Alzheimer’s disease: rapid and slow progression

Craig J. Thalhauser and Natalia L. Komarova*

Department of Mathematics, University of California Irvine, Irvine, CA 92697, USA

The variability in the progression of Alzheimer’s disease (AD) across patients has made identification of disease-delaying treatments difficult. Quantitative analysis of this variability has important implications in understanding the pathophysiology of AD and identifying disease-delaying treatments. The functional assessment staging (FAST) procedure characterizes seven stages in the course of AD from normal ageing to severe dementia. The present study applied statistical methods to analyse FAST stage durations from a dataset of 648 AD patients. These methods uncovered two distinct types of disease progression, characterized by different mean progression rates. We identified two separate distributions of FAST stage progression times differing by up to 2 years in mean duration within each stage. These results further indicate that if a patient progresses rapidly through a given FAST stage, then their further progression is also likely to be rapid. These findings support the hypothesis that progression of AD can occur via two different pathophysiological mechanisms that lead to distinct average rates of decline.

Keywords: genetic algorithm; longitudinal data analysis; rate of decline in Alzheimer’s disease

1. INTRODUCTION

It has been observed for more than 20 years that the rate of progression of Alzheimer’s disease (AD) varies to some extent from patient to patient [1], with durations in the literature reportedly ranging from a few years to two decades [2]. Many factors have been found to correlate with the rate of a patient’s cognitive deterioration, including apolipoprotein ε4 genotype [3] and other genetic factors [4, 5], brain atrophy rates [6–8], patterns of regional brain atrophy [9], ventricular enlargement [10], neuropsychological and cerebral metabolic profiles [2], vascular factors [11] and immune system factors [12].

In AD clinical trials, responder and non-responder groups are often found, and are characterized by different rates of progression of cognitive, functional, behavioural or global measures. Furthermore, rapid and slow AD progressors have been identified in many longitudinal studies. For example, in the study of Mann et al. [2], each AD patient was assigned a score calculated roughly as ‘symptom severity’ divided by ‘symptom duration’; the former quantity was assessed by means of the Disability Rating Scale score, and the latter was based on the patient’s records and information provided by family members. In the study of Bhargava et al. [13], AD patients were followed for 3 years, and those who progressed to the moderate stage were designated fast progressors, while those who remained in the early stages were designated slow progressors. In the study of Nagahama et al. [14], the rate of progression was calculated as (decline in the Mini-Mental State Examination (MMSE) score during the follow-up) × 12/(follow-up periods (months)), and then the patients with rates lower than 0.8 points per year were categorized as the slowly progressing group, while those with higher rates were categorized as the fast progression group. Another similar measure used by Doody et al. [15] was (MMSE score (expected) – MMSE score (initial))/physician’s estimate of duration (in years), with the latter estimate based on the evidence collected from the patients and their family members. Again, the patients were considered slow (0–1.9 MMSE points per year), intermediate (2–4.9 MMSE points per year) or rapid progressors (≥ 5 MMSE points per year) based on this score. Elsewhere, the rate of progression is measured by dividing the difference in the Milan Overall Dementia Assessment score by the time elapsed [16].

As with any biological variable, AD progression rates can be described by a probability distribution (rather than by one fixed number). The question arises whether one such probability distribution can explain the observed variation in progression rates. In other words, can the observed heterogeneity of deterioration patterns in AD patients be explained by sampling from different locations in a single distribution, or are there several separate distributions? If there are two (or more) distributions of the rate of decline, then it would be important to know to which distribution a given patient belongs so that the treatment effect can be accurately determined. Separate distributions of AD...
progression also imply different pathophysiological mechanisms governing them. In this article, we test the hypothesis of the existence of two separate, rapid and slow, disease courses.

2. EVIDENCE FOR DIFFERENT PATTERNS OF DECLINE

The methodology described below was developed from a mathematical analysis of a longitudinal study of AD patients seen at New York University Medical Center between 19 January 1983 and 5 May 2006. The longitudinal course of 1321 AD patients was assessed using the functional assessment staging (FAST) procedure [17,18]. Diagnosis, date of assessment, FAST stage and demographic variables were included in the assessment. In this investigation, we focus only on the integer FAST stage values (and not on the FAST substages). Out of the set of 1321 patients with an AD diagnosis, only 648 patients had at least two assessments of their FAST stage between stages 4 and 7; the histogram in figure 1a shows how many patients had two, three or and more assessments between stages 4 and 7. These 648 patients were selected for this study. The time between each patient’s first and last visit (that is, the total follow-up time) averaged 4.78 + 2.94 years.

Figure 1 presents a histogram of the time elapsed between two consecutive visits; the average of this was 3.03 + 1.59 years. The distribution of intervisit times in the longitudinal study has a bimodal shape, with a strong peak at about 2 years and another mode at about 4 years. The reason for this non-uniformity is that patients were instructed to come for an assessment every 2 years. There was no attempt to estimate the onset date of the patient’s FAST stage at the time of assessment. Thus, each patient’s FAST stage assessment can be thought of as a randomly sampled datum point during the duration of that FAST stage. For example, a patient might be diagnosed at their first visit as being in FAST stage 4, then some time later return and be found to be in FAST stage 5, but the amount of time the patient had been in FAST stages 4 and 5 prior to their first and second visits, respectively, was not determined.

The hypothesis that AD has a single mechanism governing the rate of decline is consistent with the assumption that the data come from a single distribution. If there are two mechanisms governing the rate of decline (corresponding to faster and slower speeds of progression), then there should be two different probability distributions of FAST stage durations (for each stage). Two rate distributions per FAST stage, governed by different pathophysiological mechanisms, also implies that a given AD patient will remain in the same rate distribution throughout all FAST stages.

To test the hypothesis that AD progression has multiple rate distributions, we first reconstructed the mean FAST duration per stage. This problem was addressed by methods developed for longitudinal data analysis with censoring [19–21] using a regression method similar to that previously employed [22]; see electronic supplementary material for details. In order to reproduce mean FAST stage durations similar to the original published values, we had to make an additional assumption that the patient’s first visit occurs near the beginning of a FAST stage. This is motivated by the conjecture that many patients are prompted to make their first doctor’s visit when they, or their carers, observe a change in the symptoms. To calculate the mean duration per FAST stage for a given group of patients, we split all patients into ‘transition classes’, depending on the FAST stages of their first and last visits. For a disease of four FAST stages, there are 10 possible transition classes (e.g. 4 → 4, 4 → 5, and so on). In order to ensure that the reconstruction of the mean works accurately, all transition classes of patients have to be sufficiently ‘populated’ (they must contain enough records). The number of patients in each transition class is given in table 1. Under the hypothesis of a single underlying rate distribution governing FAST stages 4–7, the reconstructed estimates of the mean durations for FAST stages 4–7 are presented in table 2. These results agree well with the originally published estimates [23].

In order to explore the data for evidence of multiple patterns of decline, we plotted the histograms of transition times for patients in different classes (figure 2).
At first glance, we can see that almost all plots are bimodal. This in itself, however, does not necessarily indicate the existence of two subgroups of patients differing in their progression patterns. To explain, we first consider $i \rightarrow i$ transitions, that is, the histograms of intervisit times where patients were diagnosed in the same FAST stage twice. We can see a pronounced peak at 2 years and another, smaller one, at 4 years. The reason for these peaks is the particular way in which these data were collected (figure 1b). For this longitudinal study, patients were instructed to come for follow-ups every 2 years, and as a result we have many patients with 2 and 4 year intervisit intervals; note that higher modes are not very pronounced. With this information, we can conclude that the intervisit times for $i \rightarrow i$ transitions do not tell us anything about the patients’ progression patterns, and only reflect the sampling patterns. The same argument applies to $i \rightarrow i + 1$ transitions. There, the frequent 2 year and 4 year intervisit times are reflected in the corresponding histogram peaks. It is the longer transitions (such as 4 $\rightarrow$ 6, 4 $\rightarrow$ 7 and 5 $\rightarrow$ 7) where we can detect another type of pattern. In the 4 $\rightarrow$ 7 and 5 $\rightarrow$ 7 transition histograms, we can see two well-proounced peaks at around 6 and 10 years. FAST stages 4 (or 5) to 7 progression usually takes longer than 2 or even 4 years (table 1), and therefore the contributions from the 2 and 4 year peaks in the sampling distribution do not affect the histogram for the 4 $\rightarrow$ 7 and 5 $\rightarrow$ 7 transitions. With 4 $\rightarrow$ 6 and 6 $\rightarrow$ 7 transitions...
transitions, we can see a blend of the two effects: while the 2 year peaks are present, there are other peaks which cannot be simply explained by the sampling timing.

We can conclude that histograms such as those presented in figure 2 are instructive, as they suggest that there may be patterns beyond the 2 year period in the sampling time distribution. The bimodality of 4 → 7 and 5 → 7 transition histograms is consistent with the existence of two subgroups of patients. However, these histograms alone are not sufficient to conclude the existence of different progression patterns. More sophisticated methods are required, which are described next.

3. STATISTICAL METHODS

To evaluate the hypothesis of two separate patterns of decline in AD, we formulated the problem as an unsupervised learning problem. We used a separation algorithm based on the genetic algorithm technique [24–27]. As an alternative, one could use a parametric technique, commonly applied for discerning the probability distribution functions underlying the presence of two or more subpopulations in a mixture model [28–30]. However, given the relative sparsity of the data (see the number of patients in each transition class; table 1), and the large number of unknown parameters that we would have to solve for (the mean and variance for each of the eight distributions, assuming two groups of patients and four FAST stages), we chose to use a non-parametric genetic algorithm methodology.

At the basis of our calculations is the notion of ‘partitions’. To create a partition, we randomly assign each patient to one of the two groups (which we call groups R and S); these groups will correspond to rapid and slow progressors after the genetic algorithm is applied. For each group, we compute the mean FAST stage durations. This process is repeated many times to create partitions with different mean FAST stage durations for groups R and S. The ‘fitness’ of each partition is defined as the smallest positive difference between mean FAST stage durations of the groups S and R, among all the FAST stages of the disease. A genetic algorithm is then applied to these partitions, which are treated as organisms, to simulate an evolutionary process. The goal was to identify the partition with the largest fitness value, which corresponds to the partition with the largest separation between mean stage durations of the R and S groups.

The algorithm, illustrated in figure 3, starts by randomly initializing a large number of R/S partitions, \(N_p\), for the patient sample and computing the fitness for each partition (in the example in figure 3, we have a sample of six patients, and create \(N_p = 4\) partitions; the fitness values, \(f\), in the figure are assigned arbitrarily for illustration purposes). After this initialization step, the partitions with the highest fitness are selected as ‘parents’ (more precisely, the probability of each partition to be selected for reproduction is proportional to its fitness). The parents then ‘mate’ to produce ‘progeny’. The new organisms are formed in the following manner. If both parents classify a particular patient the same, then that patient’s classification is unchanged in the offspring. In figure 3, such patients are circled. During mating, the remainder of each group S and R is filled at random with the patients in the corresponding groups of the parent partitions. As a result, the next generation of organisms (partitions) is created. This process of creating parents and progeny continues until all patients are consistently classified by all (or nearly all) organisms. The final set of progeny has the highest fitness, such that their FAST stage durations for groups S and R are maximally separated. Group R then comprises the list of patients identified as rapid progressors, and group S comprises the list of patients identified as slow progressors.

In order to validate this method, we created a number of artificial patient sets, consisting of 750 patients, by assuming that each patient was either a rapid or a slow progressor. Depending on whether each patient was chosen to be rapid or slow, their record was created by first randomly assigning them to a transition class, and then by drawing the durations of FAST stages 4–7 from either a ‘rapid’ or a ‘slow’ distribution. These distributions were both uniform, spanning a length of 2 years, and were shifted with respect to each other such that the average stage duration of rapid progressors was less than or equal to the average duration of slow progressors (note that other types of distributions have also been tested, with the same outcome; see electronic supplementary material).

We used six different configurations of the rapid/slow duration distributions in our simulations, which are schematically represented on top of figure 4 by two horizontal bars shifted with respect to each other. In one extreme scenario, the two distributions are identical (the leftmost panel in figure 4). This means that there is only a single distribution for each FAST stage duration. In the other extreme, the two distributions do not overlap (the rightmost panel). This means that FAST stage durations for rapid progressors are always shorter than those for slow progressors. For each configuration of distributions, we computed the probability for each patient of being a rapid progressor, \(p(\text{rapid})\). To this, we ran the separation routine 200 times, each time on the same set of patients, and recorded the number of times each particular patient was classified as rapid. The results of these simulations are presented in figure 4 in the form of histograms. Starting from the leftmost panel, we can see that if the distributions are identical (there is no difference between the rates of rapid and slow progressors), then \(p(\text{rapid})\) is 50 per cent for most patients. However, as the rapid and slow progressor distributions become more separated, an increasing fraction of patients is classified as either rapid or slow. When the two distributions are completely separated, the majority of patients are classified as either rapid or slow progressors, which corresponds to the nature of the underlying dataset. These simulations of the distributions of the durations of rapid and slow progressors show that their degree of overlap influences the shape of the classification histogram. Thus, by looking at the shape of the classification histogram, one can estimate the degree to which there are different duration distributions for rapid and slow progressors.
the simulated overlap of the probability distributions of FAST stage durations between rapid and slow progressors.

of 0.04), and the number of such cases on the
gained by the separation routine comes from transitions
whose records contain information about two consecutive
on only one stage. A similar argument holds for patients
This is because such patients’ data contain information
from two different distributions or from one distribution.

5, 6

J. R. Soc. Interface (2012)

Figure 3. Illustration using genetic algorithms to classify six patients (numbers 1–6) into rapid (R) or slow (S) progressors. The separation algorithm was applied to artificial datasets, where it was
presented as histograms that show the probability of being a rapid progressor on the

riori assumed that each patient was either a
rap ide slow progressor, and for both types the lengths of the four stages of the ‘disease’ were drawn from the corresponding
distributions. For this reason, we only show patients of
transition classes of the form i → i + 2 and higher.

4. RESULTS
We applied the separation method to the FAST staging patient dataset. The dataset is sorted 200 times and the

It is important to note that most of the information
 gained by the separation routine comes from transitions
between non-consecutive FAST stages (e.g. 4 → 6, 4 → 7,
5 → 7). Patients whose record has information only about
one FAST stage (e.g. only includes transitions 4 → 4, 5 →
5, 6 → 6 or 7 → 7) will always be classified as either rapid
or slow progressors regardless of whether the patients come
from two different distributions or from one distribution.
This is because such patients’ data contain information
on only one stage. A similar argument holds for patients
whose records contain information about two consecutive

Figure 4. Modelling rapid and slow progressors based on the simulated underlying probability distribution of FAST stage durations. The separation algorithm was applied to artificial datasets, where it was a priori assumed that each patient was either a rapid or a slow progressor, and for both types the lengths of the four stages of the ‘disease’ were drawn from the corresponding distributions. The separation algorithm was run 200 times. For each run, each patient was classified as a ‘rapid’ or a ‘slow’ progressor. From these runs, the probability of each patient being a rapid progressor, \( p(\text{rapid}) \), was computed. The results are presented as histograms that show the probability of being a rapid progressor on the \( x \)-axis (each bar has a probability range of 0.04), and the number of such cases on the \( y \)-axis. The black and grey horizontal bars at the top of each panel represent the simulated overlap of the probability distributions of FAST stage durations between rapid and slow progressors.
outcome of this procedure is shown in figure 5. The results are clear. The histogram in figure 5 most closely resembles the simulated duration data in which the rapid and slow progresses are nearly completely separated (the fifth and sixth histograms of figure 4). This means that the patients in the dataset can be classified into two groups, rapid and slow progressors, such that all FAST stages of rapid progressors are on average shorter than all FAST stages of slow progressors.

Using the aforementioned method, we computed the mean durations (years) of the rapid and slow progressor groups in each FAST stage from the patient data. These results are shown in table 3. The means of the rapid and slow progressor distributions are separated by approximately 1.5–2 years for each FAST stage. Patients who show slow progression in earlier FAST stages will also show slow progression in later FAST stages. Similarly, patients who progress rapidly through earlier FAST stages will progress rapidly through later FAST stages. This means that rapid and slow progressors come from different distributions, which implies different pathophysiological mechanisms influencing the rates of progression for these groups. We investigated the statistical significance of this result by sampling the probabilities $p_{\text{rapid}}$ for each patient 100 times and calculating the resulting mean FAST stage durations. The standard deviations yielded by this procedure are shown in the inset of figure 5; they demonstrate that the separation result is highly significant.

An important finding of the present analysis is that an AD patient’s rate of progression over the course of their disease does not change—rapid progressors remain rapid and slow progressors continue to progress slowly throughout the duration of their disease. This finding is most strongly supported by the classification results of AD patients who transitioned (jumped) across two or more FAST stages between two consecutive assessments. As mentioned above, for transitions within the same FAST stage (or between consecutive stages), the separation routine will always classify (with certainty) such patients as either rapid or slow, depending on the time between the two visits. However, for patients transitioning across two or more FAST stages (e.g. FAST stages 4–6), in the absence of two separate FAST stage duration distributions, it is possible that the progression was rapid from FAST stage 4 to 5, and that it was slow from 5 to 6, or vice versa. In other words, statistical independence of different stage durations would lead to the net averaging, and such patients would be classified sometimes as rapid, and sometimes as slow, which would lead to a peak in the middle of the histogram of figure 5. This is clearly not what we see in figure 5, where most patients can be with certainty classified as rapid or slow. This shows that FAST stage durations are not independent, and that an AD patient’s rate of progression does not change over the course of their disease.

To determine whether the rate of progression is related to demographic factors, we correlated the rate of progression group (rapid or slow) with age at baseline, sex, education and age of onset of AD (this was...
back-calculated by using the information on the estimated FAST stage durations). There were no significant correlations with these factors, which is consistent with previous reports [2,13,31–36].

5. DISCUSSION

Mathematical analysis of the patient dataset shows that there are two possible distributions of the rate of AD progression that any given patient can follow. Moreover, an AD patient’s progression type remains constant throughout the course of their dementia—slow progressors continue to progress slowly and rapid progressors continue to progress rapidly. This confirms suggestions of previous studies [16], where it is reported that the rate of progression at early stages correlates positively with the rate of progression at later stages. We did not analyse the mild cognitive impairment stages of AD (FAST stages 2 and 3), so it remains to be determined whether a fixed rate—slow or rapid—holds for these earlier stages as well.

Our analysis revealed two underlying probability distributions of stage durations for each of the FAST stages 4–7 of AD—a ‘rapid’ and a ‘slow’ one. The means of the rapid and slow duration distribution for each FAST stage are shifted with respect to each other by about 1.5–2 years (table 3).

These results suggest that there are underlying factors creating two separate distributions of rate of progression in AD. These factors are responsible for the existence of two separate, distinct patterns of decline in AD patients. While the nature of these factors is unknown, one hypothesis is that it is a genetic variation in patients, such as that associated with cerebrospinal fluid phospho-tau levels [37].

These findings are very important because the presence of two different rates of progression in AD have undoubtedly influenced the outcomes of the US Food and Drug Administration clinical trials, depending upon the proportions of rapid and slow progressors in placebo and treatment groups. There are two ways in which our findings may be relevant for future clinical trials, as follows.

In general, information about the statistics of disease progression, and, in particular, on the inhomogeneity of the patient population, is important for any drug evaluations. The knowledge that there are two subclasses of patients (rapid and slow) can help evaluate a drug’s effects. It is possible that a given treatment affects the two groups differently, and this is something that can be detected by methods similar to those implemented in this paper. In general, in order to show that a treatment delays the progression of AD, a longitudinal dataset must be collected. With this dataset, by the method described here, one can identify the two groups of progressors. Their mean stage durations must be compared with those of the untreated patients (presented in this paper). With this information, one can make better judgements about the effectiveness of disease-delaying treatments.

In addition, the existence of two patient classes can be important for individual patient evaluation. A patient can be classified as a rapid or a slow progressor in the following two ways. (i) If a patient has been evaluated at least twice and has been diagnosed in stages $i$ and $i + 1$, the patient can be added to the database, and the separation routine applied. As a result, the patient will be classified as a rapid or a slow progressor. (ii) If the length of a given FAST stage is known for a patient (e.g. from reliable informants who could recall the beginning dates of specific symptoms characterizing the onset of FAST stages $i$ and $i + 1$), then the length of FAST stage $i$ can be compared with the average stage lengths calculated here. If it is longer than the mean stage length of rapid progressors, the patient can be considered to be a rapid progressor; if it is slower than the mean stage length of the slow progressors, the patient can be regarded as slow. This information is important in establishing the disease prognosis, and can also potentially help with treatment strategies.

Obtaining a larger dataset would allow one to (i) gather better statistics and reduce the error in the calculations of the mean FAST stage durations and, more importantly, (ii) convey further research in order to identify the underlying biological causes of the rapid/slow grouping.

REFERENCES


