



Practice of Epidemiology

Empirically Derived Trajectories to Dementia Over 15 Years of Follow-up Identified by Using Mixed Membership Models

Fabrizio Lecci, Brian Junker, Lewis H. Kuller, Oscar L. Lopez, and James T. Becker*

* Correspondence to Dr. James T. Becker, Neuropsychology Research Program, Suite 830, 3501 Forbes Avenue, Pittsburgh, PA 15213 (e-mail: beckerjt@upmc.edu).

Initially submitted November 23, 2014; accepted for publication February 18, 2015.

Alzheimer disease is the most common form of dementia in the elderly, and the complex relationships among risk factors produce highly variable natural histories from normal cognition through the prodromal stage of mild cognitive impairment (MCI) to clinical dementia. We used a novel statistical approach, mixed membership trajectory models, to capture the variety of such pathways in 652 participants in the Cardiovascular Health Study Cognition Study over 22 years of follow-up (1992–2014). We identified 3 trajectories: a “healthy” profile with a peak probability of MCI between 95 and 100 years of age and only a 50% probability of dementia by age 100; an “intermediate” profile with a peak probability of MCI between 85 and 90 years of age and progression to dementia between 90 and 95 years; and an “unhealthy” profile with a peak probability of progressing to MCI between ages 75 and 80 years and to dementia between the ages of 80 and 85 years. Hypertension, education, race, and the $\epsilon 4$ allele of the apolipoprotein E gene all affected the closeness of an individual to 1 or more of the canonical trajectories. These results provide new insights into the natural history of Alzheimer disease and evidence for a potential difference in the pathophysiology of the development of dementia.

dementia; mild cognitive impairment; mixed membership; trajectory models

Abbreviations: *Apo $\epsilon 4$* , $\epsilon 4$ allele of the apolipoprotein E gene; CHS, Cardiovascular Health Study; MCI, mild cognitive impairment; MMTM, mixed membership trajectory model.

Alzheimer disease is the most common cause of dementia in the elderly; the prevalence increases exponentially between the ages of 65 and 85 years, approaching 50% in the oldest old (1, 2). After 90 years of age, the incidence of Alzheimer disease increases dramatically, from 12.7%/year in the 90–94 age group, to 21.2%/year in the 95–99 age group, and to 40.7%/year in those ≥ 100 years of age (3).

The risk of Alzheimer disease is further affected by the presence of the $\epsilon 4$ allele of the apolipoprotein E gene (*Apo $\epsilon 4$*), male sex, lower education, and having a family history of dementia (1, 4, 5). Medical risks include systemic hypertension, diabetes mellitus, and cardiovascular or cerebrovascular disease (6–11). Lifestyle factors affecting risk include physical and cognitive activity and diet (12–14). It is the interactions among these risk factors and the pathobiological cascade of Alzheimer disease that determine the likelihood of a clinical expression of Alzheimer disease as either dementia or its prodromal syndrome, mild cognitive impairment (MCI) (15).

These interactions vary as a function of an individual’s age, in part because both the risk factors and dementia itself increase the risk of death. We chose to examine a subset of these variables and how they affect the risk of developing dementia.

Our analysis is based on the data of the Cardiovascular Health Study (CHS) Cognition Study, a rich database of information obtained over more than 20 years, including detailed cognitive assessments in 1990–1991 (16), 1998–1999 (17), 2002–2003 (18), and annually thereafter. These data have been used to describe the incidence and prevalence of dementia and MCI (1, 6, 18, 19) and, more recently, to examine the patterns of progression from normal cognition to MCI (15).

Familiar approaches to assessing the importance of various risk factors for MCI or dementia involve (multinomial) logistic regression analysis, survival analysis, and the like. Such analyses (1, 6, 18, 19) have shown the importance of a range of risk factors and risk modifiers in predicting the time to

develop clinical dementia. However, these approaches do not address the widely held intuition that there are a variety of pathways or trajectories that individuals can take as part of the natural history of Alzheimer disease (20, 21).

In order to evaluate the evidence for multiple trajectories and to capture these different pathways if they are found, we have adapted the innovative, data-driven, mixed membership trajectory model (MMTM) approach of Manrique-Vallier as described by Erosheva et al. (22) and Connor (23). This technique combines features of cross-sectional grade of membership models (24) with those of longitudinal multivariate latent trajectory models (25), and it is becoming well established in the statistics and machine learning literature (26).

The present paper represents one of the first applications of MMTM methodology to neuroepidemiologic data, and specifically to the study of the development of cognitive disorders in normal aging. The technique has allowed us to identify, from the data, a small number of theoretically appealing canonical trajectories to dementia and MCI (analogous to latent class or mixture modeling). In addition, we are able to determine the extent to which classic risk factors for Alzheimer disease and MCI are associated with the closeness of an individual to each of the canonical trajectories.

METHODS

Data and subjects

The CHS began in 1989–1990 and obtained extensive clinical, radiological, and laboratory data from participants recruited from the Medicare eligibility lists in 4 US communities; details were described previously (27). In 1992–1994, 924 of the CHS participants in Pittsburgh, Pennsylvania, underwent a structural magnetic resonance imaging scan of the brain, and these individuals constitute the initial cohort of the CHS Cognition Study (6). We used data from the 652 individuals in the CHS Cognition Study cohort who were alive in 1998 and who agreed to genetic testing for *Apoε4*. The 168 subjects who had died prior to the 1998–1999 visit were older and more likely to be men and to have hypertension, diabetes, and dementia than those who were alive (Web Table 1 available at <http://aje.oxfordjournals.org/>). The 103 individuals who declined genetic testing were more likely to have hypertension than those who consented to testing but were otherwise similar (Web Table 2).

Cognitive classification

In 1998–1999, the CHS attempted to identify all participants who were demented at the time of the magnetic resonance imaging examination in 1991–1994 or who developed dementia by 1998–1999. The cohort included all participants who had magnetic resonance imaging in 1991–1994, a Modified Mini-Mental State Examination, and *Apoε4* genotyping (28). From 2002 to 2014, the surviving, nondemented participants from Pittsburgh were evaluated on an annual basis by using the standard study procedures. For those individuals who could not return to the clinic, home visits were completed. For those individuals who refused home visits, telephone interviews for cognitive status were conducted as described by Brandt et al. (29), and the informants were interviewed by

standard procedures, including the Dementia Questionnaire (30) and the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) (31).

The diagnoses of dementia and MCI were made by an adjudication committee that used all available cognitive and laboratory data from each participant (17, 28). The first step in the diagnostic process was to determine the presence of dementia using 3 sets of criteria: the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer’s Disease and Related Disorders Association (NINCDS-ADRDA) (32, 33), the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition) (DSM-IV) (34), and the *International Classification of Diseases, Tenth Revision* (ICD-10) (35). If any of those criteria were met, the specific type of dementia was identified by using multiple diagnostic criteria for Alzheimer disease (NINCDS-ADRDA (32), DSM-IV (34)), vascular dementia (State of California Alzheimer’s Disease Diagnostic and Treatment Centers (ADDTC) (36)), National Institute of Neurological Disorders and Stroke–Association Internationale pour la Recherche et l’Enseignement en Neurosciences (NINDS-AIREN) (37)), dementia with Lewy Bodies (38), and frontotemporal dementia (39).

MCI was classified following the CHS Cognition Study diagnostic criteria (17). Any cognitive deficits represented a decline from the previous level of functioning, but overall they fell within normal limits. Individuals with mild alterations on instrumental activities of daily living could be classified with MCI, and all had impairments (defined as performance >1.5 standard deviation below age/education appropriate means) in 1 or more cognitive domains (i.e., 2 or more tests abnormal) or 1 abnormal test (which could be a memory test) in at least 2 separate domains, without sufficient severity or loss of instrumental activities of daily living to constitute dementia.

The year of onset of dementia was set after review of all prior records, including the reports of informants (e.g., Informant Questionnaire on Cognitive Decline in the Elderly and Dementia Questionnaire). Once a patient was diagnosed with dementia, follow-up was limited to telephone contact and medical record review. A classification of MCI had no effect on the number or intensity of follow-up visits.

Statistical model

MMTMs are related to the category of latent class models (25, 40) that uses the data to find a small number of latent classes and then estimates the probability of each individual’s falling into each one of them. Unlike other similar methods, MMTMs allow each individual to have a weighted membership in each trajectory, and the individual is not forced into one or another of the latent classes. The latent classes are equivalent to the canonical trajectories toward MCI and dementia, and we interpret the probabilities as weights of closeness to the canonical profiles.

The outcome variable, Y , codes the diagnosis of each individual at each age:

$$Y = \begin{cases} 1 & \text{if dementia} \\ 2 & \text{if MCI} \\ 3 & \text{if normal} \end{cases} .$$

Diagnosis Y is observed at multiple ages for each individual, from 1992 to dementia, death, or the last study visit (i.e., 2012). There were a total of 3,569 observations from the 652 participants, an average of 5.46 per individual.

Mixed membership trajectory model. MMTMs assume the existence of a small number of canonical profiles describing individuals' trajectories toward specific health outcomes over time; we describe here an MMTM with $K = 3$ canonical profiles. Within the k th canonical profile ($k = 1; 2; 3$), the probability of each diagnosis y , as a function of age, $P_k(Y = y | \text{age})$, is modeled as a multinomial logistic regression on each individual study participant's age, so that $P_k(Y = y | \text{age})$ is characterized by the equations:

$$\begin{aligned} \text{logit } P_k(Y > 1 | \text{age}) &= \beta_{0k} + \beta_{1k} \text{ age} \\ \text{logit } P_k(Y > 2 | \text{age}) &= \beta_{0k} + \beta_{1k} \text{ age} - c_k. \end{aligned} \tag{1}$$

The trajectory of any specific individual is modeled as a mixture of these canonical trajectories; thus, trajectories of the individual study participants share characteristics of all of the canonical profiles to varying degrees. This allows us to 1) describe distinct general tendencies of the trajectories while 2) accounting for individual variability at the same time. For individual i ,

$$P(Y_i = y | \text{age}) = \sum_{k=1}^3 g_{ik} P_k(Y_i = y | \text{age}). \tag{2}$$

In this equation, the membership weights $g_i = (g_{i1}; g_{i2}; g_{i3})$ are nonnegative and sum to 1; an individual with weights $g_i = (0.5, 0.1, 0.4)$ would have a mixture of the trajectories, while an individual with weights $g_i = (0.0, 1.0, 0.0)$ would be a prototype for the second trajectory. Note that equation 2 implicitly contains 9 parameters: $\beta_{0k}; \beta_{1k}; c_k$, for $k = 1; 2; 3$.

Survivorship analysis. The presence of dementia is correlated with mortality; demented individuals are more likely to die than individuals of the same age without dementia (35–37). Because all subjects in our analysis are older than 70 years, any reference to the distribution of survival time refers to the conditional version, given that the subjects have already lived more than 70 years. Within each canonical profile, we modeled the random survival time exceeding the age of 70 (variable S) using the Weibull distribution: $w(s; \theta_k; \delta_k)$, with scale parameter θ_k and shape parameter δ_k , $k = 1; 2; 3$. Our goal was to understand the survival patterns and their impact on the trajectories to dementia using the strategy described by Erosheva et al. (22). We made 2 critical assumptions: 1) The canonical profiles specify both the trajectories to dementia and the mortality distributions, and 2) given the membership vectors g_i , the survival time s and the diagnosis Y are independent. Therefore, the joint model for dementia and mortality can be written as follows:

$$P(y_i, s_i | \text{age}) = \left[\sum_{k=1}^3 g_{ik} P_k(y_i | \text{age}) \right] \left[\sum_{k=1}^3 g_{ik} w(s_i; \theta_k, \delta_k) \right]. \tag{3}$$

Priors and additional predictors. We formulated the models within a Bayesian framework in which each parameter

is considered a random quantity. We specified prior distributions for the parameters in the analysis and estimated their posterior distributions based on Markov chain Monte Carlo sampling (41). We completed the Bayesian specification by choosing prior distributions:

$$\begin{aligned} \beta_{0k} &\sim N(0, 100) \quad \text{for } k = 1, 2, 3, \\ \beta_{1k} &\sim N(0, 100) \quad \text{for } k = 1, 2, 3, \\ c_k &\sim N(0, 100) \quad \text{for } k = 1, 2, 3, \\ \theta_k &\sim \text{Gamma}(1, 1) \quad \text{for } k = 1, 2, 3, \\ \delta_k &\sim \text{Gamma}(1, 0.1) \quad \text{for } k = 1, 2, 3. \end{aligned} \tag{4}$$

The prior distributions of β_{0k} , β_{1k} , and c_k are very flat (large variances) and thus are uninformative about the possible values of these regression parameters, allowing the data to determine their own estimates. Given that both θ_k and δ_k must be nonnegative, their priors can be considered diffuse but with a realistic shape to model human survival times in excess of 70 years.

We then examined the effects of a set of time-invariant covariates on the closeness of individuals to each trajectory. These 7 binary predictors were race (white), education (beyond high school), hypertension (present), *ApoE4* (present), sex (female), diabetes (present), and heart disease (present). Hypertension was defined as a systolic blood pressure greater than 160 mm, a diastolic pressure greater than 94 mm, or current use of antihypertensive medication. Diabetes mellitus was defined as a fasting glucose measurement of 140 mg/dL or greater or 200 mg/dL or greater 2 hours after a glucose load (75 g), a history of diabetes from the medical history questionnaire, or the current use of insulin or oral hypoglycemic medications. Heart disease was considered present if the participant had a history of angina, myocardial infarction, stent placement, and/or coronary artery bypass surgery. The vascular risk variables (i.e., hypertension, diabetes, heart disease) were considered present if they were identified at any time between 1998–1999 and 2002–2004 (inclusive).

We evaluated the impact of these predictors on the proximity of individuals to the 3 trajectories, by allowing the prior distribution of the membership vectors $g_i = (g_{i1}, g_{i2}, g_{i3})$ to depend on the predictors $X = (X_1, X_2, \dots, X_7)$:

$$g_i = (g_{i1}, g_{i2}, g_{i3}) \sim \text{Dirichlet}(\alpha) \quad \text{for all } i, \tag{5}$$

where

$$\begin{aligned} \alpha(X) = & (\exp(a_{01} + a_{11}X_1 + \dots + a_{71}X_7), \\ & \exp(a_{02} + a_{12}X_1 + \dots + a_{72}X_7), \\ & \exp(a_{03} + a_{13}X_1 + \dots + a_{73}X_7)). \end{aligned} \tag{6}$$

We also assume uninformative prior distributions for the coefficients in equation 6,

$$a_{jk} \sim N(0, 100) \quad \text{for } j = 0, 1, \dots, 7 \text{ and } k = 1, 2, 3. \tag{7}$$

The full Bayesian model is therefore composed of equation 3 and the priors specified in equations 4–7. From equation 6

Table 1. Mixed Membership Trajectory Model With 4 Time-Invariant Predictors, Showing Posterior Means and Standard Deviations for the Parameters of the 3 Canonical Trajectories, Cardiovascular Health Study Cognition Study, 1992–2014

Meaning	Parameter	k = 1, ^a mean (SD)	k = 2, ^b mean (SD)	k = 3, ^c mean (SD)
Intercept	β_{0k}	38.011 (0.521)	47.913 (0.161)	38.904 (0.172)
Effect of age	β_{1k}	-0.388 (0.054)	-0.531 (0.031)	-0.483 (0.027)
MCI/dementia threshold	c_k	1.799 (0.334)	4.197 (0.307)	1.647 (0.135)

Abbreviations: MCI, mild cognitive impairment; SD, standard deviation.

^a Healthy profile.

^b Intermediate profile.

^c Unhealthy profile.

and the properties of the Dirichlet distribution, we can see that

$$\log \frac{\mathbb{E}(g_{i1})}{\mathbb{E}(g_{i2})} = (a_{01} - a_{02}) + (a_{11} - a_{12})X_1 + \dots + (a_{71} - a_{72})X_7$$

$$\log \frac{\mathbb{E}(g_{i1})}{\mathbb{E}(g_{i3})} = (a_{01} - a_{03}) + (a_{11} - a_{13})X_1 + \dots + (a_{71} - a_{73})X_7$$

$$\log \frac{\mathbb{E}(g_{i2})}{\mathbb{E}(g_{i3})} = (a_{02} - a_{03}) + (a_{12} - a_{13})X_1 + \dots + (a_{72} - a_{73})X_7,$$

so that we can interpret the difference ($a_{jk} - a_{jh}$) as the effect of variable X_j on the population log odds of the event “individual i has a trajectory near profile k ” versus the event “individual i has a trajectory near profile h .”

RESULTS

Tables 1 and 2 show the estimated parameters determining the 3 canonical trajectories to MCI/dementia and the 3 canonical survival trajectories, respectively.

The canonical trajectories to MCI and dementia are shown in Figure 1, which presents the probability of being cognitively normal, having MCI, or having dementia for each of the 3 profiles: 1) healthy, 2) intermediate, and 3) unhealthy. These curves should not be viewed as if they were survival functions. Each line represents the cross-sectional probability of having a particular diagnosis at a given age. Within any given profile, the sum of the probabilities for each diagnosis for a given age equals 1.00. Refer also to Web Figure 1.

The healthy profile (Figure 1A) shows the canonical trajectory of individuals whose peak probability of transitioning to MCI occurs between 95 and 100 years of age. This group has only a 50% probability of progressing to dementia by age 100; 21% of the sample was closest to this trajectory. The intermediate profile (Figure 1B) shows the canonical trajectory of individuals having a peak probability of progressing to MCI between 85 and 90 years of age, with a peak probability of progressing to dementia between 90 and 95 years of age (29% of the sample was closest to this trajectory). Finally, the unhealthy profile (Figure 1C) shows the typical or canonical trajectory of individuals who have a peak probability of progressing to MCI between age 75 and 80 years and a peak probability of progressing to dementia between the ages of 80 and 85 years (50% of the sample). Note that the use of the terms “healthy,” “intermediate,” and “unhealthy” is to ease discussion of the findings and is not based on other data.

The results of the analysis including the 7 time-invariant predictors (refer to equation 6), as well as the survival component described in equation 3, are presented in Web Tables 3–5. The effects of sex, diabetes, and heart disease on the log ratios of the membership vectors were not statistically significant (i.e., the 95% confidence intervals for all these covariates include 0.0). Thus, the results discussed below are from the simpler model excluding those predictors.

Table 3 provides information on the influence of the 4 significant confounding factors (race, education, hypertension, *ApoE4*) on the closeness of single individuals to each canonical trajectory in the form of confidence intervals for a log-odds measure of closeness to one trajectory compared with another. From the table, we can see that race (white) had

Table 2. Mixed Membership Trajectory Model With 4 Time-Invariant Predictors, Showing Posterior Means and Standard Deviations for the Parameters of the 3 Canonical Survival Trajectories, Cardiovascular Health Study Cognition Study, 1992–2014

Meaning	Weibull Parameter	k = 1, ^a mean (SD)	k = 2, ^b mean (SD)	k = 3, ^c mean (SD)
Scale	θ_k	3.887 (0.376)	5.397 (0.616)	4.050 (0.256)
Shape	δ_k	29.059 (0.951)	29.610 (0.799)	25.217 (0.499)

Abbreviation: SD, standard deviation.

^a Healthy profile.

^b Intermediate profile.

^c Unhealthy profile.

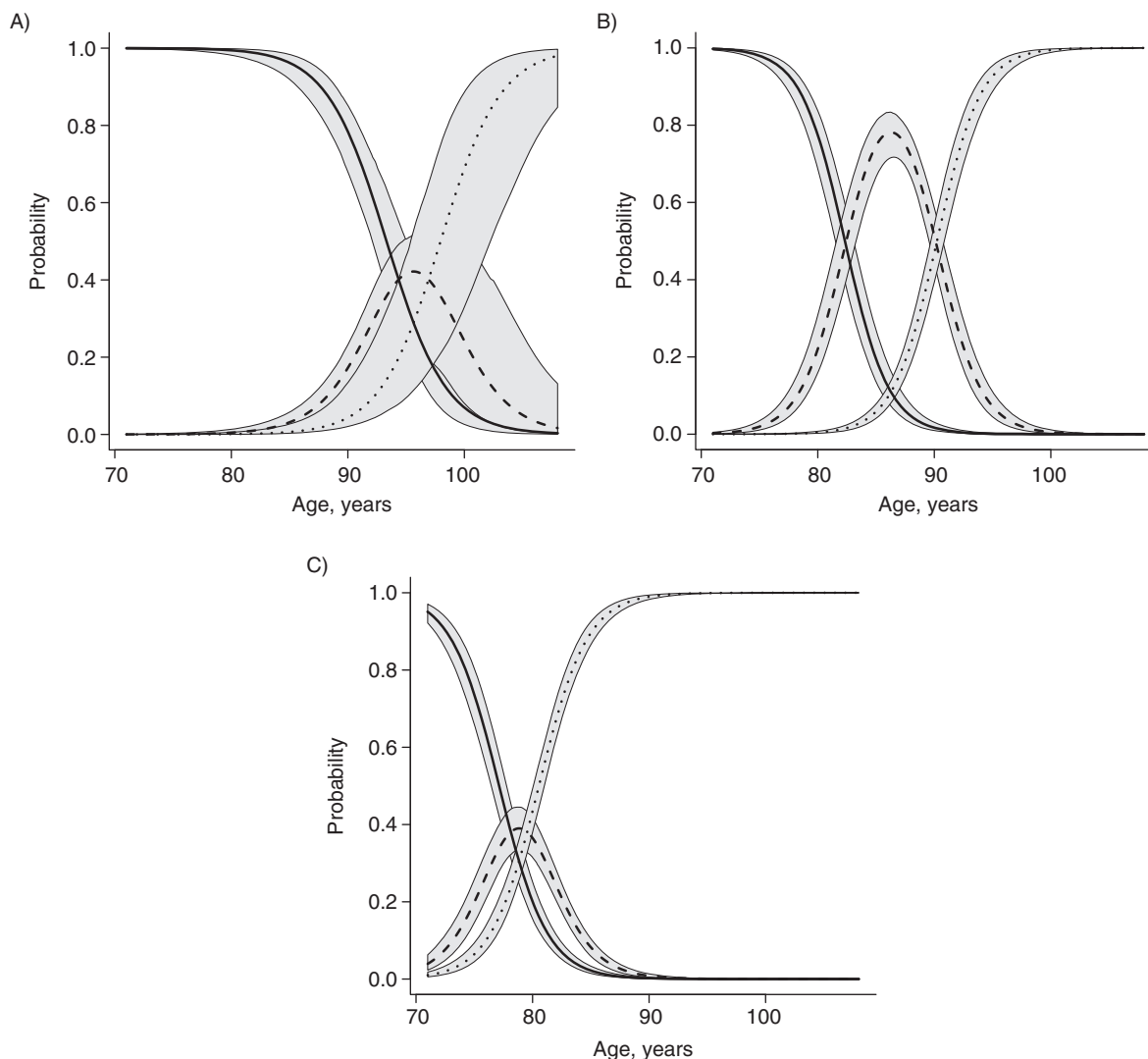


Figure 1. The probability of being cognitively normal, having mild cognitive impairment (MCI), or having dementia for each of the 3 profiles (healthy (A), intermediate (B), and unhealthy (C)), Cardiovascular Health Study Cognition Study, 1992–2014. In each plot, the solid lines represent the probability of being normal, the dashed lines represent the probability of having MCI, and the dotted lines represent the probability of having dementia. The bands around each of the probability curves are the pointwise posterior 95% credible intervals, and they describe the uncertainty of the estimation of the trajectories.

the consequence of moving individuals away from the unhealthy profile. Having more than a high school education resulted in increased closeness to the healthy profile relative to the unhealthy profile. Hypertension was associated with greater closeness to the unhealthy profile relative to the intermediate profile, while the presence of even a single copy of *ApoE4* increased the closeness of individuals to the unhealthy profile (Table 3).

Figure 2 shows the effects of varying individual parameters on the closeness to the canonical trajectories. Figure 2A shows the *predicted* average trajectories (dotted lines) of 5 study participants with different characteristics determined by the 4 time-invariant predictors. This plot refers to 5 imaginary individuals whose trajectories are estimated by using the average parameters of the model. The plot is useful to

describe what happens as the values of the predictors change. Each trajectory was obtained by using the posterior means of the parameters a_{jk} of equation 6 that determine the values of the membership vectors to be used in equation 3. An individual with white race, education beyond high school, no hypertension, and no copies of *ApoE4* has a predicted trajectory relatively close to the healthy canonical profile. As we change the values of the 4 time-invariant predictors, one by one, we see that the predicted individual trajectories move toward the unhealthy canonical profile.

Figure 2B shows the individual trajectories of 2 CHS Cognition Study participants, estimated by the posterior means of the membership vectors g_i . Their trajectories are estimated by using subject-specific estimates from the model. The trajectory closer to the unhealthy profile belongs to an individual

Table 3. Mixed Membership Trajectory Model With 4 Time-Invariant Predictors, Showing Posterior Means and 95% Credible Intervals for the Parameters Representing the Effects of Time-Invariant Predictors on the Closeness of Individual Trajectories to the Typical Profiles, Cardiovascular Health Study Cognition Study, 1992–2014

Predictor	Mean	95% Credible Interval
Race		
Healthy vs. intermediate	0.39	−0.015, 0.84
Healthy vs. unhealthy	1.21	0.82, 1.62 ^a
Intermediate vs. unhealthy	0.83	0.47, 1.26 ^a
Education		
Healthy vs. intermediate	0.26	−0.15, 0.81
Healthy vs. unhealthy	0.50	0.10, 0.92 ^a
Intermediate vs. unhealthy	0.24	−0.17, 0.66
Hypertension		
Healthy vs. intermediate	0.18	−0.22, 0.62
Healthy vs. unhealthy	−0.26	−0.60, 0.07
Intermediate vs. unhealthy	−0.43	−0.79, −0.13 ^a
<i>Apoε4</i>		
Healthy vs. intermediate	0.12	−0.31, 0.60
Healthy vs. unhealthy	−0.71	−1.12, −0.26 ^a
Intermediate vs. unhealthy	−0.83	−1.23, −0.40 ^a

Abbreviation: *Apoε4*, ε4 allele of the apolipoprotein E gene.

^a Significant effect of covariate on closeness to trajectory.

who is a black, less educated, and hypertensive and who had at least 1 copy of *Apoε4*. By contrast, the trajectory closer to the healthy profile belongs to an individual who is white, better educated, and normotensive and who had no copies of *Apoε4*.

In order to evaluate our decision to include 3 canonical trajectories in our model, we used the method of posterior predictive checking (28), replicating the original diagnoses to obtain 1,000 different simulated data sets. The model with 2 canonical profiles systematically overestimated the number of individuals who were diagnosed with MCI at least once in their life (posterior predictive $P < 0.01$), while data sets simulated from the model with 3 canonical profiles did not show a significant difference from the original data set (posterior predictive $P = 0.24$). (Refer to Web Figures 2 and 3.) We also attempted a model with 4 canonical profiles, which produced a fourth additional canonical profile that essentially duplicated the healthier one. In addition, this model significantly increased the costs of computation. Based on these results, we concluded that the model with 3 canonical profiles gave us the best tradeoff between interpretation of the data and costs of computation. In general, selection of the optimal number of canonical trajectories for MMTMs remains an open question (26).

DISCUSSION

To the best of our knowledge, we have provided for the first time a data-driven description of multiple pathways toward

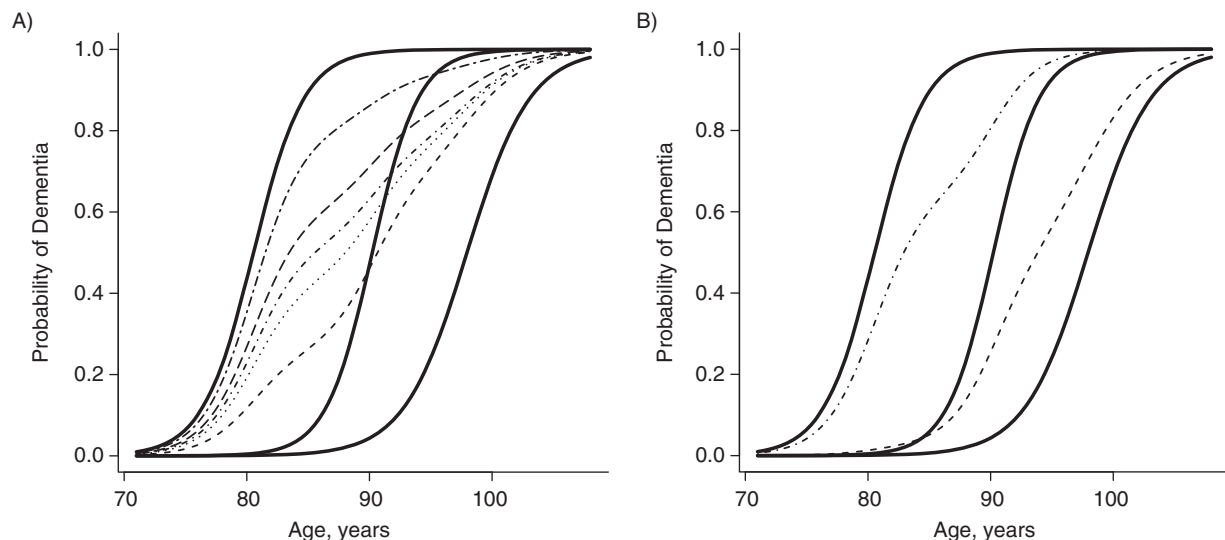


Figure 2. The effects of varying individual parameters on the closeness to the canonical trajectories, Cardiovascular Health Study Cognition Study, 1992–2014. A) Three canonical trajectories to dementia (solid lines), from top to bottom: unhealthy, intermediate, and healthy. The plot also shows the predicted average trajectories (dotted and dashed lines) of 5 individuals with different characteristics determined by the 4 time-invariant binary predictors. The trajectories were obtained by using the posterior means of the parameters a_{jk} of equation 6, which determine the values of the membership vectors to be used in equation 3. An individual with white race, education beyond high school, no hypertension, and no copies of the ε4 allele of the apolipoprotein E gene (*Apoε4*) has a predicted trajectory relatively close to the healthy canonical profile (bottom dotted line). As we change the values of the 4 time-invariant predictors one by one (in order, from bottom to top: *Apoε4*, hypertension, education, race), we see that the predicted individual trajectories move toward the unhealthy canonical profile. B) Individual trajectories of 2 Cardiovascular Health Study Cognition Study participants. Their trajectories are estimated by using subject-specific estimates from the model. The trajectory closer to the unhealthy profile belongs to an individual who was black, less educated, and hypertensive and who had at least 1 copy of *Apoε4*. The trajectory closer to the healthy profile belongs to an individual who was white, better educated, and normotensive and who had no copies of *Apoε4*.

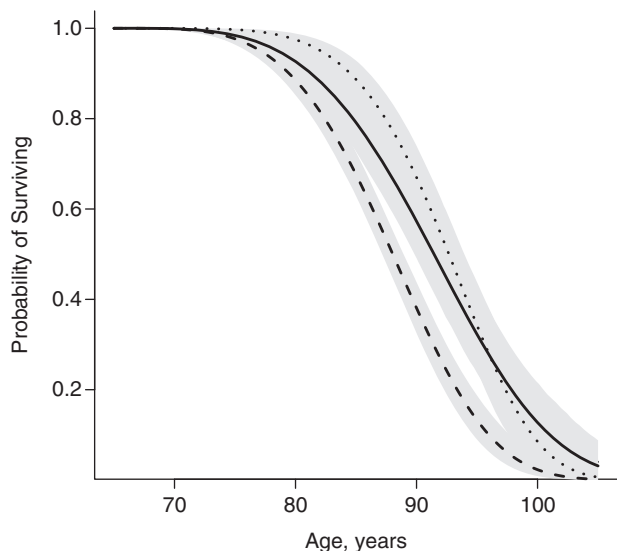


Figure 3. The survival functions for the individuals in the healthy (solid line), intermediate (dotted line), and unhealthy (dashed line) trajectories, Cardiovascular Health Study Cognition Study, 1992–2014. The gray shading indicates 95% credible intervals.

dementia among individuals over the age of 70 years, using a MMTM analysis. The results provide new insights into the natural history of Alzheimer disease, and they may also provide evidence for a potential difference in the pathophysiology of the development of dementia as a function of age.

We identified 3 separate canonical trajectories and then measured the impact of 7 time-invariant covariates, 4 of which were statistically significant, on the closeness of individuals to each of these profiles. Approximately 29% of the individuals included in this analysis are closest to what we refer to as the intermediate canonical trajectory; that is, in their membership vector (g_{i1} ; g_{i2} ; g_{i3}), the second element is larger than the other 2. These individuals have a peak probability of progressing to MCI between the ages of 85 and 90 years, with the peak probability of progressing to dementia between 90 and 95 years of age. By contrast, the unhealthy canonical trajectory revealed a peak probability of MCI before the age of 80 years, with a peak probability of dementia before the age of 85 years. Nearly one-half of the CHS Cognition Study sample was closest to this profile. Finally, almost one-quarter of the individuals in the CHS Cognition Study were estimated to be closest to the healthy profile. For this group of individuals, the peak probability of MCI occurred after the age of 95 years, with approximately 50% of the individuals with this trajectory meeting criteria for MCI. By the age of 100 years, approximately 50% of the individuals with this trajectory had been diagnosed with dementia, but the peak probability of dementia never reached 100%, even after the age of 100 years.

Thus, what we had thought might be a unique, common trajectory from normal cognition to dementia is, in fact, a collection of individual histories that can be characterized by their similarity to a small number of canonical trajectories.

From the perspective of the MMTM, the risk factors (hypertension, race, education, *ApoE4*) should not be viewed as affecting the risk of MCI and dementia, but rather as affecting the extent to which each individual's pathway to MCI and dementia is closer to one or more of the canonical trajectories. White race and higher education are associated with an increased closeness to the healthy trajectory, whereas a diagnosis of hypertension or the presence of a single copy of *ApoE4* increased the closeness of individuals to the unhealthy trajectory. We can view the results of the analysis of survivorship and of the effects of the 4 binary covariates on closeness to specific canonical profiles as an indirect validation of the 3 profile model. That is, the fact that the individuals with an unhealthy trajectory are also the ones most likely to die sooner is consistent with the observation that demented individuals have a higher risk of death (31) (Figure 3).

The subjects in the CHS Cognition Study were at least 70 years of age at the time that we began the observation window for this analysis (about 1992–1994). Thus, we cannot say whether there are additional trajectories that might be apparent had we begun observation earlier, for example, at the age of 50 or 55 years. Further, we cannot know whether earlier observation might have altered the shape of the 3 canonical trajectories that we have here. In addition, there is the very provocative possibility that the “groundwork” for these trajectories was laid years, if not decades, earlier. For example, midlife hypertension imposes significant risk for dementia 25 years later. Unfortunately, the CHS did not begin observation until the participants were at least 65 years of age. Nevertheless, what is clear is that this type of modeling can be successful only with the extensive observation time that was available in studies like the CHS Cognition Study. Studies that have relatively brief follow-up time or (perhaps more important) cognitive assessments that are less frequent than we had in the CHS Cognition Study may not be able to do these sorts of natural history analyses.

The covariates in the model do not fully explain trajectory membership, suggesting that other factors may also influence individual trajectories. These may include, for example, changes in brain structure/function, alcohol use, and access and adherence to medication regimens. Data derived from magnetic resonance imaging scans may be useful in future analyses and have already proven their worth in Cox proportional hazard modeling (42, 43). On the other hand, use of alcohol/drugs and medication usage/adherence are time varying predictors, and we had decided to delay including such variables in this first use the MMTM framework in the context of MCI and dementia.

It is of some interest that diabetes and prevalent heart disease did not affect the “closeness” of individual participants to one trajectory or another, in spite of the fact that both of these factors significantly increase risk for the development of clinical dementia (4, 6–9). In the case of diabetes, although the 95% confidence interval did include 0, the direction of the effect was away from the healthy trajectory and toward the unhealthy and intermediate trajectories. Furthermore, the upper limit of the 95% confidence interval was close to 0 (i.e., 0.06 and 0.08) suggesting the possibility that, with a larger number of subjects (and perhaps a lower age range to capture more of the diabetic participants), this effect might have been statistically significant. Moreover, vascular disease is a risk

for death, and it may be that those individuals with heart disease who are still alive at this age are the “healthiest” from the CHS cohort, and this may have minimized any impact of these conditions on the analysis. Second, it is not the case that diabetes and heart disease do not affect the risk to develop dementia. Rather, they do not appear to have an impact on the shape of or closeness to the 3 trajectories over and above the effects accounted for by the other covariates, in particular, older age. Nevertheless, it is critical to understand these relationships, and thus replication and extension of our findings with a larger cohort would be ideal.

Our results emphasize the value of MMTMs as a novel, analytical tool, and they demonstrate their ability to identify patterns within large data sets that might have otherwise gone unnoticed. Nevertheless, there are several limitations to our study. First, MMTMs cannot easily account for time-varying covariates and, thus, additional work is needed to develop computationally efficient methodologies for the incorporation of time-varying covariates into the models and to allow for the visualization of the impact of these covariates. Second, our data have a restricted age range based on the recruiting and enrollment plan of the CHS. Third, we did not follow participants after they were classified as demented and, thus, cannot account for changes from dementia back to MCI.

In spite of these limitations of the current instantiation of the MMTM technique, it is a powerful tool. MMTMs use patterns within the data to identify distinct canonical profiles; the results are not constrained by any a priori assumptions about the trajectories. These techniques are highly innovative in that they are able to 1) account for “reversing” states (i.e., moving from MCI back to normal cognition, as might be expected following use of antedementia medications, for example); 2) express each individual’s pathway as a weighted combination of the canonical trajectories; and, 3) determine the extent to which risk factors for cognitive impairment affect the “closeness” of an individual to each of the canonical trajectories. This tool has the potential for meaningful application in a variety of domains related to MCI and Alzheimer disease, and it holds the promise to reveal important data-driven insights into the natural and treated history of Alzheimer disease and dementia.

ACKNOWLEDGMENTS

Author affiliations: Department of Statistics, Dietrich College of Humanities and Social Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania (Fabrizio Lecci, Brian Junker); Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania (Lewis H. Kuller); Department of Neurology, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania (Oscar L. Lopez, James T. Becker); Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania (James T. Becker); and Department of Psychology, Dietrich School of Arts and Science, University of Pittsburgh, Pittsburgh, Pennsylvania (James T. Becker).

The research reported in this article was supported in part by contracts N01-HC-85239, N01-HC-85079 through

N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, and N01-HC-45133 and grant HL080295 from the National Heart, Lung, and Blood Institute, with an additional contribution from the National Institute of Neurological Disorders and Stroke. Additional support was provided through AG-023629, AG-15928, AG-20098, AG-027002, AG-05133, and AG-027058 from the National Institute on Aging.

A full list of principal CHS investigators and institutions can be found at <https://chs-nhlbi.org/pi>.

Conflict of interest: none declared.

REFERENCES

1. Fitzpatrick AL, Kuller LH, Ives DG, et al. Incidence and prevalence of dementia in the Cardiovascular Health Study. *J Am Geriatr Soc.* 2004;52(2):195–204.
2. Evans DA, Funkenstein HH, Albert MS. Prevalence of Alzheimer’s disease in a community population of older persons. Higher than previously reported. *JAMA.* 1989; 262(18):2551–2556.
3. Corrada MM, Brookmeyer R, Paganini-Hill A, et al. Dementia incidence continues to increase with age in the oldest old: The 90+ Study. *Ann Neurol.* 2010;67(1):114–121.
4. Tang MX, Maestre G, Tsai WY, et al. Relative risk of Alzheimer disease and age-at-onset distributions, based on *APOE* genotypes among elderly African Americans, Caucasians, and Hispanics in New York City. *Am J Hum Genet.* 1996;58(3):574–584.
5. Launer LJ, Andersen K, Dewey ME, et al. Rates and risk factors for dementia and Alzheimer’s disease: results from EURODEM pooled analyses. EURODEM Incidence Research Group and Work Groups. European Studies of Dementia. *Neurology.* 1999;52(1):78–84.
6. Kuller LH, Lopez OL, Newman A, et al. Risk factors for dementia in the Cardiovascular Health Cognition Study. *Neuroepidemiology.* 2003;22(1):13–22.
7. Irie F, Fitzpatrick AL, Lopez OL, et al. Enhanced risk for Alzheimer disease in persons with type 2 diabetes and *APOE* $\epsilon 4$: the Cardiovascular Health Study Cognition Study. *Arch Neurol.* 2008;65(1):89–93.
8. Luchsinger JA, Tang MX, Stern Y, et al. Diabetes mellitus and risk of Alzheimer’s disease and dementia with stroke in a multiethnic cohort. *Am J Epidemiol.* 2001;154(7):635–641.
9. Skoog I, Lernfelt B, Landahl S, et al. 15-year longitudinal study of blood pressure and dementia. *Lancet.* 1996;347(9009): 1141–1145.
10. Matsuzaki T, Sasaki K, Tanizaki Y, et al. Insulin resistance is associated with the pathology of Alzheimer disease: the Hisayama Study. *Neurology.* 2010;75(9):764–770.
11. Ohara T, Doi Y, Ninomiya T, et al. Glucose tolerance status and risk of dementia in the community: the Hisayama Study. *Neurology.* 2011;77(12):1126–1134.
12. Verghese J, Lipton RB, Katz MJ, et al. Leisure activities and the risk of dementia in the elderly. *N Engl J Med.* 2003;348(25): 2508–2516.
13. Erickson KI, Raji CA, Lopez OL, et al. Physical activity predicts gray matter volume in late adulthood: the Cardiovascular Health Study. *Neurology.* 2010;75(16): 1415–1422.
14. Scarmeas N, Stern Y, Tang MX, et al. Mediterranean diet and risk for Alzheimer’s disease. *Ann Neurol.* 2006;59(6): 912–921.

15. Lopez OL, Becker JT, Chang YF, et al. Incidence of mild cognitive impairment in the Pittsburgh Cardiovascular Health Study-Cognition Study. *Neurology*. 2012;79(15):1599–1606.
16. Saxton J, Lopez OL, Ratcliff G, et al. Preclinical Alzheimer disease: neuropsychological test performance 1.5 to 8 years prior to onset. *Neurology*. 2004;63(12):2341–2347.
17. Lopez OL, Jagust WJ, DeKosky ST, et al. Prevalence and classification of mild cognitive impairment in the Cardiovascular Health Study Cognition Study: part 1. *Arch Neurol*. 2003;60(10):1385–1389.
18. Lopez OL, Kuller LH, Becker JT, et al. Incidence of dementia in mild cognitive impairment in the Cardiovascular Health Study Cognition Study. *Arch Neurol*. 2007;64(3):416–420.
19. Lopez OL, Kuller LH, Mehta PD, et al. Plasma amyloid levels and the risk of AD in normal subjects in the Cardiovascular Health Study. *Neurology*. 2008;70(19):1664–1671.
20. Manrique-Vallier D. Mixed membership trajectory models. In: Airoldi E, Blei D, Erosheva E, et al, eds. *Handbook of Mixed Membership Models and Their Applications*. New York, NY: Chapman-Hall; 2014:173–188.
21. Manrique-Vallier D. *Longitudinal Mixed Membership Models with Applications to Disability Survey Data* [dissertation]. Pittsburgh, PA: Carnegie-Mellon University; 2010.
22. Erosheva EA, Fienberg SE, Joutard C. Describing disability through individual-level mixture models for multivariate binary data. *Ann Appl Stat*. 2007;1(2):346–384.
23. Connor JT. *Multivariate Mixture Models to Describe Longitudinal Patterns of Frailty in American Seniors* [dissertation]. Pittsburgh, PA: ProQuest; 2006:3275170.
24. Jack CR Jr, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol*. 2010;9(1):119–128.
25. Jack CR Jr, Knopman DS, Jagust WJ, et al. Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers. *Lancet Neurol*. 2013;12(2):207–216.
26. Petersen RC, Smith GE, Waring SC, et al. Mild cognitive impairment: clinical characterization and outcome. *Arch Neurol*. 1999;56(3):303–308.
27. Lambert JC, Schraen-Maschke S, Richard F, et al. Association of plasma amyloid beta with risk of dementia: the prospective Three-City Study. *Neurology*. 2009;73(11):847–853.
28. Lopez OL, Kuller LH, Fitzpatrick A, et al. Evaluation of dementia in the Cardiovascular Health Cognition Study. *Neuroepidemiology*. 2003;22(1):1–12.
29. Brandt J, Spencer M, Folstein M. The telephone interview for cognitive status. *Neuropsychiatry Neuropsychol Behav Neurol*. 1988;1(2):111–117.
30. Kawas C, Segal J, Stewart WF, et al. A validation study of the Dementia Questionnaire. *Arch Neurol*. 1994;51(9):901–906.
31. Jorm AF, Jacomb PA. The Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE): socio-demographic correlates, reliability, validity and some norms. *Psychol Med*. 1989;19(4):1015–1022.
32. McKhann G, Drachman D, Folstein M, et al. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*. 1984;34(7):939–944.
33. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263–269.
34. American Psychiatric Association. *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
35. World Health Organization. *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. Geneva, Switzerland: World Health Organization; 1993.
36. Chui HC, Victoroff JI, Margolin D, et al. Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. *Neurology*. 1992;42(3 pt 1):473–480.
37. Román GC, Tatemichi TK, Erkinjuntti T, et al. Vascular dementia: diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*. 1993;43(2):250–260.
38. McKeith IG, Dickson DW, Lowe J, et al. Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium. *Neurology*. 2005;65(12):1863–1872.
39. Neary D, Snowden JS, Gustafson L, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology*. 1998;51(6):1546–1554.
40. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol*. 1991;1(3):263–276.
41. Christian PR, Casella G. *Monte Carlo Statistical Methods*. New York, NY: Springer; 2014.
42. Zeifman L, Eddy W, Lopez OL, et al. Whole brain, voxel level analysis of grey matter volume and time to incident mild cognitive impairment or Alzheimer's disease. Presented at the 2014 Annual Meeting of the American Academy of Neurology, Philadelphia, PA, April 30, 2014.
43. Zeifman LE, Eddy WF, Lopez OL, et al. Voxel level survival analysis of grey matter volume and incident mild cognitive impairment or Alzheimer's disease [published online ahead of print February 26, 2015]. *J Alzheimers Dis*. (doi:10.3233/JAD-150047).