STATISTICS IN MEDICINE
Statist. Med. 2004; **23**:351–367 (DOI: 10.1002/sim.1743)

Managing and analysing data from a large-scale study on Framingham Offspring relating brain structure to cognitive function

Joseph M. Massaro[1,2,*,†], Ralph B. D'Agostino Sr[2], Lisa M. Sullivan[1,2], Alexa Beiser[1], Charles DeCarli[3], Rhoda Au[4], Merrill F. Elias[2] and Philip A Wolf[4]

[1]*Department of Biostatistics, Boston University School of Public Health, 715 Albany Street T3E, Boston, MA 02118, U.S.A.*
[2]*Statistics and Consulting Unit, Department of Mathematics and Statistics, Boston University, U.S.A.*
[3]*Department of Neurology and Center for Neuroscience, University of California at Davis, U.S.A.*
[4]*Department of Neurology, Boston University School of Medicine, U.S.A.*

SUMMARY

At the Framingham Heart Study under separate research grant funding from the National Institute of Aging, NIH, we are gathering brain structure and cognitive information on the Framingham Offspring, creating one of the largest known data sets to assess changes in brain structure associated with normative ageing and cognitive decline. Subject recruitment, data collection, data management and statistical analysis require a collaborative integrated effort on the part of the Framingham project team. Here we describe this effort, as well as the various brain structure and cognitive function parameters we are now collecting. We are currently performing analyses of data collected through 2002, and we discuss the statistical issues arising relating brain structure parameters to cognitive function. Copyright © 2004 John Wiley & Sons, Ltd.

1. INTRODUCTION

Anatomical and brain imaging studies have reported substantial differences in brain structure between men and women over a wide age range. The volume of white matter hyperintensities (WMH) seen in magnetic resonance images (MRIs) of the brain is an independent predictor of cognitive function as measured by neuropsychological (NP) test performance, especially related to frontal lobal functioning [1, 2]. The ability to extend the results of many of these studies to the general population is limited; however, due to the generally small study sample

*Correspondence to: Joseph M. Massaro, Department of Biostatistics, Boston University School of Public Health, 715 Albany Street T3E, Boston, MA 02118, U.S.A.
†E-mail: jmm@bu.edu

Contract/grant sponsor: National Heart, Lung and Blood Institute; contract/grant number: NO1-HC-25195
Contract/grant sponsor: National Institute of Aging, National Institutes of Health; contract/grant numbers: 5R01-AG16495 and 5R01-AG08122

sizes and restrictive health criteria of these studies. Given the current lack of normative data on brain structure, the Framingham Heart Study received a grant to quantify brain structure from MRIs of male and female participants of the Framingham Heart Study ranging in age from 34 to 97 years and to assess the relationship between brain structure and cognitive function.

In this paper, we discuss the management of this study and the statistical analysis and management of its data. Specifically, in Section 2, we provide a brief background of the Framingham Heart Study. In Section 3, we describe in more detail the brain MRI study and the characteristics of the study's participants. In Section 4, we outline the study management, data collection and data management protocols for the brain MRI study. In Section 5, we detail the MRI and NP measures. In Section 6, we describe issues we are encountering in analysing the relationship between MRI parameters and cognitive ability. In Section 7 we provide a summary and highlight future directions.

2. BACKGROUND OF THE FRAMINGHAM HEART STUDY

The general design and demographics of the Framingham Heart Study have been previously described [3, 4]. In brief, the Framingham Heart Study is a community-based population study of over 10 000 individuals. Extensive data have been collected from the Study participants, including major risk factors for cardiovascular disease (CVD) and cerebrovascular disease (e.g. stroke), through periodic physical and medical examinations. The occurrence of heart disease, stroke and dementia including Alzheimer's disease (AD), fulfilling standard criteria for diagnosis over the decades, has also been collected on the Framingham Study participants. The Framingham Study has a proven record of analysing longitudinal data to provide significant insights into disease detection and prevention.

The Original cohort of the Framingham Heart Study included 5209 participants from Framingham, MA who were enrolled into the study in 1948. At enrolment, the mean age was 44 years (range 28–62 years), 55 per cent were female, the majority were white and of middle socioeconomic class. In November 2001, the 26th examination of the Original cohort was completed on 558 participants (73 per cent of surviving cohort). The average age of participants undergoing this examination was 86 years (range 79–103 years old); 55 per cent were female.

The Offspring cohort included 5124 offspring of the original cohort and their spouses enrolled in 1971. At enrolment, the mean age was 36 years (range 5–70 years), 52 per cent were female. In October 2001, the 7th examination of the Offspring cohort was completed on 3539 participants (82 per cent of surviving cohort). The average age of participants undergoing this examination was 62 years (range 33–90 years old); 53 per cent were female.

As part of a large ancillary study on brain structure and cognitive function, Offspring cohort participants were recruited, starting in March 1999, to undergo MRI of the brain and to be administered an NP test battery. The MRI and NP studies allowed for establishment of quantitative measures of brain structure (e.g. volume of white matter hyperintensities [WMH]), and cognitive function, respectively. Coupled with the extensive clinical data, we have now compiled one of the largest databases of its kind allowing for comprehensive investigation of (a) changes in brain structure and cognitive performance associated with normal ageing, and (b) risk factors for Alzheimer's disease (AD) and other types of dementia. In addition,

using MRI and NP data collected from offspring's parents (under various government-funded grants) who are part of the Original cohort, we are examining genetic factors impacting both normal ageing and dementing illnesses.

3. OVERVIEW AND DESIGN OF BRAIN MRI STUDY AT FRAMINGHAM

In 1999 the National Institute of Aging (NIA), National Institutes of Health (NIH) funded a study to image the brains of the participants of the Framingham Offspring Exam 7, with data collection scheduled to end February 2004. The study is entitled the MRI, Genetics and Cognitive Precursors of AD and Dementia Project. The objectives for the short term are: (1) to assess the cross-sectional relationship between brain MRI parameters (e.g. Volume of WMH, Total Brain Volume), cognitive functioning which was measured in an extensive NP, and the Framingham Stroke Risk Profile (a composite function that generates a probability measure for future stroke based on age, gender, systolic blood pressure, diabetes, smoking, atrial fibrillation and existing CVD) [5], and (2) to assess heritability of brain MRI parameters using the Original Framingham cohort (MRI data are also being collected on approximately 275 of the original Framingham cohort, some of whom have diagnosed dementia) and siblings of participants in the MRI brain study who were also part of the Framingham Offspring study. The long-term objectives of the MRI brain study are: (1) to assess changes in MRI parameters over time and relationships between MRI parameters and changes in MRI parameters and onset (incidence) of dementia and Alzheimer's disease, and (2) to investigate whether and to what extent sub-clinical measures in MRI relate to the development of dementia/Alzheimer's disease. Assessing these long-term objectives will require collection of follow-up data to assess incident dementia and Alzheimer's disease.

The protocol for study management, data collection and data management for this current brain MRI sub-study are described in the next section.

4. STUDY MANAGEMENT, DATA COLLECTION AND DATA MANAGEMENT OF THE BRAIN MRI STUDY

At the time of initial funding of the Offspring brain MRI study, we set an objective to collect and manage data on brain structure and cognitive performance on as many Offspring as possible in order to maximize power and to enhance generalizability of results. With a potential pool of approximately 3500 participants (the number of attendees at Offspring Exam 7), we realised that this would involve a large collaborative effort on the part of many individuals. We first discuss the procedures for recruitment and data collection, followed by procedures for data and study management.

4.1. Scheduling of participants for the regularly scheduled physical exam (Exam 7)

Each Framingham Offspring participant eligible for Exam 7 was scheduled for the physical examination by a member of the Framingham Study's recruitment team. For Offspring participants in the New England area, the physical and laboratory examinations of the Offspring were conducted on site at the Framingham Heart Study and include measurement of a wide

array of risk factor data (e.g. systolic blood pressure, total cholesterol) and disease (e.g. cardiovascular disease) occurrence measures including documentation of the course and patterns of decline for older subjects with dementia. These physical examinations were carried out by Framingham physicians and medical technicians. Participants willing to be examined but unable to travel to the Framingham Heart Study clinic due to health issues and participants who lived outside of New England with no plans to visit Framingham during the time of the exam (approximately 5 per cent of the Offspring cohort live outside of the New England area, predominantly in Florida, Arizona and California) were given a physical exam at their residence by Framingham technicians.

To identify subjects hospitalised for dementia, stroke or other concurrent illnesses, a system of daily screening of medical and admission records is currently in place for all Framingham Study participants admitted to the only hospital in town, MetroWest Medical Center-Framingham Campus (hereafter referred to as MetroWest). Surveillance of hospitalisations outside of the MetroWest region and deaths of Framingham Study participants is conducted by the medical records staff at Framingham. For deaths, all hospital records prior to death event are requested as well as copies of death certificates and autopsy results (when available) in order to determine cause of death.

4.2. Recruitment of participants for the brain MRI study

At the time of scheduling the physical and laboratory exam, each participant (including participants residing outside of New England) was asked if he/she would be interested and willing to participate in the brain MRI study. Specifically, each participant examined on-site at Framingham was given the opportunity to schedule an appointment for an MRI and to be administered the NP battery. For willing participants living in New England, MRIs were performed at the MetroWest Imaging Center in Framingham, MA; each MRI session took approximately 30–45 min per participant. For participants outside of New England willing to be part of the study but unable to come to Framingham to be imaged, arrangements were made with regional MRI centres. Specifically, the Framingham team identified regional MRI centres with comparable imaging machines and provided training and continuous monitoring of the Framingham imaging protocol, the technical details of which can be found in References [6–10].

Participants imaged at MetroWest who agreed to participate in the NP portion were administered the NP battery at MetroWest by trained Framingham Heart Study neuropsychology examiners. The NP battery was usually administered before the MRI and took approximately 45 min to complete. Participants unable to complete the brain MRI (e.g. due to claustrophobia) were still asked to complete the NP battery. If a participant was unable to complete the brain MRI and NP battery on the same day, every effort was made to ensure the MRI and NP battery were completed within 6 months of one another (otherwise, the participant was not to be included in cross-sectional analysis relating cognitive function to brain structure). Trained Framingham neuropsychology technicians also scheduled and performed home visits to administer the NP battery for willing participants residing outside of the New England area. When MRI studies were conducted off-site, the MRI and neuropsychological tests were usually not performed on the same day. However, they were usually performed within the desired 6 month window. In total, through December 2002, brain MRI was performed on the same day as the neuropsychological evaluation in 97 per cent of all subjects; over 99 per cent of all subjects had an MRI within 6 months of neuropsychological testing.

Table I. Demographic characteristics for participants with and without an MRI.

Parameter	No MRI/No NP	No MRI/NP	MRI/NP	*P*-value
Per cent female	54.4 per cent	54.5 per cent	53.1 per cent	0.779
Age (years)	62.1 ± 9.7	60.9 ± 9.3	60.6 ± 9.4	<0.001
History of CVD	16.6 per cent	16.6 per cent	11.5 per cent	<0.001
History of MI	6.5 per cent	6.0 per cent	4.0 per cent	0.010
History of CHD	11.5 per cent	10.6 per cent	8.1 per cent	0.009
History of diabetes	17.1 per cent	12.8 per cent	12.0 per cent	0.001
Smoker	15.0 per cent	12.3 per cent	12.4 per cent	0.155
Taking antihypertensive Rx	40.2 per cent	37.0 per cent	30.8 per cent	<0.001
History of hypertension	52.0 per cent	50.2 per cent	42.3 per cent	<0.001
Systolic blood pressure (mmHg)	129.5 ± 19.5	128.5 ± 19.1	125.9 ± 18.5	<0.001
Diastolic blood pressure (mmHg)	74.5 ± 10.2	74.6 ± 10.0	73.8 ± 9.6	0.094
Body mass index	28.3 ± 5.1	29.6 ± 6.6	28.0 ± 5.2	<0.001
Total cholesterol (mg/dl)	198.5 ± 37.3	201.0 ± 36.9	201.1 ± 36.7	0.209
HDL cholesterol (mg/dl)	54.1 ± 17.7	53.4 ± 16.2	53.4 ± 16.8	0.548
Glucose (mg/dl)	106.4 ± 26.7	105.7 ± 35.6	103.4 ± 26.5	0.015

CVD = cardiovascular disease; MI = myocardial infarction; CHD = coronary heart disease. Numbers are percentages or mean $\pm$ standard deviation. *P*-value assesses significance of the difference between the three groups using one-way analysis of variance for continuous parameters and the χ^2 test for dichotomous parameters.

4.3. Characteristics of the brain MRI study participants

Through December 2002, 3365 Offspring participants were asked to undergo the MRI and NP testing. Approximately 20 per cent refused the brain MRI and gave no reason for refusal, approximately 5 per cent attempted to undergo the MRI but prematurely discontinued due to claustrophobia (not including several claustrophobic patients who agreed to re-attempt the MRI and were successfully imaged once a larger imaging machine became available at MetroWest), and approximately 2.5 per cent were unable to undergo to the MRI due to potential interference with implanted devices such as pacemakers. Over 70 per cent of the 3365 Exam 7 Framingham Offspring underwent an MRI through December 2002. Table I displays a breakdown of demographic characteristics for patients recruited through December 2002 who (a) did not receive an MRI or NP battery; (b) received an NP battery but not an MRI; and (c) received both an MRI and NP. As expected, participants who did not receive an MRI were generally less healthy than those participants who did.

4.4. Data management

The general flow of data is shown in Figure 1. Further details are provided below.

4.4.1. Data management of NP data. The NP battery consists of approximately 360 questionnaire items. Hardcopies of completed NP forms are sent to Framingham data management for data entry and cleaning approximately once a week. The data management cleaning protocol calls for logging, verification of study identification numbers and date sequencing. Each item in the battery is further checked for out of range values, consistency with related

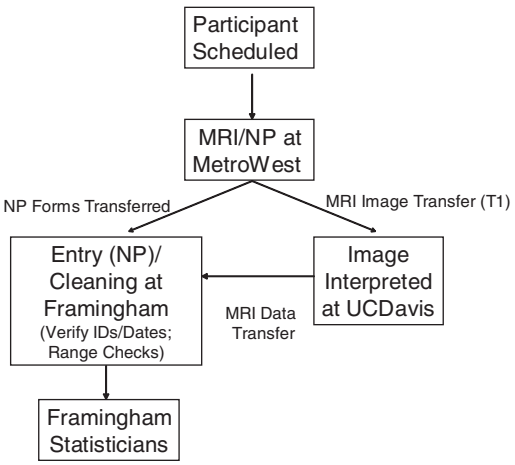

Figure 1. MRI data flow.

questions, completeness and accuracy of skip pattern execution. Data queries are generated and given to the Framingham technicians and neuropsychologists for resolution.

Data entry is carried out in the INGRES system, with periodic (usually monthly) transfers to SAS® data sets. Data checking, cleaning and query generation are performed using customised SAS® programs (in 2002, 3 years into this brain MRI study, Framingham data management acquired Clintrial™, an Oracle-based system for data entry and cleaning; however, data management for this brain MRI study was not transferred to this system because the INGRES entry screens and SAS programs were already developed and tested, and because a majority of the study's data had already been collected and entered by 2002). In general, data is entered and cleaned within 1–2 months of receipt by the Framingham data management group.

4.4.2. Data management of the MRI data. All MRI images taken at the MetroWest facility in Framingham are transferred electronically via a T1 line to Dr Charles DeCarli at the University of California at Davis (UC Davis) for reading and interpretation. All MRI images taken off-site are transferred to UC Davis on optical disks. After acquisition of the images at UC Davis, the digital information is transferred to a central laboratory directed by Dr DeCarli. Analyses and interpretation of the images are then performed at UC Davis *blinded to any subject personal identifying information* and measurements of brain structure are recorded. Complete technical details of image acquisition and interpretation can be found in References [6–10].

At UC Davis, the resulting brain structure parameters are recorded on a Microsoft EXCEL© spreadsheet with one record per participant, identified by unique Framingham Study identification numbers. An EXCEL© spreadsheet with new data is sent to the Framingham data management team approximately once a month. Upon receipt at Framingham, the EXCEL sheet is transferred to an SAS® data set. The data management team then verifies the participants' identification numbers and cross checks MRI dates on the spreadsheet with those

stored in Framingham's 'roster', an INGRES-based system in which the date the participant underwent the MRI is recorded. Each MRI parameter (there are approximately 24 such parameters as shown in Table III; further details are given in Section 5) is checked for out of range values, consistency and completeness according to algorithms specified by Dr DeCarli. Any discrepancies are resolved with UC Davis staff usually via e-mail.

4.4.3. Analysis data sets. A designated Framingham statistician (an author on this paper [AB]), oversees the combining of the MRI data, neuropsychological data and clinical risk factor data. Analysis data sets generally contain one record per patient.

4.5. Study management

Upon grant award, team members were chosen from the existing staff or were newly hired by the various managers in each department at Framingham (e.g. neuropsychology technicians, data management, statistics). In addition, a project manager was assigned. The project team, consisting of recruiters, neuropsychologists, investigators (including Principal Investigator Dr Philip A. Wolf), data managers and statisticians from Framingham and Dr DeCarli and technicians from UC Davis began to meet (and still continue to meet) once a month via conference call and occasionally in face-to-face meetings. At the first set of meetings, team member roles and responsibilities were identified, and a general agenda was set for all subsequent meetings. The agenda consisted of (1) Status of recruitment (on-site and off-site), including status of participant refusals; (2) ways to maximize recruitment (e.g. using a larger MRI machine for claustrophobic patients); (3) Status of MRI and NP data—collection, entry and cleaning/resolving any outstanding queries; (4) off-site issues (e.g. identifying MRI sites for the off-site participants; assessing the sites' ability to follow the Framingham imaging protocol); and (5) analysis issues/status of analyses/new ideas for analysis and future grants.

These team meetings have been critical in ensuring that the study runs smoothly and that all parties are abreast of the issues. Further interactions and discussions occur between individual groups (e.g. data management and neuropsychologists to resolve NP queries; data management and UC Davis to resolve MRI queries; UC Davis, recruiters and neuropsychologists to ensure off-site recruitment and imaging are carried out appropriately) often and as needed. The project manager is copied on such discussions to ensure that issues are resolved efficiently.

5. DESCRIPTION OF MRI AND NEUROPSYCHOLOGICAL MEASURES

5.1. Neuropsychological measures

The NP battery contains items that measure an array of cognitive domains. These domains and the specific NP battery items that tap each of them are outlined in Table II. Since the NP battery consists of multiple tests measuring related cognitive domains, we also derived composite scores from this battery, for analytical purposes, which met the following criteria: (1) each composite score was constructed from variables indexing a common factor; (2) the same item was not included in more than one composite; (3) the composites were theoretically meaningful. Principal components analysis was used to determine composite scores. Since 'Similarities' is highly correlated with general intelligence [11], and the Wide Range Achievement Test-Reading is an index of reading achievement and 'pre-morbid' intellect [12],

Table II. Neuropsychological test battery.

Component of neuropsychological test	Latent cognitive ability tested
Boston naming test	Confrontational naming and language
WAIS*	
Similarities	Abstract reasoning; general verbal intelligence
WMS†	
Paired associate learning	New learning and memory
Logical memory-immediate recall	Immediate recall of verbal passages
Logical memory-delayed recall	Delayed recall-verbal passages
Logical memory-delayed recognition	Recognition memory-verbal passages
Visual reproductions-immediate recall	Immediate memory-visual-spatial
Visual reproductions-delayed recall	Delayed recall-visual-spatial
Visual reproductions-delayed recognition	Recognition memory-visual-spatial
HRB‡	
Trail-making A	Attention, concentration, visual scanning, flexibility and motor speed
Trail-making B	Attention, concentration, visual scanning and motor speed; flexibility and executive function
Hooper Visual Organisation	Visual Organisation; some demands on executive function

*Wechsler adult intelligence test.
†Wechsler memory scale.
‡Halstead Reitan neuropsychological test battery.

both were excluded from the principal components. Thus, to extract factors, we performed a principal components analysis on the remaining 10 measures in Table II, followed by an orthogonal (Varimax) rotation [13]. As discussed in Reference [14], where the principal components was first performed on the population, four factors were extracted using a criterion of eigenvalues greater than 1.00 (remaining eigenvalues <0.68). Using a criterion of rotated factor pattern scores above 0.50, the following variables loaded on the four factors: Factor 1 (Hooper, 0.57; Visual Reproductions (VR)- Delayed Recognition, 0.76; VR-Delayed Recall, 0.88; VR-Immediate Recall, 0.88); Factor 2 (Logical Memory-Delayed Recognition, 0.76; LM-Delayed Recall, 0.89; LM-Immediate Recall, 0.90); Factor 3 (Trails A, 0.80; Trails B, 0.87); Factor 4 (Paired Associates Learning, 1.00). When the Paired Associates Test was excluded from the variable list and the factor analysis was repeated, three factors were identified with loadings highly similar to the initial analysis. The factors were labeled as follows: (1) Visual-Spatial Memory and Organisation; (2) Verbal Memory; and (3) Visual Scanning and Motor Speed, respectively. A composite score indexing each factor was created by averaging the standardised scores of the questionnaire items of which the factor was comprised. These composite scores are used in analyses we are conducting relating brain structure to cognitive function.

5.2. Brain structure (MRI parameters)

An array of brain structure measures were developed based on the MRI images and are outlined in Table III. See References [6–10] for complete details of the measurement techniques and descriptions of the brain regions and associated parameters. In brief, frontal lobar regions

Table III. Brain structure parameters obtained from MRI.

MRI measure	Abbreviation
Total cranial volume	TCV
Total hemispheric brain volume	TCB
Total hemispheric CSF	TCC
Segmented R lateral ventricle volume	RLC
Segmented L lateral ventricle volume	LLC
Segmented third ventricle volume	IIC
Segmented WMH volume	WMH
Segmented R temporal brain volume	RTB
Segmented L temporal brain volume	LTB
Segmented R frontal brain volume	RFB
Segmented L frontal brain volume	LFB
Segmented R parietal brain volume	RPB
Segmented L parietal brain volume	LPB
Segmented R occipital brain volume	ROB
Segmented L occipital brain volume	LOB
Segmented R temporal CSF	RTC
Segmented L temporal CSF	LTC
Segmented R frontal CSF	RFC
Segmented L frontal CSF	LFC
Segmented R parietal CSF	RPC
Segmented L parietal CSF	LPC
Segmented R occipital CSF	ROC
Segmented L occipital CSF	LOC

R = right; L = left; WMH = white matter hyperintensity; CSF = cerebral spina.

were defined as all supratemporal structures anterior to the aqueduct of Sylvius. Temporal lobe volume was traced from the anterior pole of the temporal lobe to the aqueduct of Sylvius. The superior-medial temporal lobe boundary was defined as a straight line drawn from the angle of the medial temporal lobe, where it attaches to the temporal stem, to the mid-point of the operculum (for example see References [10, 15]). The dura of the middle cranial fossa was then traced around each temporal lobe to complete the temporal lobe region. The parietal lobes were defined as the brain matter posterior to the aqueduct of Sylvius, extending to the medial transverse fissure of the striate cortex. The remaining caudal portions of the cerebral hemispheres were defined as occipital. Primary parameters used for analysis are total cranial volume (TCV, an assessment of the size of cranial vault), total hemispheric brain volume (TCB) and volume of white matter hyperintensities (WMH). For the discussion that follows, we will focus on WMH, a primary brain structure parameter.

6. ANALYTIC ISSUES

With data collected from this study through 2002, we are investigating the relationship between brain structure, other ‘standard clinical risk factors’ (e.g. age, gender, systolic blood pressure, cholesterol), and cognitive functioning. In the process we have encountered analytic issues that

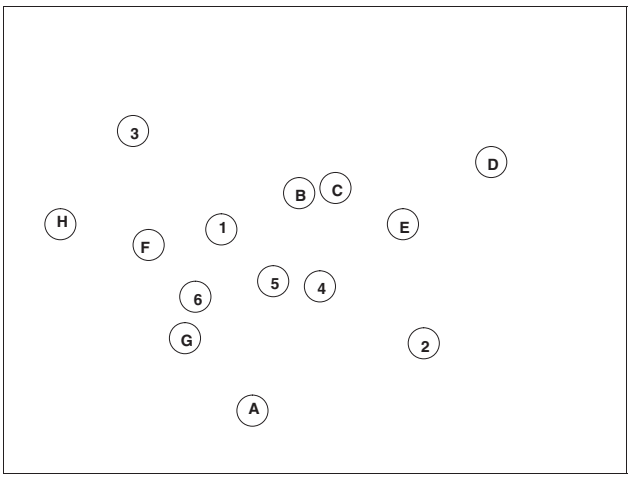

Figure 2. Trail marking test (TMT).

must be considered. Below we describe a particular example and propose several approaches for addressing the analytical issues that arise.

A hypothesis was generated that the larger the WMH in an individual, the poorer the cognitive function. WMH, or more specifically the ratio of WMH to total cranial vault (TCV), is the independent variable in this analysis. We use the WMH/TCV ratio (multiplied by 100) as opposed to WMH since WMH is a function of brain size, which is in turn a function of cranial volume. Dividing by TCV is a reasonable method for correcting for head size differences between participants. Note that we correct for TCV in a similar manner for all brain structure variables.

For cognitive function we are analysing separately as outcomes each of the various NP battery domains discussed in Table III as well as the three composite factors discussed above that were created from these domains. Here we will discuss our analysis on the outcome measure of time to complete the Trail Making Test (TMT). Time to complete TMT is the time (in seconds) the examinee takes to connect dots in numerical and alphabetical order; i.e. time it takes to connect dots in the order A-1-B-2-C-3, etc. (See Figure 2). The TMT is a test of attention and motor speed often used to examine frontal lobe function. It has also shown to activate the frontal lobe during electrophysiological measures [16]. Short-term memory impairments have been reported in patients with frontal lobe damage. This difficulty could be due to the inability in patients with frontal lobe damage to block out interference from the outside or from within and keep their attention on the task at hand. These patients' 'working memory' is not able to hold and internalise information that they would then use to direct behaviour in the absence of external cues [16]. Participants not completing the TMT due to cognitive issues in the opinion of the examiner were assigned a maximum value of 500 s (approximately 30 s longer than the maximum TMT time in the data set). Since the time to complete TMT is highly skewed, we perform our analysis using as an outcome of the natural logarithm of TMT [log(TMT)], which is more symmetric (see Figure 3). Note that extreme TMT values do exist, primarily due to participants who did not complete the

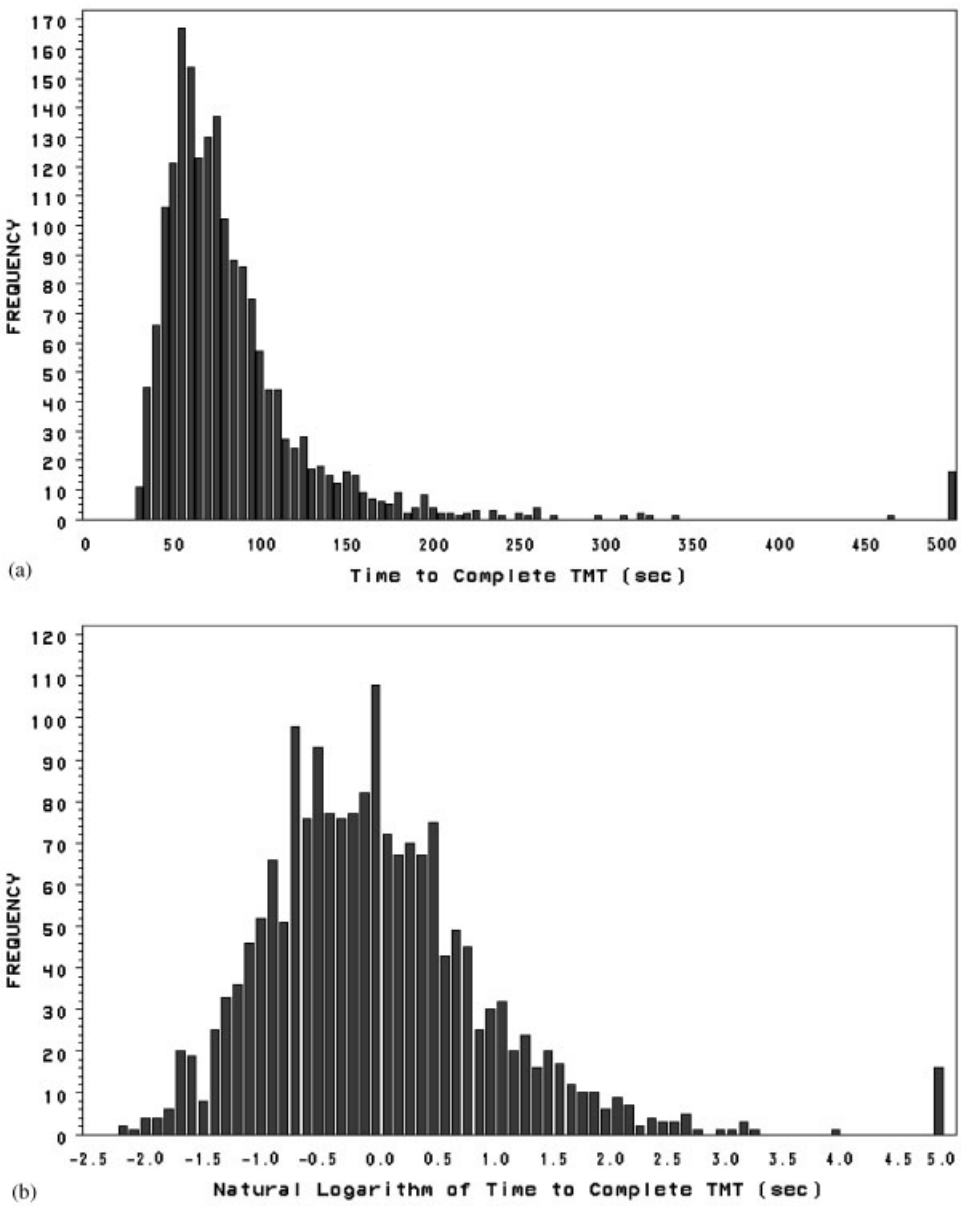

Figure 3. Distribution of (a) TMT and (b) Log(TMT) in seconds.

exam for cognitive reasons. Analyses run on log(TMT) including and excluding these extreme values yielded similar results; the results described below include extreme values.

In regressing log(TMT) on WMH/TCV ratio, we considered several approaches. For example, in the Framingham Study, when assessing the relationship between a continuous indepen-

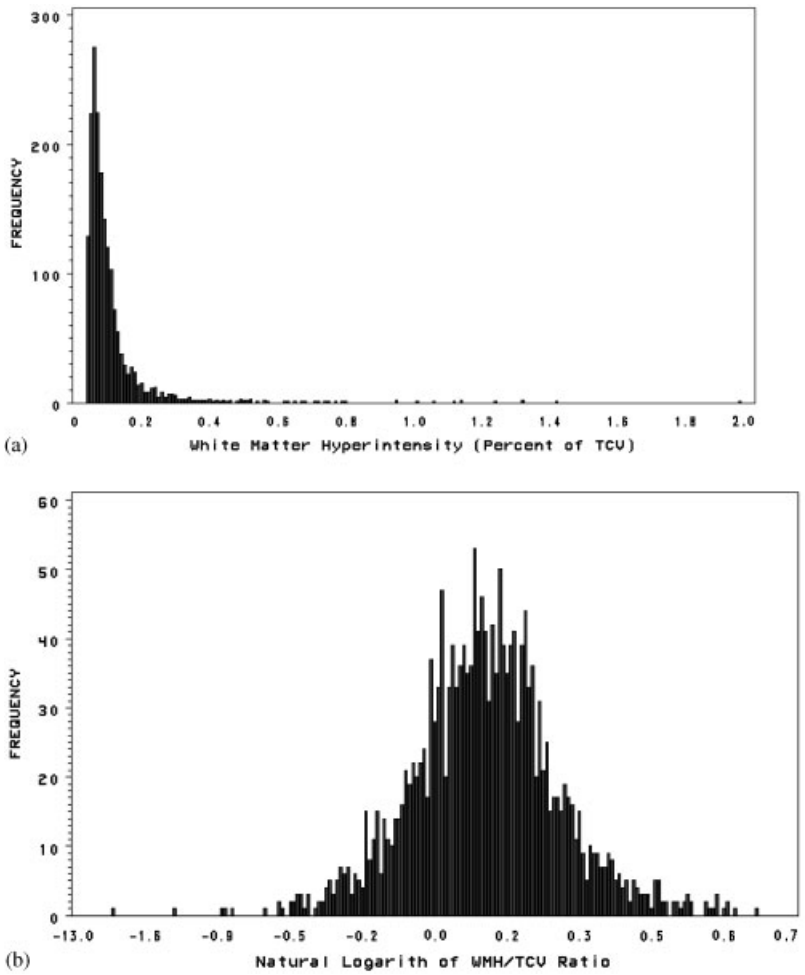

Figure 4. Distribution of (a) WMH/TCV ratio and (b) Log(WMH/TCV ratio).

dent variable and an outcome, we often dichotomise or categorise the independent variable first and then assess the difference in outcome distribution across the categories. The loss of power, if any, is minimal given the large sample size of the Framingham Study, and the interpretation of results is usually more straightforward. Despite this, we first considered treating WMH/TCV as a continuous independent variable. Figure 4(a) displays the distribution of WMH/TCV ratio. As can be seen, the distribution is highly skewed with several potentially influential observations on the high end. We then considered the natural logarithm of WMH/TCV ratio for analysis (distribution in Figure 4(b)) since it is relatively symmetric. However, while analysis relating log(WMH/TCV) to log(TMT) was justified, we were concerned about the ease with which results could be interpreted in a clinically meaningful manner. To facilitate interpretation of results, we decided to categorise WMH/TCV ratio and assess differences in log (TMT) across categories. Two categorisations were considered:

Table IV. Mean TMT and age by WMH quintile.

WMH quintile	WMH range (per cent)	N	Unadjusted geometric mean TMT (s)	Mean (SD) age (years)	Age-adjusted geometric mean TMT (s)
Q1 (lowest quintile)	$>0.00^{*} - 0.020$	368	74.5	55.1 (8.1)	83.4
Q2	$>0.020 - 0.035$	366	78.2	58.6 (8.3)	82.8
Q3	$>0.035 - 0.055$	365	79.6	60.2 (8.4)	81.3
Q4	$>0.055 - 0.09$	367	86.8	63.2 (8.5)	83.1
Q5 (highest quintile)	$>0.09 - 1.93$	361	113.5	68.8 (7.5)	99.7

*Actual minimum was 2×10^{-6} per cent.

(a) dichotomizing patients as 'LARGE' or 'NON-LARGE' WMH/TCV; and (b) categorizing patients into quintiles of WMH/TCV. A participant whose WMH/TCV ratio was over one standard deviation (SD) larger than the predicted mean for that participant's age was classified as LARGE. We will discuss this dichotomy further, but will first address the analysis assessing the relationship between log(TMT) and WMH/TCV quintiles.

There is a decreasing trend in unadjusted log(TMT) across WMH/TCV quintile (see unadjusted geometric mean TMT in Table IV). There is a highly significant difference in unadjusted mean log(TMT) values between Q5 (highest quintile) and Q1–Q4 combined ($p < 0.001$) such that participants in the highest WMH/TCV quintile take significantly longer on average to complete the TMT (i.e. have more cognitive deficit) than those in the lowest four quintiles. However, before attributing significant differences in cognitive deficit to larger WMH/TCV, we noticed the obvious increasing age trend across the WMH/TCV quintiles (Table IV). Therefore, the relationship between large WMH/TCV and longer time to complete TMT might be due in large part to the effect of age on cognitive function; i.e. once adjusting for age, the significance of the TMT–WMH/TCV relationship could decrease markedly. We then computed age-adjusted geometric mean TMT values in each quintile of WMH (See Table IV). These age-adjusted geometric means are estimates of the geometric mean TMT values in each quintile at age 61.8 years, the mean age of the analytic sample. There still exists a highly significant difference in log(TMT) between Q5 and Q1–Q4 combined after adjusting for age ($p < 0.001$). As expected, the differences in the age-adjusted TMT geometric means are smaller than the differences in raw TMT geometric means.

Since the ages of participants are so markedly different across the WMH/TCV quintiles, the appropriateness of age adjustment and the accuracy of age-adjusted results is questionable. Specifically, the mean age of participants in Q5 is 68.0 years, markedly higher than the 61.8 years at which the geometric mean TMT is estimated in each quintile in the age-adjusted analysis. Further, as can be seen from the box plots in Figure 5, the lower quartile of age for the Q5 group is above 61.8; in fact, approximately 84 per cent of the Q5 group is older than 61.8. Given this paucity of Q5 participants close to 61.8 years, the age-adjusted geometric mean TMT for Q5 could very well be an inaccurate estimate of the geometric mean TMT at age 61.8 for Q5, calling into question validity of age-adjusted results comparing Q5 with the lower quintiles on TMT.

To address this we investigated other alternatives to assess the WMH and TMT relationship in the presence of age. One such approach involved matching participants in Q5 with participants in Q1–Q4 on age in 1:1 manner. We actually took this a step further and matched

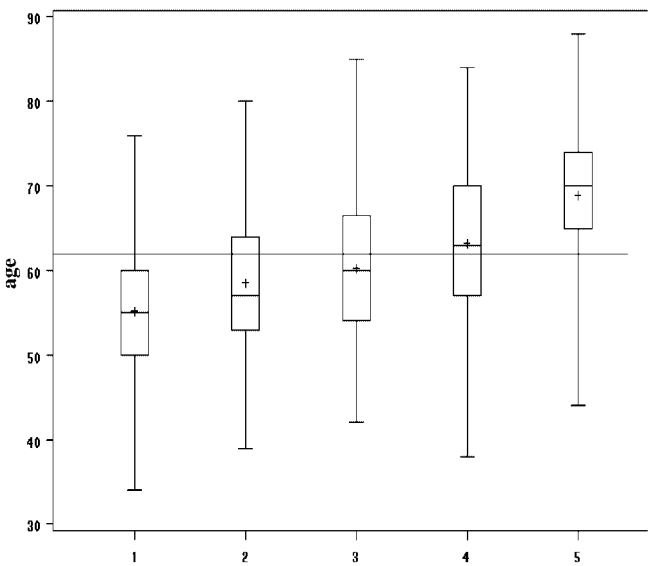

Figure 5. Ages of participants in each quintile of WMH.

Table V. Mean TMT by WMH quintile in propensity-matched sample.

WMH quintile	*N*	Unadjusted geometric mean TMT (s)	Mean (SD) age (years)	Propensity score-adjusted geometric mean TMT (s)
Q1–Q4	361	91.9	68.3 (7.2)	92.3
Q5 (highest quintile)	361	113.5	68.8 (7.5)	112.2

participants using propensity scores with Mahalanobis distance [17] on age, gender, education, systolic blood pressure and diabetes status, many of which were statistically and clinically significantly positively associated (univariately) with WMH/TCV ratio. Of the 361 'control' participants (i.e. matches with Q5), 38 per cent were from Q4, 26 per cent were from Q3, 22 per cent were from Q2 and 14 per cent were from Q1. Table V shows the geometric mean TMT scores for the matched samples. After matching, the difference in log(TMT) means between the two groups is significant at $p < 0.003$ (adjusting for the propensity score). Note further the mean age (SD) of participants in the 'matched' sample is 68.3 (7.2) in Q1–Q4 mean (SD) and 68.8 (7.5) in Q5. Thus, this method has effectively evened the comparison groups in terms of age (this held true for all demographic and clinical characteristics used in matching), allowing for a more precise comparison of TMT between Q5 and Q1–Q4.

We examined a second approach for assessing the relationship between WMH and cognitive function using regression analysis. We regressed log(WMH/TCV ratio) on age and identified participants (heretofore considered as 'LARGE') who were more than one standard deviation above the predicted mean log(WMH/TCV) for the given age. Figure 6 displays log(WMH/TCV) versus age, distinguishing participants classified in the LARGE and NON-

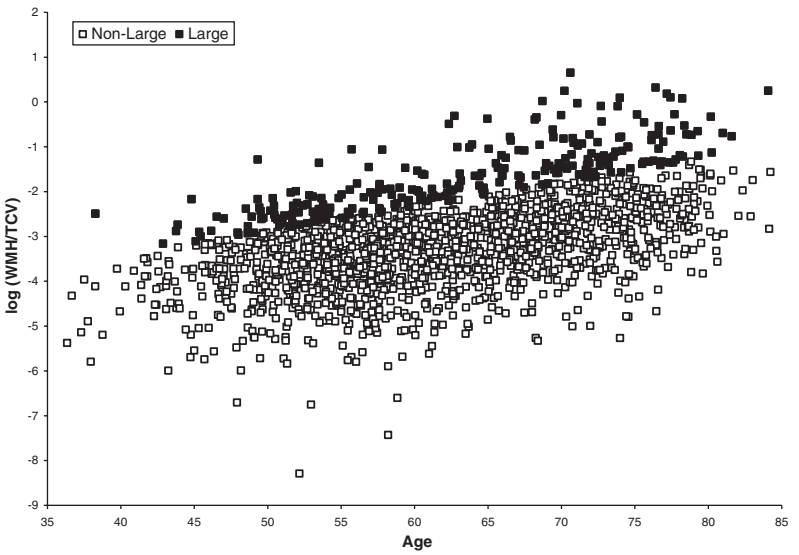

Figure 6. Large and Non-Large WMH as a function of age.

Table VI. Mean TMT by Large/Non-Large groups.

WMH	*N*	Raw geometric mean TMT	Mean (SD) age (years)	Age-adjusted geometric mean TMT
Non-Large	1576	83.7	60.2 (9.4)	83.639
Large	251	103.7	62.89 (10.0)	97.187

LARGE groups. We compared log(TMT time) between the LARGE and NON-LARGE groups. An advantage of this analysis over the quintile analysis is that it includes in the analysis persons of younger age. A disadvantage is that, especially at the younger ages, participants may be considered as LARGE despite having little white matter. (We are also investigating the use of LOESS regression to categorize participants as LARGE; conclusions so far are similar to those described below for non-LOESS, though the distribution of patients across dichotomies is more equivalent than with non-LOESS).

Table VI shows the raw and age-adjusted geometric mean TMT scores for Non-Large and Large WMH/TCV groups. The mean age of the Non-Large participants is 60.9 years as compared to 62.8 years for the participants classified as Large ($p < 0.004$). The difference in age-adjusted mean log(TMT) scores between the Non-Large and Large groups is significant at $p < 0.005$.

Owing to the significant difference in age, though not as profound as that in the quintile analysis, we again used propensity scores to match each Large group participant with up to three individuals from the Non-Large group using the propensity score method on the above-mentioned baseline covariates. The resulting adjusted difference in mean log(TMT) between the Large group and matched Non-Large group was still significantly different after adjusting

Table VII. Mean TMT by Large/Non-Large groups in propensity-matched sample.

WMH	*N*	Raw geometric mean TMT	Mean (SD) age (years)	Propensity-score adjusted geometric mean TMT
Non-Large	853	86.1	62.8 (9.5)	86.1
Large	251	103.7	63.2 (10.0)	101.5

for propensity score ($p = 0.021$). Age was no longer a confounder as the difference in age between these matched groups was 0.4 years (62.8 for Large, 63.2 for Non-Large; $p = 0.574$) (Table VII).

7. SUMMARY AND FUTURE DIRECTIONS

At the Framingham Heart Study, we are gathering brain structure and cognitive information on the Framingham Offspring, creating one of the largest known data sets to assess changes in brain structure associated with normative ageing and cognitive decline. Recruitment, data collection, data management and statistical analysis require a collaborative effort on the part of the Framingham project team. We are currently performing analysis on data collected through 2002, and have encountered a number of specific statistical issues in relating MRI measures to cognitive function.

Future analyses are planned to assess the association of polymorphisms in candidate genes with existing MRI measures, heritability of change in MRI measures after adjusting for known covariates and linkage to heritable change in MRI measures. Future longitudinal analyses are planned to assess rates of change in brain MRI measures as well as relationships between changes in MRI measures and changes in cognition over time. These analyses will be designed to adjust for important covariates such as age, educational level, dementia and family history of dementia. For each brain structure variable, analyses will compare quintiles and Large and Non-Large groups on cognitive function, incidence of dementia and incidence of Alzheimer's disease. Logistic and linear regression techniques, such as those described in the previous section, along with propensity score methods will be utilised extensively to perform these assessments.

ACKNOWLEDGEMENTS

The study isme supported by 'grant NO1-HC-25195 from the National Heart, Lung and Blood Institute, and grants 5R01-AG16495 and 5R01-AG08122 from the National Institute of Aging, National Institutes of Health.' The authors would like to thank Karen Mutalik, Bess LeRoy, Megan Young, Maureen Valentino, Katie Ball, Dr Greg Shoukamis and Maureen Gillis for their tremendous efforts in MRI and NP data collection.

REFERENCES

1. DeCarli C, Murphy DG, Tranh M, Grady CL, Haxby JV, Gillette JA, Salerno JA, Gonzales-Aviles A, Horwitz B, Rapoport SI *et al.* The effect of white matter hyperintensity volume on brain structure, cognitive performance, and cerebral metabolism of glucose in 51 healthy adults. *Neurology* 1995; **45**(11):2077–2084.

2. Boone KB, Miller BL, Lesser IM, Mehringer CM, Hill-Gutierrez E, Goldberg MA, Berman NG. Neuropsychological correlates of white-matter lesions in healthy elderly subjects. A threshold effect. *Archives of Neurology* 1992; **49**(5):549–554.
3. Dawber TR, Meadors GF, Moore FEJ. Epidemiological approaches to heart disease: the Framingham Study. *American Journal of Public Health* 1951; **41**:279–286.
4. Kannel WB, Feinleib M, McNamara PM *et al*. An investigation of coronary heart disease in families. The Framingham Offspring study. *American Journal of Epidemiology* 1979; **110**:281–290.
5. D'Agostino RB, Wolf PA, Belanger AJ, Kannel WB. Stroke risk profile: adjustment for antihypertensive medication. The Framingham Study. *Stroke* 1994; **25**(1):40–43.
6. DeCarli C, Maisog J, Murphy DG, Teichberg D, Rapoport SI, Horwitz B. Method for quantification of brain, ventricular, and subarachnoid CSF volumes from MR images. *Journal of Computer Assisted Tomography* 1992; **16**(2):274–284.
7. DeCarli C, Miller BL, Swan GE, Reed T, Wolf PA, Garner J, Jack L, Carmelli D. Predictors of brain morphology for the men of the NHLBI twin study. *Stroke* 1999; **30**(3):529–536.
8. Murphy DG, DeCarli C, Schapiro MB, Rapoport SI, Horwitz B. Age-related differences in volumes of subcortical nuclei, brain matter, and cerebrospinal fluid in healthy men as measured with magnetic resonance imaging. *Archives of Neurology* 1992; **49**(8):839–845. (Erratum in: *Archives of Neurology* 1994; **51**(1):60).
9. Murphy DG, DeCarli C, McIntosh AR, Daly E, Mentis MJ, Pietrini P, Szczepanik J, Schapiro MB, Grady CL, Horwitz B, Rapoport SI. Sex differences in human brain morphometry and metabolism: an in vivo quantitative magnetic resonance imaging and positron emission tomography study on the effect of aging. *Archives of General Psychiatry* 1996; **53**(7):585–594.
10. Carmelli D, Swan GE, DeCarli C, Reed T. Quantitative genetic modeling of regional brain volumes and cognitive performance in older male twins. *Biological Psychology* 2002; **61**(1–2):139–155.
11. Wechsler D. *The Measurement and Appraisal of Adult Intelligence* (4th edn). Williams & Wilkins: Baltimore, 1958.
12. Lezak MD. *Neuropsychological Assessment*. Oxford University Press: New York, 1983; **107**, 145–463, 556–563.
13. Cureton EE, D'Agostino RB. *Factor Analysis. An Applied Approach*. Lawrence Erlbaum Associates: Hillsdale, NJ, 1983:221–224.
14. Elias MF, Sullivan LM, D'Agostino RB, Elias PK, Beiser A, Au R, Seshadri S, DeCarli C, Wolf PA. The Framingham Stroke Risk Profile and Lowered Cognitive Performance. *Stroke*. In Press.
15. DeCarli C, Murphy DG, Gillette JA, Haxby JV, Teichberg D, Schapiro MB, Horwitz B. Lack of age-related differences in temporal lobe volume of very healthy adults. *AJNR American Journal of Neuroradiology* 1994; **15**(4):689–696.
16. Lezak MD. Domains of behaviour from a neuropsychological perspective: the whole story. *Nebraska Symposium on Motivation* 1994; **41**:23–55.
17. D'Agostino Jr RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomised control group. *Statistics in Medicine* 1998; **17**(19):2265–2281.

