Published in final edited form as:

Stat Med. 2010 March 15; 29(6): 639–648. doi:10.1002/sim.3828.

A nonstationary Markov transition model for computing the relative risk of dementia before death

Lei Yu, PhD^{1,2}, William S. Griffith, PhD³, Suzanne L. Tyas, PhD⁴, David A. Snowdon, PhD⁵, and Richard J. Kryscio, PhD^{3,6,*}

¹ Rush Alzheimer's Disease Center, Rush University Medical Center

² Department of Neurological Sciences, Rush University Medical Center

³ Department of Statistics, University of Kentucky

⁴ Department of Health Studies and Gerontology and Department of Psychology, University of Waterloo

⁵ Sanders-Brown Center on Aging and Department of Neurology, University of Kentucky

⁶ Department of Biostatistics, University of Kentucky

Abstract

This paper investigates the long-term behavior of the k -step transition probability matrix for a nonstationary discrete time Markov chain in the context of modeling transitions from intact cognition to dementia with mild cognitive impairment (MCI) and global impairment (GI) as intervening cognitive states. The authors derive formulas for the following absorption statistics: (1) the relative risk of absorption between competing absorbing states, and (2) the mean and variance of the number of visits among the transient states before absorption. Since absorption is not guaranteed, sufficient conditions are discussed to ensure that the substochastic matrix associated with transitions among transient states converges to zero in limit. Results are illustrated with an application to the Nun Study, a cohort of 678 participants, 75 to 107 years of age, followed longitudinally with up to ten cognitive assessments over a fifteen-year period.

Keywords

Nonhomogeneous Markov Chain; Transition Model; Random Effect; Absorption Statistics; Nun Study; Dementia; Mild Cognitive Impairment

1. Introduction

The natural development of a chronic disease is often expressed in terms of distinct health stages and a Markov chain is a simple yet powerful tool for modeling the progression of individuals through these stages [1,2]. Muenz and Rubinstein [3] modeled the transitions of a two-state ergodic chain using two logistic regressions where covariates are permitted to change with time. Kay [4] proposed a stochastic process to investigate the relationship between survival time and the health states with death the only absorbing state. Chen et al. [5] introduced a non-homogeneous exponential regression stochastic model to estimate the transition parameters for multi-state disease progression with no backward transition. Li and Chan [6] used a continuous Markov chain to examine the transition rate among multinomial

*Corresponding Author. Tel: (859) 257-4064 ext 295; fax: (859) 257-4665. kryscio@email.uky.edu (R. J. Kryscio).

responses all of which are transient. These approaches do not apply to the situation where the chain involves multiple competing absorbing states. Salazar et al. [7] proposed a stationary multi-state Markov model with shared random effects and multiple absorbing states. Incorporating the transition modeling strategy into a generalized linear mixed model (GLMM) facilitates the identification of risk factors associated with transitions among states and provides an expression for the joint distribution of the response vector for a subject while accounting for the dependency among observations on the same subject. Yu et al. [8] continued the investigation by incorporating the baseline distribution into the followup likelihood. The purpose of this paper is to investigate the role of time dependent risk factors such as age in Salazar's model. Briefly, the emphasis is on the long run behavior of the underlying Markov chain which is no longer assumed to be stationary. We show that in this case the fundamental matrix of the chain needs to be replaced by an infinite matrix series whose convergence status needs a closer examination before absorption statistics can be computed.

2. Motivation for Absorption Statistics

Salazar et al. [7] emphasized estimation of the one-step transition matrix because in that article we were among the first to introduce a shared random effect to correlate the observations in the vector of responses for an individual and because in the application to the BRAiNS cohort data fewer than 10% of the participants transitioned to dementia. To estimate the one-step transition matrix we wrote our own program to approximate the likelihood function, which is defined by an integral. A simulation study [7] showed that the approximation to this integral produced reasonable estimates of the unknown model parameters in the one-step transition matrix, and that these parameter estimates are robust across a spectrum of distributions for the shared random effect. Since then, several developments have occurred. The first is that with the advent of new statistical procedures, including NLMIXED in SAS [9] and GLLAMM in STATA [10], the programming involved in evaluating the one-step transition matrix by approximating an integral has been greatly simplified. In addition, recent research findings by Litieri et al. [11] validated the findings in the simulation study.

In this paper, we apply our methodology to the well-known Nun Study data. The two datasets, BRAiNS and the Nun Study, are similar in that serial cognitive assessments of cohort participants define a Markov chain having three transient states and two absorbing states; the transient states are defined differently in each cohort but the absorbing states are the same: dementia and death. In the application of our methodology to these two data sets, however, we discovered important differences that motivate the current investigation. The Nun Study cohort is older; the mean baseline age of a nun is 83, ten years older than the mean baseline age in BRAiNS. Consequently, the event rate in the Nun Study is significantly higher than that in the BRAiNS cohort, as shown in Table 1. This is true even accounting for the different follow-up periods between these two studies, reflecting the strong association of age with dementia and death.

The analysis of the Nun Study data focuses on this disparity. Recently Yu et al. [8] noted that the likelihood function proposed by Salazar et al. [7] conditions on the baseline status of the individual but that, under a shared random effects model, separating the baseline distribution from the overall model likelihood can lead to underestimation of the effects of risk factors on the one-step transitions. In the Nun Study, the attenuation of effects is substantial due to the presence of 77 participants who were demented at baseline and heterogeneous baseline diagnostics for the rest participants.

In Section 3, we review the definition of this amended likelihood and report its fit to the Nun Study data. Since events occur relatively quickly in that data set, we are able to focus on the computation of the relative risk of dementia before death, a result of more relevance to the field of dementia than the one-step transition probabilities. We end that section by reviewing the definition of this relative risk in a stationary finite Markov chain with absorbing states. Since all transitions depend heavily on age, the Markov chain for the Nun Study is nonstationary. To our knowledge there is virtually no literature for computing the absorption statistics in this case. In Section 4, we state formulae for computing the relative risk as well as the moments of the number of transitions until absorption as functions of the elements of the one-step transition matrices in the nonstationary case. Those readers more interested in the application than in the derivation of this methodology are directed to Section 5, where these absorption statistics are computed and discussed for the Nun Study data.

3. The Extended Multi-state Markov Model and its Application

The extended multi-state Markov model introduced in Yu et al. [8] provides a suitable framework for identifying the risk factors associated with the progression of healthy individuals to a chronic disease with death treated as a competing event. Let $\mathbf{Y}_i = (y_{i1}, \dots, y_{in_i})$ denote the vector of observations for subject i where y_{ij} denotes the cognitive state of the subject at visit j . Assuming the Markov property the joint distribution of this vector can be written as

$$f(y_{i1}, \dots, y_{in_i}) = f(y_{i1})f(y_{i2}|y_{i1})f(y_{i3}|y_{i2}) \dots f(y_{in_i}|y_{in_i-1}) \quad (3.1)$$

Each conditional probability $f(y_{ik}|y_{ik-1})$ can be interpreted as a particular element inside a one-step transition matrix for the Markov process. More specifically, suppose $y_{ik-1} = s$ and $y_{ik} = v$. Then $f(y_{ik} = v|y_{ik-1} = s)$, denoted by $P_{sv}(\mathbf{X}_k, \gamma)$, is simply the probability of transition for subject i from state $y_{ik-1} = s$ at $k-1$ th visit to $y_{ik} = v$ at k th visit, where s and v are elements of finite transition states within a particular multi-state system. Assume a finite stochastic process consisting of five transition states with three transient and two competing absorbing states. Given observed covariate vector $\mathbf{X}_k$ and unobserved random effect γ , we let

$$P_{sv}(\theta|\mathbf{X}_k, \gamma) = \begin{cases} \frac{1}{1 + \sum_{h=2}^5 \exp(\alpha_h + \mathbf{X}_k' \beta_h + \xi_h(s) + \gamma)} & \text{for } v=1 \\ \frac{\exp(\alpha_v + \mathbf{X}_k' \beta_v + \xi_v(s) + \gamma)}{1 + \sum_{h=2}^5 \exp(\alpha_h + \mathbf{X}_k' \beta_h + \xi_h(s) + \gamma)} & \text{for } v=2, \dots, 5 \end{cases} \quad (3.2)$$

The first equation applies for $v=1$, and the second for $v = 2, \dots, 5$. Here θ represents the vector of all unknown parameters $(\alpha|\beta|\xi(s))$, in which α is the vector of intercepts; β is the vector of unknown fixed effects for covariates of interest; and $\xi(s)$ is the vector of unknown fixed effects for the prior state s .

In addition, let $\pi_j = P(y_{i1} = j)$ represent the probability that subject i is in some state j at the baseline. The probability of the baseline status is similarly parameterized by a separate multinomial logistic regression, which gives

$$\pi_j(\phi_j | \mathbf{X}_B, \gamma) = \begin{cases} \frac{1}{1 + \sum_{h=2}^4 \exp(\tau_h + \mathbf{X}'_B \delta_h + \gamma)} \\ \frac{\exp(\tau_j + \mathbf{X}'_B \delta_j + \gamma)}{1 + \sum_{h=2}^4 \exp(\tau_h + \mathbf{X}'_B \delta_h + \gamma)} \end{cases} \quad (3.3)$$

Again the first equation applies for $j=1$, and the second for $j = 2, \dots, 4$. The vector $\phi_j \equiv (\tau_j \parallel \delta_j)$ represents the vector of unknown parameters determining the baseline probabilities. $\mathbf{X}_B$ refers to the observed covariate vector at baseline.

If N is the total number of the subjects in the study, let $\mathbf{Y} = (\mathbf{Y}_1, \dots, \mathbf{Y}_N)$ denote the entire data set. The following marginal likelihood will yield maximum likelihood estimates (MLEs) of the unknown parameters, assuming a random intercept γ distributed $N(0, \sigma^2)$

$$L(\boldsymbol{\theta}, \phi, \sigma^2 | \mathbf{Y}) = \int \prod_{i=1}^N \prod_{k=2}^{n_i} \prod_{s=1 \dots 3, v=1 \dots 5} (P_{sv}(\boldsymbol{\theta} | \mathbf{X}_k, \gamma))^{\delta_{y_{ik-1}, s} \delta_{y_{ik}, v}} \pi_{y_{i1}}(\phi | \mathbf{X}_B, \gamma) dF(\gamma) \quad (3.4)$$

where $\delta_{y_{ik-1}, s}$ and $\delta_{y_{ik}, v}$ are indicator functions valued at 1 if $y_{ik-1} = s$ and $y_{ik} = v$ and 0 otherwise. Notice that the likelihood function includes two components, with $\pi_{y_{i1}}(\phi | \mathbf{X}_B, \gamma)$

$$\prod_{k=2}^{n_i} \prod_{s=1 \dots 3, v=1 \dots 5} (P_{sv}(\boldsymbol{\theta} | \mathbf{X}_k, \gamma))^{\delta_{y_{ik-1}, s} \delta_{y_{ik}, v}}$$

representing the baseline status and representing the sequence of ensuing one-step transitions, and the two likelihood components sharing the same random effect.

We fit this likelihood function to the Nun Study data, a longitudinal study of aging and Alzheimer's disease. The cohort consists of 678 members of the School Sisters of Notre Dame religious congregation [12]. Each participant agreed to allow investigators complete access to their convent archives, participate in near-annual assessments of cognitive and physical function, and donate their brain at death. Transitions among states are modeled by a discrete-time Markov chain having three transient (intact cognition, mild cognitive impairment or MCI, and globally impairment or GI) and two competing absorbing states (death and dementia). The flowing diagram of the transitions is presented in Figure 1 [13].

The first 10 waves consisting of cognitive assessments and measures of activities of daily living are used in this investigation. The transitions among the states of the process are listed in Table 2. We let transition probabilities depend on three covariates: age, the presence/absence of an apolipoprotein E-ε4 allele (APOE4, a common genetic risk factor for Alzheimer's disease), and levels of education. The dependency among repeated follow-up waves within the same subject is captured using a random intercept. There were 177 participants being excluded from the analysis due to missing covariates or consent withdrawal.

The analysis was implemented in SAS using the NLMIXED procedure [14]. The integral is approximated numerically and the dual quasi-Newton algorithm is used to optimize the log-

likelihood function. Table 3 lists the MLEs and standard errors for the model. Transitions are significantly affected by age, level of education and the genetic factor APOE4.

Given that events occur quickly in this cohort, the absorption statistics are of particular interest in this multi-state Markov model. By modeling the transition among states with a Markov structure, the one-step transition probability matrices constructed based on the parameter estimates provide additional information with respect to the probability of absorption into dementia or death (before dementia) as well as the moments associated with the number of transitions until absorption.

Following the notation in Bhat and Miller [15], consider a finite Markov chain with S states, the first T of which are transient and the remainder absorbing. Denote the k -step transition

matrix of the chain by $\mathbf{P}^{(m,m+k)} = \mathbf{P}^{(m,m+1)} \times \dots \times \mathbf{P}^{(m+k-1,m+k)} = \prod_{i=1}^k \mathbf{P}^{(m+i-1,m+i)}$, where m is

the initial starting time. Let $\mathbf{T}$ be the set of transient states and $N_{ij}^{(m)}$, $i, j \in \mathbf{T}$ be the random variable representing the number of times the process visits j before it eventually enters an absorbing state, having initially started from state i at time m .

In the stationary case, the above transition matrix does not depend on m , and we have $\mathbf{P}^{(m,m+k)} = \mathbf{P}^k$. Also, in this case absorption is a certain event, it occurs geometrically fast, and the mean and variance of the number of visits before absorption are always finite [15]. The key component in calculating absorption statistics in that case is the fundamental matrix $(\mathbf{I} - \mathbf{Q})^{-1}$, where $\mathbf{Q}$ is the substochastic matrix describing transitions among the transient states and $\mathbf{I}$ is the T by T identity matrix. Let $\mathbf{F} = (\mathbf{I} - \mathbf{Q})^{-1} \mathbf{R}$ where $\mathbf{R}$ is the T by $(S-T)$ submatrix containing the one-step transition probabilities from a transient to an absorbing state. $\mathbf{F}$ contains the probability of eventual absorption into a given absorbing state from an initial transient state. If there are only two absorbing states, the ratio of the element in the first column to the corresponding element in the second column of $\mathbf{F}$ defines the relative risk of absorption into dementia before death for that given row or initial transient state.

These simple formulas do not apply to the Nun Study data because the one-step transition probabilities depend on age, making the underlying chain nonstationary. In the next section we provide formula for computing $\mathbf{F}^{(m)}$ (which now depends on m) and the first two

moments of $N_{ij}^{(m)}$ in the nonstationary case. These new formulas are applied to the Nun Study data in Section 5.

4. Absorption Statistics for a Nonhomogeneous Chain

Stationarity fails in the nonhomogeneous Markov chain. The transition probabilities among states change with time and as a result the time-dependent transition diagram makes it

difficult to formulate the moments associated with $N_{ij}^{(m)}$, $i, j \in \mathbf{T}$. Moreover, the absorption of transient states is not certain. Unless otherwise noted, the nonhomogeneous discrete time Markov chain considered here consists of either absorbing states or transient states at any time, with each state defined as follows.

Definition 1

A state j is absorbing if and only if $\forall m, m \in \mathbf{N}^*$, $P_{jj}^{(m,m+1)} \equiv 1$; with $\mathbf{N}^*$ being the set of non-negative integers.

Definition 2

A state i is transient if $\forall m \in \mathbb{N}^*, \exists$ a finite number of transitions that allows the process to go from i to an absorbing state with a positive probability.

Further assume that $\forall m \in \mathbb{N}^*, \exists k_m \in \mathbb{N}$ such that $P_{(i,j)}^{(m,m+k_m)} > 0$ for any $i \in T$ and some $j \in T^c$. Writing this positive probability as $h_m(i)$, it follows that there exists a $q(i)$ such that $q(i) = 1 - \inf_m h_m(i)$. Since the chain has only a finite number of states, $q = \max_{i \in T} (q(i))$ exists.

Define the norm of a real vector $v = (v_1, v_2, v_3, \dots)$ by $\|v\| = \sum_j |v_j|$, and the norm of a square matrix $A = [a_{ij}]$ by $\|A\| = \sup_i \sum_j |a_{ij}|$ [16].

Similar to a homogeneous chain, renumber the states so that $P^{(m,m+k)}$ can be expressed in the following canonical form:

$$P^{(m,m+k)} = \begin{bmatrix} Q^{(m,m+1)} & R^{(m,m+1)} \\ \mathbf{o} & I \end{bmatrix} \times \dots \times \begin{bmatrix} Q^{(m+k-1,m+k)} & R^{(m+k-1,m+k)} \\ \mathbf{o} & I \end{bmatrix} = \begin{bmatrix} Q^{(m,m+k)} & R^{**} \\ \mathbf{o} & I \end{bmatrix} \quad (4.1)$$

Here

$$R^{**} = R^{(m,m+1)} + R^{(m+1,m+2)} Q^{(m,m+1)} + \dots + R^{(m+k-1,m+k)} Q^{(m,m+k-1)}. \quad (4.2)$$

Notice that the norm of the substochastic matrix $\|Q^{(m,m+k)}\| = \left\| \prod_{i=1}^k Q^{(m+i-1,m+i)} \right\|$ is a non-increasing sequence in k for any m . Moreover if $0 < q < 1$, then $\|Q^{(m,n)}\|$ converges to zero as $n \rightarrow \infty$ [17].

In limit the elements of the submatrix R^{**} describe the probabilities of transition from transient states to absorbing states. If we let it denoted by $F^{(m)}$, we have

$$F^{(m)} = \sum_{i=0}^{\infty} [Q^{(m,m+i)}] R^{(m+i,m+1+i)} \quad (4.3)$$

Relative risk of absorption between competing absorbing states can be derived directly from the matrix $F^{(m)}$ by taking the ratio of the row elements.

The theorems below give $M^{(m)}$ and $V^{(m)}$, the matrices containing $E(N_{ij}^{(m)})$ and $\text{var}(N_{ij}^{(m)})$.

Theorem 1

If m denotes the initial starting time, then

$$\mathbf{M}^{(m)} = [E(N_{ij}^{(m)})] = \mathbf{I} + \sum_{k=1}^{\infty} \mathbf{Q}^{(m,m+k)} \quad (4.4)$$

Theorem 2

The nonstationary variance, given the initial time being m , is

$$\mathbf{V}^{(m)} = \mathbf{M}^{(m)} + 2 \left\{ \left(\sum_{i=1}^{\infty} \mathbf{Q}^{(m,m+k)} \right)_D + \left(\sum_{i=1}^{\infty} \mathbf{Q}^{(m,m+k)} \right) \left(\sum_{j=i+1}^{\infty} \mathbf{Q}^{(m+i,m+j)} \right)_D \right\} - \mathbf{M}^{(m)} \# \mathbf{M}^{(m)} \quad (4.5)$$

Here the symbol $\#$ refers to the elementwise multiplication. These moments exist if $0 < q < 1$ where $q = \max_{i \in T} (q(i))$ as in definition 2. Proofs are sketched in appendix.

The transition probabilities in the Nun Study are age-dependent and the process can therefore be treated as a nonhomogeneous Markov chain. We show that the Nun Study data meet the sufficient condition specified earlier. Based on the model MLEs $\hat{\theta}$, given a nun at age m , each row of the $\mathbf{Q}^{(m,m+1)}$ matrix can be estimated as follows:

$$\left[\frac{1}{1 + \sum_{h=2}^5 \exp(\hat{C}_h(s) + m\hat{\beta}_h^{age})}, \frac{\exp(\hat{C}_2(s) + m\hat{\beta}_2^{age})}{1 + \sum_{h=2}^5 \exp(\hat{C}_h(s) + m\hat{\beta}_h^{age})}, \frac{\exp(\hat{C}_3(s) + m\hat{\beta}_3^{age})}{1 + \sum_{h=2}^5 \exp(\hat{C}_h(s) + m\hat{\beta}_h^{age})} \right]$$

Here $\hat{C}_h(s)$ depends on the prior state s but is invariant to age, $s=1,2,3$. $\hat{\beta}_h^{age}$ are the MLEs associated with the covariate age from the Nun Study data and are positive. We have

$$\max_s \left| \frac{1 + \sum_{h=2}^3 \exp(\hat{C}_h(s) + (m+i)\hat{\beta}_h^{age})}{1 + \sum_{h=2}^5 \exp(\hat{C}_h(s) + (m+i)\hat{\beta}_h^{age})} \right| \leq \max_s \left| \frac{1 + \sum_{h=2}^3 \exp(\hat{C}_h(s) + m\hat{\beta}_h^{age})}{1 + \sum_{h=2}^5 \exp(\hat{C}_h(s) + m\hat{\beta}_h^{age})} \right| = q$$

For any m there always exists an immediate transition such that any state in the transient class will be absorbed with a positive probability at least as large as $1 - q$. As $0 < q < 1$, absorption is assured and the first and second moments of the number of transitions before absorption are finite in the Nun Study data.

5. Application: The Nun Study

Using the Nun Study data, the relative risk of dementia before death can be calculated for each initial age m by first computing the 3 by 2 matrix $\mathbf{F}^{(m)}$ from Equation (4.3) and then by taking the ratio of the element in the first column of $\mathbf{F}^{(m)}$ to the corresponding element in the second column. This yields three relative risks, one for each of the three possible initial transient states. In estimating $\mathbf{F}^{(m)}$ we substitute the MLEs for the unknown betas into the

elements of $\mathbf{P}^{(m,m+k)}$ using Equations (3.2) and (4.1). We note that while $\mathbf{F}^{(m)}$ is defined as a sum of an infinite series, it converges relatively quickly due to the involvement of exponential terms of age. Therefore the partial sum of the first few terms in this series is sufficient to approximate the final $\mathbf{F}^{(m)}$.

We adjust for the variability of both random effects as well as the estimated fixed effects in constructing approximate 90% confidence intervals for the relative risks. A smoothed bootstrap technique is adopted based on 1000 iterations [18]. Each iteration involves a random number generated from $N(0, \hat{\sigma}^2)$ and a random vector from $MVN(\hat{\beta}, \hat{\Sigma})$ with $\hat{\sigma}^2$ being the MLE for the variance of random effect γ and $\hat{\Sigma}$ the asymptotic variance covariance matrix for the fixed effects β . These estimates are then incorporated into Equation (3.2) to compute the estimated transition probabilities and accompanying $\mathbf{F}^{(m)}$. The 5th and 95th percentiles of the iterated relative risks are then used as the lower and upper bounds for the said confidence interval. The resulting estimated relative risks and confidence intervals are plotted in Figure 2.

The figure displays the estimated relative risks of developing dementia before death by APOE4 status, level of education, and the ages 75, 80, 85, 90, and 95, assuming the initial state is intact cognition. The figure shows that the relative risk for incurring dementia before death decreases with advancing age regardless of the nun's genetic status or educational level; this is most pronounced for nuns with positive APOE 4. Nuns with a negative APOE 4 and over 16 years of education have relative risks less than one. In contrast, nuns with positive APOE 4 and 12 or fewer years of education have relative risks greater than one. The estimated confidence intervals especially for this latter subgroup are wider. As suggested by one of the reviewers this could be due to the bootstrapping procedure since this procedure leads to poor estimation whenever the sample size for a subgroup is small; hence, caution is advised in interpreting these confidence intervals.

The following matrices list the mean estimates of the number of visits among transient states before absorption into death or dementia for two sets of covariate pairs: no genetic risk and high education and then genetic risk and low education. Initial age is set at 80. The means were computed using the result of Theorems 1 of the previous section; again only a few terms of the infinite series defining these matrices need to be computed. In addition, elementwise standard errors are bootstrapped in order to evaluate the variability of the mean estimates; these bootstrapped standard errors account for the variability in the beta estimates.

The first example represents a typical subject with an APOE4 allele, no more than 12 years of education and initial age being 80;

$$\mathbf{M}^{(80)} = \begin{bmatrix} 1.06(0.06) & 2.71(0.62) & 1.38(0.42) \\ 0.04(0.04) & 3.78(0.62) & 1.24(0.40) \\ 0.01(0.02) & 0.94(0.37) & 2.40(0.44) \end{bmatrix}$$

In contrast, a typical subject with no APOE4 allele, over 16 years of education and initial age being 80 has the following mean estimates for the random variable $N_{ij}^{(80)}$

$$\mathbf{M}^{(80)} = \begin{bmatrix} 3.29(0.39) & 3.04(0.28) & 0.90(0.12) \\ 2.06(0.37) & 4.16(0.28) & 0.87(0.11) \\ 1.05(0.28) & 1.60(0.26) & 1.92(0.14) \end{bmatrix}$$

The lead elements in these matrices refer to an 80 year old nun who is cognitively intact at baseline. According to these matrices, a highly educated nun with no genetic risk can expect to have an average of 3.29 transitions into the cognitively intact state before absorption, while a low educated nun with a genetic risk can expect to have 1.06 transitions into this same state before absorption. If we add the elements in the first row of $M^{(80)}$, then we obtain the number of visits an 80-year-old nun will make before absorption if she is cognitively intact at baseline (7.2 for a highly educated nun with no genetic risk and 5.2 for a low educated nun with a genetic risk). These numbers of visits translate into 9 years and 6.5 years, respectively, after accounting for the time interval between visits, which averaged 15 months. These latter figures are about the same for a nun who has mild cognitive impairment at baseline. However, the duration before absorption is much shorter if the 80-year-old nun has a global impairment at baseline (5.8 years with high education and no genetic risk, 4.3 years with low education and genetic risk).

6. Conclusion and Discussion

This paper examines the absorption statistics of the k -step transition probability matrix for a nonhomogeneous discrete-time Markov chain by focusing on (1) the relative risks of absorption between competing absorbing states and (2) the first and second moments of number of visits among transient states before absorption. These absorption statistics are important in studying the effects of risk factors associated with the progression of healthy individuals through pre-illness states to a chronic disease, with death treated as a competing event. It has been well established that these statistics always exist in a time homogeneous chain and they are directly related to the fundamental matrix. In the nonhomogeneous case, the absorption statistics are not guaranteed to hold since the fundamental matrix is replaced by an infinite matrix series. Proper assumptions about the transition matrix are required to ensure the convergence of this series. One such condition is the existence of a positive supremum bounded between 0 and 1 for the transition probabilities among transient states, independent of when (initial time m) and where (initial state $i \in T$) the process starts. Provided this condition, formulas are derived in computing the absorption statistics but partial sums can be used to evaluate the infinite series.

An alternative is to ignore baseline cases of dementia and to ignore the non stationarity. This would simplify the arithmetic since the likelihood would no longer account for baseline and would account for age only through its baseline value when estimating the long run behavior of the chain. The resulting beta estimates could be seriously biased downward in the presence of prevalent as well as incident cases of dementia implying the likelihood yields incorrect beta estimates [8]. In addition, even if the likelihood accounts for the baseline cases failure to account for the non-stationarity of the chain can lead to inaccurate statistics for the long run behavior of the chain. To illustrate consider $M^{(80)}$ for the case no APOE 4 allele and education greater than 16 years computed in Section 5. If stationarity is assumed, then its elements are inflated by a median value of 82% (range 2.6% to 352%) yielding much larger estimates of these first moments than in the non-stationary case.

The results are applied to the Nun Study. Subjects in the Nun Study do not share the same baseline state. Among 501 subjects used for analysis, 128 had intact cognition at baseline, 249 had MCI, 47 had GI and 77 were demented. The heterogeneous states at baseline for subjects in the study make it important to incorporate the baseline outcomes into the likelihood construction. A multi-state Markov transition model with multiple likelihood components sharing the same random effect is developed for this purpose. The results show that age, education and APOE4 are significant predictors of transition probabilities. The risk of dementia before death declines as age and the level of education increase, while the

presence of APOE4 increases such risk. In addition, the expected number of visits among transient states before absorption is also affected by APOE4 status.

There are other approaches for analyzing data involving competing risk events. Gooley et al. [19] proposed using cumulative incidence (CI) estimates for the probability of failure for an event subject to competing risks. Commenges et al. [20,21] adopt an intensity function approach and investigate an illness-death model to estimate the incidence and prevalence of dementia with competing risk of death. As an alternative, our extended multi-state Markov model provides a framework to specify explicitly the one-step transition matrices of the process. Thus, we can model the observations collected at successive cognitive assessments in the cohort visits and at the same time study their long-term behavior, while accounting for both observed risk factors and latent random effects. Unlike all other approaches to this problem, our approach accounts for visiting multiple transient states and for back transitions among these transient states before absorption into one of the competing absorbing states. Both of these features are characteristic of the real-life trajectories of cognitive decline.

The methodology presented here does have some limitations since it relies on a discrete time model. On the one hand a discrete time model is most sensible for a longitudinal study of aging which consists of serial assessments of cognition since it accounts for the competing risk of death when following individuals from cognitively intact to dementia with transitions to and from intervening impaired states. The reason is that all events except death are interval censored. Use of the Markov model presented here has the advantage that reverse transitions from impaired states to cognitively intact can be accounted for. But the model does assume that serial cognitive assessments are equally spaced such as in the BRAiNS cohort studied by Salazar et al [7]. In the Nun study presented here the retired nuns were located in seven different states and the same team of cognitive assessors had to make the rounds among these retirement centers. Allowing for training of the team and funding gaps the time interval between assessments was not always equal and averaged 1.3 ± 0.3 years (range 1 – 1.9 years). The effect of this lack of equality of successive assessments is unknown.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by National Institute on Aging grants R01 AG09862, K04 AG00553, and P50 AG05144, by a grant from the Kleberg Foundation, and by a University of Kentucky Research Professorship. This study would not have been possible without the support of the members, leaders, and health care providers of the School Sisters of Notre Dame religious congregation.

References

1. Chiang, CL. An introduction to stochastic processes and their applications. R. E. Krieger; Huntington, NY: 1980.
2. Agresti, A. Categorical data analysis. 2. John Wiley & Sons Inc; Hoboken: 2002.
3. Muenz LR, Rubinstein LV. Markov models for covariate dependence of binary sequences. *Biometrics*. 1985; 41:91–101. [PubMed: 4005390]
4. Kay R. A Markov model for analyzing cancer markers and diseases states in survival studies. *Biometrics*. 1986; 42:855–865. [PubMed: 2434150]
5. Li YP, Chan W. Analysis of longitudinal multinomial outcome data. *Biometrical Journal*. 2006; 48(2):319–326. [PubMed: 16708781]

6. Chen TH, Yen M, Shiu M, Tung T, Wu H. Stochastic model for non-standard case-cohort design. *Statistics in Medicine*. 2004; 23:633–747. [PubMed: 14755394]
7. Salazar JC, Schmitt FA, Yu L, Menciondo MS, 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]
8. Yu L, Tyas SL, Snowdon DA, Kryscio RJ. Effect of ignoring baseline on modeling transitions from intact cognition to dementia. *Computational Statistics and Data Analysis*. 2009; 53:3334–3343. [PubMed: 20161282]
9. Vonesh EF, Greene T, Schluchter MD. Shared parameter models for the joint analysis of longitudinal data and event times. *Statistics in Medicine*. 2007; 25:143–163. [PubMed: 16025541]
10. Rabe-Hesketh, S.; Skrondal, A. Multilevel and longitudinal modeling using Stata. Stata Press; College Station, Texas: 2005.
11. Litiere S, Alonso A, Molenberghs G. Type I and Type II error under random effects misspecification in generalized linear mixed model. *Biometrics*. 2007; 63:1038–1044. [PubMed: 17425642]
12. Snowdon DA, Greiner LH, Mortimer JA, Riley KP, Greiner PA, Markesbery WR. Brain Infarction and the Clinical Expression of Alzheimer Disease: The Nun Study. *JAMA*. 1997; 277:813–817. [PubMed: 9052711]
13. Tyas SL, Salazar JC, Snowdon DA, Desrosiers MF, Riley KP, Menciondo MS, Kryscio RJ. Transitions to mild cognitive impairments, dementia and death: findings from the Nun Study. *Am J Epidemiology*. 2007; 165(11):1231–1238.
14. SAS Institute Inc. SAS/STAT User's Guide, Version 9. SAS Institute Inc; Cary, NC: 2004.
15. Bhat, UN.; Miller, GK. Elements of Applied Stochastic Processes. 3. John Wiley & Sons, Inc; Hoboken, New Jersey: 2002.
16. Isaacson, LD.; Madsen, WR. Markov Chains Theory and Applications. John Wiley & Sons, Inc; New York: 1976.
17. Platis A, Limnios N, Du ML. Hitting time in a finite non-homogeneous Markov chain with applications. *Applied Stochastic Models and Data Analysis*. 1998; 14:241–253.
18. Efron B. Nonparametric Estimates of Standard Error, The Jackknife, the bootstrap and other methods. *Biometrika*. 1981; 68:589–599.
19. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*. 1999; 18:695–706. [PubMed: 10204198]
20. Commenges D, Joly P, Letenneur L, Dartigues JF. Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statistics in Medicine*. 2004; 23:199–210. [PubMed: 14716722]
21. Commenges D, Joly P. Multi-state model for dementia, institutionalization, and death. *Communications in Statistics-Theory and Methods*. 2004; 33(6):1315–1326.

Figure 1.
Flow diagram of one-step transitions on progression to dementia

Figure 2.
Relative Risks and 90% confidence interval of dementia before death
Initial state: intact cognition; age 75 (circle), age 80 (dot) age 85 (square) age 90 (diamond),
and age 95 (triangle)

Table 1

Comparison of events between two cohorts: BRAiNS and the Nun Study

Cohort	N	Follow-up			
		Baseline			
		Dementia	Dementia	Death*	Events
		n (%)	n (%)	n (%)	n (%)
BRAiNS	553	0 (0%)	55 (10.0%)	144 (26.0%)	199 (36.0%)
Nun Study	501	77 (15.4%)	153 (30.5%)	184 (36.7%)	414 (82.6%)

* died before conversion to dementia

Table 2

Transitions in cognitive status in the Nun Study

Prior Visit	Current Visit				
	Intact Cognition	MCI	GI	Dementia	Death
	n (%)	n (%)	n (%)	n (%)	n (%)
Intact Cognition	520 (65.8%)	179 (22.7 %)	52 (6.6 %)	5 (0.6 %)	34 (4.3 %)
MCI	159 (15.0 %)	629 (59.2 %)	123 (11.6 %)	81 (7.6 %)	71 (6.7 %)
GI	15 (4.2 %)	36 (10.0 %)	162 (45.1 %)	67 (18.7 %)	79 (22.0 %)
Dementia	0 (0%)	0 (0%)	0 (0%)	77 (27.9%)	199 (72.1%)

MCI = mild cognitive impairment; GI=global impairment

Table 3

Parameter estimates for transition probabilities in the Nun Study (Initial state: intact cognition)

Risk Factor	MCI	GI	Dementia	Death
	MLE (SE)	MLE (SE)	MLE (SE)	MLE (SE)
Age (centered)	0.1502* (0.022)	0.2289* (0.024)	0.2190* (0.027)	0.2439* (0.026)
APOE4	1.6231* (0.391)	2.0029* (0.411)	2.1163* (0.433)	1.7837* (0.434)
Educ. = 16	-1.3889* (0.584)	-1.5188* (0.607)	-1.4378* (0.637)	-1.2003 (0.643)
Educ. > 16	-2.1096* (0.582)	-2.2948* (0.605)	-2.1097* (0.636)	-1.6550* (0.641)
Prior States:				
Intact Cog.	-0.4728 (0.371)	-2.9604* (0.362)	-4.4713* (0.568)	-2.7569* (0.386)
MCI	0.2512 (0.369)	-2.7764* (0.350)	-2.3354* (0.368)	-2.6330* (0.367)

* Significant at α being 0.05; estimate of sigma = 1.99* (0.192)

MCI=mild cognitive impairment; GI=global impairment