

Joint Modeling of Transitional Patterns of Alzheimer's Disease

Wei Liu¹, Bo Zhang^{2*}, Zhiwei Zhang³, Xiao-Hua Zhou⁴

1 Department of Mathematics, Harbin Institute of Technology, Harbin, P.R. China, **2** Biostatistics Core, School of Biological and Population Health Sciences, College of Public Health and Human Sciences, Oregon State University, Corvallis, Oregon, United States of America, **3** Division of Biostatistics, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, Maryland, United States of America, **4** Department of Biostatistics, School of Public Health and Community Medicine, University of Washington, HSR&D Center of Excellence, U.S. Department of Veterans Affairs Puget Sound Health Care System, Seattle, Washington, United States of America

Abstract

While the experimental Alzheimer's drugs recently developed by pharmaceutical companies failed to stop the progression of Alzheimer's disease, clinicians strive to seek clues on how the patients would be when they visit back next year, based upon the patients' current clinical and neuropathologic diagnosis results. This is related to how to precisely identify the transitional patterns of Alzheimer's disease. Due to the complexities of the diagnosis of Alzheimer's disease, the condition of the disease is usually characterized by multiple clinical and neuropathologic measurements, including Clinical Dementia Rating (CDRGLB), Mini-Mental State Examination (MMSE), a score derived from the clinician judgement on neuropsychological tests (COGSTAT), and Functional Activities Questionnaire (FAQ). In this research article, we investigate a class of novel joint random-effects transition models that are used to simultaneously analyze the transitional patterns of multiple primary measurements of Alzheimer's disease and, at the same time, account for the association between the measurements. The proposed methodology can avoid the bias introduced by ignoring the correlation between primary measurements and can predict subject-specific transitional patterns.

Citation: Liu W, Zhang B, Zhang Z, Zhou X-H (2013) Joint Modeling of Transitional Patterns of Alzheimer's Disease. PLoS ONE 8(9): e75487. doi:10.1371/journal.pone.0075487

Editor: Xiaofeng Wang, Cleveland Clinic Lerner Research Institute, United States of America

Received: March 14, 2013; **Accepted:** August 14, 2013; **Published:** September 20, 2013

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: Dr. Wei Liu's research was partially supported by "the Fundamental Research Funds for the Central Universities" through Grant HIT.NSRIF.2012063. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The National Alzheimer's Coordinating Center database is funded by National Institute on Aging Grant U01 AG016976.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Bo.Zhang@oregonstate.edu

Introduction

Alzheimer's disease is the most common type of dementia with the loss of brain function, which affects mental processes including attention, memory, producing and understanding language, learning, reasoning, problem solving, and decision making. Patients with Alzheimer's disease often are 65 years old or older, but there have been cases of early onset Alzheimer's (diagnosed before the age of 65) [1]. Usually, symptoms of Alzheimer's disease develop gradually and get worse as the disease progresses, and eventually lead to death. Currently, there are no clinical treatments that can cure Alzheimer's disease or reverse its progression, although the safety and efficacy of more than 400 pharmaceutical treatments have been or are being investigated worldwide (approximately a quarter of these compounds are in Phase III trials). According to Alzheimer's Association, 5.4 million Americans lived with Alzheimer's disease in 2012, and the disease is now the 6th leading cause of death in the United States [2]. In this research article, we consider the Uniform Data Set from the Alzheimer's Disease Centers program established by the National Institute on Aging (NIA) [3,4]. The Alzheimer's Disease Centers program (U.S. National Institutes of Health Grant U01 AG016976) aims at characterizing individuals with mild Alzheimer's disease and cognitive impairment in comparison with nondemented aging. Currently more than 29 Alzheimer Disease

Centers are funded to collect demographic, behavioral status, cognitive testing, and clinical diagnoses information on participants. In all Alzheimer Disease Centers, the data in Uniform Data Set are obtained longitudinally in a uniform manner from participants, using standard methods and uniform diagnostic criteria to evaluate subjects. The data are then integrated in the National Alzheimer's Coordinating Center to support sharing and collaborative research.

In numerous clinical and epidemiologic studies on chronic diseases, the scientific interest lies in understanding the longitudinal transitions of primary disease measurements, and in identifying factors that govern transition probabilities. For Alzheimer's disease, the transition patterns contain important information on disease evolution and prognosis. Alzheimer's disease cannot be accurately diagnosed and described by a single measurement; clinicians need to synthesize information from various sources, including features of symptom onset and course, evaluation of mental status, and performance measurements from neurologic and neuropsychological examinations. Therefore, a major challenge in characterizing transitional patterns of Alzheimer's disease is that the disease status is represented by not one but multiple measurements. If the status of Alzheimer's disease could be ascertained by one clinical measurement, then the existing statistical methodologies for univariate transition models would be adequate for the present application. Various authors have

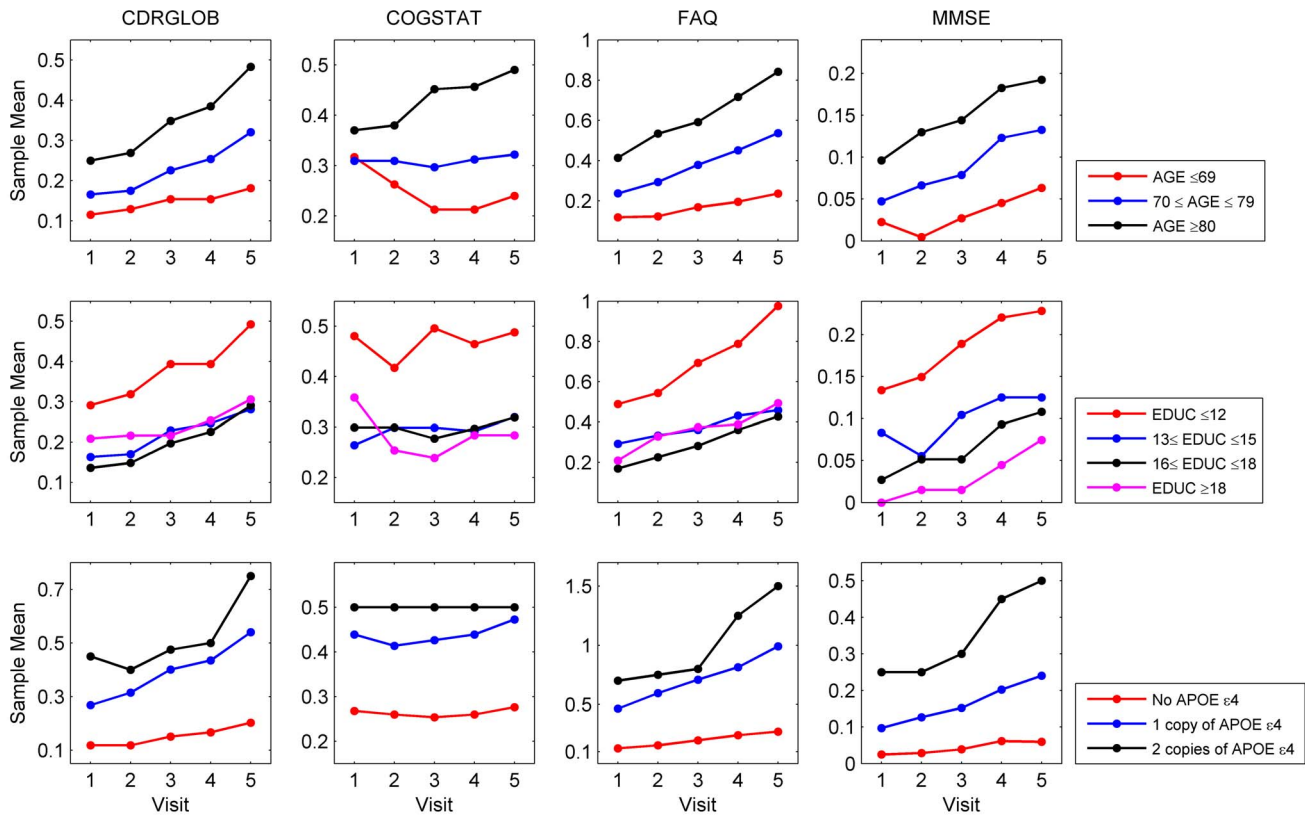


Figure 1. The Alzheimer's disease data: Sample means of four responses in each group presented by age, years of education and the number of copies APOE- $\epsilon 4$.
doi:10.1371/journal.pone.0075487.g001

proposed models for characterizing transition patterns with univariate longitudinal data: Zeger and Qaqish [5] and Wong [6] proposed Markov regression models and developed estimators based on composite likelihood functions; Albert [7] proposed a transitional model for longitudinal binary data subject to nonignorable missingness and derived an EM algorithm for parameter estimation; Lee and Daniels [8] proposed marginalized transition models for longitudinal binary data and developed Fisher-scoring algorithms for estimation. Albert and Follmann [9] developed a shared random-effects transition model that linked the propensity of transition between states to the probability of an intermittent missed observation or dropout. These methods can be used to analyze each disease measurement separately, but they are not able to account for both the association between different disease measurements and subject-specific transitional patterns.

This research article proposes a class of innovative joint random-effects transition models for elucidating the transitional patterns of multiple longitudinal measurements that are closely related to Alzheimer's disease. Multivariate transition models had been discussed in three recent articles for binary longitudinal data. A generalized estimating equation method for joint transitional models of multivariate longitudinal binary data was proposed by Zeng and Cook [10]. This approach allows modeling of covariate effects on marginal transition probabilities, as well as the association parameters. Ilk and Daniels [11] formulated marginalized random-effects models to accommodate multivariate longitudinal binary data, and Lee et al. [12] later extended these models by using a new covariance matrix with a Kronecker decomposition. All of these methods are focused on the marginal

transition patterns of multivariate binary data. In the Alzheimer's disease data, several primary measurements are multi-categorical, and investigators are more interested in subject-specific prognosis than population-averaged prognosis. The joint random-effects transition models proposed in this article involve random effects in each submodel for an individual disease measurement, thereby accounting for the heterogeneity between subjects. Meanwhile, a correlation between disease measurements is induced by the multivariate distribution of the random effects across submodels. In addition, each univariate submodel in the proposed joint random-effects transition models can be flexible enough to accommodate various types of measurements (e.g., Gaussian, multi-categorical, count, etc.). Relative to the existing literature, the benefits of the joint random-effects transition models are multifold: (i) comparing with the modeling of single process in [5,6,7], they allow flexible correlation among multiple measurements for a disease; (ii) comparing with the marginal models in [8,10,11], they offer insights on patient-specific transitional patterns of disease measurements over time; (iii) comparing with the models for binary data in [10,11], they offer flexible submodels for fitting the data with various types; and (iv) they can identify common and uncommon covariates that govern the transitional probabilities of each disease outcome. The joint random-effects transition models can help clinicians predict a patient's Alzheimer disease status over time, based on the patient's current status and other genetic or sociodemographic factors.

Methods

The National Alzheimer’s Coordinating Center: Uniform Data Set

The National Alzheimer’s Coordinating Center is responsible for developing and maintaining a database of participant information collected from the 34 past and present Alzheimer’s Disease Centers. The NIA organized the Alzheimer’s Disease Centers Clinical Task Force and defined a standardized Uniform Data Set. Alzheimer’s Disease Centers provide researchers a standard set of assessment procedures to characterize mild Alzheimer disease and mild cognitive impairment in comparison with nondemented controls. Our analysis is based on a data set concerning the status of Alzheimer’s disease at five to six annual clinic visits for 746 subjects. The status of Alzheimer’s disease for each subject was mainly characterized by four primary measurements: Clinical Dementia Rating, Mini-Mental State Examination, Composite Cognitive Score, and Functional Activities Questionnaire. At each visit, clinicians assessed the overall dementia level of the subject in term of Clinical Dementia Rating, a numeric scale used to quantify the severity of dementia [13,14]. We denote this first measurement as “CDRGLOB”. CDRGLOB is an ordinal variable with five levels (0, 0.5, 1, 2 and 3, indicating normal status without demetia and questionable, mild, moderate, and severe dementia, respectively). The second primary measurement for the evaluation of cognitive state and impairment was obtained from Mini-Mental State Examination (MMSE), which includes orientation to place and time, registration, attention and concentration, recall, language, and visual construction [15]. The integer score (from 0 to 30) for MMSE was dichotomized by the recommended cutoff score of 23 to detect impairment [15]. In addition to MMSE, another score, denoted by “COGSTAT”, was obtained from neuropsychological examinations [16] and used to assess whether or not a subject’s cognitive status was normal for the current age. The fourth primary characteristic of Alzheimer’s disease is Functional Activities Questionnaire, denoted by “FAQ”, which represents the level of functional activities of the subject in the community (as related to cognitive dysfunction) [17]. The integer score (from 0 to 30) for FAQ was categorized into five ordinal levels: 0, 1, 2, 3, and 4, representing normal, mildly difficult, moderate, serious and complete disability, respectively.

Sloane et al. [18] studied the transitional patterns of Alzheimer’s Disease in term of the rating of CDR. However, as indicated earlier, univariate modeling has limitations because it ignores the correlation among different measurements. Using novel joint random-effects transitional we attempt to address three issues in this research article: how the disease condition of the subject would be next year given his or her current diagnosis results in terms of the four primary measurements, how the four primary measurements are correlated through random effects, and how the genetic or sociodemographic factors affect the transitional probabilities. The four primary measurements of Alzheimer’s disease are all determined by the participants’ underlying condition during a clinical visit. As a result, they are expected to be substantially correlated, and understanding their correlation is a research question of interest. Moreover, patients’ age, total years of education, and the number of copies of Apolipoprotein E $\epsilon 4$ allele (APOE- $\epsilon 4$) are well known risk factors of Alzheimer’s disease [19,20,21,22]. It is essential to incorporate those genetic and sociodemographic factors into the joint random-effect transitional models as potential covariates. Hypothesized links between the four primary variables and the potential covariates are shown in Figure 1, in which the means of the four primary variables based on their first five observations are plotted by participant group.

Joint Transition Models for Bivariate Longitudinal Data

In this section, we initiate the discussion on joint random-effects transitional models by considering bivariate longitudinal data. General transitional models for multivariate longitudinal data (with three or more primary measurements) are discussed in the next section.

Consider a longitudinal study in which I participants are followed prospectively. Longitudinal profiles of the i th participant, $i = 1, 2, \dots, I$, are measured at J scheduled time point $t_1 < t_2 < \dots < t_J$, $j = 1, 2, \dots, J$, without dropout or intermittent missing values. Assume that a vector of two response variables $Y_i = (Y_{i1}, Y_{i2})'$ is measured on each participant at each time point, and longitudinal measurements of each response form a stochastic process. The processes for participant i are denoted by $Y_{i1} = (Y_{i11}, Y_{i21}, \dots, Y_{iJ1})'$ and $Y_{i2} = (Y_{i12}, Y_{i22}, \dots, Y_{iJ2})'$, where Y_{ijk} is the response variable at time point t_j for the k th process, $k = 1, 2$. Let x_{ijk} be a $p_k \times 1$ vector of covariates that may be associated with the transition of the k th process at t_j . Important covariates in x_{ijk} may include a categorical variable denoting treatment group, assessment time t_j or related time-dependent variables, and risk factors of medical conditions. The history of past s response observations and past s covariates for participant i at t_j is denoted by $H_{ijk}^{(s)} = (Y_{i,j-1,k}, Y_{i,j-2,k}, \dots, Y_{i,j-s,k})'$. The two processes may be connected through correlated random effects.

We consider several scenarios. In the first scenario, it is assumed that the bivariate longitudinal response $Y_i = (Y_{i1}, Y_{i2})'$ is a mixed-type vector consisting of a Gaussian stochastic process $Y_{i1} = (Y_{i11}, Y_{i21}, \dots, Y_{iJ1})'$ and a binary stochastic process $Y_{i2} = (Y_{i12}, Y_{i22}, \dots, Y_{iJ2})'$. We assume that Y_{ij1} follows a first-order autoregressive correlation model [23,24] given by

$$Y_{ij1} = x'_{ij1} \beta_1 + f_1^*(H_{ijk}^{(1)}; \alpha_1, \beta_1) + \varepsilon_{ij}, \tag{1}$$

where $f_1^*(H_{ij1}^{(1)}; \alpha_1, \beta_1) = \alpha_1 Y_{i,j-1,1} - \alpha_1 x'_{i,j-1,1} \beta_1$. Model (1) implies that the transition pattern for each participant is determined by a linear combination of covariates and past states; that is, participants with identical covariate values and past state values have the same expectation for the current state. This assumption may be unrealistic when there exists heterogeneity in the transition pattern across study participants due to some latent factors. In that case, a more realistic model that counts for the participant-level transition heterogeneity is given by $Y_{ij1} = x'_{ij1} \beta_1 + f_1^*(H_{ij1}^{(1)}; \alpha_1, \beta_1) + Z'_{ij1} b_{i1} + \varepsilon_{ij}$, where b_{i1} is an unobserved vector of random effects following a mean-zero multivariate normal distribution (independently of the within-participant errors ε_{ij}), and Z_{ij1} is the random-effects regression vector whose elements are selected from x_{ij1} and $H_{ijk}^{(1)}$. Although we have started by considering a specific form of $f_1^*(H_{ij1}^{(1)}; \alpha_1, \beta_1)$, it is not necessary to assume this particular form. Thus, we now replace $f_1^*(H_{ij1}^{(1)}; \alpha_1, \beta_1)$ by $f_1(H_{ij1}^{(1)}; \alpha_1)$ with an arbitrary function (say f_1) of $H_{ij1}^{(1)}$ and α_1 (the collection of all relevant parameters). Then, the random-effects transition model for the continuous process is

$$Y_{ij1} = x'_{ij1} \beta_1 + f_1(H_{ij1}^{(1)}; \alpha_1) + Z'_{ij1} b_{i1} + \varepsilon_{ij}. \tag{2}$$

The random-effects component in (2) is flexible enough to include several special cases: transition models with an random intercept

correspond to $Z'_{ij1}b_{i1} = b_{i1}$, where b_{i1} is a scalar random effect; models that characterize between-participant heterogeneity induced by current covariates, with $Z'_{ij1}b_{i1} = x'_{ij1}b_{i1}$ for a multivariate random effect vector b_{i1} or $Z'_{ij1}b_{i1} = Z'^{*}_{ij1}b_{i1}$ with Z'^{*}_{ij1} a subset of x'_{ij1} ; and transition models with between-participant heterogeneity induced by past covariates and past outcome states, with $Z'_{ij1}b_{i1} = (Y_{i,j-1,1}, x'_{i,j-1,1})b_{i1}$ or $Z'_{ij1}b_{i1} = Z'^{*}_{ij1}b_{i1}$, where Z'^{*}_{ij1} is a subset of $(Y_{i,j-1,1}, x'_{i,j-1,1})$. The random-effects transition model (2) allows a flexible transition pattern across participants but assumes a common conditional mean pattern $E(Y_{ij1}|H_{ijk}^{(1)})$. It connects the present state Y_{ij1} with the past states through both the previous state $Y_{i,j-1,1}$ and the common participant-level random effects b_{i1} .

The random-effects transition model for the binary stochastic process $Y_{i2} = (Y_{i12}, Y_{i22}, \dots, Y_{iJ2})'$, $i = 1, 2, \dots, I$, with participant-specific random effects can be constructed in an analogous manner. Suppose each binary realization y_{ij2} takes value either 1 or 0, indicating the active or inactive status of the process, then the random-effects transition model extended from the marginal first-order two-status Markov chain is

$$\text{logit}[P(Y_{ij2} = 1)|H_{ij2}^{(1)}, b_{i2}] = x'_{ij2}\beta_2 + f_2(H_{ij2}^{(1)}; \alpha_2) + Z'_{ij2}b_{i2}, \quad (3)$$

where the regression matrices x_{ij2} , Z_{ij2} , regression coefficients β_2 , and random effects vector b_{i2} are specified in a way similar to (2). In (3), $f_2(H_{ij2}^{(1)}; \alpha_2)$ is an arbitrary function of $Y_{i,j-1,2}$ and $x'_{i,j-1,2}$, namely, $f_2(H_{ij2}^{(1)}; \alpha_2) = f_2(Y_{i,j-1,2}, x'_{i,j-1,2}; \alpha_2)$. Instead of targeting marginal transition probabilities, (3) is best suited for describing individual transitional patterns based on individual characteristics and random effects.

To formulate a joint random-effects model for the Gaussian and binary mixed-type bivariate longitudinal process, we combine univariate models (2) and (3) through a joint multivariate mean-zero normal distribution for all random effects: $(b_{i1}, b_{i2})' \sim N(0, \Sigma)$. The joint model induces a correlation between the two longitudinal processes through the correlated random effects b_{i1} and b_{i2} . Usually, the correlation between processes cannot be analytically derived. An important special case of the joint model is shared-parameter models where the random effects in (2) and (3) are identical for all study participants (i.e., $b_{i1} \equiv b_{i2}$). In such models the random effects are treated as shared “parameters”. Note that it is necessary to scale the shared random effects, which means that Z_{ijk} needs to be defined carefully. Shared-parameter models make strong assumptions on the association between outcomes, though it avoids high-dimension integration in maximum likelihood estimation.

In the second important scenario, the bivariate longitudinal outcome $Y_i = (Y_{i1}, Y_{i2})'$ is assumed to be a mixed-type vector with a Gaussian stochastic process $Y_{i1} = (Y_{i11}, Y_{i21}, \dots, Y_{iJ1})'$ and a count stochastic process $Y_{i2} = (Y_{i12}, Y_{i22}, \dots, Y_{iJ2})'$. The Gaussian process is assumed to be identical to the one described by (2). The count process Y_{ij2} , $j = 1, 2, \dots, J$, takes values in the set of nonnegative integers. A natural model for Y_{ij2} would be a Poisson model with density $P(Y_{ij2} = y_{ij2}|H_{ij2}^{(1)}) = e^{-\mu_{ij2}} \mu_{ij2}^{y_{ij2}} / y_{ij2}!$, $y_{ij2} = 0, 1, 2, \dots$, where $\mu_{ij2} = E(Y_{ij2}|H_{ij2}^{(1)})$ is the intensity or rate parameter for Y_{ij2} . To study the transitional behavior of Y_{ij2} , Zeger and Qaqish [5] discussed a first-order Markov model with $\mu_{ij2} = \exp(x'_{ij2}\beta_2) \{ \max(Y_{i,j-1,2}, \kappa) / \exp(x'_{i,j-1,2}\beta_2) \}^\tau$, which is

equivalent to a log-linear Markov model given by $\log(\mu_{ij2}) = x'_{ij2}\beta_2 + \tau[\log\{\max(Y_{i,j-1,2}, \kappa)\} - x'_{i,j-1,2}\beta_2]$. The constant κ is used to prevent $Y_{i,j-1,2} = 0$ from creating an absorbing state, which would force the future states to be identically 0. When $\tau < 0$ a prior state greater than its expectation decreases the expectation for the current state and consequently Y_{ij2} and $Y_{i,j-1,2}$ are negatively correlated; whereas when $\tau > 0$ Y_{ij2} and $Y_{i,j-1,2}$ are positively correlated. Random effects can be incorporated into the log-linear Markov model of Zeger and Qaqish [5] to account for participant-level heterogeneity:

$$\log(\mu_{ij2}) = x'_{ij2}\beta_2 + \tau[\log\{\max(Y_{i,j-1,2}, \kappa)\} - x'_{i,j-1,2}\beta_2] + Z_{ij2}b_{i2}.$$

Generally, we consider a log-linear transition model for the count process Y_{ij2} given by

$$\log(\mu_{ij2}) = \log[E(Y_{ij2}|H_{ij2}^{(1)}, b_{i2})] = x'_{ij2}\beta_2 + f_2(H_{ij2}^{(1)}; \alpha_2) + Z'_{ij2}b_{i2} \quad (4)$$

where x_{ij2} , Z_{ij2} , regression coefficients β_2 , and random effects vector b_{i2} are specified in a way analogous to (2). The general form of $f_2(H_{ij2}^{(1)}; \alpha_2)$, the function that links the current state with the prior state and covariates, includes several special cases: $f_2(H_{ij2}^{(1)}; \alpha_2) = \tau[\log\{\max(Y_{i,j-1,2}, \kappa)\} - x'_{i,j-1,2}\beta_2]$ in Zeger and Qaqish [5]; $f_2(H_{ij2}^{(1)}; \alpha_2) = \log\{1 + \exp(-\alpha_0 - \alpha_1 Y_{i,j-1,2})\}$ in Wong [6]; or $f_2(H_{ij2}^{(1)}; \alpha_2)$ can be $\alpha Y_{i,j-1,2}$, which has limited utility due to the exponentially growing conditional expectation when $\alpha > 0$. Now, we can combine (2) and (4) into a random-effects joint model by assuming a joint multivariate mean-zero normal distribution for b_{i1} and b_{i2} : $(b_{i1}, b_{i2})' \sim N(0, \Sigma)$.

A joint model for two binary stochastic processes $Y_{i1} = (Y_{i11}, Y_{i21}, \dots, Y_{iJ1})'$ and $Y_{i2} = (Y_{i12}, Y_{i22}, \dots, Y_{iJ2})'$ can be specified simply by assuming each behaves as in (3):

$$\text{logit}[P(Y_{ijk} = 1|H_{ijk}^{(1)}, b_{ik})] = x'_{ijk}\beta_k + f_k(H_{ijk}^{(1)}; \alpha_k) + Z'_{ijk}b_{ik}, \quad k = 1, 2, \quad (5)$$

with $(b_{i1}, b_{i2})' \sim N(0, \Sigma)$. The transition kernels of the two processes are correlated through the matrix Σ .

Joint Transition Models for Multivariate Longitudinal Data

When there are more than two longitudinal outcomes of different types (continuous, binary, count, etc.), joint models can be specified within the framework of generalized linear mixed models in a way similar to the case of two outcomes. As before, let x_{ijk} be a $p_k \times 1$ vector and let $H_{ijk}^{(s)} = (Y_{i,j-1,k}, Y_{i,j-2,k}, \dots, Y_{i,j-s,k})'$ be the history of past s response observations and past s covariates for participant i at t_j from the k th process. An additional notation $H_{ij}^{(s)}$ is now introduced to denote the collection of past history for all K processes: $H_{ij}^{(s)} = (H_{ij1}^{(s)}, H_{ij2}^{(s)}, \dots, H_{ijK}^{(s)})'$. Conditional on random effects b_{ik} and past history $H_{ij}^{(s)}$, the current state Y_{ijk} is assumed to be independent across participants and different responses, with a density function of the form $f(y_{ijk}|H_{ij}^{(s)}, b_{ik}) = \exp\{[y_{ijk}\xi_{ijk} - \psi_k(\xi_{ijk})] / \phi_k + c_k(y_{ijk}, \phi_k)\}$, where ψ_k , ϕ_k and c_k are standard quantities in exponential families. As a standard result for exponential families, we obtain the conditional mean $\mu_{ijk} = E(Y_{ijk}|H_{ij}^{(s)}, b_{ik}) = \psi'_k(\xi_{ijk})$ and the conditional

variance $\mu_{ijk} = \text{var}(Y_{ijk}|H_{ij}^{(s)}, b_{ik}) = \psi_k''(\xi_{ijk})\phi_k$. The process-wise transition model is given by

$$h_k(\mu_{ijk}) = x'_{ijk}\beta_k + f_k(H_{ij}^{(s)}; \alpha_k) + Z'_{ijk}b_{ik}, \quad k = 1, 2, \dots, K, \quad (6)$$

for some link function h_k . In (6), x_{ijk} , Z_{ijk} , β_k , and α_k are defined analogously to in (3) and (4). It is worth noting that unlike models (3) and (4), model (6) allows the current state of each individual process to depend on the previous s states of all processes involved (i.e., f_k is a function of $H_{ij}^{(s)}$, not just $H_{ij}^{(1)}$). The extra generality translates into added flexibility with respect to the correlation between longitudinal processes. The joint modeling of Y_{ijk} ’s is finalized by assuming the random effect vectors b_{ik} to be correlated as follows:

$$(b_{i1}, b_{i2}, \dots, b_{iK})' \sim N(0, \Sigma_K). \quad (7)$$

The general formulation (6) and (7) for joint modeling of longitudinal profiles through random-effect transition models provides a general framework for studying complex transition properties for multiple outcomes which are correlated and possibly interact with each other.

Statistical Inference: Composite Conditional Likelihood

In joint random-effects transition models for multivariate longitudinal data, statistical inference is challenging for several reasons. In (6) and (7), denote by $Y_i = (Y'_{i1}, Y'_{i2}, \dots, Y'_{iK})'$ the collection of longitudinal data from the i th participant (in the Uniform Data Set of Alzheimer’s disease, $K = 4$). Let $L(\theta|y) = \prod_{i=1}^I L(\theta|y_i)$ be the full likelihood based on (6) and (7), where θ denotes the vector of all unknown parameters (including β_k ’s, α_k ’s and the parameters in the variance components). For the s th-order joint transition models, the full likelihood is

$$L(\theta|y) = \prod_{i=1}^I \int \prod_{k=1}^K \left[f(y_{i1k}, \dots, y_{iik}|b_i) \prod_{j=s+1}^J f(y_{ijk}|H_{ijk}^{(s)}, b_i) \right] f_b(b_i) db_i, \quad (8)$$

where f_b is the mean-zero multivariate normal density of the random effects b_i . There are two challenges in evaluating the likelihood (8). First, (6) only specifies the conditional distribution $f(y_{ijk}|H_{ijk}^{(s)}, b_i)$ for $j \geq s+1$, but the conditional distribution of the first s observations $f(y_{i1k}, \dots, y_{iik}|b_i)$ is not determined by (6). As a consequence, the full likelihood (8) is generally not available. Second, the number of stochastic processes, K , determines the complexity of integration in (8) (i.e., the dimension of integration needed to evaluate $L(\theta|y)$). For joint transition models with $K = 2$, direct evaluation of the full likelihood is possible. However, when $K \geq 3$, the dimension of integration in evaluating $L(\theta|y)$ increases dramatically with K , and therefore direct evaluation becomes computationally infeasible.

Here, we construct a well-defined composite conditional likelihood function to overcome the difficulties in evaluating (8). A composite likelihood function is an inference function derived by multiplying a collection of weighted conditional or marginal likelihoods [25,26]. Composite likelihood is sometimes referred to as pseudo-likelihood or quasi-likelihood, although slight differences exist in some contexts. The advantage of using composite likelihood is that it reduces the integration complexity (relative to the full likelihood), while maintaining the consistency and

asymptotic normality of parameter estimates. Composite likelihood can be a composite of conditional likelihood density functions [24,26], which have been used extensively in univariate transition models (see Diggle et al. [24], Chapter 10) as a way to avoid specification of the marginal distribution of the initial states in a process. In the present context, the conditional likelihood $L(\theta|y) = \prod_{i=1}^I f(y_{i,s+1,\bullet}, \dots, y_{iJK}|H_{i,s+1,\bullet}^{(s)}) = \prod_{i=1}^I \int \prod_{k=1}^K \left[\prod_{j=s+1}^J f(y_{ijk}|H_{ijk}^{(s)}, b_i) \right] f_b(b_i) db_i$, can be used for inference as an alternative to the full likelihood (8). Here, $y_{i,j,\bullet} = (y_{i,j,1}, \dots, y_{i,j,K})'$ and $H_{i,j,\bullet}^{(s)} = (H_{i,j,1}^{(s)}, \dots, H_{i,j,K}^{(s)})$. Although this conditional likelihood avoids specifying the marginal distributions of the initial s states, it still has the same dimension of integration as the full likelihood (8). To deal with the latter problem, we adopt the pairwise likelihood approach of Fieuwis and Verbeke [27] and Faes et al. [28], who proposed a composite marginal likelihood constructed from the marginal likelihood functions based on outcome pairs. The foregoing discussion leads to the following composite conditional likelihood:

$$L^c(\theta|y) = \prod_{i=1}^I \prod_{k_1=1}^{K-1} \prod_{k_2=k_1+1}^K f(y_{i,s+1,k_1}, \dots, y_{iik_2}, y_{i,s+1,k_2}, \dots, y_{iik_1} | H_{i,s+1,k_1}^{(s)}, H_{i,s+1,k_2}^{(s)}) \int \prod_{j=s+1}^J \left[f(y_{ijk_1}|H_{ijk_1}^{(s)}, b_{ik_1}) f(y_{ijk_2}|H_{ijk_2}^{(s)}, b_{ik_2}) \right] f_b(b_{ik_1}, b_{ik_2}) db_{ik_1} db_{ik_2}. \quad (9)$$

The composite conditional likelihood (9) focuses on the pairwise conditional likelihood, which avoids high-dimensional integration and specification of the distribution of the first s states. Instead of evaluating the high-dimensional integral in the full likelihood (8), only $L^c(\theta|y_{ik_1}, y_{ik_2})$ needs to be evaluated, thereby reducing the dimension of integration. In addition, the pairwise conditional likelihood $L^c(\theta|y_{ik_1}, y_{ik_2})$ does not require the specification of the joint distribution of the first s states of the processes.

Inference on the unknown parameters θ can be based on existing asymptotic results for composite likelihoods [25,26,29]. The maximum composite likelihood estimator $\hat{\theta}_I^c$ can be found by solving the composite score function $\nabla_{\theta} \ell^c(\theta; y_i)$, where $\ell^c(\theta; y_i) = \sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^K \log L^c(\theta|y_{ik_1}, y_{ik_2})$.

Remark 1: Under regularity conditions, the composite maximum likelihood estimator $\hat{\theta}_I^c$ is consistent:

$$\hat{\theta}_I^c \rightarrow \theta$$

as $I \rightarrow \infty$. A central limit theorem for the composite likelihood score statistic holds, which implies that $\hat{\theta}_I^c$ is asymptotically normally distributed:

$$\sqrt{I}(\hat{\theta}_I^c - \theta) \xrightarrow{D} N(0, B(\theta)^{-1} A(\theta) B(\theta)^{-1}),$$

where

$$B(\theta) = E\{-\nabla_{\theta}^2 \ell^c(\theta; y_i)\} = - \sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^K \nabla_{\theta}^2 \log L^c(\theta|y_{ik_1}, y_{ik_2}),$$

Table 1. Analysis results of the Uniform Data Set in the National Alzheimer’s Coordinating Center (four measurements): parameter estimates, sandwich estimators of variance (standard deviations) and covariance, and the *p*-values of Wald tests.

Measurement	Parameter	Estimate	SE	<i>p</i> -value	
CDRGLOB (<i>k</i> = 1)	θ_{01} Intercept	5.565	0.668	<0.001	
	γ_{11} Increment	3.933	0.197	<0.001	
	γ_{21} Increment	4.737	0.319	<0.001	
	γ_{31} Increment	8.539	0.738	<0.001	
	β_{11} Age	−0.039	0.007	<0.001	
	β_{21} Education	0.031	0.019	0.107	
	β_{31} Time	−0.045	0.038	0.235	
	β_{41} 1 APOE	−0.737	0.111	<0.001	
	β_{51} 2 APOE	−1.251	0.261	<0.001	
	α_1 Lag (Coef. of $Y_{i,j-1,1}$)	−8.866	0.359	<0.001	
	FAQ (<i>k</i> = 2)	θ_{02} Intercept	7.063	1.099	<0.001
		γ_{12} Increment	1.645	0.207	<0.001
		γ_{22} Increment	2.141	0.289	<0.001
		γ_{32} Increment	4.112	0.729	<0.001
β_{12} Age		−0.049	0.011	<0.001	
β_{22} Education		0.033	0.029	0.267	
β_{32} Time		−0.067	0.052	0.209	
β_{42} 1 APOE		−0.928	0.175	<0.001	
β_{52} 2 APOE		−1.464	0.501	0.003	
α_2 Lag (Coef. of $Y_{i,j-1,2}$)		−3.732	0.356	<0.001	
MMSE (<i>k</i> = 3)	θ_{03} Intercept	10.045	1.667	<0.001	
	β_{31} Age	−0.074	0.016	<0.001	
	β_{23} Education	0.024	0.043	0.579	
	β_{33} Time	−0.113	0.072	0.115	
	β_{43} 1 APOE	−1.717	0.255	<0.001	
	β_{53} 2 APOE	−2.682	0.485	<0.001	
	α_3 Lag (Coef. of $Y_{i,j-1,3}$)	−7.454	1.301	<0.001	
	COGSTAT (<i>k</i> = 4)	θ_{04} Intercept	5.128	0.655	<0.001
β_{14} Age		−0.047	0.008	<0.001	
β_{24} Education		0.055	0.199	0.005	
β_{34} Time		−0.071	0.033	0.029	
β_{44} 1 APOE		−0.638	0.109	<0.001	
β_{54} 2 APOE		−0.720	0.271	0.008	
α_4 Lag (Coef. of $Y_{i,j-1,4}$)		−3.396	0.441	<0.001	
Variance-covariance matrix of random effects					
Variance	σ_1^2 (CDRGLOB)	2.264	0.655		
	σ_2^2 (FAQ)	0.960	0.473		
	σ_3^2 (MMSE)	8.946	6.830		
	σ_4^2 (COGSTAT)	1.021	1.919		
Correlation	ρ_{12} (CDRGLOB, FAQ)	0.508	0.129	<0.001	
	ρ_{13} (CDRGLOB, MMSE)	0.576	0.217	0.008	

Table 1. Cont.

Measurement	Parameter	Estimate	SE	<i>p</i> -value
	ρ_{14} (CDRGLOB, COGSTAT)	0.973	0.492	0.048
	ρ_{23} (FAQ, MMSE)	0.754	0.229	<0.001
	ρ_{24} (FAQ, COGSTAT)	−0.089	0.693	0.898
	ρ_{34} (MMSE, COGSTAT)	0.211	0.917	0.818

(The increment parameters are defined as: $\gamma_{l_1,1} = \theta_{l_1,1} - \theta_{l_1,1}$, $l_1 = 1, 2, 3$; $\gamma_{12} = \theta_{0,5,2} - \theta_{02}$; $\gamma_{22} = \theta_{12} - \theta_{0,5,2}$; $\gamma_{32} = \theta_{22} - \theta_{12}$).
doi:10.1371/journal.pone.0075487.t001

is the sensitivity matrix,

$$A(\theta) = \text{var} \{ \nabla_{\theta} \ell^c(\theta; y_i) \} = E \left\{ \sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^K \nabla_{\theta} \log L^c(\theta | y_{ik_1}, y_{ik_2}) \right\} \left\{ \sum_{k_1=1}^{K-1} \sum_{k_2=k_1+1}^K \nabla_{\theta} \log L^c(\theta | y_{ik_1}, y_{ik_2}) \right\}'$$

is the variability matrix, and $G(\theta) = B(\theta)A(\theta)^{-1}B(\theta)$ is the Godambe information matrix.

The proof of Remark 1 is similar to the arguments of [25,26,29], and thus is omitted here. For testing the null hypothesis that the unknown parameters equal pre-specified values of interest, the Wald, score, and likelihood ratio test statistics can be extended to the composite conditional likelihood using existing results [26,29,30].

The joint modeling approach with the composite conditional likelihood is complicated, possibly involving mixed types of binary or multi-categorical outcomes. Although the composite likelihood asymptotic theories in Remark 1 ensure asymptotic validity, the finite sample performance of the proposed composite likelihood estimators still needs to be thoroughly examined, especially when the outcomes are binary or multi-categorical. We conducted a simulation study to investigate the finite sample performance of $\hat{\theta}_I^c$. The simulation study was designed to mimic the Alzheimer’s disease application. The simulation results, reported in Appendix (see Supporting Information), show that the composite maximum likelihood estimator $\hat{\theta}_I^c$ performs well in realistic situations similar to Alzheimer’s disease application. In addition, we compared the composite maximum likelihood estimates with estimates derived from a two-stage maximum likelihood approach.

Results

Two of the four primary measurements of Alzheimer’s disease, MMSE and COGSTAT, are dichotomous and can be handled using model (3). The other two measurements, CDRGLOB and FAQ, are both on a five-level ordinal scale. To accommodate the ordinal nature of the latter two measurements, the following random-effects transition model is constructed. For a fixed *k*, consider an ordinal longitudinal outcome $Y_{ik} = (Y_{i1k}, Y_{i2k}, \dots, Y_{iLk})$ in which Y_{ijk} , representing the state at time *j*, has *L_k* ordinal levels. Covariates and past states are linked with the

probability of the event $\{Y_{ijk} \leq l\}$ through a random-effects cumulative logit transition model:

$$\text{logit}[P(Y_{ijk} \leq l | H_{ijk}^{(s)}, b_{ik})] = \theta_{lk} + x'_{ijk} \beta_k + f_k(H_{ijk}^{(s)}; \alpha_k) + Z'_{ijk} b_{ik} \quad (10)$$

where $\{\theta_{lk}\}$ are increasing in l . In the linear predictor, the fixed-effects components $x'_{ijk} \beta_k$, random-effects component $Z'_{ijk} b_{ik}$, and past states component $f_k(H_{ijk}^{(s)}; \alpha_k)$ are the same as in (3) and (4), except that each cumulative log odds $\text{logit}(P(Y_{ijk} \leq l | H_{ijk}^{(s)}))$ has its own intercept. Each of the four primary measurements Y_{ijk} ($k=1$ for CDRGLOB; $k=2$ for FAQ; $k=3$ for MMSE; $k=4$ for COGSTAT) can be viewed as an ordinal variable with either five levels or two levels ranging from no sign of symptoms to severe dementia. Therefore, in the following analysis, we used (10) as the random-effects cumulative logit transition model for all four measurements. Precise definitions of ordinal levels have been provided in the Methods section. The vector of covariates x_{ijk} is assumed, for all four measurements, to consist of (1) participant's age at initial visit, (2) years of education, (3) clinical visit times, and (4) two dummy variables indicating 1 and 2 copies of APOE-ε4, which has been suggested as possibly associated with Alzheimer's disease. Finally, the dependence between the four outcomes is characterized by a multivariate normal distribution of $(b_{i1}, b_{i2}, b_{i4}, b_{i4})'$, as indicated earlier.

A cautionary note on the usage of (10) is as follows. Investigators have to be very cautious about the specification of $f_k(H_{ij}^{(s)}; \alpha_k)$ and Z_{ijk} . The form of $f_k(H_{ij}^{(s)}; \alpha_k)$ represents the dependence structure of current Alzheimer's disease status on the past, which can include the patients' disease status from only the last clinical visit (i.e., $Y_{i,j-1,k}$) or the last s clinical visits (i.e., $Y_{i,j-1,k}, \dots, Y_{i,j-s,k}$). The functional form of $f_k(H_{ij}^{(s)}; \alpha_k)$ can also vary beyond a linear or additive form. These two aspects have direct impacts on the specification of the random-effects covariate vector Z_{ijk} , which may consist of any fixed effects in x_{ijk} or past disease statuses $Y_{i,j-1,k}, \dots, Y_{i,j-s,k}$. A large number of covariates in Z_{ijk} will induce high dimensionality in the normal distribution of random effects, which may not be necessary. In summary, it is important to specify both $f_k(H_{ij}^{(s)}; \alpha_k)$ and Z_{ijk} carefully. In our analysis, we first included the last two clinical visits of each patient in $f_k(H_{ij}^{(s)}; \alpha_k)$ in a linear fashion, in addition to the aforementioned four covariates. Each components of $f_k(H_{ij}^{(s)}; \alpha_k)$ and x_{ijk} was considered as a possible component of Z_{ijk} , using likelihood ratio tests based on composite likelihood to eliminate unnecessary random effects. This procedure eventually led to the following first-order transition model with a random effect for $Y_{i,j-1,k}$ only:

$$\text{logit} P(Y_{ijk} \leq l | Y_{i,j-1,k}, b_{ik}) = \theta_{lk} + x'_{ijk} \beta_k + (\alpha_k + b_{ik}) Y_{i,j-1,k} \quad (11)$$

Our statistical analysis was based on the composite conditional likelihood derived in the Methods section. The SAS programs for this analysis are available upon request. Table 1 presents parameter estimates and (sandwich) standard errors (for both the regression coefficients and the variance-covariance parameters), together with p -values of Wald tests. The first-order Markov effect is remarkable for each primary measurement of Alzheimer's disease, with a high level of statistical significance ($p < 0.001$) for each α_k , which quantifies the impact of the previous stage disease status $Y_{i,j-1,k}$. The estimated α_k is negative and large in

Table 2. Analysis results of the Uniform Data Set in the National Alzheimer's Coordinating Center (three measurements): parameter estimates, sandwich estimators of variance (standard deviations) and covariance, and the p -values of Wald tests.

Response	Parameter	Estimate	SE	p -value
CDRGLOB ($k=1$)	θ_{01} Intercept	5.565	0.517	< 0.001
	γ_{11} Increment	3.932	0.202	< 0.001
	γ_{21} Increment	4.737	0.325	< 0.001
	γ_{31} Increment	8.539	0.744	< 0.001
	β_{11} Age	-0.036	0.007	< 0.001
	β_{41} 1 APOE	-0.737	0.109	< 0.001
	β_{51} 2 APOE	-1.251	0.251	< 0.001
	α_1 Lag (Coef. of $Y_{i,j-1,1}$)	-8.866	0.370	< 0.001
FAQ ($k=2$)	θ_{02} Intercept	7.063	0.858	< 0.001
	γ_{12} Increment	1.645	0.199	< 0.001
	γ_{22} Increment	2.141	0.279	< 0.001
	γ_{32} Increment	4.112	0.646	< 0.001
	β_{12} Age	-0.046	0.011	< 0.001
	β_{42} 1 APOE	-0.928	0.174	< 0.001
	β_{52} 2 APOE	-1.464	0.466	0.002
	α_2 Lag (Coef. of $Y_{i,j-1,2}$)	-3.732	0.326	< 0.001
MMSE ($k=3$)	θ_{03} Intercept	10.045	1.213	< 0.001
	β_{13} Age	-0.075	0.015	< 0.001
	β_{43} 1 APOE	-1.717	0.235	< 0.001
	β_{53} 2 APOE	-2.682	0.459	< 0.001
	α_3 Lag (Coef. of $Y_{i,j-1,3}$)	-7.454	0.849	< 0.001
Variance-covariance matrix of random effects				
Variance	σ_1^2 (CDRGLOB)	2.264	0.667	
	σ_2^2 (FAQ)	0.960	0.382	
	σ_3^2 (MMSE)	8.946	3.690	
Correlation	ρ_{12} (CDRGLOB, FAQ)	0.508	0.131	< 0.001
	ρ_{13} (CDRGLOB, MMSE)	0.576	0.227	0.011
	ρ_{23} (FAQ, MMSE)	0.754	0.239	0.001

(The increment parameters are defined as: $\gamma_{l_1,1} = \theta_{l_1,1} - \theta_{l_1,1}$, $l_1 = 1, 2, 3$; $\gamma_{12} = \theta_{0,5,2} - \theta_{0,2}$; $\gamma_{22} = \theta_{12} - \theta_{0,5,2}$; $\gamma_{32} = \theta_{22} - \theta_{12}$).
doi:10.1371/journal.pone.0075487.t002

magnitude for each measurement ($\hat{\alpha}_1 = -8.866$; $\hat{\alpha}_2 = -3.732$; $\hat{\alpha}_3 = 7.454$; $\hat{\alpha}_4 = -3.396$), indicating that patients with worse Alzheimer's disease conditions in the previous state are more likely to deteriorate. As expected, a patient's age of first-time confirmed diagnosis of Alzheimer's disease plays a significant role for all four measurements. The regression coefficient for age are generally negative and highly significant ($\hat{\beta}_{11} = -0.039$, p -value < 0.001; $\hat{\beta}_{21} = -0.049$, p -value < 0.001; $\hat{\beta}_{31} = -0.074$, p -value < 0.001; $\hat{\beta}_{41} = -0.047$, p -value < 0.001), indicating that older patients are more likely to transit into a worse state of Alzheimer's disease (as characterized by the four measurements) at the next clinical visit. This is not only consistent with our previous conjecture, but also corroborates the fact that the cognitive domains were selected for

Table 3. The estimated transition probability matrix of clinical dementia rating (CDRGLob) in Alzheimer's disease patients at the age of 55 or 85 and carrying no, one, or two APOE-ε 4 alleles.

		Y_{ijt}					
Age	ε 4 alleles	$Y_{ij-1,1}$	0	0.5	1	2	3
55	0	0	0.9740	0.0254	0.0005	0.0000	0.0000
	1	0	0.9473	0.0517	0.0011	0.0000	0.0000
	2	0	0.9148	0.0833	0.0018	0.0000	0.0000
85	0	0	0.9287	0.0698	0.0015	0.0000	0.0000
	1	0	0.8618	0.1351	0.0031	0.0000	0.0000
	2	0	0.7885	0.2062	0.0052	0.0000	0.0000
55	0	0.5	0.3084	0.6495	0.0417	0.0004	0.0000
	1	0.5	0.1758	0.7400	0.0833	0.0008	0.0000
	2	0.5	0.1132	0.7537	0.1318	0.0013	0.0000
85	0	0.5	0.1340	0.7536	0.1113	0.0011	0.0000
	1	0.5	0.0689	0.7218	0.2070	0.0023	0.0000
	2	0.5	0.0424	0.6509	0.3028	0.0039	0.0000
55	0	1	0.0053	0.2075	0.7558	0.0314	0.0000
	1	1	0.0025	0.1120	0.8220	0.0635	0.0000
	2	1	0.0015	0.0703	0.8264	0.1018	0.0000
85	0	1	0.0018	0.0839	0.8288	0.0854	0.0000
	1	1	0.0009	0.0421	0.7937	0.1633	0.0000
	2	1	0.0005	0.0256	0.7278	0.2460	0.0001
55	0	2	0.0000	0.0000	0.0043	0.9526	0.0430
	1	2	0.0000	0.0000	0.0021	0.9120	0.0859
	2	2	0.0000	0.0000	0.0012	0.8630	0.1358
85	0	2	0.0000	0.0000	0.0015	0.8838	0.1147
	1	2	0.0000	0.0000	0.0007	0.7862	0.2131
	2	2	0.0000	0.0000	0.0004	0.6879	0.3116
55	0	3	0.0000	0.0000	0.0000	0.0031	0.9969
	1	3	0.0000	0.0000	0.0000	0.0015	0.9985
	2	3	0.0000	0.0000	0.0000	0.0009	0.9991
85	0	3	0.0000	0.0000	0.0000	0.0011	0.9989
	1	3	0.0000	0.0000	0.0000	0.0005	0.9995
	2	3	0.0000	0.0000	0.0000	0.0003	0.9997

doi:10.1371/journal.pone.0075487.t003

their sensitivity to age-related changes in cognition in neuropsychologic and mental evaluations [31]. Another important factor in the transitional patterns of Alzheimer's disease is the number of copies of APOE-ε4 allele, again with a high level of statistical significance (p -value < 0.008 for all β_{k4} 's and β_{k5} 's). The results indicate that Alzheimer's patients with the APOE-ε4 allele have higher risks for worse disease symptoms and evaluations than those without the allele in the same current disease situation, and the risk increase is larger for patients with two copies of APOE-ε4 than those with only one copy. In Table 1, the effect of education level on the transition of Alzheimer's disease is statistically non-significant for all outcomes but COGSTAT, even though education level is generally considered a prognostic factor that affects the overall cognitive function in Alzheimer's patients. We attempted to investigate the effect of clinical visit times on the transition of Alzheimer's disease, which may be relevant for

Table 4. The estimated transition probability matrix of functional activities (FAQ) in Alzheimer's disease patients at the age of 55 or 85 and carrying no, one, or two APOE-ε 4 alleles.

		Y_{ijt}					
Age	ε4 alleles	$Y_{ij-1,2}$	0	1	2	3	4
55	0	0	0.9896	0.0084	0.0018	0.0002	0.0000
	1	0	0.9740	0.0209	0.0045	0.0006	0.0000
	2	0	0.9564	0.0349	0.0077	0.0010	0.0000
85	0	0	0.9601	0.0319	0.0070	0.0009	0.0000
	1	0	0.9049	0.0752	0.0175	0.0023	0.0000
	2	0	0.8478	0.1187	0.0295	0.0040	0.0001
55	0	1	0.6939	0.2276	0.0686	0.0097	0.0002
	1	1	0.4727	0.3501	0.1525	0.0243	0.0004
	2	1	0.3441	0.3869	0.2276	0.0408	0.0007
85	0	1	0.3655	0.3834	0.2131	0.0373	0.0006
	1	1	0.1855	0.3557	0.3681	0.0890	0.0016
	2	1	0.1176	0.2908	0.4461	0.1427	0.0028
55	0	2	0.0515	0.1679	0.4857	0.2881	0.0068
	1	2	0.0210	0.0790	0.3860	0.4970	0.0170
	2	2	0.0124	0.0487	0.2951	0.6151	0.0287
85	0	2	0.0136	0.0531	0.3113	0.5958	0.0262
	1	2	0.0054	0.0220	0.1662	0.7425	0.0638
	2	2	0.0032	0.0131	0.1070	0.7724	0.1043
55	0	3	0.0013	0.0054	0.0475	0.7234	0.2224
	1	3	0.0005	0.0021	0.0195	0.5581	0.4198
	2	3	0.0003	0.0013	0.0115	0.4340	0.5529
85	0	3	0.0003	0.0014	0.0126	0.4560	0.5296
	1	3	0.0001	0.0005	0.0050	0.2542	0.7401
	2	3	0.0001	0.0003	0.0030	0.1671	0.8296
55	0	4	0.0000	0.0001	0.0012	0.0758	0.9228
	0	4	0.0000	0.0001	0.0005	0.0315	0.9680
	0	4	0.0000	0.0000	0.0003	0.0187	0.9810
85	0	4	0.0000	0.0000	0.0003	0.0205	0.9792
	0	4	0.0000	0.0000	0.0001	0.0082	0.9917
	0	4	0.0000	0.0000	0.0001	0.0048	0.9951

doi:10.1371/journal.pone.0075487.t004

monitoring annual changes of a patient's condition. However, this effect is not significant in Table 1, suggesting that much of the observed temporal pattern can be explained by the previous disease status. In light of these results, education level and clinical visit time were both removed from the list of covariates in the second part of the analysis (see Table 2).

The lower portion of Table 1 presents estimates of the variance-covariance parameters, which have twofold implications. The correlation estimates of random effects characterize the dependence between four primary measurements of Alzheimer's disease. The significant positive correlations exist between the transitional probabilities of CDRGLob, FAQ and MMSE (in Table 1, $\hat{\rho}_{12} = 0.508$, p -value < 0.001; $\hat{\rho}_{13} = 0.576$, p -value = 0.008; $\hat{\rho}_{23} = 0.754$, p -value < 0.001). The correlation coefficients related to COGSTAT were all estimated with insignificant or nearly insignificant p -values (in Table 1, $\hat{\rho}_{14} = 0.973$, p -value = 0.048;

Table 5. The estimated transition probability matrix of binary cognitive status measured by MMSE in Alzheimer's disease patients at the age of 55 or 85 and carrying no, one, or two APOE-ε 4 alleles.

Age	ε 4 alleles	Y _{ij-1,3}	Y _{ij3}	
			0	1
55	0	0	0.9973	0.0027
	1	0	0.9852	0.0148
	2	0	0.9620	0.0380
85	0	0	0.9749	0.0251
	1	0	0.8745	0.1255
	2	0	0.7264	0.2736
55	0	1	0.1764	0.8236
	1	1	0.0370	0.9630
	2	1	0.0144	0.9856
85	0	1	0.0220	0.9780
	1	1	0.0040	0.9960
	2	1	0.0015	0.9985

doi:10.1371/journal.pone.0075487.t005

$\hat{p}_{24} = -0.089$, p -value = 0.898; $\hat{p}_{34} = 0.211$, p -value = 0.818). This indicates, when investigating the transitional patterns of Alzheimer's disease, the transition of COGSTAT is not correlated with other three measurements. We therefore excluded COGSTAT from the joint random-effects transition model in the second part of analysis (see Table 2), although this conclusion needs further validation and might be limited to current study cohort. Table 2 reports the results of the second part of refined analysis, which was done by fitting a joint random-effect transition model for CDRGLOB, FAQ and MMSE without patients' clinical visit times and years of education. Comparing with Table 1, the changes in parameter estimates are negligible; some standard errors have shifted in a small amount; and all parameter estimates are highly significant.

Tables 3, 4, 5 and 6 present the matrices of estimated transition probabilities for the four primary measurements, for Alzheimer's patients who are 55 or 85 years old, carrying no, one, or two APOE-ε4 alleles, with subject-specific random effects set to zero (corresponding to the median of the population). The results in these tables confirm our previous observations on the effects of age and APOE-ε4 allele number on the transitional patterns of Alzheimer's patients. Almost all (more than 99.6%) of the patients who are classified as severe dementia by CDRGLOB will stay that way in the next year's classification. The patients who are classified as in normal condition without dementia will be diagnosed to have Alzheimer's disease in the next year with a probability ranging from 2.6% (Age = 55, no APOE-ε4 allele) to 21.1% (Age = 85, two APOE-ε4 alleles), depending on the patient's age and APOE-ε4 allele number. The transition patterns among patients with very mild ($Y_{ij-1,1} = 0.5$) and mild ($Y_{ij-1,1} = 1$) dementia are diverse. The subgroup at age 55 with no APOE-ε4 allele regresses from very mild to normal condition with a probability of 30.8%, while the subgroup at age 85 with two APOE-ε4 alleles regresses with a probability of only 4.2%. The transition probabilities from moderate to severe dementia as measured by CDRGLOB range from 4.3% (Age = 55, no APOE-ε4 allele) to 31.1% (Age = 85, two APOE-ε4 alleles). The transitional patterns of functional activities evaluated by FAQ are similar to those for CDRGLOB among

Table 6. The estimated transition probability matrix of binary cognitive status measured by COGSTAT in Alzheimer's disease patients at the age of 55 or 85, having 12 or 20 years of education, and carrying no, one, or two APOE-ε 4 alleles.

Education	Age	ε 4 alleles	Y _{ij-1,4}	Y _{ij4}	
				0	1
12	55	0	0	0.9464	0.0536
		1	0	0.9036	0.0964
		2	0	0.8958	0.1042
	85	0	0	0.8311	0.1689
		1	0	0.7232	0.2768
		2	0	0.7055	0.2945
20	55	0	1	0.4312	0.5688
		1	1	0.2870	0.7130
		2	1	0.2696	0.7304
	85	0	1	0.1744	0.8256
		1	1	0.1009	0.8991
		2	1	0.0932	0.9068
20	55	0	0	0.9647	0.0353
		1	0	0.9354	0.0646
		2	0	0.9300	0.0700
		0	0	0.8828	0.1162
		1	0	0.8015	0.1985
		2	0	0.7873	0.2127
	85	0	1	0.5395	0.4605
		1	1	0.3835	0.6165
		2	1	0.3632	0.6368
		0	1	0.2461	0.7539
		1	1	0.1477	0.8523
		2	1	0.1371	0.8629

doi:10.1371/journal.pone.0075487.t006

normal patients ($Y_{ij-1,2} = 0$) and completely disable patients, but substantially different for patients with mildly difficult ($Y_{ij-1,2} = 1$), moderate ($Y_{ij-1,2} = 2$), and serious ($Y_{ij-1,2} = 3$) disability. For example, the subgroup at age 55 with no APOE-ε4 allele regresses from mildly difficult functional ability to normal condition with an amazing probability of 69.4%, while patients at age 85 with two APOE-ε4 alleles regress with a probability of only 11.8%. The transition probabilities for MMSE, shown in Table 5, depend heavily on age and APOE-ε4 allele number, with deterioration probabilities ranging from 0.3% to 27.4% and regression probabilities ranging from 0.2% to 22.0%. The transition probabilities for COGSTAT, shown in Table 6, are further stratified by two levels of education: 12 years and 20 years. There appears to be a small effect of education level on the transitional patterns, but the largest effects are still due to age and APOE-ε4 allele number. One of the benefits of our joint modeling approach is that the subject-specific random effects can be used for prediction. Once empirical Bayes estimates of the random effects are obtained for a given patient, the transition matrices in Tables 3–6 can be reproduced for this particular patient. This information can help clinicians predict future disease status of this patient.

Discussion

A new study suggests that the number of people diagnosed with Alzheimer's disease in the United States will triple over the next 40 years [32]. In 2010, 4.7 million people had the disease. By 2050, researchers predict that the number will grow to 13.8 million [32]. Even with this intimating number of prediction, the cause and progression of Alzheimer's disease are still not well understood. Without an effective treatment available for Alzheimer's disease, it is important to understand the progression patterns of the disease. Transitional probabilities based on valid statistical models can help predict disease progression for individual patients. In this research article, we proposed to use joint random-effects multivariate transition models to deliberate the transitional trajectories of Alzheimer's disease in terms of four major measurements. This is the first research article, to our knowledge, that discusses the transition of Alzheimer's disease in a multivariate fashion. The benefits of this methodology includes (i) accounting for the association between measurements, (ii) identifying factors that affect the transition of the disease, and (iii) targeting on the subject-specific transition patterns. The composite conditional likelihood approach makes it feasible to estimate these models.

The joint random-effects transition models have complex structure with several components. There are important and challenging questions, such as how to check the key assumptions. First, statistical tests can be used to compare two nests models. This can help with checking the proportional odds assumption for ordinal responses. Molenberghs and Verbeke [29] described the details on how to conduct likelihood ratio tests for the composite likelihood. Second, model selection criteria based on composite likelihood can be used to compare non-nested models. For instance, Varin and Vidoni [33] proposed an integrated and general information criterion for model selection based on composite likelihood. Gao and Song [34] proposed a composite likelihood version of the Bayes information criterion and established its consistency property for the selection. Both criteria can be adopted to conduct model selection in the joint

References

- Berchthold NC, Cotman CW (1998) Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-Roman period to the 1960s. *Neurobiology of Aging* 19: 173–189.
- Alzheimer's Association (2012) 2012 Alzheimer's disease facts and figures. *Alzheimer's and Dementia* 8: 131–168.
- Morris JC, Weintraub S, Chui HC, Cummings J, DeCarli C, et al. (2006) The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. *Alzheimer Disease & Associated Disorders* 20: 210–216.
- Beekly DL, Ramos EM, Lee WW, Deitric WD, Jacka ME, et al. (2007) The National Alzheimer's Coordinating Center (NACC) Database: The Uniform Data Set. *Alzheimer disease and associated disorders* 21: 249–258.
- Zeger SL, Qaqish B (1988) Markov regression models for time-series: A quasi-likelihood approach. *Biometrics* 44: 1019–1031.
- Wong WH (1986) Theory of partial likelihood. *Annals of Statistics* 14: 88–123.
- Albert PS (2000) A transitional model for longitudinal binary data participant to nonignorable missing data. *Biometrics* 56: 602–608.
- Lee K, Daniels MJ (2007) A class of Markov models for longitudinal ordinal data. *Biometrics* 63: 1060–1067.
- Albert PS, Follmann DA (2003) A random effects transition model for longitudinal binary data with informative missingness. *Statistica Neerlandica* 57: 100–111.
- Zeng L, Cook RJ (2007) Transition models for multivariate longitudinal binary data. *Journal of the American Statistical Association* 102: 211–223.
- Ilk O, Daniels MJ (2007) Marginalized transition random effect models for multivariate longitudinal binary data. *Canadian Journal of Statistics* 35: 105–123.
- Lee K, Joo Y, Yoo JK, Lee J (2009) Marginalized random effects models for multivariate longitudinal binary data. *Statistics in Medicine* 28: 1284–1300.
- Morris JC (1993) The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology* 43: 2412–2414.
- Morris JC (1997) Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. *International psychogeriatrics* 9 Supplement S1: 173–176.
- Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research* 12 (3): 189–198.
- Dawson JD, Anderson SW, Uc EY, Dastrup E, Rizzo M (2009) Predictors of driving safety in early Alzheimer disease. *Neurology* 72: 521–527.
- Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S (1982) Measurement of functional activities in older adults in the community. *Journal of Gerontology* 37: 323–329.
- Sloane PD, Zimmerman S, Suchindran C, Reed P, Wang L, et al. (2002) The public health impact of alzheimer's disease, 2000–2050: Potential implication of treatment advances. *Annual Review of Public Health* 23: 213–231.
- Strittmatter WJ, Saunders AM, Schmechel D, Pericak-Vance M, Enghild J, et al. (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences* 90: 1977–1981.
- Bermejo-Pareja F, Benito-León J, Vega S, Medrano MJ, Román GC (2008) Incidence and subtypes of dementia in three elderly populations of central Spain. *Journal of the Neurological Sciences* 264: 63–72.
- Koepsell TD, Kurland BF, Harel O, Johnson EA, Zhou XH, et al. (2008) Education, cognitive function, and severity of neuropathology in Alzheimer disease. *Neurology* 70: 1732–1739.
- Roe CM, Xiong C, Grant E, Miller JP, Morris JC (2008) Education and reported onset of symptoms among individuals with Alzheimer disease. *Archives of Neurology* 65: 108–111.
- Pinheiro JC, Bates DM (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer-Verlag.
- Diggle PJ, Heagerty P, Liang KY, Zeger SL (2002) *Analysis of Longitudinal Data (Second Edition)*. New York: Oxford University Press.

random-effects transition models, though Gao and Song's method is targeting high-dimensional data. Third, the joint models can be decomposed into marginal distributions (for separate outcomes) and the copula (for the dependence between different outcomes). The marginal models can be checked using existing techniques for a univariate outcome such as residual plots [35]. It appears more difficult to verify assumptions concerning the dependence structure of multiple outcomes (multivariate normal assumption for the random effects). The robustness of joint modeling of multivariate longitudinal profiles on the misspecification of random-effects distribution is an area for future research.

Supporting Information

File S1 Appendix.
(PDF)

Acknowledgments

The authors sincerely thank the Editor and three anonymous referees for their insightful and valuable comments, which have led to substantial improvement in this article. The authors acknowledge Sarah Monsell, MS, Lilah Besser, MSPH, Maggie Dean, BS, and the staff of the National Alzheimer's Coordinating Center for assistance in providing access to the database and preparation of data sets. Xiao-Hua Zhou, Ph.D., is presently a Core Investigator, Research Career Scientist (RCS OS-196), and Biostatistics Unit Director at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. Zhiwei Zhang, Ph.D., is presently a Mathematical Statistician in the Division of Biostatistics, Office of Surveillance and Biometrics, Center for Devices and Radiological Health, Food and Drug Administration. The views expressed in this article are those of the authors and do not necessarily represent the views of either the Department of Veterans Affairs or Food and Drug Administration.

Author Contributions

Conceived and designed the experiments: WL BZ ZZ XHZ. Analyzed the data: WL BZ ZZ. Wrote the paper: WL BZ ZZ XHZ.

25. Lindsay B (1988) Composite likelihood methods. *Contemporary Mathematics* 80: 220–239.
26. Varin C, Reid N, Firth D (2011) An overview of composite likelihood methods. *Statistica Sinica* 21: 5–42.
27. Fieuws S, Verbeke G (2006) Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* 62: 424–431.
28. Faes C, Aerts M, Molenberghs G, Geys H, Teuns G, et al. (2008) A high-dimensional joint model for longitudinal outcomes of different nature. *Statistics in Medicine* 27: 4408–4427.
29. Molenberghs G, Verbeke G (2005) *Models for Discrete Longitudinal Data*. New York: Springer-Verlag.
30. Geys H, Molenberghs G, Ryan L (1999) Pseudo-likelihood modeling of multivariate outcomes in developmental studies. *Journal of the American Statistical Association* 94: 734–745.
31. Weintraub S, Salmon D, Mercaldo N, Ferris S, Graff-Radford NR, et al. (2009) The Alzheimer's Disease Centers' Uniform Data Set (UDS): the neuropsychological test battery. *Alzheimer Disease & Associated Disorders* 23: 91–101.
32. Hebert LE, Weuve J, Scherr PA, Evans DA (2013) Alzheimer disease in the United States (2010–2050) estimated using the 2010 census. *Neurology*, in press.
33. Varin C, Vidoni P (2005) A Note on Composite Likelihood Inference and Model Selection. *Biometrika* 92: 519–528.
34. Gao X, Song PXX (2010) Composite likelihood Bayesian information criteria for model selection in high-dimensional data. *Journal of the American Statistical Association* 105: 1531–1540.
35. Johnson VE, Albert JH (1999) *Ordinal Data Modeling*. New York: Springer-Verlag.