Published in final edited form as:

J Appl Stat. 2011 ; 38(10): 2313–2326. doi:10.1080/02664763.2010.547567.

A comparison of non-homogeneous Markov regression models with application to Alzheimer's disease progression

R. A. Hubbard^{1,2,*} and X.H. Zhou²

¹Group Health Research Institute, Biostatistics Unit, 1730 Minor Ave, Suite 1600, Seattle, WA 98101, USA

²Department of Biostatistics, University of Washington, Campus Mail Stop 357232, Seattle, WA, 98195, USA

Abstract

Markov regression models are useful tools for estimating the impact of risk factors on rates of transition between multiple disease states. Alzheimer's disease (AD) is an example of a multi-state disease process in which great interest lies in identifying risk factors for transition. In this context, non-homogeneous models are required because transition rates change as subjects age. In this report we propose a non-homogeneous Markov regression model that allows for reversible and recurrent disease states, transitions among multiple states between observations, and unequally spaced observation times. We conducted simulation studies to demonstrate performance of estimators for covariate effects from this model and compare performance with alternative models when the underlying non-homogeneous process was correctly specified and under model misspecification. In simulation studies, we found that covariate effects were biased if non-homogeneity of the disease process was not accounted for. However, estimates from non-homogeneous models were robust to misspecification of the form of the non-homogeneity. We used our model to estimate risk factors for transition to mild cognitive impairment (MCI) and AD in a longitudinal study of subjects included in the National Alzheimer's Coordinating Center's Uniform Data Set. Using our model, we found that subjects with MCI affecting multiple cognitive domains were significantly less likely to revert to normal cognition.

Keywords

Alzheimer's disease; interval censoring; Markov process; mild cognitive impairment; non-homogeneous; panel data

1 Introduction

Markov regression models are frequently used to estimate the effect of covariates on transitions between multiple states. They have been applied in many studies of health and disease, including studies of stroke [20], cancer [25], and diabetic retinopathy [18]. The continuous time Markov model is particularly useful because it allows for interval censoring and unequally spaced observations. These are both common features of data in longitudinal observational studies of health and disease, where subjects may only be observed at discrete follow-up visits which may occur at irregular times.

*Corresponding author: hubbard.r@ghc.org, phone: 206-287-2066, fax: 206-287-2871.

Alzheimer's disease (AD) is an example of a multi-state disease process in which scientific interest centers on identifying risk factors for transitions between disease states. In this example, cognitive functioning may be classified as normal cognition, mild cognitive impairment (MCI), and AD. In clinical studies of AD, MCI has been recognized as an important transitional state. Subjects in this state experience increased rates of conversion to AD with an estimated 10 – 15% of patients with MCI converting to AD each year [21, 4]. However, recent research has highlighted the volatility of MCI [15, 13, 17, 16]. One study has estimated the proportion of subjects with MCI reverting to a cognitively normal state within two years at 31% [17]. Understanding characteristics of patients with MCI at increased risk of converting to AD and characteristics of those more likely to revert to normal cognition can aid in appropriate targeting of therapeutic agents. Identifying risk factors for conversion to MCI is of special interest because it is believed that disease-modifying therapies may have greater efficacy in subjects who have not yet developed AD and therefore have not experienced neuronal loss. In this context, the objective is to estimate the association between patient-level characteristics and the rate of conversion from MCI to AD and reversion from MCI to normal cognition.

Markov process models are ideal for use in studies of AD because they estimate the rate of transition between multiple disease states while allowing for the possible reversibility of some states, such as MCI, and accounting for the competing risk of death. Previous studies have demonstrated the extent of bias induced by ignoring the competing risk of death when estimating the rate of progression to MCI and AD [10]. Markov process models are also most appropriate to the observational nature of studies of AD in which cognitive status is typically assessed at periodic clinic visits, giving rise to interval censored data. Moreover, if subjects transition among multiple disease states between clinic visits the complete disease history will not be available. Continuous time Markov process models can easily accommodate both interval censoring and failure to observe the complete disease history.

In many studies of disease, age dependent increases in the rate of transition between states are a prominent feature. However, limited methods are available for estimating disease processes characterized by time-varying transition rates using continuous time Markov models. In the context of Markov process models, variation in transition intensities with respect to time is referred to as temporal non-homogeneity. A simple method for accommodating temporal non-homogeneity is to stratify the population into age groups and estimate transition rates separately within groups. This method has been previously applied to studies of AD [26, 10]. Recently a new piecewise model for transition intensities for transition to AD was proposed in which piecewise intensities are modeled as explicit functions of age and other time-dependent covariates [24]. Disadvantages of piecewise approaches are that they require a large study population to obtain precise estimates because transition rates must be estimated separately in each time period. Additionally, transition rates must be assumed constant within periods.

In the case of irreversible disease, maximum penalized likelihood has been used to estimate smoothly time-varying transition intensities [6, 5]. While this approach is appealing because it makes few parametric assumptions about the shape of the time-varying transition intensities, it is limited by the assumption that the disease process under study is irreversible. Irreversible Markov process models have limited usefulness in studies including MCI due to the volatility of MCI resulting in many MCI subjects returning to normal cognition [15, 13, 17, 16].

In this research we developed a non-homogeneous Markov regression model that allows for reversible states via the method of time-transformation [11] and compared this method to alternative Markov regression models. This work builds upon our previous research by

allowing for estimation of covariate effects on time-varying transition intensities and by providing useful guidance on the relative performance of several different Markov regression models. In Section 2 we introduce Markov regression models for multi-state data and discuss the proposed time-transformed model. In Section 3.1, we present simulation studies to compare our non-homogeneous Markov regression model to other existing models under a variety of data generating mechanisms. Finally, we applied these models to data on AD progression from the National Alzheimer's Coordinating Center in Section 3.2.

2 Non-homogeneous Markov regression models

2.1 Continuous time Markov process models

We assume that transitions between disease states follow a first-order Markov process. We define the current state of subject i at observation j occurring at time t_{ij} as $X(t_{ij})$, taking realized value x_{ij} . We further define the probability of transitioning between two states as

$$p_{x_{i(j-1)}x_{ij}}(t_{i(j-1)}, t_{ij}) \doteq P\{X(t_{ij})=x_{ij}|X(t_{i(j-1)})=x_{i(j-1)}\}.$$

The Markov process is fully characterized by its transition intensities,

$$q_{ab}(t) = \lim_{\Delta t \rightarrow 0} p_{ab}(t, t+\Delta t)/\Delta t, a \neq b, \text{ and} \\ q_{aa}(t) = - \sum_{b \neq a} q_{ab}(t) [7].$$

We denote by $\mathbf{P}(t_1, t_2)$ and $\mathbf{Q}(t)$ the $s \times s$ matrix of transition probabilities $p_{ab}(t_1, t_2)$ and matrix of transition intensities $q_{ab}(t)$, respectively.

Under temporal homogeneity, transition intensities are independent of the time since the process origin (in the AD context, the age of the subject), and transition probabilities depend only on the elapsed time between successive observations, $p_{ab}(t_1, t_2) = p_{ab}(0, t_2 - t_1)$, for $a, b = 1, \dots, s$. In this case transition probabilities can be related directly to transition intensities via the relationship

$$\mathbf{P}(t_1, t_2) = \exp(\mathbf{Q}(t_2 - t_1)) = \sum_{r=0}^{\infty} (\mathbf{Q}(t_2 - t_1))^r / r!. \quad (1)$$

This allows us to formulate the likelihood for the transition intensities as

$$L^{(m)}(\mathbf{Q}) = \prod_{i=1}^m \left[P\{X(t_{i1})=x_{i1}\} \prod_{j=2}^{n_i} \{\exp(\mathbf{Q}(t_{ij} - t_{i(j-1)}))\}_{x_{i(j-1)}x_{ij}} \right], \quad (2)$$

2.2 Time-transformation model

One method for accommodating non-homogeneity due to age or other time dependence of transition intensities is the method of time-transformation [11]. This method makes use of the assumption that there exists an alternative time scale on which the process is homogeneous and then jointly estimates the transformation of the time scale leading to homogeneity and the transition intensity matrix.

We denote the transition intensity matrix for the transformed homogeneous process $\mathbf{Q}_0$ and new time scale $h(t; \boldsymbol{\theta})$. The assumption of homogeneity on the time scale $h(t; \boldsymbol{\theta})$ implies that the non-homogeneous transition intensity matrix for the process is given by $\mathbf{Q}(t) = \mathbf{Q}_0 dh(t; \boldsymbol{\theta})/dt$. By carrying out estimation of model parameters on the alternate time scale, we are able to make use of equation (2) to formulate the likelihood.

The time-transformation method has been previously applied to studies of multi-state disease with time-varying transition intensities [11, 12]. However, in many studies of disease evolution, we are primarily interested in the effects of covariates that act on transition rates rather than the baseline transition intensities themselves. We therefore extend the existing methodology by incorporating covariate effects in a non-homogeneous Markov regression model that makes use of the time-transformation framework to accommodate non-homogeneity. In this model, we include covariate effects on transition rates by introducing a fixed effects regression model for each element of the baseline transition intensity matrix, $\mathbf{Q}_0$. Let q_{0abi} be the ab^{th} element of $\mathbf{Q}_{0i}$, the baseline transition intensity matrix for subject i . We specify a regression model of the form

$$q_{0abi} = q_{0ab} f(\mathbf{W}_{abi}; \boldsymbol{\alpha}_{ab}). \quad (3)$$

where $\mathbf{W}_{abi} = (W_{abi1}, \dots, W_{abip})'$ is a p -dimensional vector of covariates for subject i , $\boldsymbol{\alpha}_{ab}$ is a vector of regression parameters, and $f(\cdot; \cdot)$ is any positive function. We do not need to use the same covariates for all transition intensities but may wish to constrain some regression parameters to be equal in order to limit the dimensionality of the model. If $\mathbf{W}_{abi}$ includes time-varying covariates, then we must assume that covariates are constant on the interval between observations. This model is appropriate for non-time-varying covariates such as demographic variables and risk factors such as medications that are altered at the time of clinic visits.

Using the above non-homogeneous Markov regression model, we can construct the likelihood for the transition intensities, regression parameters, and time-transformation parameters as

$$L^{(m)}(\mathbf{Q}_0, \boldsymbol{\theta}, \boldsymbol{\alpha}) = \prod_{i=1}^m \left[P\{X(t_{i1}) = x_{i1}\} \prod_{j=2}^{n_i} \{\exp(\mathbf{Q}_{0i}(h(t_{ij}; \boldsymbol{\theta}) - h(t_{i(j-1)}; \boldsymbol{\theta})))\}_{x_{i(j-1)}x_{ij}} \right], \quad (4)$$

where m is the number of subjects and n_i is the number of observations for the i th subject.

Maximum likelihood estimators (MLEs) for covariate effects, elements of the baseline transition intensity matrix and time-transformation parameters will be consistent and asymptotically normal under the usual regularity conditions [8]. Details of the large sample properties for the time-transformation method are provided in Appendix A.

Closed form MLEs for $\mathbf{Q}_0$, $\boldsymbol{\theta}$, and $\boldsymbol{\alpha}$ are not available when exact transition times are not observed. However, estimates can be obtained via numerical maximization of the likelihood. Previous work has proposed the Fisher scoring algorithm for the homogeneous Markov process [14]. This method has been extended to the case of joint estimation of the baseline intensity matrix and the parameters of the time-transformation [11]. We use a modification of this algorithm to carry out estimation of the vector of regression parameters. An extension of the Fisher scoring algorithm to the case of estimation for the time-transformed Markov regression model is provided in Appendix B.

2.3 Piecewise homogeneous model

An alternative approach for addressing non-homogeneity in Markov processes is to divide the observation period into discrete intervals and assume that the process is temporally homogeneous within intervals. That is, we create a discrete partition of time from the process origin, $(t^{(0)}, t^{(1)}, \dots, t^{(p)})$ and assume that for $t^{(k-1)} < t_{i-1} < t_i \leq t^{(k)}$, $P(t_{i-1}, t_i) = \exp\{Q_k(t_i - t_{i-1})\}$, where Q_k is the transition intensity matrix applicable to observations in the k^{th} time interval. This method is quite flexible because it makes no assumptions about the type of variation in Q_k across the observation period. However, the complete transition intensity matrix must be estimated within each partition of the follow-up period. This method thus requires a large sample size in order to achieve precise estimates of the transition rates.

To estimate covariate effects in the piecewise homogeneous model, we assume covariates act in a common fashion on transition intensities across the observation period. That is,

$q_{abi}^{(k)} = q_{ab}^{(k)} f(W_{abi}; \alpha_{ab})$, where $q_{ab}^{(k)}$ is the ab th element of Q_k . The assumption that α_{ab} is independent of k could easily be relaxed. Scientific information about the process under study should be used to inform decisions about the reasonableness of the assumption of constant covariate effects across the follow-up period.

3 Applications

We carried out simulation studies to compare the performance of our proposed time-transformed Markov regression model to a homogeneous model and a piecewise homogeneous model. The objective of these simulations is to evaluate the performance of these models both when model assumptions are satisfied and when the data generating mechanism and the fitted model differ. Models were evaluated at a variety of sample sizes representative of those likely to be encountered in the study of multi-state diseases. We compared bias and efficiency of the three models under each simulated scenario. To demonstrate the practical application of these methods to a study of multi-state disease, we then applied our proposed model to data on AD progression from the National Alzheimer's Coordinating Center (NACC).

3.1 Simulation study

We evaluated the performance of three models: a homogeneous Markov regression model, a piecewise homogeneous model, and our proposed non-homogeneous Markov regression model under several different data generating mechanisms. The aim of these simulation studies was to evaluate the relative performance of each model under sample sizes likely to be encountered in studies of multi-state disease and to compare the ability of each to estimate covariate effects of risk factors for transitions among disease states both when the underlying model for transitions among disease states was correctly specified and when it was misspecified. For each model we evaluated bias, mean asymptotic standard errors based on the inverse Fisher information, and 95% confidence interval coverage probabilities for MLEs for covariate effects, α .

The number of subjects was varied across $m = 500, 1000$, and 2000 , and each subject was observed for $n = 4$ observation times. We chose to examine properties of model estimators for small numbers of observations per subject such as would be feasible to observe in a longitudinal observational study of a disease process. We assumed that the time between observations was distributed uniformly on $(0.2, 1)$. Results are based on 1000 simulated data sets.

We first generated data from a homogeneous Markov process with the following transition intensity matrix:

$$Q_i = \begin{pmatrix} -(0.1+0.2\alpha_1^{w_i}) & 0.2\alpha_1^{w_i} & 0.1 \\ 0.2\alpha_2^{w_i} & -(0.1+0.2\alpha_2^{w_i}) & 0.1 \\ 0 & 0 & 0 \end{pmatrix},$$

where w_i is a binary covariate. In our simulation study we set $\alpha_1 = 2$ and $\alpha_2 = 2$. The covariate in this example represents a risk factor that accelerates progression to state 2 as well as reversion to state 1. At the first observation time, subjects were equally likely to be observed in state 1 or 2.

Results for the three models when data arose from this homogeneous process are presented in Table 1. Under homogeneity all three models performed extremely similarly. Of note, no appreciable loss of efficiency was observed for either of the two non-homogeneous models relative to the homogeneous model. These results suggest that covariate effects can be estimated well using a non-homogeneous Markov regression model even when the underlying process is in fact homogeneous without sacrificing precision of the estimates.

We next simulated data from a piecewise homogeneous process. In these simulations, data arose from a model in which

$$Q_i(t) = \begin{pmatrix} -(0.1+0.2\alpha_1^{w_i}) & 0.2\alpha_1^{w_i} & 0.1 \\ 0.2\alpha_2^{w_i} & -(0.1+0.2\alpha_2^{w_i}) & 0.1 \\ 0 & 0 & 0 \end{pmatrix}, t \leq 2 = \begin{pmatrix} -(0.2+0.1\alpha_1^{w_i}) & 0.1\alpha_1^{w_i} & 0.2 \\ 0.1\alpha_2^{w_i} & -(0.2+0.1\alpha_2^{w_i}) & 0.2 \\ 0 & 0 & 0 \end{pmatrix}, t > 2.$$

This corresponds to a process in which the covariate, w , exerts the same effect on transition intensities throughout the observation period but subjects transition to the absorbing state more rapidly later in the observation period and move between states 1 and 2 more rapidly earlier in the observation period. In our simulations, $\alpha_1 = 2$ and $\alpha_2 = 2$. We note that under this data generating mechanism the time-transformed model is misspecified because it assumes that $Q(t) = Q_0 dh(t; \theta)/dt$. That is, all transition intensities are assumed to vary by the same multiplicative factor. This simulation allowed us to investigate whether covariate effects could be estimated using the time-transformed Markov regression model under this form of model misspecification.

Under the piecewise homogeneous data generating mechanism the piecewise model achieved the lowest percent bias across all three sample sizes. Efficiency of the three models was similar. At the smallest sample size, 500 subjects, percent bias did not differ substantially between the three models, suggesting that with relatively few observations there is little loss of performance associated with misspecification of temporal variations in the baseline transition intensities.

We next simulated data from a time-transformed non-homogeneous Markov process. Under this model we used the transition intensity matrix presented above for the homogeneous Markov process, but assumed that observations arose from this process after transforming the natural time-scale via the transformation function $h(t; \theta) = t\theta^t$ with $\theta = 1.4$.

At the smallest sample size, we found that there was no advantage to using one of the non-homogeneous Markov processes over using a homogeneous model. At this small sample size the piecewise non-homogeneous model returned extremely biased parameter estimates. As the sample size increased, we were able to better estimate α using the time-transformed

Markov process, and this model achieved the lowest percent bias for sample sizes of 1000 and 2000. The homogeneous process fit the data poorly at larger sample sizes and was substantially biased. Although the time-transformed model performed best at sample size 2000, the piecewise model was only slightly more biased. This suggests that for larger sample sizes, sufficient numbers of transitions are observed for this model to approximate variations in transition intensities.

3.2 Analysis of risk factors for changes in cognitive status in the NACC UDS

The National Alzheimer's Coordinating Center's (NACC) Uniform Data Set (UDS) is an ongoing longitudinal database of subjects seen at one of the National Institute on Aging's 29 funded Alzheimer's Disease Centers (ADC) located throughout the USA. Longitudinal follow-up began in 2005. Subjects seen at the ADCs represent a clinical case series of individuals who are either referred to the clinic for evaluation of dementia, self-refer to the clinic, or are recruited by clinics to participate in dementia research. Subjects are assessed on a variety of neuropsychological outcomes and risk factors including assessments of cognitive functioning, co-morbidities, and demographics at baseline and annual follow-up visits. Additional information on the structure of the UDS is available in [19] and [2].

We defined disease states using clinical diagnoses of cognitive functioning and AD. The cohort for this study was defined as all subjects age 50 and older included in the NACC UDS with normal cognition; a diagnosis of MCI affecting either the memory domain alone or memory domain and other cognitive domains; or a diagnosis of possible or probable AD. In this analysis, we defined AD as a diagnosis of possible or probable AD because definitive diagnosis of AD is possible only through neuropathological examination, which is not available for the majority of subjects. We also required that data from at least two clinic visits was available for each subject. At the time of analysis, 2724 subjects met inclusion criteria. Baseline characteristics for these subjects are presented in Table 4. The objective of our analysis was to characterize risk factor effects on the rate of transition among cognitive functioning states defined by normal cognition, MCI, and AD while accounting for the competing risk of death and age-related variation in the underlying transition rates.

We explored the use of non-homogeneous Markov regression models for estimating the rate of transition between cognitive states for the NACC UDS data. In our models we treated death as an absorbing state but allowed transitions between all other states. Follow-up visits for subjects were scheduled at approximately one year intervals. In our sample, mean follow-up time between observations was 1.05 years with minimum elapsed time between observations of 0.05 years and maximum of 2.58 years. Numbers of transitions between successive states are given in Table 5.

We first modeled the disease process using piecewise and time-transformed non-homogeneous models. We fit two time-transformed models allowing for variations in the effective time scale via a power transformation of the form $h(t) = t\theta^t$ and using a non-parametric locally weighted smoother [11]. The power model for non-homogeneity was investigated as a simple alternative that reflects the expectation of monotonically increasing rates of transition with respect to age. We contrast this with the non-parametric model which allows for more flexible variation in the rate of process evolution but uses a larger number of parameters to do so. In the piecewise model, we fit separate homogeneous processes for subjects less than the median age (76.7 years) and those older than the median age.

We assessed the fit of the models by comparing goodness-of-fit using the Bayesian Information Criterion (BIC). On the basis of BIC, the best fitting model was the non-parametric time-transformed model. Its BIC was 19.25 points lower than the power time-transformed model and 111.63 points lower than BIC for the piecewise homogeneous

model. BIC for both time-transformation models was smaller than for the piecewise model indicating that the increased flexibility provided by the piecewise model does not provide an improvement in model fit in this application.

We used the non-parametric time-transformed Markov regression model identified above as best fitting the UDS data to investigate the association between demographic covariates and transition rates among cognitive functioning states. In this model, we estimated the effect of covariates on baseline transition intensities using equation (3) with

$f(W_{abi}; \alpha_{ab}) = \exp(W_{abi}' \alpha_{ab})$. We investigated effects of history of dementia in a first degree relative and at least 16 years of education on transitions from MCI to normal and MCI to AD as well as the reverse of these transitions. For transitions from MCI to normal and MCI to AD we additionally investigated the effect of MCI subtype, classified as single or multiple domain MCI.

Results of this regression model are presented in Table 6. Regression parameter estimates reported are transition intensity ratios, multiplicative effects of covariates on baseline transition rates. Our model indicated a significantly decreased transition rate from MCI to normal cognition for subjects with multi-domain MCI when compared to those with single domain MCI. Other covariates investigated were not significantly associated with transition intensities.

The effect of covariates on transition intensities does not have great clinical relevance. We therefore further investigated whether this effect of multi-domain MCI on reversion to normal cognition translated into an effect on the one year probability of normal cognition or AD for subjects with MCI, a measure of greater clinical interest. Probability of transition from MCI to normal cognition and MCI to AD for subjects with single domain MCI and those with multi-domain MCI are presented in Figure 1. The observed decreased rate of transition from MCI to normal cognition for subjects with multi-domain MCI translates into a decreased one year transition probability from MCI to normal and a slightly increased probability of conversion from MCI to AD over most of the age range. The difference in conversion from MCI to AD is very modest and confidence bands are substantially overlapping, while the difference in transition from MCI to normal cognition is more marked

4 Discussion

We developed a non-homogeneous Markov regression model using the method of time-transformation and compared Markov regression models for estimating the effect of covariates on transitions between disease states via simulation studies. The proposed time-transformed Markov regression model provided relatively unbiased estimates of covariate effects even under misspecification of the form of the non-homogeneity in the process.

Under homogeneity, we found that the time-transformed model performed similarly to homogeneous and piecewise homogeneous models. No loss of efficiency was observed for either the time transformed model or the piecewise homogeneous model relative to the homogeneous model. This suggests that if the disease process is suspected to be non-homogeneous, there is no loss of efficiency associated with fitting a non-homogeneous process to the data.

When data were simulated from a non-homogeneous process, we found that the model that correctly specified the form of the temporal non-homogeneity tended to perform best. That is, we found that the time-transformed model slightly outperformed the piecewise model when data arose from a time-transformed Markov model and, similarly, that the piecewise

homogeneous model performed best when data arose from a piecewise homogeneous model. More importantly, we found that, for large sample sizes, both non-homogeneous models performed adequately even under misspecification. However, at a sample size of 500 we found that the piecewise model returned biased parameter estimates when the data arose from a time-transformed non-homogeneous process. Under this data generating mechanism the homogeneous model provided substantially biased estimates at sample sizes of 1000 and 2000.

Overall, our simulation study indicated that estimates of covariate effects can be substantially biased if an underlying temporal trend in transition rates is ignored and a homogeneous model is fit to non-homogeneous data. However, both the piecewise model and the time-transformation model adequately fit the data in most scenarios investigated. These results indicate that it is important to account for temporal non-homogeneity when investigating risk factor effects. However, the precise method of addressing non-homogeneity is less important and should be selected to best reflect our understanding of the expected behavior of the process. The time-transformation model may be preferred for smaller sample sizes because it requires estimation of fewer parameters. Estimation may be feasible using this model in cases where the piecewise model is impractical due to the large number of model parameters that must be estimated. Conversely, the piecewise model would be preferred in cases where it is unrealistic to believe that all transition rates vary over time at the same rate.

We used a non-homogeneous Markov regression model to estimate the effect of risk factors on rates of transition among cognitive functioning states using data from the NACC UDS. Past research addressing non-homogeneity in transitions between cognitive functioning states while allowing for reversibility of MCI has used either piecewise homogeneous models [26, 10, 24] or discrete time models [23]. Piecewise models are limited by the assumption of homogeneity within age strata. Additionally a large number of subjects or lengthy series of observations is necessary in order to estimate all transitions reliably. Discrete time models are limited by the necessity of equally spaced observation times and the assumption of at most only a single transition between successive observations.

Our non-parametric time-transformed Markov regression model indicated a decreased rate of transition from MCI to normal cognition for subjects with multi-domain MCI. Multi-domain MCI has been implicated as a more advanced pre-dementia stage than other MCI subtypes [3, 22, 1, 9]. Our data indicated that these subjects were less likely to experience improvements in cognition than subjects with an MCI confined to the memory domain. Several recent studies have also noted that subjects with MCI affecting multiple cognitive domains are less likely to revert to normal cognition [17, 16].

There are limitations to the data used in the application presented above. The UDS represents data for subjects seen in a clinical, academic setting. Our results are not necessarily comparable to those that would be obtained in a community based study. Additionally, because of the aims of the UDS, subjects with less severe disease were intentionally over-sampled [19]. As a result, our estimates of transition rates may differ from those that would be observed in the general population. Another limitation of these analyses is the possibility of misclassification in AD diagnoses. Because a definitive diagnosis is not possible prior to death we have used clinical classifications of AD. Our model therefore captures only transitions among clinically defined states.

Acknowledgments

This research was supported by National Institute on Aging grant U01AG016976. The authors also acknowledge the support of the National Alzheimer's Coordinating Center. Dr. Zhou is presently a Core Investigator and

Biostatistics Unit Director at the Northwest HSR&D Center of Excellence, Department of Veterans Affairs Medical Center, Seattle, WA. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

1. Alexopoulos P, Grimmer T, Preneczky R, Domes G, Kurz A. Progression to dementia in clinical subtypes of mild cognitive impairment. *Dementia and Geriatric Cognitive Disorders*. 2006; 22:27–34. [PubMed: 16679762]
2. Beekly D, Ramos E, Lee W, Deitrich W, Jacka M, Wu J, Hubbard J, Koepsell T, Morris J, Kukull W. The NIA Alzheimer's Disease Centers. The National Alzheimer's Coordinating Center (NACC) database: The Uniform Data Set. *Alzheimer Disease and Associated Disorders*. 2007; 21:249–258. [PubMed: 17804958]
3. Bozoki A, Giordani B, Heibebrink J, Berent S, Foster N. Mild cognitive impairments predict dementia in nondemented elderly patients with memory loss. *Archives of Neurology*. 2001; 58:411–416. [PubMed: 11255444]
4. Bruscoli M, Lovestone S. Is MCI really just early dementia? A systematic review of conversion studies. *International Psychogeriatrics*. 2004; 16:129–140. [PubMed: 15318760]
5. Commenges D, Joly P. Multi-state model for dementia, institutionalization, and death. *Communications in Statistics - Theory and Methods*. 2004; 33:1315–1326.
6. Commenges D, Joly P, Letenneur L, Dartigues J. Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statist. Med*. 2004; 23:199–210.
7. Cox, D.; Miller, H. *The Theory of Stochastic Processes*. New York: Chapman and Hall; 1965.
8. Cramer, H. *Mathematical Methods of Statistics*. Princeton: Princeton University Press; 1946.
9. Gabryelewicz T, Styczynska M, Luczywek E, Barczak A, Pfeffer A, Androsiuk W, Chodakowska-Zebrowska M, Wasiaik B, Peplonska B, Barcikowska M. The rate of conversion of mild cognitive impairment to dementia: predictive role of depression. *International Journal of Geriatric Psychiatry*. 2007; 22:563–567. [PubMed: 17136705]
10. Harezlak J, Gao S, Hui S. An illness-death stochastic model in the analysis of longitudinal dementia data. *Statist. Med*. 2003; 22:1465–1475.
11. Hubbard R, Inoue L, Fann JR. Modeling a non-homogeneous Markov process via time-transformation. *Biometrics*. 2008; 64:843–850. [PubMed: 18047532]
12. Hubbard R, Inoue L, Diehr P. Joint modeling of self-rated health and changes in physical functioning. *Journal of the American Statistical Association*. 2009; 104:873–885.
13. Ishikawa T, Ikeda M. Mild cognitive impairment in a population-based epidemiological study. *Psychogeriatrics*. 2007; 7:104–108.
14. Kalbeisch JD, Lawless JF. The analysis of panel data under a Markov assumption. *J. Am. Statist. Ass.* 1985; 80:863–871.
15. Larrieu S, Letenneur L, Orgogozo JM, Fabrigoule C, Amieva H, Le Carret N, Barberger-Gateau P, Dartigues JF. Incidence and outcome of mild cognitive impairment in a population-based prospective cohort. *Neurology*. 2002; 59:1594–1599. [PubMed: 12451203]
16. Loewenstein D, Acevedo A, Small B, Agron J, Crocco E, Duara R. Stability of different subtypes of mild cognitive impairment among the elderly over a 2-to 3-year follow-up period. *Dementia and Geriatric Cognitive Disorders*. 2009; 27:418–423. [PubMed: 19365121]
17. Manly J, Tang M, Schupf N, Stem Y, Vonsattel J, Mayeux R. Frequency and course of mild cognitive impairment in a multiethnic community. *Annals of Neurology*. 2008; 63:494–506. [PubMed: 18300306]
18. Marshall G, Jones R. Multistate models and diabetic retinopathy. *Statist. Med*. 1995; 14:1975–1983.
19. Morris J, Weintraub S, Chui H, Cummings J, DeCarli C, Ferris S, Foster N, Galasko D, Graff-Radford N, Peskind ER, Beekly D, Ramos E, Kukull W. The Uniform Data Set (UDS): Clinical and cognitive variables and descriptive data from Alzheimer disease centers. *Alzheimer Disease and Associated Disorders*. 2006; 20:210–216. [PubMed: 17132964]

20. Pan S, Wu H, Yen A, Chen T. A Markov regression random-effects model for remission of functional disability in patients following a first stroke: A Bayesian approach. *Statist. Med.* 2007; 26:5335–5353.
21. Petersen R, Smith G, Waring S, Ivnik R, Tangalos E, Kokmen E. Mild cognitive impairment - clinical characterization and outcome. *Archives of Neurology.* 1999; 56:303–308. [PubMed: 10190820]
22. Rasquin S, Lodder J, Visser P, Lousberg R, Verhey F. Predictive accuracy of MCI subtypes for Alzheimer's disease and vascular dementia in subjects with mild cognitive impairment: A 2-year follow-up study. *Dementia and Geriatric Cognitive Disorders.* 2005; 19:113–119. [PubMed: 15591801]
23. Salazar J, Schmitt F, Yu L, Mendiondo M, Kryscio R. Shared random effects analysis of multi-state Markov models: application to a longitudinal study of transitions to dementia. *Statist. Med.* 2007; 26:568–580.
24. van den Hout A, Matthews F. A piecewise-constant Markov model and the effects of study design on the estimation of life expectancies in health and ill health. *Statistical Methods in Medical Research.* 2009; 18:145–162. [PubMed: 18445698]
25. Yen A, Chen T. Mixture multi-state Markov regression model. *Journal of Applied Statistics.* 2007; 34:11–21.
26. Yesavage J, O'Hara R, Kraemer H, Noda A, Taylor J, Ferris S, Gely-Nargeot M, Rosen A, Friedman L, Sheikh J, Derouesne C. Modeling the prevalence and incidence of Alzheimer's disease and mild cognitive impairment. *Journal of Psychiatric Research.* 2002; 36:281–286. [PubMed: 12127595]
27. Albert A. Estimating the infinitesimal generator of a continuous time, finite state Markov process. *Annals of Mathematical Statistics.* 1962; 33:727–753.
28. Billingsley, P. *Statistical inference for Markov Processes.* Chicago: University of Chicago Press; 1961.

Modeling risk factors for Alzheimer's disease progression using a non-homogeneous Markov process

Appendix A: Large sample properties

Asymptotics for continuous time Markov processes were developed by Billingsley (1961) [28]. However, for these results to hold, we require that the process be ergodic. Because many of the processes of interest in studies of human health include an absorbing state representing death, we formulate asymptotics for transition intensity and regression parameter estimates following the method of Cramer (1946) [8].

Theorem 0.1 *Let ψ denote the functionally independent elements of $\mathbf{Q}_0$, θ , and α and $L^{(m)}(\psi)$ denote the likelihood based on m independent realizations of a time transformed homogeneous Markov process with s states. Each series takes realized values $(x_{i1}, \dots, x_{im})$, $i = 1, \dots, m$ with positive, non-decreasing time transformation function $h(t; \theta)$ and true*

parameter values ψ_0 . If $\frac{\partial \log L^{(m)}(\psi)}{\partial \psi_{ij}}$, $\frac{\partial^2 \log L^{(m)}(\psi)}{\partial \psi_{ij} \partial \psi_k}$, and $\frac{\partial^3 \log L^{(m)}(\psi)}{\partial \psi_{ij} \partial \psi_k \partial \psi_l}$ satisfy the conditions of [8] for all $\psi_{ij}, \psi_k, \psi_l \in \mathbf{Q}_0$; $f(\mathbf{W}; \alpha)$ and $h(t; \theta)$ have third order partial derivatives with respect to α and θ respectively that are bounded by integrable functions; and $P(X(t) = a) > 0$ for some $t > 0$ and all $a \in \{1, \dots, s\}$, then $\{\hat{\psi} : L^{(m)}(\hat{\psi}) > L^{(m)}(\psi), \forall \psi \in \Psi\}$ is a consistent estimator for ψ_0 and

$$m^{1/2}(\hat{\psi}^{(m)} - \psi_0) \sim N(\mathbf{0}, I^{-1}(\psi_0)).$$

Proof Following the method of Cramer (1946) [8], $\hat{\psi}$ achieves asymptotic normality if $\partial \log L^{(m)}(\psi)/\partial \psi_k$, $\partial^2 \log L^{(m)}(\psi)/\partial \psi_k \partial \psi_{k'}$, $\partial^3 \log L^{(m)}(\psi)/\partial \psi_k \partial \psi_{k'} \partial \psi_{k''}$ exist for all $\psi_k, \psi_{k'}, \psi_{k''} \in \psi$ and are bounded by integrable functions. We can see that the log-likelihood

$$\log L^{(m)}(\psi) = \sum_{i=1}^m \left\{ \log P(X(t_{i1})=x_{i1}) + \sum_{j=2}^n \log \left\{ e^{\mathbf{Q}_{0i}(h(t_{ij};\theta)-h(t_{i(j-1)};\theta))} \right\}_{x_{i(j-1)}, x_{ij}} \right\}$$

has third order partial derivatives with respect to the elements of $\mathbf{Q}_{0i}$ by rewriting the matrix exponential according to its definition as a power series. We can then see that

$$\begin{aligned} \frac{\partial}{\partial q_{ab}} e^{\mathbf{Q}_{0i}(h(t_{ij};\theta)-h(t_{i(j-1)};\theta))} \\ &= \sum_{s=0}^{\infty} \frac{\partial}{\partial q_{ab}} \frac{(\mathbf{Q}_{0i}(h(t_{ij};\theta)-h(t_{i(j-1)};\theta)))^s}{s!} \\ &= \sum_{s=0}^{\infty} \frac{(h(t_{ij};\theta)-h(t_{i(j-1)};\theta))^s}{s!} \frac{\partial}{\partial q_{ab}} \mathbf{Q}_{0i}^s \\ &= \sum_{s=0}^{\infty} \frac{(h(t_{ij};\theta)-h(t_{i(j-1)};\theta))^s}{s!} \frac{\partial}{\partial q_{ab}} \sum_{r=0}^{s-1} \mathbf{Q}_{0i}^r \mathbf{J}_{ab} \mathbf{Q}_{0i}^{s-r-1}, \end{aligned}$$

where $\mathbf{J}_{ab}$ is a matrix of the same dimension as $\mathbf{Q}_{0i}$ with all elements zero except for a one in the a^{th} row and b^{th} column.

If additionally $h(t; \theta)$ and $f(\mathbf{W}; \alpha)$ have bounded third order partial derivatives, then the conditions of [8] are satisfied and the MLEs are asymptotically normal and consistent for fixed n as m tends to infinity.

As first noted by Albert (1962) [27], if $P(X(t) = a) = 0$ for all $t > 0$ for some state a then the likelihood contains no information about transitions from state a to any other state. The information matrix is singular in this case. We thus require positive probability that all states are occupied at some time in order to ensure non-singularity of the information matrix.

Appendix B: Maximum likelihood estimation using the Fisher scoring algorithm

Closed form maximum likelihood estimators (MLEs) for $\mathbf{Q}_0$, θ , and α are not available when exact transition times are not observed. However, estimates can be obtained via numerical maximization of the likelihood. Kalbeisch (1985) [14] proposed the Fisher scoring algorithm for the homogeneous Markov process. Hubbard (2008) [11] extended this method to the case of joint estimation of the baseline intensity matrix and the parameters of the time transformation. We further extend the algorithm to include joint estimation of the vector of regression parameters.

Let ψ be the vector of functionally independent elements of $\mathbf{Q}_0$, θ , and α . Given initial estimates, $\psi^{(0)}$, at the $(k+1)^{st}$ step an estimate for the MLEs is

$$\hat{\psi}^{(k+1)} = \hat{\psi}^{(k)} + \hat{\mathcal{J}}_m(\hat{\psi}^{(k)})^{-1} \mathbf{S}_m(\hat{\psi}^{(k)}),$$

with the algorithm being iterated until convergence. In the above, $\mathcal{Q}_m(\boldsymbol{\psi})$ is the expected information matrix and $\mathbf{S}_m(\boldsymbol{\psi})$ is the score function for m subjects.

For subject i the a^{th} element of the score function is given by

$$\mathbf{S}^{(a)}(\boldsymbol{\psi}) = \frac{\partial}{\partial \boldsymbol{\psi}_a} \sum_{j=2}^{n_i} \log(p_{x_{i(j-1)}, x_{ij}}(h(t_{ij}) - h(t_{i(j-1)}))) = \sum_{j=2}^{n_i} \frac{1}{(p_{x_{i(j-1)}, x_{ij}}(h(t_{ij}) - h(t_{i(j-1)})))} \frac{\partial}{\partial \boldsymbol{\psi}_a} p_{x_{i(j-1)}, x_{ij}}(h(t_{ij}) - h(t_{i(j-1)})).$$

The ab^{th} element of the expected information matrix $\{\mathcal{Q}(\boldsymbol{\psi})\}_{uv}$ is given by

$$\{\mathcal{Q}(\boldsymbol{\psi})\}_{ab} = - \sum_{j=2}^n E \left[\frac{1}{p_{x_{i(j-1)}, x_{ij}}^2 \{h(t_{ij}) - h(t_{i(j-1)})\}} \frac{\partial}{\partial \boldsymbol{\psi}_a} p_{x_{i(j-1)}, x_{ij}} \{h(t_{ij}) - h(t_{i(j-1)})\} \times \frac{\partial}{\partial \boldsymbol{\psi}_b} p_{x_{i(j-1)}, x_{ij}} \{h(t_{ij}) - h(t_{i(j-1)})\} \right].$$

The advantage of using the Fisher scoring algorithm over other Newton-Raphson numerical maximization techniques is that the expected information matrix can be estimated using only the expressions for the first derivatives of the log-likelihood. The ab^{th} element of the estimated expected information matrix based on m subjects is given by

$$\left\{ \widehat{\mathcal{J}}_m(\widehat{\boldsymbol{\psi}}^{(k)}) \right\}_{ab} = - \sum_{i=1}^m \sum_{j=2}^{n_i} \left[\frac{1}{p_{x_{i(j-1)}, x_{ij}}^2 \{h(t_{ij}) - h(t_{i(j-1)})\}} \frac{\partial}{\partial \boldsymbol{\psi}_a} p_{x_{i(j-1)}, x_{ij}} \{h(t_{ij}) - h(t_{i(j-1)})\} \times \frac{\partial}{\partial \boldsymbol{\psi}_b} p_{x_{i(j-1)}, x_{ij}} \{h(t_{ij}) - h(t_{i(j-1)})\} \right].$$

By using the Fisher scoring method to obtain MLEs we also obtain estimates of the inverse Fisher information which provides an estimate of the variance. Once we have obtained estimates for $\mathbf{Q}_0$, $\boldsymbol{\theta}$, and $\boldsymbol{\alpha}$ and their variance-covariance matrix we can also derive other measures of clinical interest such as transition probabilities or mean time to first transition to a state of clinical interest. Variance estimates for derived measures such as conversion probabilities are available via the delta method.

Figure 1.
One year probability of conversion from MCI to normal cognition (left) and MCI to AD (right) with pointwise 95% confidence bands (dashed) for subjects with multi-domain MCI (black) and those with single domain MCI (grey).

Table 1

Model performance when data arise from a homogeneous Markov process with a sample size of m subjects. Performance was evaluated using percent bias, estimated standard errors (SE), empirical standard errors (ESE), and 95% confidence interval coverage probabilities (95% CP) for regression parameter estimates.

m	Fitted model	Percent Bias		SE		ESE		95% CP	
		a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
500	Homogeneous	0.75	0.60	0.1888	0.1890	0.1930	0.1964	95.00	93.60
1000	Homogeneous	1.00	0.44	0.1333	0.1331	0.1357	0.1337	95.00	95.20
2000	Homogeneous	0.65	0.63	0.0941	0.0940	0.0941	0.0928	94.90	95.20
500	Piecewise	0.84	0.69	0.1889	0.1891	0.1935	0.1965	94.80	93.70
1000	Piecewise	1.07	0.50	0.1333	0.1332	0.1358	0.1339	95.10	95.00
2000	Piecewise	0.69	0.68	0.0941	0.0941	0.0941	0.0928	94.90	95.10
500	Time-transformed	0.83	0.69	0.1889	0.1891	0.1933	0.1965	95.00	93.60
1000	Time-transformed	1.06	0.49	0.1333	0.1332	0.1357	0.1337	95.10	95.10
2000	Time-transformed	0.68	0.66	0.0941	0.0941	0.0941	0.0928	94.80	95.20

Model performance when data arise from a piecewise homogeneous Markov process with a sample size of m subjects. Performance was evaluated using percent bias, estimated standard errors (SE), empirical standard errors (ESE), and 95% confidence interval coverage probabilities (95% CP) for regression parameter estimates.

Table 2

m	Fitted model	Percent Bias		SE		ESE		95% CP	
		α_1	α_2	α_1	α_2	α_1	α_2	α_1	α_2
500	Homogeneous	1.99	1.12	0.2152	0.2159	0.2083	0.2153	96.50	95.80
1000	Homogeneous	1.49	2.74	0.1520	0.1519	0.1519	0.1509	95.20	95.00
2000	Homogeneous	2.50	1.53	0.1073	0.1072	0.1065	0.1023	93.90	95.70
500	Piecewise	0.64	1.34	0.2172	0.2180	0.2122	0.2170	96.40	95.90
1000	Piecewise	0.95	0.33	0.1533	0.1532	0.1549	0.1538	94.50	95.30
2000	Piecewise	0.14	0.89	0.1081	0.1081	0.1071	0.1032	95.10	96.30
500	Time-transformed	1.74	0.89	0.2151	0.2158	0.2088	0.2163	96.50	95.40
1000	Time-transformed	1.30	2.55	0.1519	0.1518	0.1522	0.1514	95.20	95.10
2000	Time-transformed	2.34	1.37	0.1072	0.1071	0.1069	0.1027	93.80	95.80

Table 3

Model performance when data arise from a non-homogeneous time-transformed Markov process with a sample size of m subjects. Performance was evaluated using percent bias, estimated standard errors (SE), empirical standard errors (ESE), and 95% confidence interval coverage probabilities (95% CP) for regression parameter estimates.

m	Fitted model	Percent Bias		SE		ESE		95% CP	
		a_1	a_2	a_1	a_2	a_1	a_2	a_1	a_2
500	Homogeneous	0.20	1.15	0.8140	0.8147	0.4319	0.4346	94.40	94.30
1000	Homogeneous	5.70	5.91	0.2794	0.2792	0.2617	0.2656	92.90	93.30
2000	Homogeneous	8.15	8.32	0.1563	0.1562	0.1583	0.1608	91.20	91.30
500	Piecewise	14.78	13.94	0.4782	0.4783	0.9451	0.9514	96.60	96.20
1000	Piecewise	3.64	3.45	0.2377	0.2376	0.2564	0.2612	96.40	96.20
2000	Piecewise	1.44	1.25	0.1552	0.1551	0.1530	0.1538	95.70	95.30
500	Time-transformed	3.04	2.17	0.4388	0.4390	0.3351	0.3396	97.10	96.10
1000	Time-transformed	2.83	2.64	0.2204	0.2203	0.2137	0.2179	96.40	96.70
2000	Time-transformed	1.28	1.07	0.1493	0.1492	0.1463	0.1474	96.10	95.80

Table 4

Baseline characteristics for 2724 subjects meeting study inclusion criteria.

	Mean	(SD)
Age (years)	76.2	(8.9)
Education (years)	14.9	(3.4)
	N	(%)
Female	1546	(56.8)
Dementia in first degree relative	1654	(61.7)
Cognitively normal	1195	(43.9)
MCI	435	(16.0)
Single domain *	263	(60.5)
Multi-domain *	172	(39.5)
Poss. or prob. AD	1094	(40.2)

* Percentages for MCI sub-types are computed out of all subjects with MCI.

Table 5

Number of transitions between cognitive functioning states at successive clinic visits. Rows represent starting states and columns represent state at subsequent visit.

	Normal	MCI	AD	Death
Normal	1103	48	9	43
MCI	43	281	86	29
AD	5	12	922	178

Table 6

Multiplicative effect of covariates on transition intensities in non-parametric time-transformed Markov regression model between clinically defined states. Effects reported are transition intensity ratios (QR). Family history is defined as dementia reported in a first degree relative.

	95% CI		
	QR	LCL	UCL
Normal → MCI			
Family history	1.808	0.963	3.397
Education	1.501	0.823	2.739
MCI → Normal			
Family history	1.419	0.730	2.760
Multi-domain MCI	0.358	0.176	0.729
Education	1.522	0.820	2.824
MCI → AD			
Family history	0.995	0.668	1.483
Multi-domain MCI	0.787	0.525	1.180
Education	0.994	0.671	1.473
AD → MCI			
Family history	2.149	0.515	8.968
Education	0.702	0.189	2.605