Data from a longitudinal study provided measurements of cognition to screen for Alzheimer’s disease

Larry J. Brant a,*, Shan L. Sheng a, Christopher H. Morrell a, b, Alan B. Zonderman a

a Gerontology Research Center, National Institute on Aging, 5600 Nathan Shock Drive, Baltimore, MD 21224, USA
b Mathematical Sciences Department, Loyola College of Maryland, 4501 North Charles Street, Baltimore, MD 21205, USA

Accepted 7 January 2005

Abstract

Background: This article presents a computerized method to help predict individuals at risk for developing Alzheimer’s disease (AD). This would be a valuable tool for clinicians in developing treatment plans for potential AD patients. Using the initial level and rates of change in visual memory performance, such a method could predict potential AD patients in a fast and inexpensive manner. A longitudinal case–control study of 52 female and 145 male participants was performed in a gerontology research center using premorbid tests of visual memory and neurologic examinations to identify individuals with and without dementia and AD.

Methods: The classification method for each individual starts on the second examination and proceeds to compute that person’s risk of AD one examination at a time based on all the follow-up information of the remaining individuals.

Results: By performing a crossvalidation study, the optimal combination of sensitivity and specificity derived from a receiver operating characteristic (ROC) curve showed 65% of the Alzheimer cases and 75% of the noncases were correctly classified for females, while 65 and 60% of cases and noncases, respectively, were correctly classified for males.

Conclusion: Longitudinal measurements of cognition can be useful in detecting the presence of AD.

Keywords: Alzheimer’s disease; Classification; Disease screening; Longitudinal data; Random-effects models; Visual Memory Performance

1. Introduction

Alzheimer’s disease (AD) is a common cause of dementia, affecting approximately 4 million Americans, and is a leading cause of death in the United States [1]. In the next 50 years, the prevalence of AD is projected to nearly quadruple [2]. At present, there is no effective treatment for AD, and there has been only limited success in providing treatments for delaying cognitive loss in individuals with clinically diagnosed AD. Because memory changes are usually the first cognitive declines seen in AD [3,4], a method of identifying individuals with patterns of change in cognition that might be suggestive of preclinical AD seems important [5,6]. Many other studies have addressed the topic of how to identify preclinical AD patients, including major concepts such as mild cognitive impairment (MCI) [7,8] and cognitive impairment but no dementia (CIND) [9–13]. Patients with MCI are considered to be at higher risk for developing clinically probable AD, and often these patients are candidates for treatment intervention [8]. Also, when individuals with MCI are followed longitudinally, they develop to clinically probable AD faster than do age-matched controls. In a 5-year study of persons with CIND, Tuokko and colleagues [13] found that persons with memory impairment at the beginning of the study were at a three times greater risk to develop dementia. Further, the authors suggest that the rate of decline in cognition may be a better predictor of dementia than the level of cognitive impairment. The Benton Visual Retention Test (BVRT) is one test of cognition that has been used by numerous investigators to evaluate potential AD patients [14–19].

The BVRT is a test of cognition that requires the reproduction of geometric figures [20]. Because the BVRT requires spatial conception, immediate recall, and visuomotor reproduction, it is often used in conjunction with other neuropsychologic tests as an indicator of dementia. Zonderman and colleagues [21] demonstrated that immediate visual memory performance can have a prognostic long-term importance for as many as 22 years prior to the development of disease. In practice, screening for dementia usually begins with the patient being evaluated according to a set of cognitive test scores. Those selected as cognitively impaired are further examined by a clinician who makes the final diagnosis with
regard to dementia. The use of neuropsychologic tests thus provides a low-cost and effective first screening procedure, and when evidence of dementia is suggested from these tests, the additional use of higher cost diagnostic procedures and clinical evaluation can result in a clinical diagnosis of AD.

In a study of screening instruments for dementia using neuropsychologic tests, Jacqmin-Gadda et al. [18] concluded that the most discriminate combination of cognitive tests includes the Mini-Mental State Examination subscores of “orientation to time” and “recall three objects,” the BVRT, and the Isaacs’ Set Test of verbal fluency. The authors used the pooled results from these three tests at an initial and four subsequent follow-up examinations over an 8-year period to predict dementia using a logistic regression model.

In clinics, medical practitioners often collect cognitive test results sequentially or one examination at a time for use in the screening for dementia. Ideally, after each set of cognitive test scores, the clinician would make a decision about whether or not to do a more extensive examination of the patient. In this article, we use a sequential classification approach considering one examination at a time, rather than a prediction approach based on a single measurement and a probability of developing disease within a fixed time period. Following a method of classification based on random-effects models and posterior probabilities [22], we develop a procedure for predicting the development of preclinical AD based on longitudinal changes in BVRT in a group of individuals who showed no signs of dementia at the start of the observation period. For the sake of simplicity in demonstrating our procedure, we consider a single cognitive test, the BVRT. Using information on initial BVRT performance as well as the rate of change over time in those who did and did not develop AD, a method of discrimination into preclinical diagnostic groups is given that is based on posterior probabilities of group membership. The relative size of the estimated probabilities for each individual provides clinicians with a preliminary indication that a cognitive change is occurring in the individual that may be suggestive of the early onset of AD.

2. Materials and methods

2.1. Data

Data for this study are from participants of the Baltimore Longitudinal Study of Aging (BLSA), a longitudinal study of normal human aging that was begun in 1958, and is administered by the National Institute on Aging [23]. Participants in the study tend to be White (95%) and well-educated (over 75% have at least a bachelor’s degree). All participants who enter the BLSA are given a careful health screening to ensure that they are in excellent health with no known diseases. Participants in the study visit the Gerontology Research Center in Baltimore approximately every 2 years for 3 days of biomedical and psychologic examinations, including BVRT measurements and neurologic evaluation.

There are virtually no participants (<5) who refuse participation in the psychologic studies. All participants are monitored to obtain information regarding their health status, especially information related to neurologic and other disease events. This monitoring continues over time regardless of the collection of BVRT measurements. In the case of death, information is collected from the individual’s personal physician regarding the cause of death, and autopsy information is obtained when available. Diagnoses of AD are made without considering BVRT scores. Studies of risks for dementia began in 1986 on active BLSA participants 70 years of age and older. In 1988, the sample was expanded to include all active participants 60 and older who were evaluated for the presence of AD. A previous study revealed that incidence rates for AD in the BLSA are similar to published rates from other studies [24].

This study includes only BLSA participants who were determined to be free of AD before 1988, and who were actively enrolled in the study at that time. Since 1988, 20 female and 57 male subjects were clinically diagnosed with AD. Individuals with only one BVRT measurement or who had evidence of AD before the first BVRT measurement where excluded from the analysis because these participants did not have adequate time to develop a BVRT trend before diagnosis. The remaining AD cases were matched by age of initial examination (±3 years), number of repeated BVRT measurements (±1 measurement), and the length of follow-up (±2 years) to 32 female and 88 male controls who showed no evidence of any type of dementia throughout the study period. Within each gender, both cases and controls tended to have the same level of education, with about 70% of the female and 80% of the male participants having attained at least a bachelor’s degree. Separate disease classifications were performed for female and male participants. Table 1 gives a description of the follow-up information for the AD and the control subjects.

2.2. Neuropsychological and neurologic evaluations

The BVRT is a neuropsychologic test of memory for geometric designs requiring spatial conception, immediate recall, and visuomotor reproduction. The test consists of 10 stimulus cards where each card is displayed for 10 sec before being removed from sight. The task is to reproduce the design on each card with no time limit, and the recorded test score is the total number of errors in the reproduction of all 10 designs. On average, the test takes about 7 min to complete. The BVRT was initiated in the BLSA in 1960, and was administered every 6 years until 1991. Since 1991, the BVRT has been measured every 2 years.

In 1986, neurologic examinations were begun by trained neurologists to identify subjects who developed dementia and AD since the start of the study [24]. Subjects were screened for signs of dementia at each visit based on a battery of neuropsychologic tests, a neurologic examination, and medical imaging when necessary to rule out strokes and
other brain pathology. Additional medical records, laboratory tests, and questionnaires were obtained for all subjects identified with cognitive problems. Using all available information, a clinical diagnosis of dementia and/or AD was given using NINCDS/ADRDA [25] and DSM-III-R [26] by a multidisciplinary panel consisting of a neurologist, several neuropsychologists, nurse practitioners, neuropsychologic testers, and a neuropathologist. BVRT scores were not used in the clinical diagnosis.

2.3. Statistical analysis

The primary outcome variable was the BVRT score, which is only included in the analysis before any clinical diagnosis of AD. The last BVRT score included is from the BLSA examination before the diagnosis or first recognized date of AD for the AD cases and up to the last completed BLSA examination for the noncases or normal subjects. The task is then to classify each of the 197 subjects as either normal (free of AD) or having AD based on repeated preclinical BVRT measurements.

There are two stages in the classification procedure. The first stage consists of modeling the BVRT scores as a function of the subjects’ starting age, time in the study, and determined diagnosis (normal or AD) for all the subjects, except the one being classified, using a linear mixed-effects (LME) model [27]. The final form of the LME model used for the classification is presented in an Appendix (available on the journal’s website at www.elsevier.com/jce). This LME model is used to describe longitudinal trends in the BVRT measurements as a function of, in addition to time in the study, each individual’s starting age, and the individual’s clinical diagnosis of normal or AD. The LME model contains terms that differentiate age-specific longitudinal changes between the normal and AD groups. The second step then uses the longitudinal BVRT measurements from the individual being classified to calculate posterior probabilities sequentially, starting with the second repeated measurement, for membership into either the normal or AD group. The Appendix also shows the general form of the Bayes formula used to calculate the posterior probabilities for an individual with respect to an AD classification. The individual is classified as an AD case once the calculated posterior probability exceeds some diagnostic criteria or cutoff value whose predictive ability is evaluated using a receiver operating characteristic (ROC) curve [28]. Classification results are tested as being different for those derived by chance alone using a Mantel-Haenszel test for matched samples [29].

Female and male subjects are evaluated separately.

3. Results

3.1. Cognitive levels and statistical modeling

To illustrate the individual variability in BVRT scores, the average BVRT scores were calculated and ordered from all the repeated BVRT measurements taken before a clinical diagnosis for each subject in the Alzheimer’s and normal groups, separately. Fig. 1 shows the observed repeated BVRT error scores for five systematically selected Alzheimer’s cases (upper graph) and five systematically selected normal cases (lower graph) with the subjects chosen with the minimum, the first quartile, the median, the third quartile, and the maximum averages in the two diagnostic groups. As a reference value, the graph also shows the BVRT mean number of errors for the subjects 60 years and older, and this corresponds closely with the average value reported by Benton [30].

Using the LME model, average longitudinal patterns of change in BVRT scores were estimated for both female and male subjects in the Alzheimer’s and normal groups using the BVRT values obtained on all the subjects before any clinical diagnosis of AD. Fig. 2 gives the modeled longitudinal trends in BVRT scores for 60-year-old subjects 18 years prior to the clinical diagnosis. Note that the decline in visual memory performance represented by an increasing number of errors in the AD group (upper graph) is more pronounced in female than in male subjects. Female subjects at age 60 start with higher BVRT scores and increase more rapidly than male subjects of similar age over an 18-year follow-up period. The lower graph in the figure shows that normal female subjects have similar visual memory performance than the corresponding normal male subjects.

3.2. Disease classification

To classify a subject, the longitudinal data from the remaining subjects is used to fit the mixed-effects model.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Description of the dementia data (mean ± standard deviation) by diagnosis group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male</td>
</tr>
<tr>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>Number of participants</td>
<td>88</td>
</tr>
<tr>
<td>Age at diagnosis (years)</td>
<td>79.5 ± 6.9</td>
</tr>
<tr>
<td>Length of follow-up (years)</td>
<td>14.8 ± 8.4</td>
</tr>
<tr>
<td>Number of repeated BVRT Measurements per individual</td>
<td>4.1 ± 2.0</td>
</tr>
<tr>
<td>Interval between repeated PSA Measurements (years)</td>
<td>4.8 ± 2.4</td>
</tr>
</tbody>
</table>
This model is then used to classify the subject by calculating the posterior probability of being either normal or a potential AD case for every BVRT measurement made at and after the second measurement for that subject. In addition, classification results are obtained for all subjects using different diagnostic criteria or cutoff probability values for an AD classification. As the cutoff value needed for an AD classification decreases, the proportion of correctly classified AD cases (sensitivity) increases. Correspondingly, however, the proportion of normal subjects incorrectly classified as AD (1 − specificity) also increases. Fig. 3 gives the ROC curves used to evaluate the cutoff probability that optimizes the errors in classification between the AD and normal groups for men and women separately. The diagnostic criteria giving a point on the ROC curve located nearest to the top left corner of the plot is taken as the cutoff value that jointly optimizes the errors in classification represented by sensitivity and 1−specificity. For both female and male subjects, this occurs for a cutoff probability value of 0.42. Fig. 4 illustrates the classification approach for a 73-year-old male at the start of testing with four BVRT measurements taken before a clinical diagnosis of AD. The classification is made at age 84.4 (posterior probability of AD = 0.51), 2.4 years before the clinical diagnosis at age 86.8.

Using the diagnostic criteria for the BVRT scores that optimizes the classification error rates for the female and male subjects, Table 2 gives the classification results for all 197 subjects in the study. The table shows that overall 37 of 52 (71.2%) of the female subjects and 90 of 145 (62.1%) of the male subjects were correctly classified. For the AD cases, 13 of 20 (65.0%) of the female cases and 37 of 57 (64.9%) of the male cases were correctly classified (sensitivity), while for the normal subjects 24 of 32 (75.0%) for the female and 53 of 88 (60.2%) for the male subjects were correctly classified (specificity). By the Mantel-Haenszel test for matched samples, the classification rates using our proposed method are better than those expected by chance alone for both female \((P = .011)\) and male \((P = .007)\) subjects. Also, AD cases that were correctly classified had a mean lead time between the classification and the actual diagnosis of 2.7 years (95% confidence interval 1.25–4.15) for female cases and 6.7 years (95% confidence interval 4.33–9.07) for male cases.
4. Discussion

The use of longitudinal measurements of visual memory performance to screen for Alzheimer’s disease in a group of community-dwelling adults with no prior evidence of AD suggests that a systematic prospective screening can be useful in detecting the presence of AD. In this article, the BVRT modeling and the AD classification process were carried out separately for the female and male participants. The results show that 71.2% of the female and 62.1% of the male subjects were correctly classified into the normal and AD groups (Table 2). The results show a higher specificity for female subjects (75.0%) than for male subjects (60.2%), while the sensitivity of the method is approximately the same for male (64.9%) and female subjects (65.0%). Although the sensitivity and specificity of these results are somewhat low, the use of another cognitive screening test in addition to the BVRT would likely improve the sensitivity and specificity of this classification method. The procedure identified AD cases nearly 3 years earlier than the clinical diagnosis for female cases and nearly 7 years before the clinical diagnosis for male cases. This amount of lead time would be affected by the frequency of BVRT testing and would be even greater if the BVRT tests were administered more frequently. Thus, for this study, more examinations and a longer follow-up period for female subjects could have resulted in a greater sensitivity and a longer lead time of classifying AD cases before the actual clinical diagnosis for females subjects.

Other authors have observed sex differences in cognitive performance for healthy individuals as well [17,30,31]. Generally, females tend to excel at word-related tasks, object location memory, and perceptual speed, while males excel at mathematical problem solving and visuospatial tasks. At the beginning of our follow-up period, the BVRT scores for visual performance were similar for female and male subjects in both the Alzheimer and normal groups, respectively (Fig. 2). However, for the AD group, the female error rate increases about twice as fast as the male error rate with an average increase of about three errors compared to about 1.5 errors per 5 years of follow-up. For normal individuals, females change at nearly the same rate as males. Also, for...
both female and male subjects in the study population, the distribution of the level of education was similar between the AD and normal groups. Thus, because all analyses and classifications were done separately for each of the sexes, no sex differences should exist in the study population to affect the classification results.

Both age and education have been shown to be related to BVRT performance in both normal subjects and normal subjects with memory complaints [17]. These two variables should be considered in any modeling of cognitive scores and the subsequent classification for AD. This can be done by entering variables for both age and education in the LME model, and thus, the prediction process becomes a function of these two variables. In this article, participants tended to have similar education levels for female and male subjects

considered separately. Age at first examination was controlled for in the model. Thus, in implementing this method in standard practice in any population, researchers should develop their own data bases of AD cases and normal patients with subsequent cognitive measures and collect information on patient characteristics that relate cognitive change and AD. When relevant to the study population, the investigator could also consider variables representing genetic profiles and measures of physiologic function, such as hearing and vision. This information can then be used in developing the LME model of cognitive level and change to get a method of classification for the population of interest.

In clinical practice, it continues to be recognized that there is no single test that can be used in the diagnosing of dementia [32]. The use of more than one test in the screening for preclinical AD is likely to improve the classification results from any classification method similar to the one presented in this article. It is also desirable to use cognitive tests that require only a short time to administer. The BVRT used in this study takes, on average, 7 min to complete.

The data presented in this article were collected from participants of the BLSA, a convenience sample of mostly college-educated and predominantly Caucasian individuals. Thus, caution must be taken when trying to apply the classification results from this special population to another population or the general population. Although the trends in BVRT scores may not be completely generalized to the general population because of this homogeneity and lack of representativeness, the method of classification is applicable to any population.

With regard to the practice of screening for AD, medical practitioners can collect measures of cognitive performance sequentially or one examination at a time, and so the classification approach presented in this article would be quite natural. The model used in the classification can be extended to a multivariate classification model where more than one cognitive test or potential predictor of disease might be used in the classification process. For example, three different tests of cognition might be used such as the three tests (MMSE, BVRT and Isaac’s Set Test of verbal fluency) found by Jacqmin-Gadda [18] to be the most discriminant combination of screening tests for dementia. Finally, the results of this study suggest that there is an important window of opportunity for identifying Alzheimer’s disease early enough that clinicians could implement any potentially available interventions or secondary preventive measures.

Table 2
Classification results for Alzheimer’s Disease based on longitudinal BVRT scores

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>AD</td>
</tr>
<tr>
<td>Normal</td>
<td>24</td>
<td>7</td>
</tr>
<tr>
<td>AD</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>32</td>
<td>20</td>
</tr>
</tbody>
</table>

References


Appendix

The final LME model for the prediction of AD disease using BVRT measurements is

$$\text{BVRT}_{ij} = \beta_1 \text{Age}_i + \beta_{21} \text{Normal}_i + \beta_{22} \text{AD}_i + b_{1i} + (\beta_{31} \text{Normal}_i + \beta_{32} \text{AD}_i + b_{2i}) \times t_{ij} + (\beta_{41} \text{Normal}_i + \beta_{42} \text{AD}_i + b_{3i}) \times t_{ij}^2 + \epsilon_{ij},$$

where BVRT$_{ij}$ denotes the $j$th measurement of BVRT on the $i$th individual. The variables Normal$_i$ and AD$_i$ denote the diagnostic group to which individual $i$ belongs. Age$_i$ is the age at first measurement for individual $i$. The variable $t_{ij}$ represents the time (in years) since the first BVRT measurement. In this model the normal and AD groups have a different time$^2$ coefficient, as well as different time coefficients and different intercepts or initial values. The random effect $b_{1i}$ allows each individual to have his or her own initial BVRT value, and the random effects $b_{2i}$ and $b_{3i}$ allow each individual to have his or her own coefficient of the time and time$^2$ terms, respectively.

If $\mathbf{y}_k$ represents the vector of all BVRT measurements on the $k$th individual and $f_{c(k)}(\mathbf{y}_k)$ represents the corresponding LME marginal distribution of $\mathbf{y}_{k|\text{group } c}$ for diagnostic group $c$ (normal or AD) for individual $k$, the LME marginal distribution $f_{c(k)}(\mathbf{y}_k)$ is determined from the longitudinal data in a crossvalidation procedure. The procedure for classifying individual $k$ is to first fit the LME model in each diagnostic group $c$, omitting individual $k$ from the mixed-effects regression analysis, and to use the resulting parameter and variance-covariance estimates to compute the marginal distribution for the normal and AD diagnostic groups for this individual $k$. Next, using prior probabilities $p_c$ estimated by the observed proportions of the individuals in each diagnostic group, the Bayes formula for calculating the posterior probability that individual $k$ belongs to group $c$ ($=1$ for normal and $=2$ for AD) is given by

$$p_{kc} = p_c f_{c(k)}(\mathbf{y}_k)(p_1 f_{1(k)}(\mathbf{y}_k) + p_2 f_{2(k)}(\mathbf{y}_k)).$$

More complete details of this classification procedure applied to the study of prostate disease can be found in the recent article by Brant and colleagues [22].