A New Algorithm for Predicting Time to Disease Endpoints in Alzheimer's Disease Patients

Qolamreza R. Razlighi[a], Eric Stallard[b], Jason Brandt[c,d], Deborah Blacker[e], Marilyn Albert[c],
Nikolaos Scarmeas[a], Bruce Kinosian[f], Anatoliy I. Yashin[b] and Yaakov Stern[a,*]

[a] Cognitive Neuroscience Division, Department of Neurology, Columbia University College of Physicians and
Surgeons, New York, NY, USA
[b] Social Science Research Institute and Center for Population Health and Aging, Duke University, Durham,
NC, USA
[c] Departments of Psychiatry and Behavioral Science, and Neurology, Johns Hopkins University School of Medicine,
Baltimore, MD, USA
[d] The Copper Ridge Institute, Sykesville, MD, USA
[e] Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School,
Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA
[f] Philadelphia VA Medical Center, Philadelphia, PA, USA

Abstract.
Background: The ability to predict the length of time to death and institutionalization has strong implications for Alzheimer's disease patients and caregivers, health policy, economics, and the design of intervention studies.
Objective: To develop and validate a prediction algorithm that uses data from a single visit to estimate time to important disease endpoints for individual Alzheimer's disease patients.
Method: Two separate study cohorts (Predictors 1, $N = 252$; Predictors 2, $N = 254$), all initially with mild Alzheimer's disease, were followed for 10 years at three research centers with semiannual assessments that included cognition, functional capacity, and medical, psychiatric, and neurologic information. The prediction algorithm was based on a longitudinal Grade of Membership model developed using the complete series of semiannually-collected Predictors 1 data. The algorithm was validated on the Predictors 2 data using data only from the initial assessment to predict separate survival curves for three outcomes.
Results: For each of the three outcome measures, the predicted survival curves fell well within the 95% confidence intervals of the observed survival curves. Patients were also divided into quintiles for each endpoint to assess the calibration of the algorithm for extreme patient profiles. In all cases, the actual and predicted survival curves were statistically equivalent. Predictive accuracy was maintained even when key baseline variables were excluded, demonstrating the high resilience of the algorithm to missing data.
Conclusion: The new prediction algorithm accurately predicts time to death, institutionalization, and need for full-time care in individual Alzheimer's disease patients; it can be readily adapted to predict other important disease endpoints. The algorithm will serve an unmet clinical, research, and public health need.

Keywords: Alzheimer's disease, full-time care, grade of membership model, nursing home, prediction algorithm, time to death

INTRODUCTION

The ability to predict the length of time from disease onset to need for full-time care, institutionalization, or death in individual patients with Alzheimer's disease

*Correspondence to: Yaakov Stern, 630W 168th St, P&S Box 16, New York, NY 10032, USA. Tel.: +1 212 342 1350; Fax: +1 212 342 1838; E-mail: ys11@columbia.edu.

(AD) has implications for patients and their caregivers. It also has strong implications for the design of intervention studies, health policy, and economics. A new model of the progression of AD based on data collected semiannually for up to 10 years in 252 patients from the Predictors of Severity in Alzheimer's Disease Study (Predictors 1 cohort) was presented in [1]. That model was a longitudinal Grade of Membership (L-GoM) model—a form which has many advantages over other models of the progression of AD [2–6], including the widely-used Cox-based model in [7]. Cox models rely on data from only a single visit to model disease progression as opposed to incorporating data over all visits as occurs in the L-GoM model. Moreover, Cox models have the incorrect implicit assumption that covariates used for prediction are constant over time, when in fact 5 of the 7 covariates in our original model [1] changed significantly over time. In this paper, we develop and validate in a separate (Predictors 2) cohort a prediction algorithm based on the L-GoM model that uses data from a single visit to accurately predict time to need for full-time care, institutionalization, and death in individual AD patients. Similar algorithms could be readily developed for other important disease endpoints.

L-GoM modeling addresses the shortcomings of traditional AD staging [3, 8], which typically views AD as progressing through discrete stages. For example, Green and colleagues' recent review [8] of current models concluded: "It is widely acknowledged that a single symptom, such as cognition, is not able to characterize AD progression, and that AD is heterogeneous in presentation and disease course; for example, across the main symptom domains of cognition, function, and behavior. Recent evidence suggests that future modeling initiatives should incorporate a multivariable approach, and that a latent variable analytic approach to characterizing AD progression is a promising avenue for advances in the statistical development of modeling methods."

Application of L-GoM modeling showed that the clinical status of any AD patient at any point in time in the Predictors 1 data was best characterized by a combination of four latent variables that summarized multiple aspects of disease status using three dimensions [1]. Further, the L-GoM model contains transition matrices which control how any particular clinical status progresses from visit to visit.

The original L-GoM model and the derived prediction algorithm made use of information that is obtained in a standard clinical evaluation of a patient with AD. The initial evaluation typically includes measures of mental status (such as the modified Mini-Mental Sta-

tus used here) and function (e.g., the Blessed Dementia Rating Scale activities of daily living items). Assessment of extrapyramidal signs and psychiatric and behavioral features are required to differentiate AD from Lewy body dementia. These measures, along with basic demographic information, provide the primary variables for the prediction algorithm. We also included a measure of dependence, because of its relevance to caregivers and increased use in clinical trials, although we demonstrate that it is not required for accurate prediction. Thus the variables used for prediction should not represent an undue burden to collect in a clinical setting.

In this paper, we develop a prediction algorithm based on the modeled Predictors 1 cohort data. The algorithm estimates the time to three important disease endpoints: need for full-time care, institutionalization, and death. We validate the prediction algorithm using out-of-sample data from the Predictors 2 cohort.

MATERIALS AND METHODS

Data

The Predictors study was designed to investigate the natural progression of AD in order to develop improved prediction models [11]. All subjects were diagnosed with probable AD based on the NINCDS-ADRDA criteria. This diagnosis was confirmed in up to 96% of available postmortem diagnostic evaluations [12, 13]. All participants had relatively mild dementia at the time of recruitment into the study, e.g., nearly 95% scored 30+ on our modified Mini-Mental Status (mMMS) exam (roughly equivalent to 16+ on the standard Mini-Mental Status Examination).

The study comprised two distinct cohorts, designated Predictors 1 and Predictors 2. Predictors 1 consisted of 103 men and 149 women; Predictors 2 consisted of 106 men and 148 women, all recruited in the United States at three study sites: Columbia University Medical Center, Massachusetts General Hospital, and Johns Hopkins Medical Center. In both cohorts, patients were re-assessed every 6 months using the same methodology.

The L-GoM model in [1] was developed from Predictors 1 based on data collected in the first 21 waves of follow-up, and validated out-of sample on Predictors 2 based on the first 16 waves of follow-up using similar data. In this paper, we take the Predictors 1 L-GoM model as given, extend its estimated parameters to cover two additional endpoints, also using Predictors 1, and validate the out-of sample results on Predictors 2.

The L-GoM model included 16 sets of covariates obtained from the clinical assessments of patients with AD and two additional sets obtained postmortem, as follows: 1) Cognition (mMMS; 6 items plus total score and test-completion indicator); 2) Functional capacity (Part 1 of the Blessed Dementia Rating Scale; 11 items and total score); 3) Behaviors (5 items: verbal outbursts, difficulty sleeping, amount of sleep, wandering, physical treats); 4) Psychopathological symptoms (3 items: delusions, hallucinations, illusions); 5) Motor signs (1 item); 6) Seizures (3 items); 7) Alcohol use (4 items); 8) Cardiovascular disease risk factors/signs (9 items); 9) Vision impairment indicator (1 item); 10) Dependence Scale (13 items and total score); 11) Residence status (1 item with 6 levels: home, retirement home, nursing home, hospital, rehabilitation center, and other); 12) Equivalent institutional care (1 item with 3 levels: limited home care, adult home care, health related facility; assessed in conjunction with the Dependence Scale); 13) Neurologist's estimation of AD duration at first examination (1 item); 14) Other durations (3 items: time since first seeking help, since start of seizures, and since entering a nursing home); 15) Age at first examination (1 item); 16) Socio-demographics (9 items); 17) 6-month mortality endpoint (1 item); 18) Death indicator, cause of death, and autopsy status (3 items; for death records only).

Because L-GoM was developed for categorical data analysis, all covariates were recorded as discrete outcomes [9, 14]. The 18 sets of covariates comprised 80 distinct covariate items with 248 distinct responses.

Endpoints

Date of death was determined from family report or other sources such as the National Death Index. Date of nursing home entry was determined from family report. Because individuals may require full-time care but remain in their homes, we also used the equivalent institutional care level of full-time care (health related facility) from the Dependence Scale assessment as an endpoint [15]. As with nursing home entry, we dated this as the first interval at which the subject was rated at this level on the Dependence Scale. There were cases where a subject who was not in a nursing home, or did not require full-time care, subsequently missed several follow-up visits and when he/she returned to the study was in a nursing home or full-time care status, so the exact interval of the transition could not be determined. To address this issue, we created two different endpoint covariates for nursing home and full-time care: early transition and late transition, which were the two most

extreme possibilities for the actual reality. The final reported results were averages of the separate results from the two endpoint covariates.

Prediction algorithm

Figure 1 provides a geometrical illustration of the 3-dimensional L-GoM model. At all times, each AD patient is located inside a bounded continuous 3-dimensional latent state space, shown as an inverted tetrahedron. The vertices of the tetrahedron represent the four identified latent variables or pure types (PT-1 to PT-4), which are ordered by increasing severity of symptoms. Each patient's proximity to each vertex at each visit can be scaled to the range 0–100%, with the four proximities constrained to sum to 100%. The proximities constitute the "grade of membership scores" associated with GoM models [9]. Although the proximity measures are unobserved latent variables, they can be estimated from the observed patient characteristics using L-GoM [10]. Once the location of a patient within the tetrahedron is determined, the L-GoM model also predicts the change in the location from one 6-month interval to the next. This is done using a set of transition matrices.

Because the grade of membership for (or proximity to) the most severe state (PT-4) is almost always zero or close to zero at the initial visit, most patients start at the top of the tetrahedron. Over time, the patients move to state-space locations associated with more severe impairment, and this movement generates for each patient at each subsequent visit a new location inside the tetrahedron, eventually reaching the bottom vertex (the most severe state). The prediction algorithm is based on the fact that once the initial position of each patient in the tetrahedron is known, the transition matrices can be used to estimate the changes in grades of membership from one visit to the next. This makes it possible to use the model to predict all later outcomes using data from the initial visit.

To use L-GoM for predicting the specific endpoints in the current paper, initial GoM scores, transition matrices, and probability loading vectors are required. The transition matrices and probability loading vectors are the same for every subject and were obtained from fitting the L-GoM model to the Predictors 1 study cohort (see [1]). Therefore, for a new subject, only the initial GoM scores need to be computed. Maximum likelihood estimation was used to determine the initial GoM scores using the probability loading vectors for the initial examination (see Supplementary Material for detailed description). The initial GoM scores, in
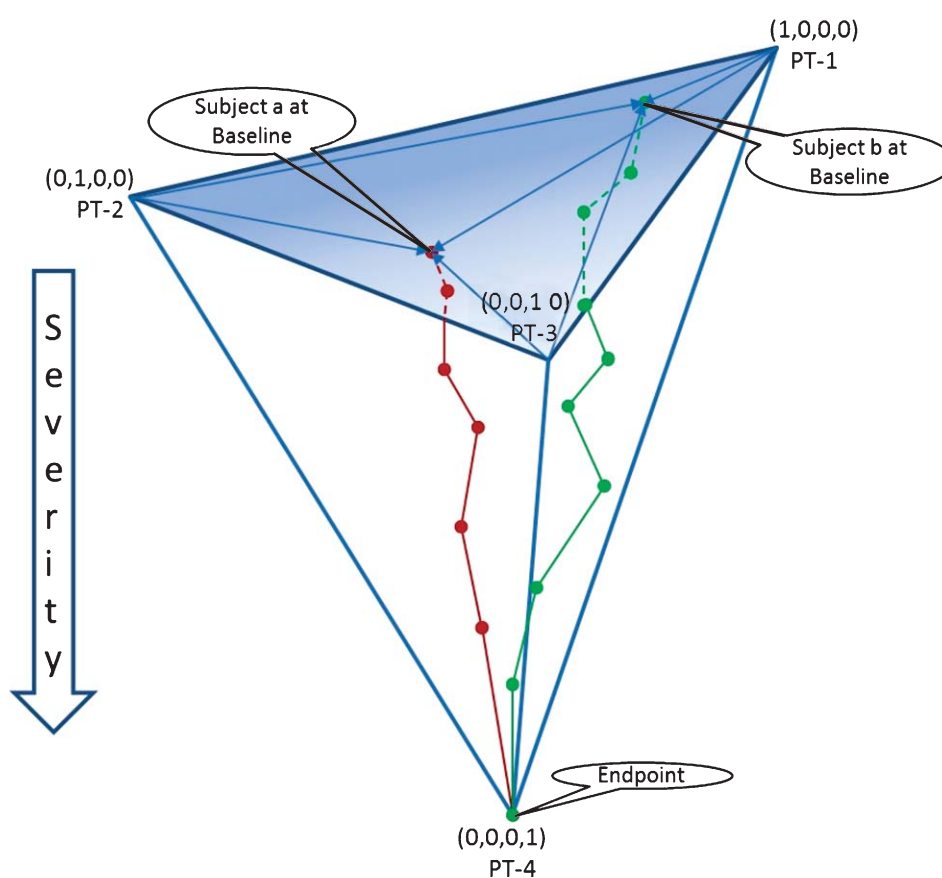
Fig. 1. A representation of the longitudinal Grade of Membership model. See text for explanation. PT, pure type.

combination with the transition matrices and the probability loading vectors for each endpoint, are sufficient to compute the probability of reaching each endpoint at each subsequent interval. The complements of these probabilities are the survival probabilities for not reaching the endpoint, and their product yields the cumulative survival probability for each endpoint (i.e., survival curve) in each patient. The areas under each survival curve generate the expected patient-specific times to reach the corresponding endpoints. Thus, the new prediction algorithm generates a subject-specific estimate of the time to reach each endpoint.

Validation

For each endpoint, the average predicted survival curve was compared to the observed Kaplan-Meier (KM) survival curve along with its 95% confidence interval. If the predicted survival curve was within the corresponding confidence interval, then we deemed the prediction satisfactory.

Because the L-GoM model in [1] was fitted separately for men and women, we evaluated the prediction accuracy for each gender separately, as well as jointly. Furthermore, to assess the calibration of the prediction algorithm for extreme patient profiles, we divided the subjects into quintiles based on the predicted times to each endpoint. For mortality, the first quintile contained the patients estimated to have the shortest survival, and the fifth quintile, the longest; we used similar definitions for the other two endpoints. We compared the predicted survival curves to the observed KM survival curves within each quintile and for each endpoint.

RESULTS

Demographics

The average age (standard deviation) at intake examination was 71.4 (9.4) years for men and 74.5 (9.0) years for women in Predictors 1; and 75.4 (7.5) years
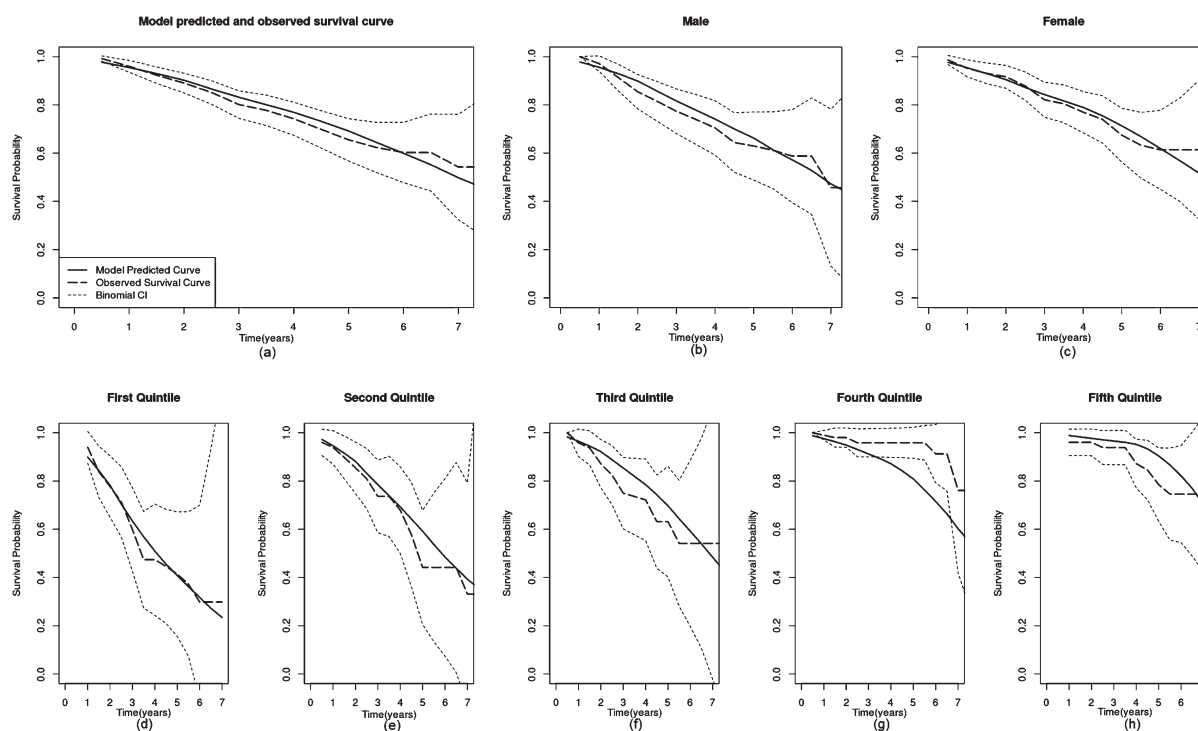
Fig. 2. Comparison of the predicted and observed survival curve a) for all subject, b) for male subjects, c) for female subjects, and d–h) for five quintiles ordered by predicted life expectancy.

and 77.3 (8.2) years, respectively, in Predictors 2. The estimated average duration of AD at intake was 4.8 (2.7) years for men and 4.3 (2.4) years for women in Predictors 1; and 4.6 (2.3) years and 4.3 (2.3) years, respectively, in Predictors 2. On average, the Predictors 2 cohort was 3–4 years older at intake. The average estimated duration of AD in the two cohorts was 4.3–4.8 years at intake.

Validation

Figure 2a shows the predicted and observed survival curves for mortality for males and females jointly; Figs. 2b and 2c show the same results separately. The model-predicted curves were closely fitted to the observed data and the deviations were all within the confidence intervals. Figures 2d to 2h show the predicted and actual survival curves for mortality quintiles, from shortest to longest predicted survival. Again, the model-predicted curves were closely fitted to the observed data, except the fourth quintile.

Figure 3 shows similar predicted and observed survival curves for nursing home admissions, and Fig. 4 shows similar survival curves for the onset of need for full-time care. There were very limited occasions where the predicted survival probabilities were not within the confidence intervals of the observed probabilities.

To test the robustness of the algorithm to missing data, we removed the Dependence Scale (13 items and total score), and re-evaluated the accuracy of the predictions in their absence. The results showed negligible change in the accuracy and calibration of the predictions.

DISCUSSION

We developed an algorithm to predict the time to need for full-time care, institutionalization, and death in AD patients. This algorithm was based on an L-GoM model of AD progression using data from a cohort of 252 patients followed semiannually for 10 years (Predictors 1). We tested the algorithm on a separate, similar validation population of 254 patients followed for similar amounts of time (Predictors 2). The results demonstrated that the algorithm-predicted times-to-outcome closely matched the actual observed times, and hence, that the algorithm produced highly accurate out-of-sample predictions using data from only the initial visit—a testament to the power of the under-
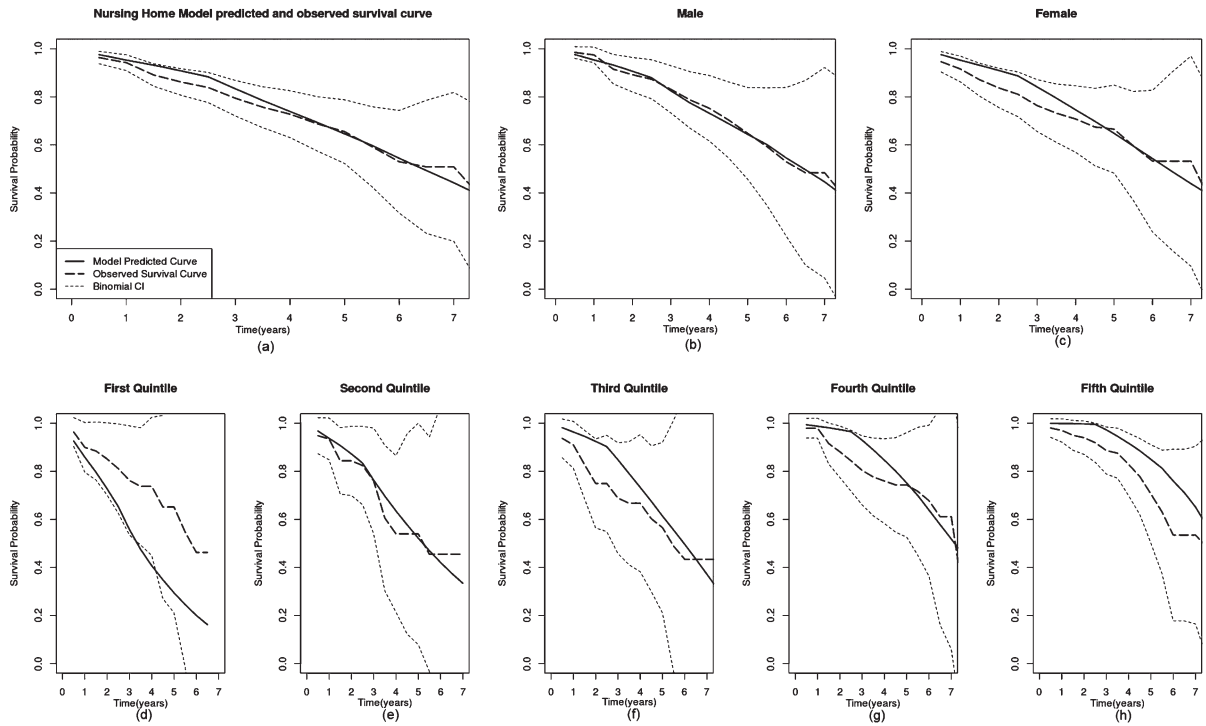
Fig. 3. Comparison of the predicted and observed survival curve for entering nursing home a) for all subjects, b) for male subjects, c) for female subjects, and d–h) for five quintiles ordered by predicted time to nursing home admission.
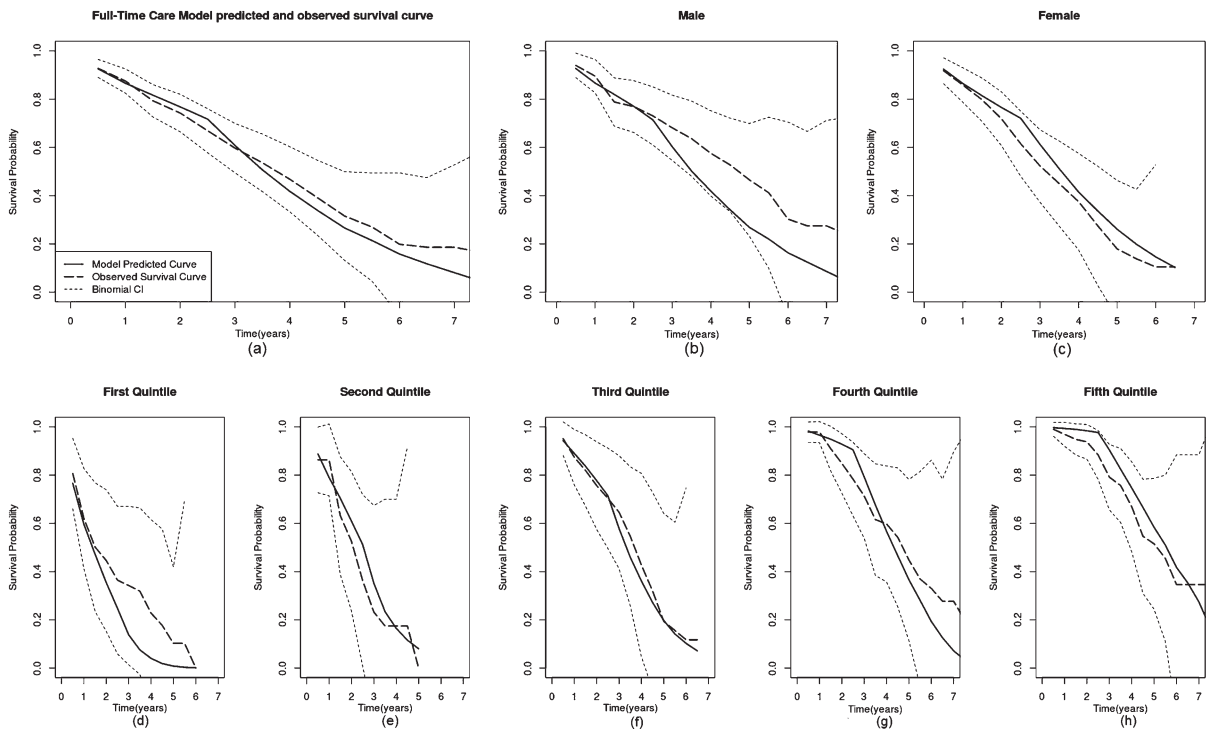

Fig. 4. Comparison of the predicted and observed survival curve for need for full-time care a) for all subjects, b) for male subjects, c) for female subjects, and d–h) for five quintiles ordered by predicted time to need for full time care.

lying L-GoM model. We know of no other prediction algorithm that approaches this level of performance.

The underlying L-GoM model for prediction uses information such as mental status, some measures of cognition and function, presence of extrapyramidal and motor signs, psychiatric and behavioral features, and basic demographic information. This information is typically collected in an initial visit when evaluating patients with AD in a standard clinical setting. Thus this implementation of the algorithm is feasible despite the number of data points that it uses.

The algorithm was designed to use data acquired at a patient's first visit. This is the most common situation for treating physicians: follow-up data do not yet exist but the physician needs some basis for prognosis. Similarly, when selecting patients for a clinical trial or simply modeling a trial, one would typically work with data acquired on a single visit. We found that including 1–2 additional visits did not appreciably increase prediction accuracy.

The algorithm employs a rigorous and parsimonious L-GoM model of AD progression that incorporated all of the data collected at all patient visits in Predictors 1 and used those data to generate a robust model of disease progression. Although the number of distinct patients was modest, the combination of 80 covariates and 21 visits yielded a large comprehensive data set comprising 166,223 distinct data points. The "effective sample sizes" for this design ranged from 1,623–1,879 depending on the dimension of the model [1].

Because the underlying model was fitted to the complete series of longitudinal data in Predictors 1, it provided transition matrices that allow one to predict progression in new data using covariate information from only the initial visit. We empirically demonstrated here that the initial GoM scores, in combination with the transition matrices and the probability loading vectors for each endpoint, were sufficient to accurately represent the heterogeneous nature of the disease progression; thus, even when we estimated the initial GoM scores from only the initial visit data the results were accurate. Incorporating the data from the first 2 or 3 consecutive intervals did not significantly improve predictions. From a practical standpoint, requiring data from multiple visits would diminish the utility of a prediction algorithm.

Previous approaches have typically relied on much more limited data for calibration and prediction. For example, a widely-used prediction algorithm [7] based on the Cox model relied on data from only a single visit in combination with subsequent observation of times to full-time care and/or death to model disease progression and conduct subsequent predictions for new patients.

A related limitation of the Cox model in [7] is the incorrect implicit assumption that covariates used for prediction were constant over time, when in fact 5 of the 7 covariates changed significantly. An alternative form of the Cox model allows for time-dependent covariates, but this form has limited use in prediction because the future values of the covariates are unknown and are not modeled [16]. This differs from L-GoM where the future values of the covariates are explicitly modeled.

Other progression models based on Markov processes are not adequate for AD progression because they incorrectly assume that each patient is always in one of a small number of discrete health states. In fact, health states in AD are continuous and multidimensional [8]. L-GoM captures this continuous multidimensional aspect of patient status.

The new algorithm generates a specific time to endpoint for each subject, easily calculated as the area under the survival curve. It was not feasible to compare actual to predicted survival times because not all patients reached the study endpoints during the Predictors 2 follow-up period. Restricting such comparisons to only the available endpoint times would introduce a censoring bias because patients with better survival would be systematically underrepresented. Therefore, we presented comparisons only of observed verus predicted survival curves. By subdividing the patients into quintiles, we were able to demonstrate that the predictions remain calibrated and accurate for patients across a broad range of conditions.

Further refinement of the algorithm will be useful. Notably, ApoE genotype is associated with rate of AD progression [17] but was not included in the L-GoM model/algorithm because this information was not universally available in Predictors 1.

Although the 16 sets of covariates in the algorithm may initially seem to be incompatible with standard clinical practice, most variables are derived from instruments that are typically collected at a patient's initial visit, including assessments of cognition, function, behavior, and motor signs. The algorithm is robust to missing data and thus can be readily applied to any data set that has a reasonably broad subset of the covariates. For example, here we demonstrated that predictions showed negligible change in accuracy and calibration when the entire Dependence Scale was treated as missing data.

We focused here on need for full-time care, institutionalization, and mortality for initial validation,

because they are most often used as endpoints in prediction models. However, L-GoM describes the progression of all covariates over time, so that the algorithm could also generate predictions about reaching specific levels of impairment in activities of daily living, specific dependence scores, etc.

The complexity of the model and the optimization process for computing initial GoM scores for any new subject preclude utilizing a simple equation for prediction; however, a computerized calculator that provides estimated times to any measured endpoint based on data collected in a single visit can easily be implemented.

Criteria for institutionalization might vary in different countries or across different cultures. All of the participants in the Predictors 1 and 2 study cohort were recruited in the United States, so forward application of the algorithm to data collected in other countries might need to be adjusted. However, need for full-time care should be equivalent across different cultures and countries.

In summary, we developed a new prediction algorithm based on the L-GoM model of AD progression [1]. This algorithm was validated in a separate data set and shown to make accurate predictions of mortality and need for full-time or nursing home care using data from only a single patient visit. The algorithm shows great promise for use in clinical practice, research, and pharmaceutical trials.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

Supplementary figures and tables are available in the electronic version of this article: http://dx.doi.org/10.3233/JAD-131142.

REFERENCES

[1] Stallard E, Kinosian B, Zbrozek AS, Yashin AI, Glick HA, Stern Y (2010) Estimation and validation of a multiattribute model of Alzheimer disease progression. *Med Decis Making* **30**, 625-638.

[2] Neumann PJ, Araki SS, Arcelus A, Longo A, Papadopoulos G, Kosik KS, Kuntz KM, Bhattacharjya A (2001) Measuring Alzheimer's disease progression with transition probabilities: Estimates from CERAD. *Neurology* **57**, 957-964.

[3] Green C (2007) Modelling disease progression in Alzheimer's disease: A review of modelling methods used for cost-effectiveness analysis. *PharmacoEconomics* **25**, 735-750.

[4] Stern Y, Liu X, Albert M, Brandt J, Jacobs DM, Del Castillo-Castaneda C, Marder K, Bell K, Sano M, Bylsma F, Lafleche G, Tsai WY (1996) Application of a growth curve approach to modeling the progression of Alzheimer's disease. *J Gerontol A Biol Sci Med Sci* **51**, M179-M184.

[5] Eisdorfer C, Cohen D, Paveza GJ, Ashford JW, Luchins DJ, Gorelick PB, Hirschman RS, Freels SA, Levy PS, Semla TP (1993) Empirical evaluation of the global deterioration scale for staging Alzheimer's disease. *Am J Psychiatry* **150**, 190-194.

[6] Caro JJ, Getsios D, Migliaccio-Walle K, Raggio G, Ward A (2001) Assessment of health economics in Alzheimer's disease (AHEAD) based on need for full-time care. *Neurology* **57**, 964-971.

[7] Stern Y, Tang MX, Albert MS, Brandt J, Jacobs DM, Bell K, Marder K, Sano M, Devanand D, Albert SM, Bylsma F, Tsai WY (1997) Predicting time to nursing home care and death in individuals with Alzheimer disease. *JAMA* **277**, 806-812.

[8] Green C, Shearer J, Ritchie CW, Zajicek JP (2011) Model-based economic evaluation in Alzheimer's disease: A review of the methods available to model Alzheimer's disease progression. *Value Health* **14**, 621-630.

[9] Woodbury MA, Clive J (1974) Clinical Pure types as a fuzzy partition clinical pure types as a fuzzy partition. *J Cybernetics* **4**, 111-121.

[10] Stallard E (2007) Trajectories of morbidity, disability, and mortality among the US elderly population: Evidence from the 1984-1999 NLTCS. *N Am Actuar J* **11**, 16-53.

[11] Stern Y, Folstein M, Albert M, Richards M, Miller L, Bylsma F, Lafleche G, Marder K, Bell K, Sano M (1993) Multicenter study of predictors of disease course in Alzheimer disease (the "predictors study"). I. Study design, cohort description, and intersite comparisons. *Alzheimer Dis Assoc Disord* **7**, 3-21.

[12] Zhu CW, Scarmeas N, Torgan R, Albert M, Brandt J, Blacker D, Sano M, Stern Y (2006) Clinical features associated with costs in early AD: Baseline data from the Predictors Study. *Neurology* **66**, 1021-1028.

[13] Steinerman JR, Irizarry M, Scarmeas N, Raju S, Brandt J, Albert M, Blacker D, Hyman B, Stern Y (2008) Distinct pools of beta-amyloid in Alzheimer disease-affected brain: A clinicopathologic study. *Arch Neurol* **65**, 906-912.

[14] Woodbury MA (1994) A general model for statistical analysis using fuzzy sets: Sufficient conditions for identifiability and statistical properties. *Inf Sci Appl* **1**, 149-180.

[15] Stern Y, Albert SM, Sano M, Richards M, Miller L, Folstein M, Albert M, Bylsma FW, Lafleche G (1994) Assessing patient dependence in Alzheimer's disease. *J Gerontol* **49**, M216-M222.

[16] Fisher LD, Lin DY (1999) Time-dependent covariates in the Cox proportional-hazards regression model. *Annu Rev Public Health* **20**, 145-157.

[17] Cosentino S, Scarmeas N, Helzner E, Glymour MM, Brandt J, Albert M, Blacker D, Stern Y (2008) APOE epsilon 4 allele predicts faster cognitive decline in mild Alzheimer disease. *Neurology* **70**, 1842-1849.