Published in final edited form as:

Biom J. 2011 November ; 53(6): . doi:10.1002/bimj.200900259.

Multi-Stage Transitional Models with Random Effects and Their Application to The Einstein Aging Study

Changhong Song¹, Lynn Kuo^{2,*}, Carol A. Derby³, Richard B. Lipton³, and Charles B. Hall³

¹Hoffmann-La Roche Inc., Nutley, NJ, 07110, U.S.A

²Department of Statistics, University of Connecticut, Storrs, CT, 06269, U.S.A

³Department of Neurology, Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA

Summary

Longitudinal studies of aging often gather repeated observations of cognitive status to describe the development of dementia and to assess the influence of risk factors. Clinical progression to dementia is often conceptualized by a multistage model of several transitions that synthesizes time-varying effects. In this study, we assess the influence of risk factors on the transitions among three cognitive status: cognitive stability (normal cognition for age), memory impairment, and clinical dementia. We have developed a shared random effects model that not only links the propensity of transitions and to the probability of informative missingness due to death, but also incorporates heterogeneous transition between subjects. We evaluate four approaches using generalized logit and four using proportional odds models to the first order Markov transition probabilities as a function of covariates. Random effects were incorporated into these models to account for within-subject correlations. Data from the Einstein Aging Study are used to evaluate the goodness-of-fit of these models using the Akaike information criterion. The best fitting model for each type (generalized logit and proportional odds) is recommended and their results are discussed in more details.

Keywords

Generalized Logit; Multi-Stage; Proportional Odds; Random Effects; Transition

1 Introduction

In aging studies, longitudinal data collection and analysis are carried out to identify risk factors for age-related cognitive decline and dementia. The natural history of dementia is often conceptualized as a three stage process: a stage of relative cognitive stability (normal cognition for age), a stage of accelerated cognitive decline, and a stage of clinically diagnosable dementia (Morris *et al.*, 2001). The intermediate state of accelerated cognitive decline is often called Mild Cognitive Impairment (MCI). MCI is regarded as a state of high risk for dementia, particularly Alzheimer disease (c.f. Bennett *et al.*, 2005). Two forms of MCI are often distinguished. Amnesic MCI is characterized by memory impairment in the absence of dementia. Non-amnesic MCI is characterized by impairment in cognitive domains other than memory. Herein, we focus on amnesic intermediate states as those have

*Corresponding author: lynn@stat.uconn.edu, Phone: +001 860 486 2951, Fax: +001 860 486 4113.

Conflict of Interest

The authors declare no conflict of interest.

the strongest association with the future onset of dementia, most commonly Alzheimer's disease. Operational definitions of amnesic MCI vary widely (Gauthier *et al.*, 2006) and may include clinical complaints, informant reports, clinician impression, and cognitive test performance in various combinations. For the purpose of this manuscript, we define our memory-impaired intermediate state (MI) by performance below an empirical cut score on a specific memory test (≥ 24 on the sum of three trials of free recall test of the Buschke Grober Free and Cued Selective Reminding Test, Grober and Buschke, 1987). The third stage, clinical dementia was defined by consensus conference using standard DSM-IV criteria (American Psychiatric Association, 1994). Because death is an important cause of drop-out in this longitudinal aging study, and rates are too high to ignore, we added a fourth stage 'death without dementia' to handle the informative missingness due to death. When two adjacent data collection points are observed in a longitudinal study, eight transitions are considered among these four stages of transitions as shown in Figure 1. They are normality to normality, normality to memory impairment, normality to dementia, normality to death without dementia, memory impairment to normality, memory impairment to memory impairment, memory impairment to dementia, and memory impairment to death without dementia. Death without dementia is considered as an absorbing stage. A subject will not transit to other stages, if (s)he is in the death stage. Dementia can be considered as another absorbing stage in this kind of study, as it is extremely rare for a person diagnosed with dementia to be determined not to have dementia at a later date; it has happened only twice during the 17 years of the Einstein Aging Study. As our primary interest in this study is in risk factors for transitions between normal aging, memory impairment, and dementia, we do not model the transition between dementia and death but simply treat dementia as an absorbing stage. Note the dementia here may include death after onset of dementia. Therefore, we have denoted the fourth stage by 'death without dementia' to differentiate it from death with dementia included in the third stage. In the followings, we will call the fourth stage 'death without dementia' by 'death' for short.

From a clinical point of view, identifying risk factors correlated with these transitions is important. Emerging evidence suggests that risk factors may have different effects at particular stage of illness (Tyas *et al.*, 2007). Preventing dementia onset through risk factor modification may require an understanding of the stage of illness and the influence of particular risk factors on particular stages. Interventions to prevent future dementia may be different for a person in the stage of cognitive normality than for someone in a stage intermediate between normality and dementia.

Some methodologies have been proposed to study age-related dementia and Alzheimer's disease. They include multi-stage studies by Harezlak *et al.* (2003), Commenges *et al.* (1998), and Salazar *et al.* (2007). Harezlak *et al.* (2003) propose an illness-death stochastic model that can estimate disease incidence and mortality rates simultaneously for the analysis of longitudinal dementia data. Commenges *et al.* proposed an illness-death model for modeling the incidence and the prevalence of Alzheimer disease or dementia. Salazar *et al.* proposed a shared random effects model for the analysis of multi-stage transitions for dementia. Tyas *et al.* (2007) applied a polytomous logistic regression model to the Nun Study to address whether known risk factors for dementia increase the risk of mild cognitive impairments or of the further progress to dementia. There are also studies based on continuous outcomes. For example, Hall *et al.* (2000, 2001, 2003) developed a change point method for modeling the stages of cognitive decline. Commenges *et al.* (1998) studied age-specific risks for analyzing dementia data. Jacqmin-Gadda *et al.* (2006) jointly modeled cognitive function and dementia incidence by combining a random changepoint model with a survival model. Yu and Ghosh (2010) have extended the Jacqmin-Gadda model to include the competing risk of mortality.

We focus here on transitional models which assess the influence of risk factors on the transitions between stages of cognitive decline. For each subject, we collected longitudinal data on his/her cognitive status and related explanatory variables including gender, age, education, medical conditions, and self-reported depression status. Some of the explanatory variables can be time varying. In particular, we consider firstorder Markov transitional models where the conditional distribution of an observed outcome at the t^{th} time point given the entire history depends only on the $(t - 1)^{th}$ time point's observed outcome (including covariates).

For longitudinal categorical variables, it is important to account for the within-subject correlations. Both random effects models (Zeger and Karim, 1991; Breslow and Clayton, 1993), and marginal models (Liang and Zeger, 1986; Zeger and Liang, 1986) have been applied for this purpose. Examples of random effects models for longitudinal categorical data include the latent autoregressive model by Albert *et al.* (2002) and shared random effects model by Albert and Follmann (2003) and Salazar *et al.* (2007). Examples of marginal approaches include the two-stage Markov chain transitional model by Albert and Waclawiw (1998), and marginal regression model by Martus *et al.* (2004).

In this study, we focus on multilevel transitional models with random effects. Given the four stages, death is essentially a competing risk for the other stages. So we model death and change in cognitive function simultaneously. Basically, we use a logistic regression model with random effects to fit death or 'non-death' (the other three stages) outcomes with the same past state. Then conditioning on being alive, two types of models for the transitional distribution (called conditional models) between cognitive stages are explored: generalized logit models (Agresti, 1990, 2002) and proportional odds models (McCullagh, 1980). Generalized logit models are extensions of logistic regression models when the binary response is changed to a multi-categorical response. The models are also called polytomous logit models or multinomial logit models. Proportional odds models are applied to ordinal response data. They imply the difference between the cumulative logit for a response evaluated at two covariates is proportional to the distance between the two covariates and independent of the response category. The first generalized logit model is a fixed effects model, which does not have a latent variable. The second is a shared random effects model, which assumes a common random effect for all observations within the same subject. The third is a previous-stage dependent random effects model, which has a scale parameter that measures the subject's tendency to stay at the normal stage or the MI stage and adjust for the variance of the random effects depending on the previous stage. The fourth model is a general random effects model where each subject has a multivariate vector of random effects with individual components depending on the cognitive status of the previous time. A multivariate normal distribution is used for this vector of random effects. The third and fourth models are more flexible than the first two. Proportional odds models are defined similarly. The generalized logit models are expected to be more flexible in general. While the proportional odds models make stronger assumptions on the cumulative logit, they can be more powerful than the generalized logit models because fewer parameters need to be estimated. Both have potential to be useful in practice and a major purpose of this work is to compare the two approaches.

We apply our methodology to the Einstein Aging Study (EAS). EAS study design and methods are described elsewhere (Lipton *et al.*, 2003). Briefly, potential participants aged 70 and older were recruited through systematic sampling from Medicare or voter registration lists for Bronx County, New York, beginning in 1993. Participants were excluded who reported severe sensory loss or medical conditions (e.g., significant tremor, arthritis) that would interfere with completion of a neuropsychological assessment, or were non-English speakers, or who were institutionalized. At baseline and annually, participants visit the EAS

clinic and undergo standardized assessments which include neuropsychological evaluation, neurological examination, medical history, physical measures (blood pressure, anthropometrics, fasting blood sampling), demographic and behavioral questionnaires. Dementia diagnosis is determined by consensus conference using DSM-IV criteria (American Psychiatric Association 1994). The data are from 1682 participants recruited between 1993 and 2007. The Einstein Aging Study was approved by the institutional review board at the Albert Einstein College of Medicine of Yeshiva University.

Among the conditional models, we wished to evaluate whether random effects models dependent on previous stage are superior to the more commonly used shared random effects transitional models for explaining heterogeneity among transitions within subject. Models were tested in the EAS cohort using a list of covariates which have been associated with dementia risk. In particular, we focus on *six* covariates of interests: gender, years of education, age, and self reported history of myocardial infarction, stroke, and diabetes status, each of which has been shown in at least one study to be associated with dementia (Launer, 2002; Stampfer, 2006). When we fit the full models, we include main effects of all seven covariates (including the intercept). For ease of interpretation, no interaction effects are considered. Model goodness-of-fit is evaluated using the Akaike information criterion (AIC) given by Akaike (1973, 1974).

Section 2 describes the models. Sections 3 and 4 address methods of parameter estimation and model selection. Section 5 describes the results. Section 6 contains concluding comments.

2 Transitional Models

Let Y_{it} , $t = 1, \dots, T$ denote the outcome status of the i^{th} subject at the t^{th} time with J possible stages 1, 2, ..., J , where the last one, J , is an absorbing (death) stage. For brevity, we define the history for the i^{th} subject at the t^{th} time to be $H_{it} = \{(Y_{ik}), k = 1, \dots, t-1\}$. Therefore the joint probability mass function of the longitudinal data collected at n_i time points for the i^{th} subject is given by

$$p_i(Y_{i1}, \dots, Y_{in_i}) = p_{i1}(Y_{i1}) \prod_{t=2}^{n_i} p_{it}(Y_{it}|H_{it}),$$

where p_{it} denotes the conditional probability of the response at the t^{th} time given all its history. These probabilities are allowed to depend on covariates. We assume the evolution of the stages satisfies a first order Markov chain so

$$p_{it}(Y_{it}|H_{it}) = p_{it}(Y_{it}|Y_{i,t-1}).$$

Therefore we need to define possible models depending on the covariates for these first order transitional probabilities. Suppose the i^{th} subject is at stage j at the $(t-1)^{th}$ observed epoch and moves to stage k at the t^{th} observed epoch, then we define the transitional probability from stage j to k as

$$p_{it}(k|j) = P(Y_{it}=k|Y_{i,t-1}=j, x_{i,t-1}).$$

Here $p_{it}(k|j)$ denotes the probability of being stage k conditional on previous stage j for subject i at time t . Note $x_{i,t-1}$, a time dependent covariate, denotes the vector of covariates of

the i^{th} subject collected at the previous time $(t-1)^{th}$ epoch. Essentially, we use the covariates observed in the last visit as the predictors for current cognitive status. Some investigators may prefer to use covariates observed at the present visit. We are using the covariates observed at the last visit for better predictive power and easier interpretation.

The framework above represents two different kinds of transitions. One is when $k = J$, that is when the transition is to death and then the sequence of transitions terminates. The other is for $k = J-1$, that includes all living stages. So we first model whether the outcome is a death or not using logistic regression models, and then model the transition probabilities among cognitive stages conditioning on being alive. To be specific, for each Y_{it} , we introduce indicator variable Z_{it} , where $Z_{it} = 1$ if Y_{it} is J (for death), and $Z_{it} = 0$ if $Y_{it} = J-1$ (non death). We will specify models for $p_{it}(J|j) = P(Y_{it} = J | Y_{i,t-1} = j, x_{i,t-1}) = P(Y_{it} = J | Y_{i,t-1} = j, x_{i,t-1}, Z_{it} = 1)$ using logistic regression with specific choices given in Section (2.2). So if $Z_{it} = 1$, we use this probability for death, the transition process terminates. However, if $Z_{it} = 0$, the subject may continue to make transitions. So we define the conditional transitional probability from stage j ($J-2$) to k ($J-1$) as

$$\tilde{p}_{it}(k|j) = P(Y_{it} = k | Y_{i,t-1} = j, x_{i,t-1}, Z_{it} = 0).$$

We have dropped Z from the left hand side for easier notation. Then for these j, k indices, we have

$$\begin{aligned} p_{it}(k|j) &= P(Y_{it} = k | Y_{i,t-1} = j, x_{i,t-1}, Z_{it} = 0) P(Z_{it} = 0 | Y_{i,t-1} = j, x_{i,t-1}) \\ &= \tilde{p}_{it}(k|j) P(Z_{it} = 0 | Y_{i,t-1} = j, x_{i,t-1}). \end{aligned}$$

So in the followings, we first describe the eight possible models for the conditional models $p_{it}(k|j)$ in Section 2.1. Then we describe the logistic regression models for $P(Z_{it} = 0 | Y_{i,t-1} = j, x_{i,t-1})$ in Section 2.2.

2.1 Conditional Models

Two kinds of models, generalized logit and proportional odds, can be used to model the above conditional transitional probabilities $p_{it}(k|j)$.

2.1.1 Generalized Logit Models—We propose four types of generalized logit models.

1. Fixed effects model: For the i^{th} subject with the previous stage j , the fixed effects model can be defined as

$$\log \left(\frac{\tilde{p}_{it}(k|j)}{\tilde{p}_{it}(1|j)} \right) = \alpha_{jk} + \beta_{jk}^T x_{i,t-1}, \quad k=2, \dots, J-1, \quad (1)$$

where we use the probability of transition to the first category $p_{it}(1|j)$ as a reference, and β_{jk} is a vector of parameters that model the effect of the covariate vector on the transition from j to k . To avoid any identifiability problems, we define $\alpha_{j1}^T = 0$ and $\beta_{j1}^T = 0$ for each j for all the generalized logit models in this study. Equivalently, the transitional probability for subject i from stage j to k from the $t-1^{th}$ time point to the t^{th} time point in this model is

$$\tilde{p}_{it}(k|j) = \frac{\exp\{\alpha_{jk} + \beta_{jk}^T x_{i,t-1}\}}{\sum_{h=1}^J \exp\{\alpha_{jh} + \beta_{jh}^T x_{i,t-1}\}}, \quad k=1, \dots, J-1.$$

2. Random effects model: For longitudinal categorical data, observations from the same subject are usually correlated. Failure to account for this correlation may result in incorrect inferences. In transitional models, some subjects may be more likely to make transitions than others, which should be accounted for in the model. The correlation for the observations within the same subject is modeled by a subject specific random effect. Specifically, we introduce a shared random effect u_i for the observations within the same i^{th} subject. The model is defined to be

$$\log \left(\frac{\tilde{p}_{it}(k|j)}{\tilde{p}_{it}(1|j)} \right) = \alpha_{jk} + \beta_{jk}^T x_{i,t-1} + u_i, \quad k=2, \dots, J-1, \quad (2)$$

where $u_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$. Note that using equation (2) and $\sum_{k=1}^{J-1} \tilde{p}_{it}(k|j) = 1$, we can easily show

$$\tilde{p}_{it}(k|j) = \frac{\exp\{\alpha_{jk} + \beta_{jk}^T x_{i,t-1} + u_i\}}{1 + \sum_{h=2}^J \exp\{\alpha_{jh} + \beta_{jh}^T x_{i,t-1} + u_i\}}, \quad k=2, \dots, J-1,$$

and

$$\tilde{p}_{it}(1|j) = \frac{1}{1 + \sum_{h=2}^J \exp\{\alpha_{jh} + \beta_{jh}^T x_{i,t-1} + u_i\}},$$

so we can consider this as a model having one subject-specific random effect for all categories except the reference category for each j . Consequently, a subject with a larger random effect is less likely to stay at normal and less likely to back transition to normal from MI.

3. Previous-stage dependent random effects model: In this model, we also define a shared random effect for observations within the same subject, but we add a scale parameter $\theta_j, j=2, \dots, J-2$ that depends on the previous stage. We also set $\theta_1 = 1$ to avoid the identifiability problem. The model is defined as

$$\log \left(\frac{\tilde{p}_{it}(k|j)}{\tilde{p}_{it}(1|j)} \right) = \alpha_{jk} + \beta_{jk}^T x_{i,t-1} + \theta_j u_i, \quad k=2, \dots, J-1. \quad (3)$$

The scale parameter θ_j can adjust for the magnitude of the variance of the random effects as a function of the previous stage, which is likely to be caused by the heterogeneity for transitions from different stages. The sign for θ_j gives information regarding the heterogeneity in the transitions of different individuals. To illustrate the information we get from the sign of θ_j , we use this aging study as an example. We have $\theta_1 = 1$ by definition, and the parameter θ_3 is not meaningful because dementia is an absorbing stage. So, only θ_2 is estimated for this study. When $\theta_2 > 0$, subjects with high risk of transition from normal to MI or dementia are also at high risk to remain at MI or become demented rather than returning to normal. On the contrary, if $\theta_2 < 0$, subjects with a high risk of transition from normal to MI or

dementia tend to be at low risk to remain at MI or become demented rather than returning to normal. A simulated graph with two different values of σ^2 in a generalized logit model $\log\{p_t(k|j)/p_t(1|j)\} = \alpha_{jk} + \beta_j u$, $k = 2, 3$ is shown in Figure 2, where α 's were chosen to have similar trends as the estimated α 's given in Table 3. We assume the time t ranges from 1 to 20 to get a sequence of 20 observations for each model. Different values of u can be viewed as different subject-specific random effects. As we can see, when u is positive, a subject tends to stay at his/her current stage. When u is negative, the subject tends to make transitions among different stages before she/he gets to the dementia stage. Note the shared random effects model defined in (2) is a special case of the previous stage dependent random effects model in (3) with $\beta_j = 1$ for all $j = 2, \dots, J-2$.

4. General previous-stage dependent random effects model: In this model we extend the univariate subject-specific random effects u_i as modeled in (3) above to a multivariate version u_{ij} with $J-2$ dimension. Let u_{ij} denote the j^{th} component of this random effect, then we have

$$\log\left(\frac{\tilde{p}_{it}(k|j)}{\tilde{p}_{it}(1|j)}\right) = \alpha_{jk} + \beta_{jk}^T x_{i,t-1} + u_{ij}, \quad k=2, \dots, J-1, \quad (4)$$

where the vector of $\mathbf{u}_i = \{u_{ij}\}_{j=1, \dots, J-2}$ has a multivariate normal distribution, i.e.

$\mathbf{u}_i \stackrel{i.i.d.}{\sim} N(0, \Sigma)$. Here the random effects are not only subject specific, but also previous state specific. For a fixed subject i , the random effect is extended from a univariate to a multivariate vector with each component corresponding to previous state effects respectively. This random effects vector shares the same multivariate normal distribution across subjects.

Instead of depending on the previous stage, the random effects could depend on both j and k (or $k-j$ or k only). In particular the model depending on $k-j$ could be interesting and more flexible. It includes 4 random effects, each corresponding to remaining in the stage ($k-j=0$), progressing toward the next stage ($k-j=1$), progressing directly from normal to dementia ($k-j=2$), and backward transition from MI to normal ($k-j=-1$). However, we are not attempting to fit a model such as this, because the number of transitions observed in each subject is small. Too many random effects may lead to identifiability problem, imprecision in integral computation, and convergence problems.

Model (1) is nested within model (2) which is further nested within model (3). Although model (3) is a special case of model (4), the two marginal models are not nested. Model (3) has a univariate unknown random effect u_i for repeated observations for subject i . These random effects across subjects share the same distribution $N(0, \sigma^2)$ with σ^2 together with other fixed effect parameters estimated by maximizing the numerically approximated likelihood to the integrated likelihood with the random effects integrated out. Moreover, model (3) has scale parameters $\sigma_2, \dots, \sigma_{J-2}$ shared by all the subjects to be estimated by the maximum of the integrated likelihood function with u_i integrated out. So the scale parameters σ_j are quite different from the extra variance/covariance parameters in model (4). These scale parameters effectively differentiate the transition probabilities among different previous states. Therefore, model (3) would be appropriate when transitional forces from previous states show distinct characteristics. Model (4) has a multivariate random effect that needs to be integrated out, with hyperparameters of variance and covariance estimated by all subjects similarly.

2.1.2 Proportional Odds Models—The proportional odds model uses cumulative

probability to model the outcome. Define $w_{it}(k|j) = P(Y_{it} \geq k | Y_{i,t-1} = j) = \sum_{h=k}^{J-1} \tilde{p}_{it}(h|j)$ to be the cumulative probability for the i^{th} subject with response value larger than or equal to k . We consider four types of models similar to those for the generalized logit models:

1. Fixed effects model: Consider

$$\text{logit}(w_{it}(k|j)) = \alpha_{jk} + \beta_j^T x_{i,t-1},$$

where α_{jk} , decreasing in k for all the models in this subsection, is the intercept for transition from stage j to k , and β_j is the common regression coefficient. Note the exponential of β_j is the odds of transition from lower stages to higher stages. The transitional probability for subject i from stage j to k can be calculated by

$$\tilde{p}_{it}(k|j) = \frac{1}{1 + \exp(-\alpha_{jk} - \beta_j^T x_{i,t-1})} - \frac{1}{1 + \exp(-\alpha_{j,k+1} - \beta_j^T x_{i,t-1})}.$$

2. Shared random effects model: Similar to the generalized logit model, we incorporate random effects to account for within-subject correlations for longitudinal categorical data. Consider

$$\text{logit}(w_{it}(k|j)) = \alpha_{jk} + \beta_j^T x_{i,t-1} + u_i, \quad k=1, \dots, J-1,$$

where $u_i \stackrel{i.i.d.}{\sim} N(0, \sigma^2)$. In both the generalized logit model and proportional odds model, a large u_i means that the subject is more likely to transit to higher categories from normal and remain at the present category or higher from MI.

3. Previous-stage dependent random effects model: Consider

$$\text{logit}(w_{it}(k|j)) = \alpha_{jk} + \beta_j^T x_{i,t-1} + \theta_j u_i, \quad k=1, \dots, J-1,$$

where the parameter θ_j has a similar meaning as model (3) in generalized logit models.

4. General previous-stage dependent random effects model: Similarly to model (4) in generalized logit models, we consider subject specific random effects u_{ij} for the repeated transitions for the i^{th} subject with previous stage j , and

$$\text{logit}(w_{it}(k|j)) = \alpha_{jk} + \beta_j^T x_{i,t-1} + u_{ij}, \quad k=1, \dots, J-1,$$

where the vector of $\mathbf{u}_i = (u_{i1}, \dots, u_{i,J-1})$ and $\mathbf{u}_i \stackrel{i.i.d.}{\sim} N(0, \Sigma)$. Similarly to that in the corresponding generalized logit model, the subject-specific random effect is a vector of $J-1$ different components corresponding to contributions from different previous states.

2.2 Logistic Regression Models for Death/Survival Transitions

We use a logistic regression model to fit the outcomes of death/no death with the same previous stage. In particular, we have, for given $j = J-2$

$$q(1|j) = P(Z_{it}=1|Y_{i,t-1}=j, x_{i,t-1}) = \frac{\exp\{\alpha_{jJ} + \beta^T x_{i,t-1} + \tau_j u_i\}}{1 + \exp\{\alpha_{jJ} + \beta^T x_{i,t-1} + \tau_j u_i\}}. \quad (5)$$

The interpretation of this model is similar to (3), except we are using the same slope for different last stage. This model is simpler and more meaningful from the clinical point of view. τ_j are parameters that characterize correlations between death probabilities from previous stages. Note the shared random effect between equations (2, 3, or 4) and (5) links the propensity of transition to the probability of death to model their correlation and also allows for heterogeneous transitions among subjects.

So if $Z_{it} = 1$, we use the above probability for death, the transition process terminates. However, if $Z_{it} = 0$, the subject may continue to make transitions. So we define the transitional probability from stage $j = J-2$ to $k = J-1$ as

$$p_{it}(k|j) = P(Y_{it}=k|Y_{i,t-1}=j, x_{i,t-1}, Z_{it}=0)P(Z_{it}=0|Y_{i,t-1}=j) = \tilde{p}_{it}(k|j)q(0|j).$$

3 Parameter Estimation

We illustrate the estimation algorithm for model (3), the previous stage dependent random effects model. Other models can be estimated similarly. For model (2), we set all τ_j 's to 1, and for model (1), we set all random effects to zero. For model (4), we change the random effects density to a bivariate normal density and drop τ_j 's.

Due to the first order Markov assumption, the likelihood for the i^{th} subject can be calculated by

$$L(\theta; \mathbf{y}_i) = P(Y_{i1}=y_{i1}) \prod_{t=2}^{n_i} p_{it}(y_{it}|y_{i,t-1}),$$

where $\theta = (\alpha, \beta, \tau, \gamma, \delta, \dots, \alpha_{J-2,J}, \tau_{J-2}, \gamma_{J-2}, \delta_{J-2})$ is defined to be the vector of all the parameters, and n_i is the number of observed time points for the i^{th} subject. The overall likelihood function can be calculated by

$$L(\theta; \mathbf{y}) = \prod_{i=1}^I L(\theta; \mathbf{y}_i) = \prod_{i=1}^I \left\{ P(Y_{i1}=y_{i1}) \prod_{t=2}^{n_i} [p_{it}(y_{it}|y_{i,t-1}, z_{it}=0)q(0|y_{i,t-1})]^{1-z_{it}} q(1|y_{i,t-1})^{z_{it}} \right\},$$

where I denotes the number of subjects in the study, and all the quantities defined in the parentheses except the initial stage $P(Y_{i1}=y_{i1})$ are defined in section 2. In fact, we can simplify the above expression by setting $z_{it} = 0$ for all $t = n_i - 1$. However, it does not save much space so we leave it as is. The parameters for the initial stage and transitions are different. Therefore for estimating parameters related to the transitions, we ignore the likelihood contributed from the initial stage. The random effects u_i are incorporated into the model through $p_{it}(y_{it}|y_{i,t-1}) = p_{it}(y_{it}|y_{i,t-1}, \alpha, \beta, \tau, \gamma, \delta, \dots, \alpha_{J-2,J}, \tau_{J-2}, \gamma_{J-2}, \delta_{J-2}, u_i)$. Then the marginal likelihood for the i^{th} subject can be calculated by

$$m_i(\theta) = \int \left\{ \prod_{t=2}^{n_i} p_{it}(y_{it}|y_{i,t-1}, \alpha, \beta, \theta_2, \dots, \theta_{J-2}, \tau_2, \dots, \tau_{J-2}, u_i) \right\} p(u_i|\sigma) du_i,$$

where $p(u_i|\sigma)$ denotes the random effect density. The overall marginal likelihood is calculated by

$$m(\theta) = \prod_{i=1}^I m_i(\theta).$$

The parameter estimation is implemented by maximizing the marginal likelihood $m(\theta)$ over θ . As special cases of non-linear mixed models (NLMM), parameter estimation for multilevel transitional models can be handled using techniques for NLMM. In particular, all transitional models we propose can be implemented through programming in SAS PROC NLMIXED. (SAS Institute, Inc., Cary, NC, <http://www.sas.com>.) The parameter estimation is implemented by maximizing an approximation to the marginal likelihood integrated over random effects. An adaptive Gaussian quadrature method (Pinheiro and Bates, 1995) and a first order Taylor series method are two principal procedures for approximating the integrated likelihood. Several options are given for the maximization algorithm. The default one is a dual quasi-Newton algorithm. Other methods include trust region method, Newton-Raphson method with line search, Newton-Raphson method with ridging, double-dogleg method, conjugate gradient methods, and Nelder-Mead simplex method. PROC NLMIXED produces parameter estimates along with their approximate standard errors computed from the second derivative matrix of the likelihood function upon successful convergence of the optimization algorithm. It also produces prediction results using parameter estimates and the empirical Bayes estimates of the random effects. Details for them are given in SAS online documentation (SAS Institute Inc.). In this paper, we have used the default methods: the adaptive Gaussian quadrature method for approximating the integrated likelihood and the quasi Newton method for the maximum likelihood estimation.

In this paper, we show all the models can be estimated with PROC NLMIXED in SAS by coding the components of the likelihood directly as described in Section 2. As pointed out by Wolfinger (www.support.sas.com/rnd/app/papers/nlmixedsugi.pdf, 2010): “PROC NLMIXED is best suited for models with a single random effect, although you can also successfully compute integrals in two and three dimensions as well. In addition, you can use PROC NLMIXED as a general optimization tool by omitting random effects altogether. Problems which are badly scaled or sufficiently noisy will not perform well with PROC NLMIXED.”

4 Model Evaluation Criterion

We use Akaike information criterion (AIC; Akaike, 1973, 1974) to evaluate all the models in this study. For a model with K parameters, AIC is defined to be

$$AIC = -2l(\hat{\theta}) + 2K,$$

where $l(\theta)$ is defined to be the logarithm of the maximum of the marginal likelihood $m(\theta)$. The smaller the AIC, the better the model fits. Because the marginal likelihood is not available analytically for the random effect models, the SAS PROC NLMIXED procedure approximates it using an adaptive Gaussian quadrature procedure which is its default setting.

The number K is the number of unknown parameters $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_{J-2}, \beta_2, \dots, \beta_{J-2}, \beta_J)$ estimated from the model that includes all the regression coefficients, scalar parameters β_j and the variance-covariance parameters Σ for the random effects.

The recent work by Greven & Kneib (2009) has pointed out theoretical difficulties with the marginal AIC when comparing mixed effects models with and without random effects for the simple case of a linear mixed model. Their difficulty may carry over to our more complex setting; more research in the future is needed to clarify this point.

5 Results

5.1 The data

Between 1993 and 2007, the Einstein Aging Study recruited 1682 subjects. These subjects have been followed annually, and cognitive status was evaluated at each clinic visit. Due to intermittent missing visits for some subjects, we only consider consecutive transitions made within two years. Therefore, we have observed 833 subjects who have had at least one stage transitions with a total of 2152 transitions.

Table 1 summarizes the frequency counts of the observed stage evolutions by discrete covariates and the mean values of other (continuous) covariates. The covariates include (1) gender, (2) years of education, (3) age at previous visit, (4) self-reported history of myocardial infarction, (5) self-reported history of stroke, and (6) self-reported history of diabetes. We have 1790 stage evolutions starting from normality and 362 evolutions starting from the MI stage.

Older and less well educated persons were more likely to transition from normal to MI or dementia ($p < 0.001$ and $p < 0.001$), as were persons with self-reported history of stroke ($p = 0.041$). Women were less likely to transit to MI but more likely to transition to dementia, either from normality ($p < 0.001$) or MI ($p = 0.008$), than men. The p-values for age at previous visit and years of education were derived using PROC LOGISTIC with types of transitions as response categories and age/years of education as predictors. The p-values for categorical factors are calculated using large sample chi-square tests by SAS PROC FREQ with cross tabulation for transition types and categorical factors.

5.2 Model comparisons

Table 2 shows the model fitting statistics AIC for both generalized logit models and proportional odds models. To avoid the complexity of model comparisons due to different parameter spaces, we only evaluate the full models with the main effects of all seven covariates.

For generalized logit models, the previous-stage dependent random effects model has the best model fit, followed by the general random effects model. The fixed effects model and shared random effects model do not fit the data as well.

For proportional odds models, the previous-stage dependent random effects model also has the smallest AIC, followed by the general previous stage dependent random effects model, the fixed effects model, and the shared random effects model.

The previous-stage dependent random effects proportional odds model has only a slightly better AIC than the corresponding generalized logit model. This indicates that there is not strong evidence against the proportional odds assumption. We present results for both the generalized logit model and proportional odds model.

It is interesting to observe both models (1) and (2) in both cases have essentially the same likelihood values. This indicates that including the random effects u_i does not help in modeling the transition probabilities. This can be confirmed by looking up the estimate of the random effect variance in model (2) generalized logit model which is 0.003, close to 0, and γ_2 is negative in model (3). Similar results hold for the proportional odds model.

5.3 Generalized logit models

For the generalized logit models, the data analysis here is based on the previous stage dependent random effects model since it has the best AIC among the four models. Table 3 illustrates the parameter estimates and p -values of the factors by transition types in the reduced multilevel model. An overall p -value is calculated for each factor by testing whether all the four transition slopes for this factor can simultaneously be zeroes. Backward elimination is used to obtain a reduced multivariate model. Factors with the largest overall p -values are dropped sequentially, and the final model with only significant transition parameters includes age at previous visit, gender, and self reported history of stroke and diabetes. A positive parameter estimate indicates that a factor is predictive of the declining transition, and a negative parameter estimate shows a protective factor for the transition.

In the generalized logit models, the odds for the corresponding transitions for each subject can be calculated by exponentiating the parameter estimates. For example, the odds of transiting from normal to MI versus normal to normal (no transition) for a study participant of a particular age, gender, and stroke history is calculated by $\exp\{-3.161 + 0.146 * (\text{Age at Previous Visit (years)} - 75) - 0.633 * \text{Gender} + 0.740 * \text{Stroke}\}$. When other factors are the same, the odds ratio of transition from normal to MI versus staying at normal for a participant who is 5 years older is calculated by $\exp(0.146 * 5) = 2.1$. The odds ratio of transition to MI for female versus male is $\exp(-0.633) = 0.5$. Participants with a self reported history of stroke are $\exp(0.740) = 2.1$ times more likely to make the normal to MI transitions.

In terms of transitions from normality to dementia versus staying at normal, significant factors include age at previous visit, and gender. The odds ratios of making the transitions for participants who are 5 years older are $\exp(0.207 * 5) = 2.8$. This shows that age is an important predictor of normal to dementia transitions. Female relative to male is $\exp(1.347) = 3.8$ more likely to make the normal to dementia transitions.

For the transitions from MI, age is significantly predictive of the transition from MI to dementia versus MI to normal. The other factors do not significantly predict or protect transitions from MI, which may be related to small sample size and insufficient power.

The scale parameter γ_2 is estimated to be -1.238. The negative value indicates a tendency for transitions to go back and forth between stages. This may stem from heterogeneity in the natural history of preclinical dementia or cognitive instruments of imperfect reliability.

Age, self reported history of stroke, and self reported history of diabetes are significant predictors of death.

5.4 Proportional odds models

Table 4 shows the parameter estimates and p -values for all the parameters in the best reduced multilevel proportional odds model. Similar to Table 3, the overall p -value for each factor is calculated. The backward elimination procedure is applied for variable selection. The factors in the reduced model include age at previous visit, and self-reported history of myocardial infarction, stroke, and diabetes.

For the proportional odds model, the odds ratio of each factor for the corresponding transition type is calculated by taking the exponential of the parameter estimate. Age at previous visit is an important predictor of transition to MI or to dementia. When other factors are the same, a subject at the normal stage who is 5 years older than others is $\exp(0.150 * 5) = 2.1$ times as likely to transition to dementia than to MI or staying at normal. The subject also has the same chance of making the MI or to dementia transition versus staying at normal. If a subject is already in MI, then the odds ratio of transition to dementia versus staying at MI or back to normal is $\exp(0.093 * 5) = 1.6$. Similarly, this person is 1.6 times as likely to be in the MI or dementia stage than the normal stage at the next clinic visit.

Self-reported history of myocardial infarction is predictive for transition from normality to MI or to dementia. The odds ratio of getting to worse stages versus better stages from the normal stage is $\exp(0.730) = 2.1$. At the MI stage, self-reported history of myocardial infarction is not a statistically significant factor.

At the normal stage, self-reported history of stroke is predictive of transitions to dementia versus staying at MI or normal, or transitions to MI or dementia versus staying at normal. The odds ratio is $\exp(0.692) = 2.0$. However, if a person is already in the MI stage, self-reported history of stroke is protective of transitions to worse stages relative to better stages with the odds ratio $\exp(-1.318) = 0.3$.

The point estimates of the scale parameter $\sigma^2 = -1.342$ and $\sigma^2 = 1.227$. The results are consistent with that of the generalized logit model.

6 Discussion

In this study, we evaluated seven possible risk factors potentially associated with the transitions among three different cognitive states. We applied the likelihood-based longitudinal transitional model to the data from the Einstein Aging Study. Compared to other methodologies, the inference and model evaluation are better developed for likelihood-based methods. For the transitional probabilities in the likelihood, we assume they satisfy the first order Markov property. This allows us to consider only models for the adjacent time points. The first order Markov model is usually the first model for investigators to explore because it is much simpler than higher order Markov models. It is also supported by the investigators' knowledge that dependence between successive responses on the same individual is much stronger than between responses further apart. Much of the manuscript is devoted to comparing different variance component models. Those variance components absorb much of the variability that would be modeled by a higher order Markov model. The interpretation of higher order Markov parameters would be conditional on the random effects and therefore would not have a straightforward interpretation. Furthermore, we do not believe that the data can support modeling both higher order Markov parameters along with the variance components we have tried to estimate, as the higher order Markov parameters would significantly decrease the analysis sample size and the power of the analysis.

Wilson et al. (2003) and our reviewers pointed out the importance of accounting for death in analyzing risk factors for cognitive decline. Death is a competing risk, so we model the transition to death from the normal or memory impaired stages, and the propensity to transition across the cognitive stages simultaneously. We have developed a method for implementing this. The random effects are incorporated into a multilevel model to account for within-subject correlations, and correlation between death and cognitive transitions. Two possible modeling approaches for the cognitive transition probability, generalized logit and proportional odds, were evaluated in this study. Each model has its own strengths. The

generalized logit model allows us to make inferences for all the possible transitions, while the proportional odds model assumes an overall trend of transitions from better stages to worse stages. As a result, the generalized logit model can capture information specific for each transition and factor, but at the expense of more parameters and less power. The proportional odds model can only show the trend. It has the advantages of model simplicity and natural interpretations that the odds for lower stages to transit to higher stages are the same as characterized by the common odds. It combines information with fewer parameters and has higher power. In a model with the same number of covariates, the generalized logit model has twelve more parameters than the proportional odds model, which adds to model complexity. But in a data set where the proportional odds assumption is not valid or when we are interested in gaining insight into the possible differences between transitions starting from different stages, the generalized logit would be a better choice. For example, in our study women are less likely to transit to MI from normal, but more likely to transit to dementia from normal. The generalized logit model detected the differential effect while the proportional odds model did not. Both models showed that stroke was a risk factor for the transition from normal to MI. However, the generalized logit model did not detect the effect of stroke on transitions from the MI stage, while the proportional odds model showed a significant protective effect, possibly due to its greater power. This effect may reflect recovery from stroke.

Given that our data were collected annually between 1993 and 2007 at each clinical visit, our likelihood-based method is direct in modeling the observed data and describing the dynamic changes of the observed outcomes with their risk factors. Consequently, it is easy to interpret the results. We think a non-homogeneous Markov model for the time-to-event recurrent analysis is possible, however, its definition and interpretation would be less straightforward. First, one would need to incorporate interval censoring because the transition time is not observed directly. Second, a subject usually has several different times-to-event recorded. Transition times are from state 1 to 2, state 1 to 3, state 2 to 1 and state 2 to 3, so it is likely that even more careful modeling for the correlation structure that incorporates between-variation among different transitions, and within-variation among recurrent events of the same transitions would be required. Third, although much of the variability that is captured in the variance component models described in this manuscript could be modeled using a non-homogeneous Markov model, the interpretation of the parameters would not be straightforward and we do not believe that the data are sufficiently rich to adequately model both the variance components and higher order Markov parameters.

In summary, we propose a likelihood-based transition model to dynamically model multi-stage discrete outcomes of a clinical study. The conditional transition probabilities being alive are modeled by generalized logit models or proportional odds models. We have incorporated the subject specific random effects in the transition probabilities to explain the correlations among the longitudinal data. A simple random effects model given in model (2) is often used in these types of analysis, however, it could be misleading. Due to the random effects that may influence the transition probabilities in different directions from different previous states, we have found it is desirable to extend the random effects model to scale multiples of random effects to incorporate this kind of heterogeneity in transitions. Our model selection among various variance components models using AIC further supports this. We show all the models can be estimated with PROC NLMIXED in SAS by coding the components of the likelihood directly as described in Section 2. Interpretation for the parameter estimates in transitional models is directly related to the between stage transitions, which is of great importance from the clinical point of view. The generalizability of our results is dependent on the data set analyzed. We think our methods are easy to follow and

can be applied to a wide class of longitudinal data sets with multiple-stage outcomes and time varying covariates.

Acknowledgments

The Einstein Aging Study is supported by the National Institutes on Aging #5P01-AG003949. We would like to thank the Editor, Associate Editor, and two referees for their helpful comments which have improved this article.

References

- Agresti, A. Categorical Data Analysis. New York: John Wiley & Sons, Inc; 1990.
- Agresti, A. Categorical Data Analysis. 2. New York: John Wiley & Sons, Inc; 2002.
- Akaike, H. Information theory and an extension of the maximum likelihood principle. In: Petrov, BN.; Csaki, F., editors. Secon International Symposium on Information Theory; Budapest: Akademiai Kiado; 1973. p. 267-281.
- Akaike H. A new look at the statistical model identification. IEEE Transactions on Automatic Control. 1974; 19:716–723.
- Albert PS, Follmann DA. A random effects transition model for longitudinal binary data with informative missingness. Statistica Neerlandica. 2003; 57:100–111.
- Albert PS, Follmann DA, Wang SA, Suh EB. A latent autoregressive model for longitudinal binary data subject to informative missingness. Biometrics. 2002; 58:631–642. [PubMed: 12229998]
- Albert PS, Waclawiw MA. A two state Markov chain for heterogeneous transitional data: a quasilielihood approach. Statistics in Medicine. 1998; 17:1481–1493. [PubMed: 9695193]
- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. DSM-IV; Washington, DC: 1994.
- Beal SL, Sheiner LB. Estimating population kinetics. Critical Reviews in Biomedical Engineering. 1982; 8:195–222. [PubMed: 6754254]
- Beal SL, Sheiner LB. Heteroskedastic nonlinear regression. Technometrics. 1988; 30:327–338.
- Bennett DA, Schneider JA, Bienias JL, Evans DA, Wilson RS. Mild cognitive impairment is related to Alzheimer disease pathology and cerebral infarctions. Neurology. 2005; 64:834–841. [PubMed: 15753419]
- Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. Journal American Statistical Association. 1993; 88:9–25.
- Buschke H. Selective reminding for analysis of memory and learning. Journal of Verbal Learning and Verbal Behavior. 1973; 12:543–550.
- Carey V, Zeger SL, Diggle P. Modelling multivariate binary data with alternating logistic regressions. Biometrika. 1993; 80:517–526.
- Cole BF, Bonetti M, Zaslavsky AM, Gelber RD. A multistate Markov chain model for longitudinal, categorical quality-of-life data subject to non-ignorable missingness. Statistics in Medicine. 2005; 24:2317–2334. [PubMed: 15977292]
- Commenges D, Letenneur L, Joly P, Alioum A, Dartigues JF. Modeling age-specific risk: application to dementia. Statistics in Medicine. 1998; 17:1973–1988. [PubMed: 9777690]
- Gauthier S, Reisberg B, Zaudig M, Petersen RC, Ritchie K, Broich K, Belleville S, Brodaty H, Bennett D, Chertkow H, Cummings JL, De Leon M, Feldman H, Ganguli M, Hampel H, Scheltens P, Tierney MC, Whitehouse P, Winblad B. International Psychogeriatric Association Expert Conference on mild cognitive impairment. Mild cognitive impairment. Lancet. 2006; 367(9518): 1262–1270. [PubMed: 16631882]
- Greven S, Kneib T. On the behaviour of marginal and conditional Akaike information criteria in linear mixed models. Biometrika. 2010; 97:773–789.
- Grober E, Buschke H. Genuine memory deficits in dementia. Developmental Neuropsychology. 1987; 3:13–36.
- Hall CB, Lipton RB, Sliwinski M, Stewart WF. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. Statistics in Medicine. 2000; 9:1555–1566. [PubMed: 10844718]

- Hall CB, Ying J, Kuo L, Lipton RB. Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics and Data Analysis*. 2003; 42:91–109.
- Hall CB, Ying J, Kuo L, Silwinski M, Buschke H, Katz M, Lipton RB. Estimation of bivariate measurements having different change points, with application to cognitive aging. *Statistics in Medicine*. 2001; 20:3695–3714. [PubMed: 11782027]
- Harezlak J, Gao S, Hui SL. An illness-death stochastic model in the analysis of longitudinal dementia data. *Statistics in Medicine*. 2003; 22:1465–1475. [PubMed: 12704610]
- Huber, PJ. The behaviour of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*; 1967. p. 221–233.
- Jacqmin-Gadda H, Commenges D, Dartigues JF. Random change point model for joint modeling of cognitive decline and dementia. *Biometrics*. 2006; 62:254–260. [PubMed: 16542253]
- Launer LJ. Demonstrating the case that AD is a vascular disease: epidemiologic evidence. *Aging Research Reviews*. 2002; 1:61–77. [PubMed: 12039449]
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
- Lipton RB, Katz MJ, Kuslansky G, Sliwinski MJ, Stewart WF, Crystal HA, Buschke H. Screening for dementia by telephone using the memory impairment screen. *Journal of the American Geriatrics Society*. 2003; 51:1382–1390. [PubMed: 14511157]
- Martus P, Stroux A, Junemann AM, Korth M, Jonas JB, Horn FK, Ziegler A. GEE approaches to marginal regression models for medical diagnostic tests. *Statistics in Medicine*. 2004; 23:1377–1398. [PubMed: 15116348]
- McCullagh P. Regression models for ordinal data (with discussions). *Journal of the Royal Statistical Society, Series B*. 1980; 42:109–142.
- Morris JC, Storandt M, Miller JP, McKeel DW, Price JL, Rubin EH, Berg L. Mild cognitive impairment represents early-stage of Alzheimer's disease. *Archives of Neurology*. 2001; 58:397–405. [PubMed: 11255443]
- Pinho JC, Bates DM. Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model. *Journal of Computational and Graphical Statistics*. 1995; 4:12–35.
- Rooij M, Kroonenberg PM. Multivariate multinomial logit models for dyadic sequential interaction data. *Multivariate Behavioral Research*. 2003; 38:463–504.
- Rubin D. Inference and missing data. *Biometrika*. 1976; 63:581–592.
- Salazar JC, Schmitt FA, Yu L, Mendiondo MM, Kryscio RJ. Shared random effects analysis of multi-state Markov models: application to a longitudinal study of transitions to dementia. *Statistics in Medicine*. 2007; 26:568–580. [PubMed: 16345024]
- SAS INSTITUTE INC. SAS/STAT: The NLMIXED Procedure. 2008. <http://support.sas.com/onlinedoc/912/docMainpage.jsp>
- Stampfer MJ. Cardiovascular disease and Alzheimer's disease: common links. *Journal of Internal Medicine*. 2006; 260:211–223. [PubMed: 16918818]
- Tyas SL, Salazar JC, Snowdon DA, Desrosiers MF, Riley CP, Mendiondo MS, Kryscio RJ. Transitions to mild cognitive impairments, dementia, and death: findings from the nun story. *American Journal of Epidemiology*. 2007; 165:1231–1238. [PubMed: 17431012]
- Wilson RS, Beckett LA, Bienias JL, Evans DA, Bennett DA. Terminal decline in cognitive function. *Neurology*. 2003; 60:1782–1787. [PubMed: 12796531]
- Wolfinger, RD. Fitting nonlinear mixed models with the new NLMIXED procedure. SAS Institute Inc.; Cary, NC: 2010. www.support.sas.com/rnd/app/papers/nlmixedsugi.pdf
- Yesavage JA, Brink TL, Ross TL, Lum O, Huang V, Adey MB, Leirer VO. Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*. 1983; 17:37–49. [PubMed: 7183759]
- Yu B, Ghosh P. Joint modeling for cognitive trajectory and risk of dementia in the presence of death. *Biometrics*. 2010; 66(1):294–300. [PubMed: 19432791]

- Zeger SL, Karim MR. Generalized linear models with random effects; A Gibbs sampling approach. *Journal of the American Statistical Association*. 1991; 86:79–86.
- Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986; 42:121–130. [PubMed: 3719049]

Fig. 1.

Possible transitions for the four stages and their regression coefficients γ_{ij} to be estimated. We have used γ_{jk} to denote $(\gamma_{jk}, \gamma_{jk})$ for $j = 1, 2$ and $k = 1, 2, 3, 4$. We are using the normality stage as a reference, so γ_{11} , and γ_{21} , are not estimated.

Fig. 2.

Simulated transitions for the generalized logit model $\log \frac{\tilde{p}_t(k|j)}{\tilde{p}_t(1|j)} = \alpha_{jk} + \theta_j u$, $k = 2, 3$ with $\alpha_{12} = -1$, $\alpha_{13} = -3$, $\alpha_{22} = 1$, and $\alpha_{23} = -1.5$. The three figures on the left are simulated with $\theta = 1$ and $u = -1, 1, 2$. The three figures on the right are derived with $\theta = -1$ and $u = -1, 1, 2$.

Table 1

Summary of observed stage evolutions by covariates.

	Transitions From Normality Stage			Normal to Normal			Normal to MI			Normal to Dementia			Normal to Death			Total		
	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE
Total	1560	100.0		163	100.0		23	100.0		44	100.0		1790	100.0		1790	100.0	
Female	961	61.6		80	49.1		20	87.0		22	50.0		1083	60.5		1083	60.5	
Myocardial Infarction	136	8.7		24	14.7		3	13.0		6	13.6		169	9.4		169	9.4	
Diabetes	187	12.0		27	16.6		3	13.0		8	18.2		225	12.6		225	12.6	
Stroke	111	7.1		17	10.4		3	13.0		10	22.7		141	7.9		141	7.9	
	Mean	SE		Mean	SE		Mean	SE		Mean	SE		Mean	SE		Mean	SE	
Age at Previous Visit	78.8	0.1		81.9	0.4		83.6	1.3		82.2	0.9		79.2	0.1		79.2	0.1	
Years of Education	13.6	0.1		12.7	0.3		11.4	0.7		12.8	0.5		13.4	0.1		13.4	0.1	
	MI to Normal			MI to MI			MI to Dementia			MI to Death			Total					
	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE	N	%	SE
Total	121	100.0		160	100.0		61	100.0		20	100.0		362	100.0		362	100.0	
Female	50	41.3		68	42.5		39	63.9		10	50.0		167	46.1		167	46.1	
Myocardial Infarction	16	13.2		15	9.4		9	14.8		4	20.0		44	12.2		44	12.2	
Diabetes	18	14.9		28	17.5		7	11.5		7	35.0		60	16.6		60	16.6	
Stroke	17	14.0		16	10.0		4	6.6		3	15.0		40	11.0		40	11.0	
	Mean	SE		Mean	SE		Mean	SE		Mean	SE		Mean	SE		Mean	SE	
Age at Previous Visit	81.3	0.5		81.5	0.4		83.5	0.7		86.2	1.5		82.0	0.3		82.0	0.3	
Years of Education	12.7	0.4		12.6	0.3		12.8	0.5		12.0	0.6		12.6	0.2		12.6	0.2	

Table 2

Goodness-Of-Fit for generalized logit and proportional odds models.

Models	Parameters	-2LogLike	AIC
Generalized logit models			
(1) Fixed effects	36	2170.3	2242.3
(2) Shared random effects	37	2170.2	2244.2
(3) Previous stage dependent random effects	39	2126.8	2204.8
(4) General previous stage dependent random effects	41	2125.5	2207.5
Proportional odds model			
(1) Fixed effects	24	2199.6	2247.6
(2) Shared random effects	25	2199.6	2249.6
(3) Previous stage dependent random effects	27	2153.9	2207.9
(4) General previous stage dependent random effects	29	2152.1	2210.1

Table 3

Parameter estimates in the reduced multilevel generalized logit model.

Factor	Transition Type	Base Transition	Estimate	OR	95% CI	p-value
Intercept	Normal to MI	Normal to Normal	-3.161			<0.001
	Normal to Dementia	Normal to Normal	-6.899			<0.001
	Normal to Death	Normal to Alive	-4.927			<0.001
	MI to MI	MI to Normal	0.803			0.033
	MI to Dementia	MI to Normal	-1.278			<0.001
	MI to Death	MI to Alive	-4.627			<0.001
Age at Previous Visit (Continuous)	Normal to MI	Normal to Normal	0.146	1.16	(1.10,1.21)	<0.001
	Normal to Dementia	Normal to Normal	0.207	1.23	(1.13,1.34)	<0.001
	MI to MI	MI to Normal	0.034	1.03	(0.96,1.11)	0.349
	MI to Dementia	MI to Normal	0.124	1.13	(1.04,1.23)	0.004
	Normal/MI to Death	Normal/MI to Alive	0.127	1.13	(1.07,1.20)	<0.001
Gender (F vs M)	Normal to MI	Normal to Normal	-0.633	0.53	(0.34,0.84)	0.007
	Normal to Dementia	Normal to Normal	1.347	3.85	(1.08,13.7)	0.038
	MI to MI	MI to Normal	0.162	1.18	(0.54,2.55)	0.682
	MI to Dementia	MI to Normal	0.954	2.60	(1.06,6.37)	0.038
Stroke (Yes vs No)	Normal to MI	Normal to Normal	0.740	2.82	(1.25,6.35)	0.050
	Normal to Dementia	Normal to Normal	1.066	3.60	(0.86,15.1)	0.127
	MI to MI	MI to Normal	-1.325	0.31	(0.10,1.01)	0.037
	MI to Dementia	MI to Normal	-1.312	0.26	(0.05,1.23)	0.086
	Normal/MI to Death	Normal/MI to Alive	1.220	3.39	(1.64,6.98)	0.001
Diabetes	Normal/MI to Death	Normal/MI to Alive	0.938	2.55	(1.33,4.89)	0.005
2			0.502			0.033
2			-1.238			0.007
			1.304			

Table 4

Parameter estimates in the reduced multilevel proportional odds model.

Factor	Transition Type	Base Transition	Estimate	OR	95% CI	p-value
Intercept	Normal to (MI or Dementia)	Normal to Normal	-3.456			<0.001
	Normal to Dementia	Normal to (Normal or MI)	-5.833			<0.001
	Normal to Death	Normal to Alive	-4.911			<0.001
	MI to (MI or Dementia)	MI to Normal	0.814			0.009
	MI to Dementia	MI to (Normal or MI)	-2.298			0.001
	MI to Death	MI to Alive	-4.613			<0.001
Age at Prev. Visit (Continuous)	Normal to (MI or Dementia)	Normal to Normal	0.150	1.16	(1.11,1.21)	<0.001
	Normal to Dementia	Normal to (Normal or MI)	"	"		
	MI to (MI or Dementia)	MI to Normal	0.093	1.10	(1.03,1.17)	0.007
	MI to Dementia	MI to (Normal or MI)	"	"		
	Normal/MI to Death	Normal/MI to Alive	0.125	1.13	(1.07,1.20)	<0.001
Myocardial Infarction (Yes vs No)	Normal to (MI or Dementia)	Normal to Normal	0.730	2.08	(1.11,3.87)	0.022
	Normal to Dementia	Normal to (Normal or MI)	"	"		
	MI to (MI or Dementia)	MI to Normal	0.257	1.29	(0.48,3.47)	0.610
	MI to Dementia	MI to (Normal or MI)	"	"		
Stroke(Yes vs No)	Normal to (MI or Dementia)	Normal to Normal	0.692	2.00	(1.00,3.97)	0.049
	Normal to Dementia	Normal to (Normal or MI)	"	"		
	MI to (MI or Dementia)	MI to Normal	-1.318	0.27	(0.08,0.86)	0.027
	MI to Dementia	MI to (Normal or MI)	"	"		
	Normal/MI to Death	Normal/MI to Alive	1.204	3.33	(1.62,6.84)	0.001
Diabetes (Yes vs No)	Normal/MI to Death	Normal/MI to Alive	0.943	2.57	(1.34,4.91)	0.004
2			0.511			0.048
2			-1.342			0.011
			1.227			

Note: The " signs in the above indicate that the estimates are the same as given in the line above for each coefficient.