BACKGROUND

A government sponsored cohort study of adults aged 65 years and older was conducted to observe the incidence of cardiovascular disease (especially heart attacks and congestive heart failure) and cerebrovascular disease (especially strokes) in the elderly over an 11 year period, and to relate the incidence of those diseases to various risk factors measured in the population on a regular basis. By such an observational study, greater insight into the natural history of chronic diseases in the elderly would be obtained. This is of particular importance, because there is increasing evidence that some of the associations observed between cardiovascular or cerebrovascular disease and various risk factors in middle aged adults are not observed in older adults. Possible mechanisms for such disparities include:

- effects due to survivorship: those people particularly susceptible to getting diseases from specific risk factors have already died, and are thus not present in an older cohort, and

- effects due to increased risks from other diseases: the elderly might have increased risk of diseases that rarely affect middle aged adults and that participants are actually protected from those diseases by characteristics that are associated with increased risk of the diseases prevalent in middle aged people. An example of such a mechanism might be that the increased risk of infectious diseases in the elderly and the protective effect of greater energy reserves against infection combine to produce a tendencyfor greater weight to be associated with greater longevity in the elderly.

In this study, elderly, generally healthy, adults were randomly selected from Medicare rolls. Agreement to participate was high, and thus the sample can be regarded as a fairly accurate representation of healthy older Americans. At the time of study enrollment, and on annual visits over the length of the study, the participants' data regarding various behavioral (e.g., smoking, alcohol consumption), functional (e.g., ability to perform routine tasks), and clinical (e.g., blood pressure, laboratory tests) measures are recorded. In addition, all serious medical events (e.g., hospitalizations, heart attacks) are investigated and categorized based on standardized, study-wide definitions.

Magnetic resonance imaging (MRI) is a relatively new imaging modality that has been found to be of great utility in diagnosing brain tumors, strokes, and other brain abnormalities. In imaging older people, however, a number of brain changes have been observed for which the medical community is uncertain of the clinical significance. In particular, it has been observed that brains tend to atrophy (shrink) with age, that the white matter in the brain (regions of the brain that appear white rather than gray on gross examination) tends to show up as ``brighter'' on MRI in older people, and that there is increasing incidence with age of regions of the brain that look like areas of dead tissue (infarct-like lesions) even in persons having no history of clinical stroke. It has not yet been established whether these changes should be viewed as part of the normal aging process, whether these changes are merely signs of other disease such as heart disease, or whether these changes are indicative of separate disease processes. Thus, approximately three years into the follow-up of this study, participants were asked to submit to magnetic resonance imaging (MRI) of their brains.

For this analysis, we shall focus on the measure of brain atrophy from MRI. We are interested in determining whether brain atrophy represents some form of primary central nervous system disease or whether it is a sign of some other disease process such as cardiovascular disease or merely results from the ``normal'' aging process.

The data to be analyzed for this assignment is a subset of the thousands of variables on a subset of the thousands of participants in this study. The questions to be addressed are:

1. What associations exist between the prevalence of global brain atrophy, white matter changes, and infarct-like lesions detected on MRI and the available data on participant demographics (age, sex, weight, height), behavior (smoking, alcohol consumption, physical activity), disease status (self-perceived health status and pre-existing cardiovascular, cerebrovascular, or metabolic disease), and various clinical and laboratory measures of organ system functioning (e.g, blood pressure, liver function, kidney function, lung function, mental functioning)?

2. What associations exist between the prevalence of MRI changes and mortality?

3. Are any of the MRI changes predictive of mortality beyond the predictive capabilities of the available non-MRI variables?

4. Do the changes on MRI reflect underlying risk factors for death separate from other known disease processes (such as cardiovascular disease, kidney disease, diabetes, high blood pressure), or are the MRI changes merely signs of those disease processes?

AVAILABLE DATA

The data are stored in the file mriproject.txt in ASCII format. Each line corresponds to the observations on one of the 735 participants. When data is missing for a particular variable, `NA' is recorded. The descriptions of the variables are as follows:

- *ptid*= Participant identification number
- *mridate*= The date on which the participant underwent MRI scan in MMDDYY format.
- *age*= Participant age at time of MRI (years)
- *male*= Indicator of whether participant is male (0= female,1=male)
- *race*= Indicator of participant's race (1= white, 2= black, 3= Asian, 4= other)
- *weight*= Participant weight at time of MRI (pounds)
- *height*= Participant height at time of MRI (centimeters)
- *packyrs*= Participant smoking history in pack years (1 pack year = smoking 1 pack of cigarettes per day for 1 year). A participant who never smoked has 0 pack years.
- *yrsquit*= Number of years since quitting smoking. A current smoker will have nonzero packyrs and 0 for yrsquit. A never smoker will have 0 for both packyrs and yrsquit.
- *alcoh*= Average alcohol intake for the participant for the two weeks prior to MRI (drinks per week, where 1 drink = 1 oz. whiskey, 4 oz. wine, or 12 oz. beer).
- *physact*= Physical activity of the participant for the week prior to MRI (measured in 1,000 kcal)
- *chf*= Indicator of whether the participant had been diagnosed with congestive heart failure prior to MRI (0= no, 1= yes). Congestive heart failure is a condition in which the heart muscle becomes too weak to pump blood properly.
- *chd*= Indicator of whether the participant had been diagnosed with coronary heart disease prior to MRI (0= no, 1= diagnosis of angina, 2= diagnosis of myocardial infarction). Coronary heart disease is the condition in which the arteries which supply blood to the heart muscle (the coronary arteries) become blocked. In such a situation, the heart muscle does not get sufficient oxygen and may die. If the blockage is not complete, a patient will occasionally suffer chest pain (angina), especially with exercise. If the blockage is complete and not treated promptly, some part of the heart muscle may die: a myocardial infarction (MI) or heart attack.
- *stroke*= Indicator of whether the participant had been diagnosed with a cerebrovascular event prior to MRI (0= no, 1= diagnosis of a transient ischemic attack, 2= diagnosis of a stroke). Cerebrovascular disease refers to narrowing of the blood vessels that supply the brain (cerebrum). In mild cases of the disease, a patient may experience a short period of weakness or paralysis of one side of his/her body and/or difficulties with speech, but then he/she will recover completely. Such an event is called a transient ischemic attack (or TIA), where ischemia means a condition in which tissue is deprived of oxygen. In severe cases of the disease, a blood vessel in the brain may be completely blocked, or it may rupture causing bleeding into the brain. In either of these events, a portion of the brain is deprived of oxygen and dies (a cerebral infarct is when the vessel is blocked, a cerebral hemorrhage is when the vessel ruptures, and both of these conditions are more popularly called a stroke). When the brain tissue dies in certain key areas of the brain, the patient might lose use of half of their body, and depending upon the side of the brain affected, may lose speech.
- *diabetes*= Indicator of whether the participant had been diagnosed with diabetes prior to MRI (0= no, 1= yes). Diabetes is a disease in which a patient does not regulate his/her blood glucose in a normal fashion. Glucose is the main energy source for our bodies, and in diabetes, the cells lose the ability to take glucose from the blood. Persons with diabetes are at high risk of blindness, kidney disease, heart disease, and other diseases of the circulation system.
- *genhlth*= An indicator of the participant's view of his/her own health (1= excellent, 2= very good, 3= good, 4= fair, 5= poor).
- *ldl*= A laboratory measure of a certain kind of cholesterol in the participant's blood at the time of MRI. LDL (low density lipoprotein) is often referred to as ``bad cholesterol'', because persons with high levels of LDL have been found to have higher risk of heart disease and cerebrovascular disease. (HDL- high density

lipoprotein-is the ``good cholesterol".) Typical ranges for LDL tend to be age dependent, with measurements between 100 and 189 mg/dl reported as typical for persons over 70.

- *alb*= A laboratory measure of a certain kind of protein in the participant's blood at the time of MRI. Albumin is made by the liver, and persons with poor liver function or poor nutritional status will have low levels of albumin. Most often, albumin is used as a marker for normal liver function. Typical ranges for albumin are 3.2 to 5.5 g/l.
- *crt*= A laboratory measure of creatinine in the participant's blood at the time of MRI. Creatinine is a waste product of muscles that is excreted by the kidneys. In persons with kidney disease, creatinine is not excreted appropriately, and it builds up in the blood. Hence high levels of creatinine are taken as indication of kidney disease. ``Normal'' levels of creatinine are approximately 0.5 to 1.2 mg/dl.
- *plt*= A laboratory measure of the number of platelets circulating in the particpant's blood at the time of MRI. Platelets are usually the first step in blood clotting. A wide variety of diseases will cause a decrease in the body's ability to form blood cells, and thus low platelet levels are often an indication of chronic disease or infections. Some diseases also cause platelet counts to be too high. Typical ranges for platelet counts are 140 to 440 thousand platelets per cubic millimeter.
- *sbp*= A measurement of the participant's systolic blood pressure in his/her arm at the time of MRI. The systolic blood pressure is the maximum pressure generated during a contraction of the heart muscle. Persons with high blood pressure have been found to be at increased risk for heart disease, cerebrovascular disease, and kidney disease. The ``normal'' range for systolic blood pressure is 110 to 140 mm Hg.
- *aai*= The ratio of systolic blood pressure measured in the participant's ankle at the time of MRI to the systolic blood pressure measured in the particpant's arm. Typically, we measure blood pressure in the arm. However, in patients with severe hardening of the arteries, the arteries in the legs may become so blocked as to restrict blood flow to the lower extremities. Thus, measuring the ankle blood pressure relative to the arm blood pressure is a marker of extent of arterial disease: A low ankle : arm index suggests more severe peripheral arterial disease. A person with no peripheral arterial disease might be expected to have aai=1.
- *fev*= A measure (in liters per second) of forced expiratory volume in the participant at the time of MRI. This measures the volume of air that can be forcibly exhaled within 1 second. Normal FEV measurements depend upon the size of the lungs, which in turn is usually proportional to body size.
- *dsst*= A measure of cognitive function (ability to think) for the participant at the time of MRI. In this Digit Symbol Substitution Test, the participant is asked to replace each of a number of numerical digits with specific symbols. The highest possible score on the test is 100.
- *atrophy*= A measure of global brain atrophy detected on MRI. In persons with shrinking (atrophy) of the brain, certain fluid filled cavities in the cerebrum (the ventricles) become larger. From the MRI exam, a measurement of the degree of ventricular enlargement relative to predicted ventricular size was made. These measurements were then rescaled to be a number between 0 and 100, with 0 indicating no ventricular enlargement and 100 indicating the most severe degree of atrophy.
- *whgrd*= A measure of white matter changes detected on MRI. When dissected, a brain appears as regions of gray matter and regions of white matter that appear in characteristic areas. MRI can detect the difference between white matter and gray matter quite well. With aging, there is often seen some changes in the appearance of the white matter on MRI scans. The radiologists performing the MRI were asked to assign a number between 0 and 9 to each participant on the basis of the extent of such white matter changes. A grad of 0 means no changes, and a grade of 9 means marked changes.
- *numinf*= A count of the number of distinct regions identified on MRI scan which were suggestive of infarcts (dead areas of the brain due to oxygen deprivation). It should be noted that the clinical impact of brain lesions is very dependent upon the location of the lesion. However, in this dataset, we have no information about location of the lesions.
- *volinf*= A measure of the total volume (in cubic centimeters) of the infarct-like lesions found on MRI scan.
- *obstime*= The total time (in days) that the participant was observed on study between the date of MRI and death or September 16, 1997, whichever came first.
- *death*= An indicator that the participant was observed to die while on study. If death=1, the number of days recorded in *obstime* is the number of days between that participant's MRI and his/her death. If death=0, the number of days recorded in *obstime* is the number of days between that participant's MRI and September 16, 1997.