the specificity and sensitivity calculated with information in the database, both methods provide relative accurate results. However, According to the time consumed and the difficulty to apply, MMSE is the preferred method as it can be easily applied by well-trained caregivers within 15 minutes.

For better understanding about what are the demographical factors that may influence the accuracy of the two measurements, we used statistical analysis to identify the potential risk factors and their combined effects on the accuracy of the two scales. From the statistical result, we concluded that Age and education level are influential for the results of MMSE.

Gender is the risk factor for CDR. Subjects with higher age and lower education level should be aware of the more probability of getting incorrect unfavorable outcomes from MMSE method. Also, males may have less accuracy in identifying the onset of Alzheimer's disease through CDR.

Similar analysis was applied on determine the effects of risk factors on the actual cognitive progression process. The logistic regression model is brought up. Based upon the specificity and sensitivity of the logistic model, we figured that although the accuracy is relative acceptable, this method tend to be more aggressive, i.e. high risk of getting false alarms. The solution would be to combine logistic regression and MMSE, which would result in better prediction accuracy.

With all the information integrated in the agent-based simulation model, we can visualize the change of cognitive impairment status of patients over time, as well as explore the different performance of various populations.

The main contribution of our research is providing a guidance about which cognitive impairment measurement scale to choose based upon everyone's individual set of demographical characteristics. The logistic regression model we brought up is also a good choice for prediction of onset of Alzheimer's disease as it considers each subject's individual demographical characteristics, as well as a relative accurate result. The agent-based simulation model we built would be helpful to explore the change of overall impairment status in various populations. In the future, it can be also applied to discover the effects of proposed treatments that reduce the probability of getting dementia in the whole population.

References

[1] National Institute on Aging. “Alzheimer's Disease Fact Sheet.” Publication Date: Augus 2016.

[2] David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 128.

[3] Walker, SH; Duncan, DB (1967). "Estimation of the probability of an event as a function of several independent variables". Biometrika. 54: 167–178. doi:10.2307/2333860.

[4] National Alzheimer’s Coordination Center, “NACC Uniform Data Set Coding Guidebook for Initial Visit Packet”. Retrieved Version 2.0, February 2008.

[5] National Alzheimer’s Coordination Center, “NACC Uniform Data Set Data Element Dictionary for Initial Visit Packet”. Retrieved Version 2.0, February 2008.

[6] National Alzheimer’s Coordination Center, “NACC Derived Variables – Description of NACC Derived Variables to be used in data analysis” Retrieved or NACC Derived Variables computed From UDS V2.0, December 2013.

[7] R language and environment, Hornik, Kurt (November 26, 2015). "R FAQ". The Comprehensive R Archive Network. 2.1 What is R?. Retrieved 2015-12-06.

[8] Grimm, Volker; Railsback, Steven F. (2005). Individual-based Modeling and Ecology. Princeton University Press. p. 485. ISBN 978-0-691-09666-7.[9] Hughes, Charles P., et al. "A new clinical scale for the staging of dementia." The British journal of psychiatry 140.6 (1982): 566-572.[10] Morris, John C. "The Clinical Dementia Rating (CDR): current version and scoring rules." Neurology (1993).

[11] Morris, John C. "Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type." International psychogeriatrics 9.S1 (1997): 173-176.

[12] Berg, Leonard, et al. "Mild senile dementia of the Alzheimer type: 2. Longitudinal assessment." Annals of neurology 23.5 (1988): 477-484.

[13] Folstein, MF; Folstein, SE; McHugh, PR (1975). ""Mini-mental state". A practical method for grading the cognitive state of patients for the clinician". Journal of Psychiatric Research 12 (3): 189–98.

[14] Withers, Eileen, and John Hinton. "Three forms of the clinical tests of the sensorium and their reliability." The British Journal of Psychiatry (1971).

[15] Dementia can cooperate short time. HALSTEADH,. A psychometricstudyof senility.J. Ment. Sci. 89, 363, 1943

[16] Pangman, Verna C., Jeff Sloan, and Lorna Guse. "An examination of psychometric properties of the mini-mental state examination and the standardized mini-mental state examination: implications for clinical practice." Applied Nursing Research 13.4 (2000): 209-213.

[17] Juva, K., et al. "Staging the severity of dementia: comparison of clinical (CDR, DSM-III-R), functional (ADL, IADL) and cognitive (MMSE) scales." Acta neurologica scandinavica 90.4 (1994): 293-298.

[18] Tuijl, Jolien P., et al. "Screening for cognitive impairment in older general hospital patients: comparison of the Six-Item Cognitive Impairment Test with the Mini-Mental State Examination." International journal of geriatric psychiatry 27.7 (2012): 755-762.

[19] Tombaugh, Tom N., and Nancy J. McIntyre. "The mini-mental state examination: a comprehensive review." Journal of the American Geriatrics Society 40.9 (1992): 922-935.

[20] Juva, K., et al. "Staging the severity of dementia: comparison of clinical (CDR, DSM-III-R), functional (ADL, IADL) and cognitive (MMSE) scales." Acta neurologica scandinavica 90.4 (1994): 293-298.

[21] Steenland, Kyle, et al. "Statins and cognitive decline in older adults with normal cognition or mild cognitive impairment." Journal of the American Geriatrics Society 61.9 (2013): 1449-1455.

[22] Forsell, Y., et al. "Clinical staging of dementia in a population survey: comparison of DSM-III-R and the Washington University Clinical Dementia Rating Scale." Acta psychiatrica scandinavica 86.1 (1992): 49-54.

[23] Breteler, Monique MB. "Vascular risk factors for Alzheimer’s disease:: An epidemiologic perspective." Neurobiology of aging 21.2 (2000): 153-160.

[24] Uhlmann, Richard F., and Eric B. Larson. "Effect of Education on the Mini-Mental State Examination as a Screening Test for Dementia." Journal of the American Geriatrics Society 39.9 (1991): 876-880.

[25] Borson, Soo, et al. "Simplifying detection of cognitive impairment: comparison of the Mini-Cog and Mini-Mental State Examination in a multiethnic sample." Journal of the American Geriatrics Society 53.5 (2005): 871-874.

[26] Pezzotti, Patrizio, et al. "The accuracy of the MMSE in detecting cognitive impairment when administered by general practitioners: a prospective observational study." BMC family practice 9.1 (2008): 1.

[27] Grut, Michaela, et al. "Accuracy of the Mini-Mental Status Examination as a screening test for dementia in a Swedish elderly population." Acta Neurologica Scandinavica 87.4 (1993): 312-317.

[28] Newman, Anne B., et al. "Dementia and Alzheimer's disease incidence in relationship to cardiovascular disease in the Cardiovascular Health Study cohort." Journal of the American Geriatrics Society 53.7 (2005): 1101-1107.

[29] Kotagal, Vikas, et al. "Factors associated with cognitive evaluations in the United States." Neurology 84.1 (2015): 64-71.

[30] Sattler, Christine, et al. " Cognitive activity, education and socioeconomic status as preventive factors for mild cognitive impairment and Alzheimer's disease." Psychiatry research 196.1 (2012): 90-95.

[31] Hall, Kathleen S., et al. "Low education and childhood rural residence Risk for Alzheimer's disease in African Americans." Neurology 54.1 (2000): 95-95.

[32] Fratiglioni, Laura, Stephanie Paillard-Borg, and Bengt Winblad. "An active and socially integrated lifestyle in late life might protect against dementia."The Lancet Neurology 3.6 (2004): 343-353.

[33] Zhang, Mingyuan, et al. "The prevalence of dementia and Alzheimer's disease in Shanghai, China: impact of age, gender, and education." Annals of neurology 27.4 (1990): 428-437.

[34] Jun, J. B., Sheldon H. Jacobson, and J. R. Swisher. "Application of discrete-event simulation in health care clinics: A survey." Journal of the operational research society 50.2 (1999): 109-123.

[35] Brailsford, S. C., et al. "An analysis of the academic literature on simulation and modelling in health care." Journal of simulation 3.3 (2009): 130-140.

[36] Günal, Murat M., and Michael Pidd. "Discrete event simulation for performance modelling in health care: a review of the literature." Journal of Simulation 4.1 (2010): 42-51.

[37] Brandeau, Margaret L., François Sainfort, and William P. Pierskalla, eds.Operations research and health care: a handbook of methods and applications. Vol. 70. Springer Science & Business Media, 2004.

[38] Yi, Wei, and Linet Özdamar. "A dynamic logistics coordination model for evacuation and support in disaster response activities." European Journal of Operational Research 179.3 (2007): 1177-1193.

[39] Wilson, Robert S., et al. "Cognitive activity in older persons from a geographically defined population." The Journals of Gerontology Series B: Psychological Sciences and Social Sciences 54.3 (1999): P155-P160.

[40] Wilson, Robert S., Lisa L. Barnes, and David A. Bennett. "Assessment of lifetime participation in cognitively stimulating activities." Journal of Clinical and Experimental Neuropsychology 25.5 (2003): 634-642.

[41] Wilson, Robert S., and David A. Bennett. "Cognitive activity and risk of Alzheimer's disease." Current Directions in Psychological Science 12.3 (2003): 87-91.

[42] Scarmeas, Nikolaos, et al. "Influence of leisure activity on the incidence of Alzheimer's disease." Neurology 57.12 (2001): 2236-2242.

[43] Norton, Maria C., et al. "The design and progress of a multidomain lifestyle intervention to improve brain health in middle-aged persons to reduce later Alzheimer's disease risk: The Gray Matters randomized trial." Alzheimer's & Dementia: Translational Research & Clinical Interventions 1.1 (2015): 53-62.

[44] Wilson, Robert S., et al. "Loneliness and risk of Alzheimer disease." Archives of general psychiatry 64.2 (2007): 234-240.

Appendix I

Appendix I: Netlogo Codes to accomplish the agent-based simulation model.

```
globals [


  group-Female-AD
  group-Male-AD

  group-Female-norm
  group-Male-norm

  patches-Female
  patches-Male

  MCI?
  high-pre?
  high-fat?

  Female-ave
  Male-ave
  time
   ]

turtles-own [gender
  Age
  MMSE
  EduLev
  EduYr
  smoking
  Hyper
  Hycho
  Dia
  LivSit
  LivDep
  AD
  Hyper-delay
  Hycho-delay
  Dia-delay
  T-Hyper
  T-Hycho
  T-Dia
  Pre-Hyper
  Pre-Hycho
  P_hyper
  P_hycho
  P_dia
  P_ls
  Pre-Dia
  LeisureScore
  CognitiveScore
   my-group-site
   num-close-tie
   num-patient-group
    ls ld u s k R
    P_mmse
    CogScoreChange
    CognitiveIntensity
    CognitiveFrequency
    hp da NumPatch NumPatchPrevious]


;----------------------------
;-----Set up procedure------
;----------------------------
to setup
  clear-all
  setup-patch-label
  setup-turtles
  reset-ticks
end

;--------------------------------
to setup-patch-label

  ask patch (min-pxcor + 2) 7 [
    set plabel ("Female")
    set plabel-color black]
  ask patch (min-pxcor + 1) 0 [
    set plabel ("Male")
    set plabel-color black]

  set group-Female-AD patches with [group-Female-AD?]
  set group-Female-norm patches with [group-Female-norm?]
  ask patch (min-pxcor + 1) 6[
    set plabel ("AD")
    set plabel-color white]

  ask patch (min-pxcor + 2) 3[
    set plabel ("Normal")
    set plabel-color white]

  set group-Male-AD patches with [group-Male-AD?]
  set group-Male-norm patches with [group-Male-norm?]
  ask patch (min-pxcor + 1) -1[
    set plabel ("AD")
    set plabel-color white]

  ask patch (min-pxcor + 2) -5[
    set plabel ("Normal")
    set plabel-color white]

  ask patches [set pcolor black]
  set patches-Female patches with [pycor >= 0 and pycor <= 7]
```

```
    ask patches-Female [set pcolor 107]
  set patches-Male patches with [pycor >= -7 and pycor <= 0]
    ask patches-Male [set pcolor 117]
    ask patches with [pycor = 0 or pycor = 4 or pycor = -4]
    [set pcolor white]
end;---------------------------------
to setup-turtles
  set-default-shape turtles "square"
  create-turtles total-num-patients
  [ ;set color 84 + random 5
    set size 1
    set gender random 2
    set color black
    setup-color
    set smoking random 1
    set Age random-normal 74.11 7.947
    set Hyper-delay random-normal 2.87 2.82
    ask turtles with [gender = 0]
    [set Hycho-delay random-normal 3.014 2.783
      set Dia-delay random-normal 4.24 2.554]
    ask turtles with [gender = 1]
    [set Hycho-delay random-normal 2.586 2.725
      set Dia-delay random-normal 2.115 2.747]
  ]

    setup-EduYr
    setup-Hypertension
    setup-Hypercho
    setup-Diabetes
    setup-LivSit
    setup-LivDep

   ask turtles [ find-group ]
   initialize
end
;-------------------------
to find-group
 if gender = 0
      [ ifelse AD = 1 [
         set color red
         set my-group-site one-of group-Female-AD
           move-to my-group-site
         ]
      [ set my-group-site one-of group-Female-norm
           move-to my-group-site
         ]]

 if gender = 1
      [ ifelse AD = 1 [
         set color red
         set my-group-site one-of group-Male-AD
           move-to my-group-site
         ]
      [ set my-group-site one-of group-Male-norm
           move-to my-group-site
         ]]
end

to setup-color
  ask turtles with [gender = 0 ]
  [ set shape "square"]
  ask turtles with [gender = 1]
  [ set shape "square"];;0 = female
end
;---------------------------------
to setup-EduYr
  ask turtles with [gender = 0] ;female
  [ set EduLev random-normal Ave-Edu.Level 0.9402
    ifelse Edulev <= 1.5
    [ set EduYr 6]
    [ ifelse Edulev <= 2.5
      [ let p random-float 1
       ifelse p <= 0.4 [set EduYr 7]
       [ ifelse p <= 0.8 [set EduYr 8]
         [set EduYr 9]]]
      [ifelse Edulev <= 3.5
       [let p random-float 1
        ifelse p <= 0.02 [set EduYr 10]
        [ifelse p <= 0.05 [set EduYr 11]
          [set EduYr 12]
        ]]
       [ifelse Edulev <= 4.5
        [let p random-float 1
        ifelse p <= 0.1 [set EduYr 13]
        [ ifelse p <= 0.4 [set EduYr 14]
          [ifelse p <= 0.5 [set EduYr 15]
            [set EduYr 16]
          ]]]
        [ifelse Edulev <= 5.5
          [ let p random-float 1
            ifelse p <= 0.05 [set EduYr 17]
            [ifelse p <= 0.85 [set EduYr 18]
              [ifelse p <= 0.9 [set EduYr 19]
                [set EduYr 20]
              ]]]
          [ let p random-float 1
            ifelse p <= 0.6667 [set EduYr 21]
            [set EduYr 22]
          ]]]]]]
  ask turtles with [gender = 1] ;male
  [ set EduLev random-normal 4.219 0.8575
    ifelse Edulev <= 1.5
    [ set EduYr 6]
    [ ifelse Edulev <= 2.5
      [ let p random-float 1
       ifelse p <= (1 / 6) [set EduYr 7]
       [ ifelse p <= 0.5 [set EduYr 8]
         [set EduYr 9]]]
      [ifelse Edulev <= 3.5
       [let p random-float 1
        ifelse p <= 0.05 [set EduYr 10]
        [ifelse p <= 0.15 [set EduYr 11]
          [set EduYr 12]
        ]]
       [ifelse Edulev <= 4.5
        [let p random-float 1
        ifelse p <= 0.15 [set EduYr 13]
        [ ifelse p <= 0.4 [set EduYr 14]
          [ifelse p <= 0.45 [set EduYr 15]
            [set EduYr 16]
          ]]]
        [ifelse Edulev <= 5.5
          [ let p random-float 1
            ifelse p <= 0.1 [set EduYr 17]
            [ifelse p <= 0.6 [set EduYr 18]
              [ifelse p <= 0.65 [set EduYr 19]
                [set EduYr 20]
              ]]]
          [ let p random-float 1
            ifelse p <= 0.7 [set EduYr 21]
            [ifelse p <= 0.85 [set EduYr 22]
              [set EduYr 23]
            ]]]]]]]
end

;----------------------------
to setup-Hypertension
  ask turtles[
    let p random-float 100
    ifelse p <= 53.169
```

```
    [let q random-float 100
     ifelse q <= 85.5615 [set Hyper 1]
     [set Hyper 2]]
   [set Hyper 0]
  ]
end



;-----------------------------
to setup-Hypercho
  ask turtles with [gender = 0]; female
  [ let p random-float 100
    ifelse p <= 56.88
     [let q random-float 100
      ifelse q <= 84.466 [set Hycho 1]
      [set Hycho 2]]
    [set Hycho 0]
  ]
  ask turtles with [gender = 1]; male
  [ let p random-float 100
    ifelse p <= 61.30
    [let q random-float 100
      ifelse q <= 84.466 [set Hycho 1]
      [set Hycho 2]]
    [set Hycho 0]
  ]
end

;-----------------------------
to setup-Diabetes
  ask turtles with [gender = 0]; female
  [ let p random-float 100
    ifelse p <= 7.6087
    [let q random-float 100
      ifelse q <= 81.3953 [set Dia 1]
      [set Dia 2]]
    [set Dia 0]
  ]
  ask turtles with [gender = 1]; male
  [ let p random-float 100
    ifelse p <= 12.3288
     [let q random-float 100
      ifelse q <= 81.3953 [set Dia 1]
      [set Dia 2]]
    [set Dia 0]
  ]
end
;-----------------------------
to setup-LivSit
  ifelse Initial_Setting1? = True [
       ask turtles with [gender = 0];female
       [ let p random-float 100
         ifelse p <= 40.58 [ set LivSit 1]
         [ ifelse p <= 85.87 [ set LivSit 2]
           [ifelse p <= 97.46 [ set LivSit 3]
             [ifelse p <= 98.55 [ set LivSit 4]
             [set LivSit 5]
         ]]]]
       ask turtles with [gender = 1]; male
       [ let p random-float 100
         ifelse p <= 10.96 [ set LivSit 1]
         [ ifelse p <= 95.89 [ set LivSit 2]
           [ ifelse p <= 99.32 [ set LivSit 3]
             [set LivSit 4]
         ]]]
  ]
  [ ask turtles
    [ let p random-float 1
      ifelse p <= %_living_alone [set LivSit 1] [
       ifelse p <= (%_living_with_family + %_living_alone) [set
LivSit 2]
        [ set LivSit 3]
    ]]
  ]
End

;-----------------------------
to setup-LivDep
  ifelse Initial_Setting2? = True [
    ask turtles with [gender = 0]
    [ let p random-float 100
      ifelse p <= 83.333 [ set LivDep 1]
      [ ifelse p <= 98.188 [ set LivDep 2]
        [set LivDep 3]
      ]]
    ask turtles with [gender = 1]
    [ let p random-float 100
      ifelse p <= 75.6849 [ set LivDep 1]
      [ ifelse p <= 97.9452 [ set LivDep 2]
        [set LivDep 3]
      ]]
]
  [ ask turtles
    [ let p random-float 1
      ifelse p <= %_Living_independently [set LivDep 1]
      [ ifelse p <= %_Living_independently
+ %_Living_complex_dependent [set LivDep 2]
        [ set LivDep 3]]]]

end

;-----------------------------
to initialize
  update-labels
  ask turtles [set k 0
    spread-out-horizontal]
end
;-----------------------------
to update-labels
  ask group-Female-AD [ set plabel count turtles-here
    set plabel-color red]
  ask group-Male-AD [ set plabel count turtles-here
    set plabel-color red]
  ask group-Female-norm [ set plabel count turtles-here
    set plabel-color red]
  ask group-Male-norm [ set plabel count turtles-here
    set plabel-color red]
end
;-----------------------------
to spread-out-horizontal
  set heading 90
  fd 2
  set heading 0
  while [any? other turtles-here] [
    ifelse k < 2 [
      fd 1
      set k k + 1]
    [ set xcor xcor + 1
      set ycor ycor - 2
      set k 0
    ]
  ]
;;  set heading 270
;;  fd 1
;;  while [any? other turtles-here] [
;;    ifelse can-move? 2 [
;;      fd 1]
;;    [ set ycor ycor + 1
;;      set xcor 0
```

```
;;     fd 1
;;   ]
;;  ]
end
;------------------------------
to-report group-Female-AD?
  let group-interval 3
  report
  (pxcor = 0) and
  (pycor > 4 and pycor <= 7) and
  (pycor mod group-interval = 2)
end
to-report group-Male-AD?
  let group-interval 3
  report
  (pxcor = 0) and
  (pycor >= -3 and pycor < 0) and
  (pycor mod group-interval = 0)
end
to-report group-Female-norm?
  let group-interval 3
  report
  (pxcor = 0) and
  (pycor > 0 and pycor <= 3) and
  (pycor mod group-interval = 1)
end
to-report group-Male-norm?
  let group-interval 3
  report
  (pxcor = 0) and
  (pycor >= -7 and pycor < -4) and
  (pycor mod group-interval = 2)
end

;--------------------------------
;-----------Go Procedure--------
;--------------------------------
to go
  turtles-progress
  AD-progress
  ask turtles  [find-group]
  initialize
  ask turtles [
    set Pre-Hyper Hyper
    set Pre-Hycho Hycho
    set Pre-Dia Dia  ]
  tick
  if ticks >= 84
  [
  stop
  ]
end

to turtles-progress
    Age-progress
    Hypertension-progress
    Hypercholestro-progress
    Diabetes-progress
    LivDep-progress
    ifelse time = 7
    [ LivSit-progress
      set time 1]
    [set time (time + 1)]
end

to Age-progress
  ask turtles [set Age Age + (1 / 12)]
end

to Hypertension-progress
  ask turtles with [Hyper = 0]
  [let p random-float 10000
    if p <= 42.593 [
      let q random-float 100
      ifelse q <= 85.5615 [set Hyper 1]
      [set Hyper 2]]
  ]
;  ask turtles with [Hyper = 1 and Pre-Hyper = 0]
;  [ let p random-float 100
;    if p < 21.3
;   [set T-Hyper ticks ]]
;  ask turtles with [T-Hyper > 0]
;  [
;   if ticks - T-Hyper > (12 * Hyper-delay)
;  [set AD 1]
;  ]
end
to Hypercholestro-progress
 ask turtles with [gender = 0 and Hycho = 0]
  [let p random-float 10000
    if p <= 58.8604 [
      let q random-float 100
      ifelse q <= 84.466 [set Hycho 1]
      [set Hycho 2]]
  ]
; ask turtles with [Hycho = 1 and gender = 0 and Pre-Hycho = 0]
;  [ let p random-float 1
;    if p < 0.2391
;    [ set T-Hycho ticks]]

 ask turtles with [gender = 1 and Hycho = 0]
  [let p random-float 10000
    if p <= 45.798 [
      let q random-float 100
      ifelse q <= 84.466 [set Hycho 1]
      [set Hycho 2]]
  ]
; ask turtles with [Hycho = 1 and gender = 1 and Pre-Hycho = 0]
;  [ let p random-float 1
;    if p < 0.2464
;    [ set T-Hycho ticks]]
; ask turtles with [T-Hycho > 0]
;  [
;   if ticks - T-Hycho > (12 * Hycho-delay)
;  [set AD 1]
;  ]
end


to Diabetes-progress
 ask turtles with [gender = 0 and Dia = 0]
  [let p random-float 10000
    if p <= 8.35559
      [let q random-float 100
      ifelse q <= 81.3953 [set Dia 1]
      [set Dia 2]]
  ]
; ask turtles with [Dia = 1 and gender = 0 and Pre-Dia = 0]
;  [ let p random-float 1
;    if p < 0.02174 [ set T-Dia ticks]]

 ask turtles with [gender = 1 and Dia = 0]
  [let p random-float 10000
    if p <= 7.75129
      [let q random-float 100
      ifelse q <= 81.3953 [set Dia 1]
      [set Dia 2]]
  ]
; ask turtles with [Dia = 1 and gender = 1 and Pre-Dia = 0]
;  [ let p random-float 1
```

```
;    if p < 0.05797 [ set T-Dia ticks]]
;
; ask turtles with [T-Dia > 0]
; [
;   if ticks - T-Dia > (12 * Dia-delay)
; [set AD 1]
; ]
end

to LivDep-progress
  ask turtles with [gender = 0]
[  ifelse LivDep = 1
   [ let p random-float 1000
     if p <= 6.7880 [set LivDep 2]
    ]
 [ifelse LivDep = 2
   [ let p random-float 1000
     if p <= 16.3548 [set LivDep 3]
    ]
   [ let p random-float 1000
     if p <= 19.9405 [set LivDep 4]
    ]]
]
  ask turtles with [gender = 1]
[  ifelse LivDep = 1
   [ let p random-float 1000
     if p <= 5.1651 [set LivDep 2]
    ]
 [ifelse LivDep = 2
   [ let p random-float 1000
     if p <= 9.8865 [set LivDep 3]
    ]
   [ let p random-float 1000
     if p <= 8.09524 [set LivDep 4]
    ]]
]

end

to LivSit-progress
 ;;;Female
 ask turtles with [gender = 1 and LivSit = 1]
 [ let p random-float 100
   ifelse p <= 90.53 [set LivSit  LivSit]
   [ ifelse p <= 91.72 [set LivSit  (LivSit + 1)]
     [ifelse p <= 95.86 [set LivSit  (LivSit + 2)]
       [ifelse p <= 97.19 [set LivSit  (LivSit + 3)]
         [set LivSit  (LivSit + 4)]]]]]

 ask turtles with [gender = 1 and LivSit = 2]
 [ let p random-float 100
   ifelse p <= 3.58 [set LivSit  (LivSit - 1)]
   [ ifelse p <= 97.85 [set LivSit  (LivSit)]
     [ifelse p <= 98.57 [set LivSit  (LivSit + 1)]
       [ifelse p <= 99 [set LivSit  (LivSit + 2)]
         [set LivSit  (LivSit + 3)]]]]]

 ask turtles with [gender = 1 and LivSit = 3]
 [ let p random-float 100
   ifelse p <= 13.4 [set LivSit  (LivSit - 2)]
   [ ifelse p <= 14.43 [set LivSit  (LivSit - 1)]
     [ifelse p <= 96.91 [set LivSit  (LivSit)]
       [ifelse p <= 98.97 [set LivSit  (LivSit + 1)]
         [set LivSit  (LivSit + 2)]]]]]

  ask turtles with [gender = 1 and LivSit = 4]
 [ let p random-float 100
   ifelse p <= 2.5 [set LivSit  (LivSit - 3)]
   [ ifelse p <= 5 [set LivSit  (LivSit - 1)]
     [ifelse p <= 95 [set LivSit  (LivSit)]
         [set LivSit  (LivSit + 1)]]]]

  ask turtles with [gender = 1 and LivSit = 5]
 [ let p random-float 100
   ifelse p <= 14.89 [set LivSit  (LivSit - 4)]
   [ ifelse p <= 19.15 [set LivSit  (LivSit - 2)]
     [ifelse p <= 23.4 [set LivSit  (LivSit - 1)]
        [set LivSit  (LivSit)]]]]

 ;;;;Male
 ask turtles with [gender = 0 and LivSit = 1]
 [ let p random-float 100
   ifelse p <= 88.53 [set LivSit  LivSit]
   [ ifelse p <= 95.41 [set LivSit  (LivSit + 1)]
     [ifelse p <= 98.62 [set LivSit  (LivSit + 2)]
       [ifelse p <= 99.54 [set LivSit  (LivSit + 3)]
         [set LivSit  (LivSit + 4)]]]]]

 ask turtles with [gender = 0 and LivSit = 2]
 [ let p random-float 100
   ifelse p <= 2.28 [set LivSit  (LivSit - 1)]
   [ ifelse p <= 99.10 [set LivSit  (LivSit)]
     [ifelse p <= 99.38 [set LivSit  (LivSit + 1)]
       [ifelse p <= 99.65 [set LivSit  (LivSit + 2)]
         [set LivSit  (LivSit + 3)]]]]]

 ask turtles with [gender = 0 and LivSit = 3]
 [ let p random-float 100
   ifelse p <= 4.48 [set LivSit  (LivSit - 2)]
   [ ifelse p <= 8.96 [set LivSit  (LivSit - 1)]
     [ifelse p <= 97.01 [set LivSit  (LivSit)]
        [set LivSit  (LivSit + 1)]]]]

  ask turtles with [gender = 0 and LivSit = 4]
 [ let p random-float 100
   ifelse p <= 7.14 [set LivSit  (LivSit - 1)]
    [set LivSit  (LivSit)]]

   ask turtles with [gender = 0 and LivSit = 5]
 [ let p random-float 100
   ifelse p <= 16.67 [set LivSit  (LivSit - 3)]
   [set LivSit  (LivSit)]]

end



to AD-progress
  ask turtles with [gender = 1];male
  [ if LivSit = 1 [set ls 0]
    if LivSit = 2 [set ls 1.51265]
    if LivSit = 3 [set ls 1.05007]
    if LivSit = 4 [set ls 1.09824]
    if LivSit = 5 [set ls 1.60590]
    if LivDep = 1 [set ld 0]
    if LivDep = 2 [set ld 1.71508]
    if LivDep = 3 [set ld 3.10540]
    if LivDep = 4 [set ld 2.82090]
    set num-close-tie count patches in-radius R with [pcolor =
yellow]
    set num-patient-group count turtles in-radius R
    set s -0.99185 - 0.01668 * Age - 0.07453 * EduYr + ls + ld +
0.05446 * Dia + 0.51238 * Hyper - 0.16588 * Hycho
    set u ((exp s)/(1 + exp s))
;   if u >= 0.5 [set AD 1]
;    if AD = 0
;    [ if ((u >= 0.5) and (MMSE < 25)) [set AD 1]]

;    ifelse AD = 1 [set AD 1]
;    [ ifelse ((u >= 0.5) and (MMSE < 27)) [set AD 1]
```

```
;     [set AD 0]]
  ]

  ask turtles with [gender = 0];female
  [ if LivSit = 1 [set ls 0]
    if LivSit = 2 [set ls 0.92305]
    if LivSit = 3 [set ls 0.93984]
    if LivSit = 4 [set ls 1.18757]
    if LivSit = 5 [set ls 0.80800]
    if LivDep = 1 [set ls 0]
    if LivDep = 2 [set ld 2.94373]
    if LivDep = 3 [set ld 3.92581]
    if LivDep = 4 [set ld 4.24581]
    set num-close-tie count patches in-radius R with [pcolor =
yellow]
    set num-patient-group count turtles in-radius R
    set s -8.11811 + 0.02570 * Age + 0.16443 * EduYr + ls + ld -
0.93669 * Dia - 0.70128 * Hyper + 0.60903 * Hycho
    set u ((exp s)/(1 + exp s))
  ]

  ask turtles [
    set CognitiveFrequency (0.017 * Age - 0.06 * EduYr + 0.167 *
gender)
    set CognitiveIntensity (0.003 * Age + 0.022 * EduYr + 0.087 *
gender)
    set CognitiveScore (0.5 * CognitiveIntensity + 0.5 *
CognitiveFrequency)
    if Hyper = 0 [set hp 1]
    if Dia = 0 [set da 1]
    set LeisureScore (0.4 * gender + 0.3 * hp + 0.7 * da + 0.22 *
Edulev)
  ]

  ask turtles
  [ if CognitiveScore > 1
    [ set u (u * 0.6667)
    set NumPatch (NumPatch + 1)]
    if CognitiveScore > 2
    [ set u (u * 0.6667)
    set NumPatch (NumPatch + 1)]
    if CognitiveScore > 3
    [ set u (u * 0.6667)
    set NumPatch (NumPatch + 1)]
    if CognitiveScore > 4
    [ set u (u * 0.6667)
    set NumPatch (NumPatch + 1)]
    if LeisureScore > 1
    [ set u (u * 0.89)
    set NumPatch (NumPatch + 1)]
    if LeisureScore > 2
    [ set u (u * 0.89)
    set NumPatch (NumPatch + 1)]
    if LeisureScore > 3
    [ set u (u * 0.89)
    set NumPatch (NumPatch + 1)]
    if u > 0.5 [set AD 1]]
end

to output
file-open "AD.txt"  ;; Opening file for writing
file-write (sum [AD] of turtles) / total-num-patients
file-close
file-open "MMSE.txt"  ;; Opening file for writing
ask turtles
 [ file-write MMSE]
file-close
file-open "u.txt"  ;; Opening file for writing
ask turtles
 [ file-write u]
file-close
file-open "AGE.txt"  ;; Opening file for writing
ask turtles
 [ file-write Age]
file-close
file-open "EduYr.txt"  ;; Opening file for writing
ask turtles
 [ file-write EduYr]
file-close
file-open "ls.txt"  ;; Opening file for writing
ask turtles
 [ file-write LivSit]
file-close
file-open "ld.txt"  ;; Opening file for writing
ask turtles
 [ file-write LivDep]
file-close
file-open "Dia.txt"  ;; Opening file for writing
ask turtles
 [ file-write Dia]
file-close
file-open "Hyper.txt"  ;; Opening file for writing
ask turtles
 [ file-write Hyper]
file-close
file-open "Hycho.txt"  ;; Opening file for writing
ask turtles
 [ file-write Hycho]
file-close
file-open "gender.txt"  ;; Opening file for writing
ask turtles
 [ file-write gender]
file-close
file-open "CognitiveScore.txt"  ;; Opening file for writing
ask turtles
 [ file-write CognitiveScore]
file-close
file-open "LeisureScore.txt"  ;; Opening file for writing
ask turtles
 [ file-write LeisureScore]
file-close
end
```

Appendix II

Appendix II: Overview of the NACC database.