# A Bayesian Approach to Multistate Hidden Markov Models: Application to Dementia Progression

Jonathan P. Williams, Curtis B. Storlie, Terry M. Therneau, Clifford R. Jack Jr & Jan Hannig

Taylor & Francis
Taylor & Francis Group

Check for updates

# A Bayesian Approach to Multistate Hidden Markov Models: Application to Dementia Progression

Jonathan P. Williams[a,b], Curtis B. Storlie[a], Terry M. Therneau[a], Clifford R. Jack Jr[a], and Jan Hannig[b]

[a]Mayo Clinic, Rochester, MN; [b]Department of Statstics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, NC

## ABSTRACT

People are living longer than ever before, and with this arises new complications and challenges for humanity. Among the most pressing of these challenges is of understanding the role of aging in the development of dementia. This article is motivated by the Mayo Clinic Study of Aging data for 4742 subjects since 2004, and how it can be used to draw inference on the role of aging in the development of dementia. We construct a hidden Markov model (HMM) to represent progression of dementia from states associated with the buildup of amyloid plaque in the brain, and the loss of cortical thickness. A hierarchical Bayesian approach is taken to estimate the parameters of the HMM with a truly time-inhomogeneous infinitesimal generator matrix, and response functions of the continuous-valued biomarker measurements are cut-point agnostic. A Bayesian approach with these features could be useful in many disease progression models. Additionally, an approach is illustrated for correcting a common bias in delayed enrollment studies, in which some or all subjects are not observed at baseline. Standard software is incapable of accounting for this critical feature, so code to perform the estimation of the model described below is made available online. Code submitted with this article was checked by an Associate Editor for Reproducibility and is available as an online supplement.

## 1. Introduction

People are living longer and with this arises new complications and challenges for humanity. Among the most pressing of these challenges is of understanding the role of aging in the development of dementia. Such is the initiative of the Mayo Clinic Study of Aging (MCSA), a large prospective study with the goal of understanding the natural history of dementia and particularly Alzheimer's disease.

This article is motivated entirely by the MCSA and how the resulting data can be used to draw inference on the role of aging in the development of dementia. The goal is to create a model of progression to dementia which can accommodate: (1) a wide variation in age (the dominant variable under consideration), (2) significant fluctuation in the time between subject visits, (3) different amount of information available for each subject (e.g., missing visits and/or clinical data), and (4) subject specific covariates.

The main contribution of this work is to provide an innovative statistical analysis of this important and unique dataset via a continuous-time, discrete-state hidden Markov model (HMM) estimated within the Bayesian paradigm. Additionally, we demonstrate the existence of and provide solutions for various methodological gaps in the analysis of disease progression for studies like the MCSA. First, we provide an approach for correcting a common bias in *delayed enrollment* studies which has been overlooked in the literature. Second, we introduce a methodological framework for estimating the

strength and persistence of a separate *death rate bias* specific to death rates, which could be present in any study relying on enrollment of subjects. Our final methodological innovation is a proposed Bayesian approach to estimating the biomarker regions most associated with high/low burden states in a manner that does not require the specification of cut-points.

The term *delayed enrollment*, here, is used to describe a study with a given baseline (age 50 in the case of the MCSA) such that some or all subjects are *not* observed at baseline. We demonstrate empirically that the effects of this bias cannot be ignored, and existing software is not equipped to handle this feature.

We formalize the discrete-state space exhibited in Figure 1 in which many of the states are defined by continuous biomarkers. The previous work of Jack et al. (2016) defined a state space similar to Figure 1, but in which the high/low burden biomarker states were defined by practitioner chosen, hard biomarker cut-points. Hard cut-points for discretizing continuous measurements of biological processes are practically and philosophically problematic, and have to be chosen more or less arbitrarily.

Moreover, we illustrate a general and effective framework for fitting a continuous-time, discrete-state HMM within the Bayesian paradigm, and the infinitesimal generator matrix of the underlying Markov process is allowed to be truly time-inhomogeneous (as a function of an individual's age). Time must be treated as continuous because, as in much of medical research, subjects are often observed irregularly in time.

---

Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/JASA*.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JASA*.
These materials were reviewed for reproducibility.

Our final contribution is that in addition to the effect of age, the effects of the covariates gender, number of years of education, and presence of an APOE-$\varepsilon$4 allele on the infinitesimal transition rates are also estimated. The importance of these variables has been well documented in the medical literature but their effect on aging has not been studied in this context (i.e., how they affect the transition rates between states in Figure 1). In addition to the new insights these features bring to the medical community, flexible software to fit the models described below is provided at *https://jonathanpw.github.io/software.html*.
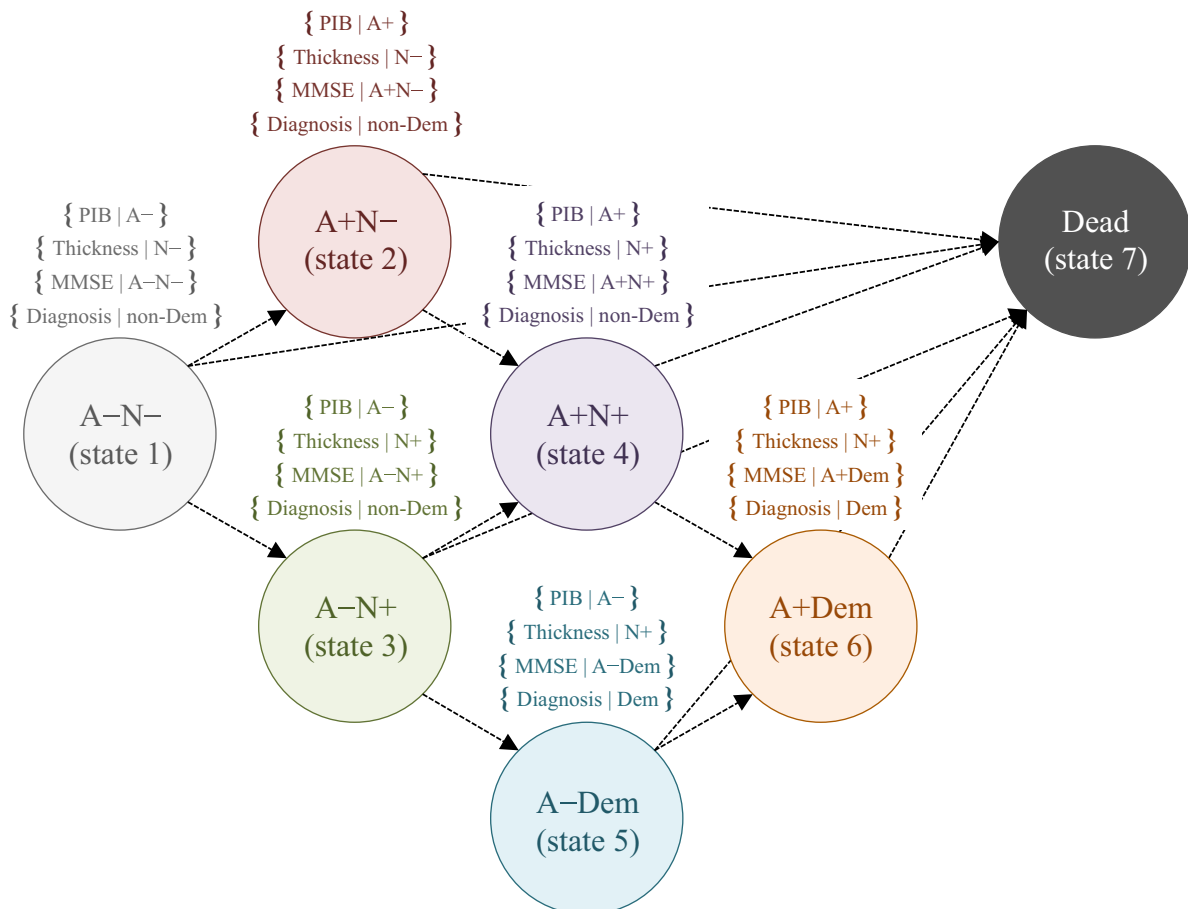
Our analysis builds on the work of Jack et al. (2016) with more sophisticated modeling which allows for deeper insights. They found that a Markov model of disease progression for dementia is indeed a natural approach, that almost all rates are log-linear, and at age 50 nearly everyone is in state A$-$N$-$ (i.e., low Amyloid burden and low cortical thickness loss burden which is state 1 in Figure 1) but that soon begins to change.

Most implementations of a continuous-time, discrete-state HMM, including Jack et al. (2016) estimate parameters in a maximum likelihood fashion. However, as mentioned in Jack et al. (2016) optimization becomes exceedingly difficult as more parameters are introduced in the model. Convergence time for standard methods may become impractical, and analytical gradient formulas for use in more efficient optimization procedures can become intractable. Additionally, it is often difficult/awkward to fit prior information into an optimization-based frequentist approach (e.g., via constrained optimization or penalty functions which require tuning parameters), and deriving confidence intervals becomes a challenge. Further, as the model becomes more complex to better capture reality in our application, prior information becomes necessary for practical identifiability of HMM parameters which makes Bayesian methodology a natural approach. For these reasons, we propose a hierarchical Bayesian framework with model estimation via Markov chain Monte Carlo (MCMC). Using MCMC for the estimation of complicated models requires creative proposal strategies, but is extremely flexible for a variety of model specifications. Moreover, credible regions become a convenient way to represent uncertainty.

Accounts of continuous-time, discrete-state HMM are given by Lange and Minin (2013), Bureau, Shiboski, and Hughes (2003), Jackson et al. (2003), Titman and Sharples (2008), Jack et al. (2016), and within a Bayesian framework (Zhao et al. 2016). Further, Satten and Longini (1996) is a very complete account of how to implement such a model and is recommended reading for anyone not familiar with the methodology.

More relevant to the present application, the work of Jackson et al. (2003) uses an HMM to model the state misclassification error of a disease and includes age as a covariate on the transition rates. However, they make a very restrictive assumption that the transition rates are constant between subject observation times. The work of Jack et al. (2016) is seemingly the first in



**Figure 1.** State space. Emitted response variables are displayed in brackets above the respective hidden state. A+ corresponds to high amyloid burden, and N+ corresponds to high neurodegenerative burden. States 1–4 are all non-demented.

the literature to estimate the transition rates as a function of age, in a truly continuous-time fashion for a multistate model of dementia, however, Yu et al. (2010) also treated transition probabilities as a function of age, in a discrete-time fashion.

The organization of the rest of the article is as follows. Section 2 describes the MCSA dataset and the HMM methodology. An illustration of the *death rate bias* and the *delayed enrollment bias* is then given in Section 3, and a simulation study is provided in Section 4 which implements our Bayesian estimation procedure on synthetic data generated to resemble the MCSA dataset. Finally, a detailed analysis of the MCSA data is presented in Section 5, accompanied by a discussion and interpretation of the biological findings. This article also has accompanying supplementary material containing details of the computations and further MCSA analysis results.

## 2. Methodology

### 2.1. Description of the Mayo Clinic Study of Aging data

The MCSA has enrolled a large age/sex stratified random sample from Olmsted County, Minnesota. Subjects are followed forward approximately every 15 months, and clinical visits collect information on major aspects of the disease. The MCSA study began in 2004, currently has 4742 subjects, and is still ongoing (at the time of this writing) and enrolling new subjects. Estimates of quantities such as expected time in a given state, probability of ever entering a given state, or the fraction of the population that will pass through a high amyloid burden state on the way to dementia, are all of clinical interest.

It is known in the medical community that amyloid protein buildup in the brain, and significant neurodegeneration are strongly associated with dementia. Accordingly, amyloid buildup, as measured by Pittsburgh compound B (PIB) from a positron emission tomography (PET) scan, and neurodegeneration, as measured by cortical thickness (thickness) from magnetic resonance imaging (MRI), are continuous outcomes measured during the regular clinical visits for approximately 50% of the subjects.

With regard to dementia, high amyloid burden is a notion which refers to a build-up of amyloid plaques in the brain significant enough to effect pathways and lead to neurodegeneration, but precise measurements of the extent of amyloid protein would require autopsy (PIB measurements serve as a proxy for measuring this extent). Likewise, high neurodegeneration burden refers to a state of loss of neurons and synapses denoted by atrophy of the cerebral cortex in Alzheimer's-sensitive areas.

For the (approximately 50% of) subjects who were not chosen to undergo regular brain scans less clinical information is available. However, the Mini-Mental State Exam (MMSE) is almost always observed. The MMSE is a questionnaire-based test administered by a medical professional to assess cognitive impairment on an integer scale out of 30 points (Xu et al. 2015). Furthermore, baseline data such as age, sex, clinical and genetic markers is always recorded in the data.

Finally, at the time of observation all subjects are determined to be either cognitively unimpaired, or to be demented. This represents a substantial amount of information for making inference on the underlying cognitive state of a subject. However, diagnosing dementia is not an exact science, and so the observed label is not without error.

### 2.2. The HMM State Space and Emitted Response Variables

A simplistic formalization of the biology is to theorize a seven-state model to describe cognitive health in relation to dementia. Figure 1 illustrates such states and depicts the allowed transitions with directional arrows. A notable feature of the state space is that an individual must be in a high neurodegeneration burden state (i.e., N+) to develop dementia, but not necessarily in a high amyloid burden state (i.e., A+). In fact, the transition from A+N+ (state 4) to A+Dem (state 6) is identified as Alzheimer's disease, a particular type of dementia. Isolating this Alzheimer's disease transition is not possible using the previous state space of Jack et al. (2016).
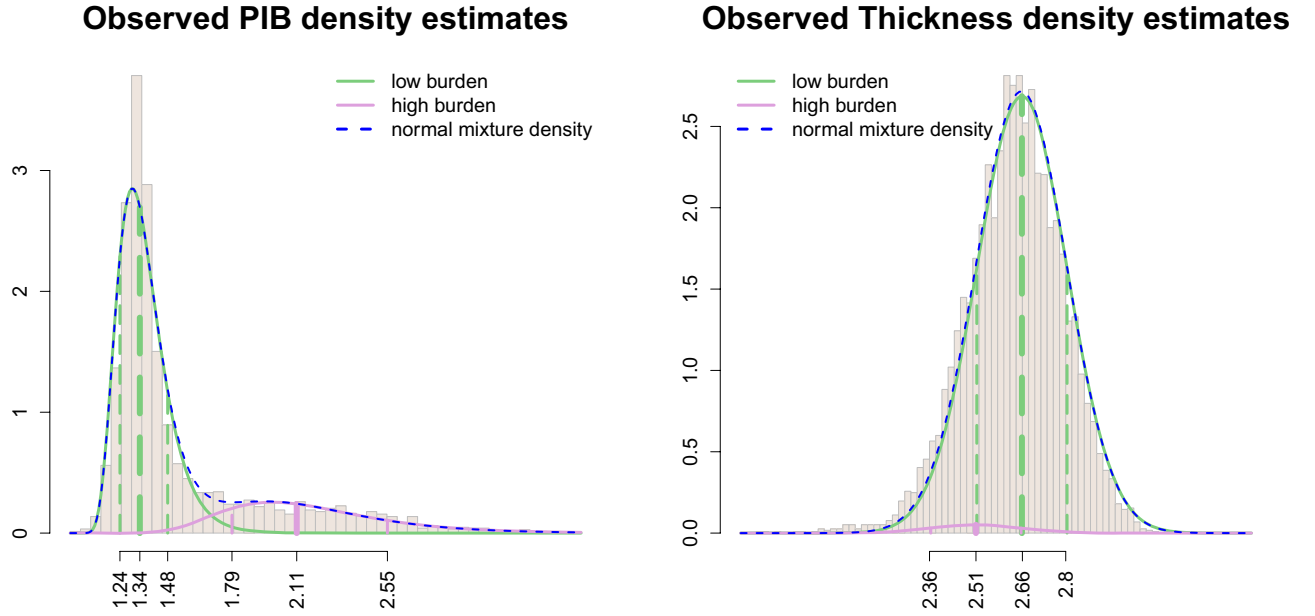
Time must be treated as continuous because patient visit times are irregular, and the underlying sequence of states visited for an individual is hidden by uncertainty (even when PIB/thickness are available, they are only proxies for the true level of amyloid/neurodegeneration burden). Moreover, the states of high amyloid burden and neurodegenerative burden are not precisely defined and are best treated as hidden. It is worth remarking that amyloid build-up and neurodegeneration each develop on a continuum, but the time spent in any intermediate states, not explicitly represented by the state space in Figure 1, is believed to be relatively short and thus ignorable in these data.

The PIB and thickness values associated with amyloid buildup and neurodegeneration, respectively, are used as emitted response variables (in a traditional HMM sense) to make inference on an underlying sequence of states visited. States 2, 4, and 6 which correspond to increased amyloid burden will emit PIB values from a distribution corresponding to A+, while states 1, 3, and 5 will emit PIB from the distribution corresponding to A−, and similarly for neurodegenerative burden (N+ or N−).

The prior distributions for these response distribution parameters was chosen to correspond to biomarker values which are consistent with the medical community's most up-to-date understanding of the biology. Using a Gaussian distribution for thickness and for $\log(\text{PIB}-1)$ appears to be quite reasonable, as there is evidence that the error from the PIB measurements follows more closely to a constant coefficient of variation than to a constant variance. Figure 2 displays histograms of the observed response data along with the respective normal mixture densities resulting from the posterior mean estimates from the full HMM described in the coming sections.

The MMSE score serves as an additional emitted response, and a separate Gaussian emission distribution for each of the first six states is assumed (deceased subjects do not emit cognitive test scores). Figure 3 overlays the estimated six-component normal mixture density function on top of a histogram of the observed MMSE scores from the study subjects' visits.

Lastly, a simple misclassification response model is used to allow for a probability of a dementia diagnosis given the underlying state is a dementia state (i.e., states 5 or 6), and given it is

## Observed PIB density estimates



## Observed Thickness density estimates



**Figure 2.** Observed response data for PIB which is a measure of amyloid buildup from a PET scan, and (cortical) thickness which is associated with neurodegeneration. Note that the response densities for PIB correspond to the data transformation, $\log(PIB - 1)$. The component density estimates correspond to the posterior mean estimates from Section 5. The blue dashed lines represent the normal mixture density estimates.

## Distribution of observed MMSE



**Figure 3.** Observed response data for MMSE test scores associated with the six non-death states in the state space. The component density estimates here correspond to the posterior mean estimates from Section 5. The blue dashed line represents the normal mixture density estimate.

not a dementia state. Death is the only state in the state space which is known without error, and the exact time of death is known, as well.

### 2.3. Continuous-Time Transition Probabilities

This section serves to specify the HMM in the context of the state space illustrated in Figure 1. For $r, s \in \{1, 2, 3, 4, 5, 6, 7\}$ and $h, t \geq 0$, the probability of transitioning from state $r$

at time $h$ to state $s$ at time $h + t$ is denoted by $P_{r,s}(h, t) = P(S(h + t) = s | S(h) = r)$. Assuming these probabilities are differentiable functions in $t$ and that the Markov process is time-homogeneous, it can be shown that they satisfy the Kolmogorov forward equations (Karlin and Taylor 1981),

$$\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q}, \qquad (1)$$

where $\boldsymbol{Q}$ is called the transition rate matrix, and $\boldsymbol{P}(t)$ is the matrix with components

$$P_{rs}(t) := P_{rs}(h = 0, t). \qquad (2)$$

Note that $h$ can be taken to be 0 in (2) because the probabilities are assumed for now to be time-homogeneous. The off-diagonal components of $\boldsymbol{Q}$ are interpreted as the change in transition probabilities for an infinitesimal amount of time into the future, that is,

$$q_{rs} = \lim_{t \downarrow 0} \frac{P(S(t) = s | S(0) = r)}{t}, \quad r \neq s, \qquad (3)$$

with diagonal elements $q_{rr} = -\sum_{s \neq r} q_{rs}$.

The forward equations in (1) have the matrix exponential solution, $\boldsymbol{P}(t) = e^{t\boldsymbol{Q}}$. However, as discussed in Section 1, the transition rates will be expressed as a function of a subject's age at the time of transition. That is, $\boldsymbol{Q} = \boldsymbol{Q}(t)$ which violates the time-homogeneity of the Markov process. A simple work around is to discretize the effect of age and assume that the transition rates only change when a subject's integer age changes. Doing so implies that subjects' transition rates, $\boldsymbol{Q}$, remain constant between birthdays and yields

$$\boldsymbol{P}(h, t) = e^{(\lfloor h+1 \rfloor - h)\boldsymbol{Q}(\lfloor h \rfloor)} \cdot e^{\boldsymbol{Q}(\lfloor h+1 \rfloor)} \cdots e^{\boldsymbol{Q}(\lfloor h+t \rfloor - 1)}$$
$$\cdot e^{(h+t-\lfloor h+t \rfloor)\boldsymbol{Q}(\lfloor h+t \rfloor)}, \qquad (4)$$

for $\lfloor h \rfloor \neq \lfloor h+t \rfloor$, where $h$ represents the subject's current age, $t$ is the time (in years) into the future, and $\lfloor \cdot \rfloor$ is the floor function.

Observe that with expression (4) all transition probabilities can be computed, as long as the components of $Q(t)$ are specified. As mentioned above, the transition rates will be modeled as a function of age, gender, presence of an APOE-$\varepsilon 4$ allele, and number of years of education. Specifically, denoting each of the 13 nonzero transition rates illustrated in Figure 1 by $q_l$ for $l \in \{1, \ldots, 13\}$,

$$\log(q_l) = \beta_0^{(l)} + \beta_1^{(l)} \cdot \text{age} + \beta_2^{(l)} \cdot \text{male} + \beta_3^{(l)} \cdot \text{educ} + \beta_4^{(l)} \cdot \text{apoe4}, \quad (5)$$

where,

$$Q = \begin{pmatrix} -q_1 - q_2 - q_3 & q_1 & q_2 & 0 & 0 & 0 & q_3 \\ 0 & -q_4 - q_5 & 0 & q_4 & 0 & 0 & q_5 \\ 0 & 0 & -q_6 - q_7 - q_8 & q_6 & q_7 & 0 & q_8 \\ 0 & 0 & 0 & -q_9 - q_{10} & 0 & q_9 & q_{10} \\ 0 & 0 & 0 & 0 & -q_{11} - q_{12} & q_{11} & q_{12} \\ 0 & 0 & 0 & 0 & 0 & -q_{13} & q_{13} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (6)$$

Overall death rates over age 50 are log-linear (see Figure 4) which makes the log-linear function of age a natural starting place. This functional form was also argued in Jack et al. (2016) to be reasonable for all of the rates except the rate from A−N− (state 1) to A+N− (state 2). They compared a log-linear rate to that obtained from log-cubic splines. We came to the same conclusion and accordingly, a cubic spline is used for estimating only the rate of transition from state $1 \rightarrow 2$, with knots at ages 55, 65, 75, 90, and boundary knots at 50 and 120.

### 2.4. Likelihood Function

If the states were known for each subject at each observation, then the contribution to the likelihood function from each



**Log death rate**

**Figure 4.** Annualized natural logarithm of Minnesota overall population death rates. Solid line corresponds to female and dashed line corresponds to male.

subject consists of a product of matrix exponentials, that is, apply (4) to get the probability of being in a given state for each observation time. However, the underlying state sequences are not observed. Within an HMM, rather, responses emitted from the underlying process (conditional on the true state of the process at a given point in time) are used to inform of the underlying state. In this application, there are four emitted responses (i) log(PIB−1), (ii) thickness, (iii) MMSE (see Figures 2 and 3), and (iv) dementia diagnosis (binary). Denote the observations for each of these four responses by $y_{i,k} = [y_{i,k,1}, y_{i,k,2}, y_{i,k,3}, y_{i,k,4}]'$,

respectively, for the $i$th subject's $k$th clinical visit (observation). Further, the $i$th subject has $n_i$ clinical visits, and has for each visit $k \in \{1, \ldots, n_i\}$ an (unknown) state $s_{i,k}$.
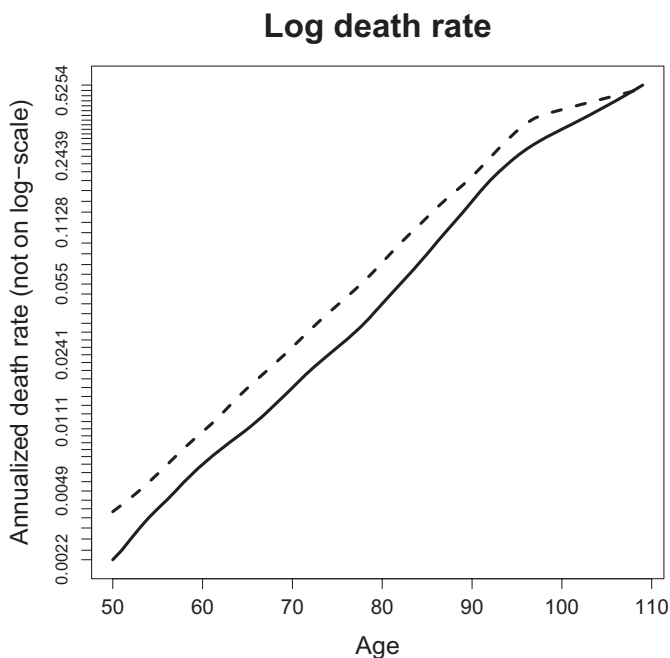
Thus, the likelihood contribution from the $i$th subject at the $k$th visit can be expressed as

$$f(y_{i,k}) = \sum_{s_{i,k}=1}^{7} f(y_{i,k}, s_{i,k}) = \sum_{s_{i,k}=1}^{7} P(s_{i,k}) \cdot f(y_{i,k}|s_{i,k}),$$

where the sum is taken over all possible states (since the true state sequence is unknown). If a given response $y_{i,k,j}$ is missing (e.g., missing PIB scan), then the missing value is integrated out of the likelihood (i.e., the response density of the missing value contributes a product of 1 to the likelihood function). It is assumed that missing response values are predominantly missing at random (MAR). While it could be suspected that missing clinical visits might be associated with deteriorating mental health, conditional on the underlying state, $s_{i,k}$, it is reasonable to assume that no systematic pattern of missingness persists. Furthermore, the approximately 50% of subjects chosen to undergo regular brain imaging were assigned randomly, so a large portion of missing PIB and thickness values are missing completely at random (MCAR).

Making the standard assumption that the responses are conditionally independent given an underlying state sequence $s_{i,1}, \ldots, s_{i,n_i}$, and applying the Markov property for the state sequence gives,

$$f(y_{i,1}, \ldots, y_{i,n_i}) = \sum_{s_{i,1}=1}^{7} \sum_{s_{i,2}=1}^{7} \cdots \sum_{s_{i,n_i}=1}^{7} P(s_{i,1})$$

$$\times P(s_{i,2}|s_{i,1}) \cdots P(s_{i,n_i}|s_{i,n_i-1}) \cdot \prod_{k=1}^{n_i} f(y_{i,k}|s_{i,k})$$

$$= \sum_{s_{i,1}=1}^{7} P(s_{i,1}) f(y_{i,1}|s_{i,1}) \cdot \sum_{s_{i,2}=1}^{7} P(s_{i,2}|s_{i,1})$$

$$\times f(y_{i,2}|s_{i,2}) \cdots \sum_{s_{i,n_i}=1}^{7} P(s_{i,n_i}|s_{i,n_i-1}) f(y_{i,n_i}|s_{i,n_i})$$

$$= \pi_0' D_{(i,1)} \cdot P(t_{i,1}, t_{i,2} - t_{i,1})$$
$$\times D_{(i,2)} \cdots P(t_{i,n_i-1}, t_{i,n_i} - t_{i,n_i-1}) D_{(i,n_i)} \cdot 1, \tag{6}$$

where $\pi_0$ is the baseline state probability vector, $D_{(i,k)}$ is a diagonal matrix with diagonal components $f(y_{i,k}|s_{i,k})$ for each $s_{i,k} \in \{1, \ldots, 7\}$ in the state space, $P(t_{i,k-1}, t_{i,k} - t_{i,k-1})$ is the transition probability matrix given in (4) with $t_{i,k}$ denoting the (continuous) age of subject $i$ at visit $k$, and $1$ is a column vector of ones. Note that for Equation (6) it is assumed that $t_{i,1}$ is equal to the baseline age of 50 for the MCSA. Relaxing this assumption along with two other subtle but important features of the MCSA data must to be addressed, and results in slight modifications of this likelihood function. These topics will be addressed in Sections 2.4.1, 3.1, and 3.2.

Next, to complete the specification of the likelihood, the form of the response density $f(y_{i,k})$ and the baseline state probability vector $\pi_0$ must be specified. It is assumed that the four responses, $y_{i,k,1} = \log(\text{PIB} - 1)$, $y_{i,k,2} = \text{thickness}$, $y_{i,k,3} = \text{MMSE}$, and $y_{i,k,4} = \text{Dementia}$, are conditionally independent given state $s_{i,k}$ and subject specific covariates. As illustrated in Figure 2, the transformed PIB measurements are assumed to be generated according to two normal random variables with different means. One of the means, say $\mu_{A-}$, corresponds to the distribution of transformed PIB measurements for an individual in a low amyloid burden state, and the other, say $\mu_{A+}$, corresponds to a high amyloid burden state. Specifically,

$$f(y_{i,k,1}|s_{i,k}) = \text{N}(y_{i,k,1}|\mu_{A-}, \sigma_{\text{pib}}) \cdot I_{\{s_{i,k}\in\{1,3,5\}\}}$$
$$+ \text{N}(y_{i,k,1}|\mu_{A+}, \sigma_{\text{pib}}) \cdot I_{\{s_{i,k}\in\{2,4,6\}\}}, \tag{7}$$

where $I_A$ is the indicator function which is equal to 1 if $A$ and 0 otherwise. The variance of both Gaussians is assumed to be equal to aid in identifiability of the two groups. The density function for thickness and MMSE are defined analogously.

For the thickness response variable,

$$f(y_{i,k,2}|s_{i,k}) = \text{N}(y_{i,k,2}|\mu_{N-}, \sigma_{\text{thick}}) \cdot I_{\{s_{i,k}\in\{1,2\}\}}$$
$$+ \text{N}(y_{i,k,2}|\mu_{N+}, \sigma_{\text{thick}}) \cdot I_{\{s_{i,k}\in\{3,4,5,6\}\}}, \tag{8}$$

and for the MMSE response variable,

$$f(y_{i,k,3}|s_{i,k}) = \text{N}(y_{i,k,3}|\mu, \sigma_{\text{mmse}}), \tag{9}$$

with

$$\mu = \sum_{j=1}^{6} \alpha_j \cdot I_{\{s_{i,k}=j\}} + \alpha_7 \cdot \text{age} + \alpha_8 \cdot \text{male} + \alpha_9 \cdot \text{educ}$$
$$+ \alpha_{10} \cdot \text{apoe4} + \alpha_{11} \cdot \text{ntests}.$$

The first four covariates are the same as those in Section 2.3, and "ntests" is the number of times a subject has taken the MMSE by a given clinical visit. It is observed in the medical practice that scores on the MMSE may improve as an individual becomes familiar with the exam, and so the "ntests" covariate is included to control for this effect.

The probability mass function for misdiagnosis of dementia is

$$P(y_{i,k,4}|s_{i,k}) = \text{Bernoulli}(y_{i,k,4}|p_0) \cdot I_{\{s_{i,k}\in\{1,2,3,4\}\}}$$
$$+ \text{Bernoulli}(1 - y_{i,k,4}|p_1) \cdot I_{\{s_{i,k}\in\{5,6\}\}}, \tag{10}$$

where $y_{i,k,4} = 1$ if the subject was diagnosed with dementia, and $y_{i,k,4} = 0$ if not. Accordingly, $p_0$ and $p_1$ are misclassification probabilities, with $p_0$ the probability of an incorrect diagnosis of dementia, and $p_1$ the probability of an incorrect non-diagnosis of dementia.

Finally, the baseline state probability vector is $\pi_0 = [\pi_{0,1}, \pi_{0,2}, \pi_{0,3}, \pi_{0,4}, 0, 0, 0]'$ where $\pi_{0,j} = P(s_{i,1} = j)$ for $j \in \{1, 2, 3, 4\}$, for all subjects, $i$. As the MCSA did not enroll demented or deceased individuals, subjects would necessarily have been in states 1–4 at baseline. Accordingly, $\sum_{j=1}^{4} \pi_{0,j} = 1$, and the last three components of $\pi_0$ must be zero.

### 2.4.1. Treating Time of Death as Known Without Error

One feature common to many population-based studies is that death is observed without error and time of death is known exactly, so the likelihood must be modified to account for this more precise information (Satten and Longini 1996). Suppose that subject $i$ transitions to death at time $t_{i,n_i}$. The final term in the subject's contribution to the likelihood can be re-expressed as follows. For a state $s_{i,n_i-1}$ at time $t_{i,n_i-1}$, and $\varepsilon \in (0, t_{i,n_i} - t_{i,n_i-1})$, let,

$$B(\varepsilon) := \{s_{i,n_i}(t_{i,n_i}) = 7, s(t_{i,n_i} - \varepsilon) < 7\},$$

where $s(t_{i,n_i} - \varepsilon)$ represents the underlying state of the $i$th subject $\varepsilon$ amount of time before $t_{i,n_i}$. Then $B := \bigcap_\varepsilon B(\varepsilon)$ is the event that the $i$th subject dies precisely at time $t_{i,n_i}$. Further,

$$P(B(\varepsilon)|s_{i,n_i-1}) = \sum_{s(t_{i,n_i}-\varepsilon)=1}^{6} P\big(s(t_{i,n_i} - \varepsilon)|s_{i,n_i-1}(t_{i,n_i-1})\big)$$
$$\cdot P\big(s_{i,n_i}(t_{i,n_i}) = 7|s(t_{i,n_i} - \varepsilon)\big).$$

Thus, dividing both sides by $\varepsilon$ and taking the limit as $\varepsilon \to 0$ gives a proportional likelihood function value of

$$\sum_{s(t_{i,n_i})=1}^{6} P(s(t_{i,n_i})|s_{i,n_i-1}) \cdot Q_{s(t_{i,n_i}),7}(\lfloor t_{i,n_i} \rfloor) \tag{11}$$

evaluated at the event $\{B|s_{i,n_i-1}\}$. The quantity in (11) is interpreted as the average probability of being in each of the first six states the instant prior to death, each weighted by the probability of transitioning to death at the next instant (given by the instantaneous transition rates $Q_{s,7}(\lfloor t_{i,n_i} \rfloor)$). Note that the response functions are not needed/defined when $s_{i,n_i} = 7$ because death is assumed observed without error.

## 3. Population-Based Study Challenges for an HMM

The purpose of this section is to demonstrate the existence of and provide solutions for two critical methodological gaps in the analysis of disease progression for studies like the MCSA. After describing these two overlooked sources of bias in the literature, Section 3.3 serves to demonstrate empirically on a simple synthetic dataset the impact that ensues when the *delayed enrollment bias* is not properly addressed. We argue that the effects of this bias cannot be ignored, and existing software is not equipped to handle this feature.

## 3.1. The Death Rate Bias

A common feature of many human population based studies (*delayed enrollment* or not), is the following *death rate bias*. In addition to not representing those members of the population who would have enrolled if they were not already dead (which is tied into the *delayed enrollment bias*), it is often the case that people are much less likely to enroll in a study if they are very sick and/or dying. As a result, the death rate for the sub population of individuals who are most likely to enroll is probably smaller than the overall population death rate, for at least the first few years after enrollment into the study. Consistent with the difference between this healthier sub-population and the overall population, this phenomenon will lead to a reduced estimate of the death rate and higher likelihoods of other paths in the state space. This represents a bias with respect to the true parameter values of the overall population.

Our proposed approach to correct for the *death rate bias* is to explicitly estimate the bias on the non-dementia to death rate. This can be done in a linear fashion by including two additional parameters in the log-linear rate Equation (5). The first, say $c \leq 0$, will be for estimating the baseline effect of the *death rate bias*, and the second, say $d \geq 0$, will be for estimating the linear slope at which the *death rate bias* vanishes for every integer year in the study (since the time effect is discretized annually). That is, Equation (5) for only the non-dementia to death rate ($l \in \{3, 5, 8, 10\}$) becomes,

$$\log(q_l) = \beta_0^{(l)} + \beta_1^{(l)} \cdot \text{age} + \beta_2^{(l)} \cdot \text{male} + \beta_3^{(l)} \cdot \text{educ}$$
$$+ \beta_4^{(l)} \cdot \text{apoe4} + g(\text{iyears}), \quad (12)$$

where "iyears" is integer years enrolled, and $g(\text{iyears}) := \min\{c + d \cdot \text{iyears}, 0\}$. The *death rate bias* term is only allowed to decrease the non-dementia to death rate. Further, the smallest root of $g$ is the duration for which the *death rate bias* persisted in the study.

The coefficients in Equation (12) become identifiable due to strong prior information available for the overall population death rate. We provide evidence that this bias exists in the MCSA data by appealing to the fact that $c$ and $d$ are both estimated to be nonzero (see Figure 9), and by demonstrating in Section 4 that all parameters in Equation (12) can be accurately estimated in a similar, truth known synthetic data setting.

## 3.2. The Delayed Enrollment Bias

The *delayed enrollment bias* occurs in situations in which the first observation time for the $i$th subject, $t_{i,1}$, is not necessarily equal to the baseline age of 50 years old (in the case of the MCSA). In this case, the probability of transitioning from the (unknown) underlying state at baseline to the (unknown) underlying state at the first observation must be accounted for in the likelihood function. That is, the baseline state probability vector $\boldsymbol{\pi}_0$ in (6) is replaced with,

$$\boldsymbol{\pi}(t_{i,1}) = \begin{bmatrix} v_1(t_{i,1}), v_2(t_{i,1}), v_3(t_{i,1}), v_4(t_{i,1}), 0, 0, 0 \end{bmatrix}'$$
$$\times \frac{1}{\sum_{j=1}^4 v_j(t_{i,1})}, \quad (13)$$

where $v(t_{i,1})' = \boldsymbol{\pi}_0' P(50, t_{i,1} - 50)$. The last three components are set equal to zero due to the fact that demented and dead subjects are not enrolled into the study.

If the initial probabilities from baseline to enrollment age were not conditioned on the underlying states being non-demented and non-dead, then the transition rates (especially those to dementia and death) will exhibit strong downward bias (with respect to the true population parameter values) due to the fact that neither demented nor dead subjects are enrolled. We refer to this bias as the *delayed enrollment bias*, and it will be illustrated empirically in Section 3.3. Relevant standard software such as the msm package in R (Jackson 2011) does allow for specification of a common baseline in a *delayed enrollment* study (after manually adding censored observations in the dataset at baseline), but unfortunately does not offer the ability to perform conditioning (rescaling) on the initial state probability vector, as in (13).

## 3.3. Demonstrating the Effect of a Delayed Enrollment Bias

The cardiac allograft vasculopathy (CAV) dataset from the msm package was collected from a study of the progression of CAV, which is a common cause of death after heart transplant (Sharples et al. 2003). The state space as described in Jackson (2011) includes four states labeled "no CAV" (state 1), "mild/moderate CAV" (state 2), "severe CAV" (state 3), and "death" (state 4), and forward-only transitions are assumed (patients can only get worse). Observed remissions in the state of a patient are considered a result of misclassification error. In the CAV data, baseline is defined as time of heart transplant. All patients are observed at baseline, and in fact at baseline all patients are in state 1 because CAV does not develop immediately.

In this simple dataset, the only response variable is the observed state at each visit which is assumed to follow a categorical distribution. In particular, given a patient is in state 1, 2, or 3 there is a nonzero probability of observing an adjacent state. Additionally, as is the case for the MCSA, death and time of death are known without error. Formally, given a true state (row) the response probability mass function takes the form given in Table 1, where the probabilities $p_1$, $p_2$, $p_3$, and $p_4$ are each interpreted as the probability a of particular misclassification. Note that the rows must sum to one.

Time is treated as continuous and discretized annually for the transition rate matrix. Integer years since heart transplant, and gender are included as covariates on the transition rates for the CAV data,

$$\log(q_l) = \beta_0^{(l)} + \beta_1^{(l)} \cdot \text{iyears} + \beta_2^{(l)} \cdot \text{sex}, \quad (14)$$

**Table 1.** Misclassification response function.

| | | Observed state | | | |
|---|---|---|---|---|---|
| | | No CAV | Mild | Severe | Death |
| True state | No CAV | $1 - p_1$ | $p_1$ | 0 | 0 |
| | Mild | $p_2$ | $1 - p_2 - p_3$ | $p_3$ | 0 |
| | Severe | 0 | $p_4$ | $1 - p_4$ | 0 |
| | Death | 0 | 0 | 0 | 1 |

where

$$\boldsymbol{Q} = \begin{pmatrix} -q_1 - q_2 & q_1 & 0 & q_2 \\ 0 & -q_3 - q_4 & q_3 & q_4 \\ 0 & 0 & -q_5 & q_5 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

To replicate the CAV data in a "known truth" simulation, first the HMM parameter estimates were obtained using the msm package on the original CAV data. As suggested in Jackson (2011), the "BFGS" quasi-Newton optimization algorithm was specified to estimate the HMM parameters with the "msm" function. The parameter estimates from the CAV dataset are then used as the "true" values to generate new datasets.

Important features of the MCSA as distinct from the CAV data are that not all subjects are in state 1 at baseline, and virtually none of the subjects are observed at baseline. Thus, to more closely resemble the MCSA data, each patient in the simulated CAV data is generated with a baseline state according to the following distribution,

| $s_{i,1}$ | No CAV | Mild | Severe | Death |
|---|---|---|---|---|
| $P(s_{i,1})$ | 0.95 | 0.04 | 0.01 | 0 |

See the supplementary materials for complete detail on how synthetic data was generated.

A sample size of 2000 was generated for 100 simulated datasets. Figure 5 shows estimates of the log-rate intercept coefficients from Equation (14) using the posterior mean from the proposed Bayesian approach, and the MLE obtained via msm. The boxplots are over the 100 estimates of each parameter and demonstrate that in a simple idealized setting the Bayesian and MLE estimates are very similar.

In the synthetic data example above, all patients are observed at baseline; this is the type of study for which the msm package was designed. To illustrate the bias that ensues when subjects are not observed at baseline, the 100 datasets are generated once more from the same random generator seeds. However, instead of beginning with the initial observations at baseline (zero years after heart transplant), the time of initial observation is generated, with probability 0.75, from a Gaussian distribution with a mean of 5 integer years and a SD of 1, that is, the initial observations remain at baseline with probability 0.25. Additionally, if a patient transitions to death prior to the generated initial observation time, then the patient is not included in this *delayed enrollment* study. This is a critical point because it illustrates a root cause of the bias: the *delayed enrollment* study is less likely to include patients with immediate adverse reactions to heart transplant. If the sample were truly representative, then all 2000 patients would be represented. However, studies typically only sample from the living. The average sample size of the 100 synthetic datasets for the *delayed enrollment* study is 1686.

In Section 3.2, the procedure for accounting for the *delayed enrollment bias* in the likelihood function was described. It amounts to evaluating the transition rate matrix, $\boldsymbol{Q}$, (here, annually) from baseline to initial observation, and then computing the conditional probabilities for the initial states of enrollment. This conditioning feature is not available in the msm package which was not designed for a *delayed enrollment* study. Figure 6 illustrates the effect of ignoring this feature and estimating as

though enrollment is *not* conditional on being alive, that is, MLE (msm). The estimates from Bayes and MLE are analogous to those in Figure 5, however, they do explicitly account for the initial probability according to (13). Without accounting for the *delayed enrollment* effect in the likelihood function certain estimates are significantly biased downward, suggesting slower rates of transition. The biases become more extreme as fewer patients are observed at baseline; recall that about 25% of the patients are still observed at baseline in this example.

The bias also filters into other HMM parameter estimates; see the supplementary materials for the full results of these two simulation setups. The objective of this synthetic *delayed enrollment* study example is to demonstrate one of the crucial reasons why the analysis of the MCSA data requires methodological developments which are not readily available in standard software. The msm package, as well as other similar software, are not flawed, rather they were simply not designed for this type of application.
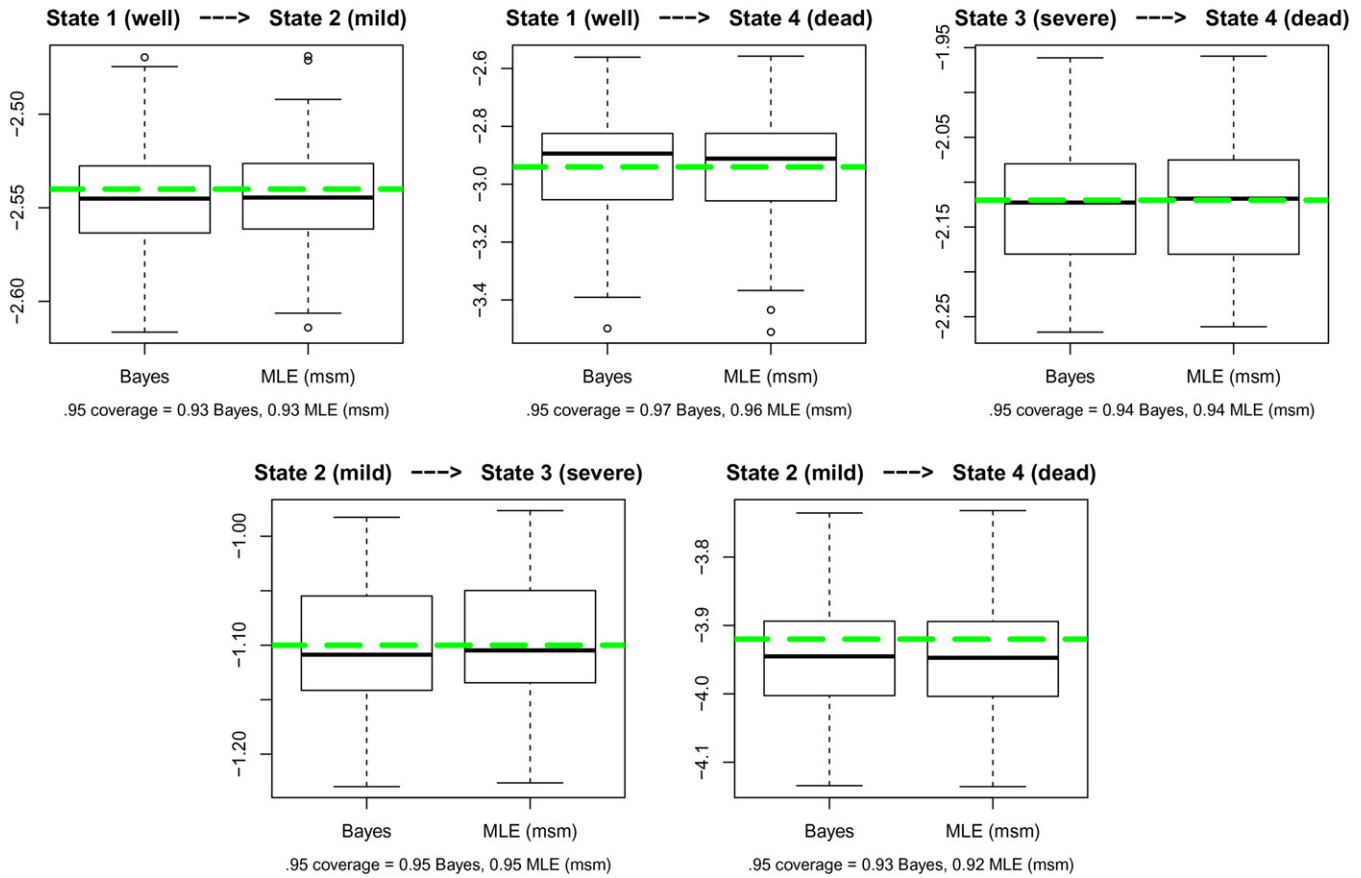
## 4. Synthetic Mayo Clinic Study of Aging Data

Section 3.3 presented the results of the proposed estimation procedure in a simple idealized simulation example. In this section, a more realistic simulation is presented which is intended to closely replicate the MCSA data generating process with respect to sample size, the frequency of clinical visits, and the proportion of biomarker measurements available. The three objectives are to (i) provide evidence that the synthetic data reasonably resembles the real data in an effort to verify that the data generating mechanisms of the real data are sufficiently understood, (ii) validate the estimation procedure by demonstrating that credible regions concentrate around the true parameter values, and (iii) demonstrate the reliability of the estimates with respect to the true parameter values.

The standard R package for estimating an HMM with arbitrary continuous-time observations is msm (Jackson 2011). While msm is a well written and powerful package, it does not offer Bayesian estimation options, and as discussed previously it cannot correct for the biases on the transition rates that arise from a *delayed enrollment* sample.

The same techniques used to generate data resembling the CAV dataset are applied here, just with more features to simulate such as the additional response functions and a *death rate bias*. See the supplementary materials for the particulars of how the synthetic data was generated. Before diving right into the estimation results, a few estimation details are discussed, mainly relating to prior information.

In the state space of Figure 1, there are a number of assumptions which can reasonably be made about the biology of this process. For instance, the rate of transition to the N+ state with high amyloid burden should be at least as fast as with low amyloid burden. That is, the rate parameter from $\boldsymbol{Q}$ for transitioning from A+N− to A+N+ should be at least as big as the rate parameter for transitioning from A−N− to A−N+. Similar constraints should be true for transitioning from low to high amyloid burden with respect to high/low neurodegeneration burden (i.e., having N+ cannot lower the rate of transitioning to A+), and those with dementia die at a rate no less than those without dementia. These constraints are a mathematical

**Figure 5.** Intercept coefficient estimates for a traditional study in which all subjects are enrolled at a common baseline time. "Bayes" corresponds to the Bayesian posterior mean estimates, and "MLE (msm)" corresponds to maximum likelihood estimates computed from the "msm" function. Green dashed lines represent the true values. Coverage is the proportion of 0.95 probability credible intervals (confidence intervals for the MLE) which contain the true parameter value.

formulation of the assumption that having a larger burden may not escalate disease progression, but it certainly cannot help. A table with the full list of rate constraints is given in the supplementary materials.

In addition to these constraints, it is reasonable to assume that all of the age coefficients on the rate parameters (see Equation (5)) are nonnegative. That is, becoming older will not slow down an individual's rates of progression. Next, constraints are placed on the response variable parameters to help with identifiability. The mean $\log(\text{PIB} - 1)$ measurement associated with the low amyloid burden state should not exceed that of the high amyloid burden state. Analogously, the mean thickness measurement associated with high neurodegeneration burden should be less than or equal to that for low neurodegeneration burden (low thickness is associated with high burden). Lastly, the mean MMSE scores should be monotone non-increasing in the state ordering $\{4, 5, 6\}$, the mean MMSE score for state 1 should be no smaller than that for states 2 and 3, and the mean MMSE score for states 2 and 3 should be no smaller than that for state 4. No constraint is placed between the mean scores for states 2 and 3.
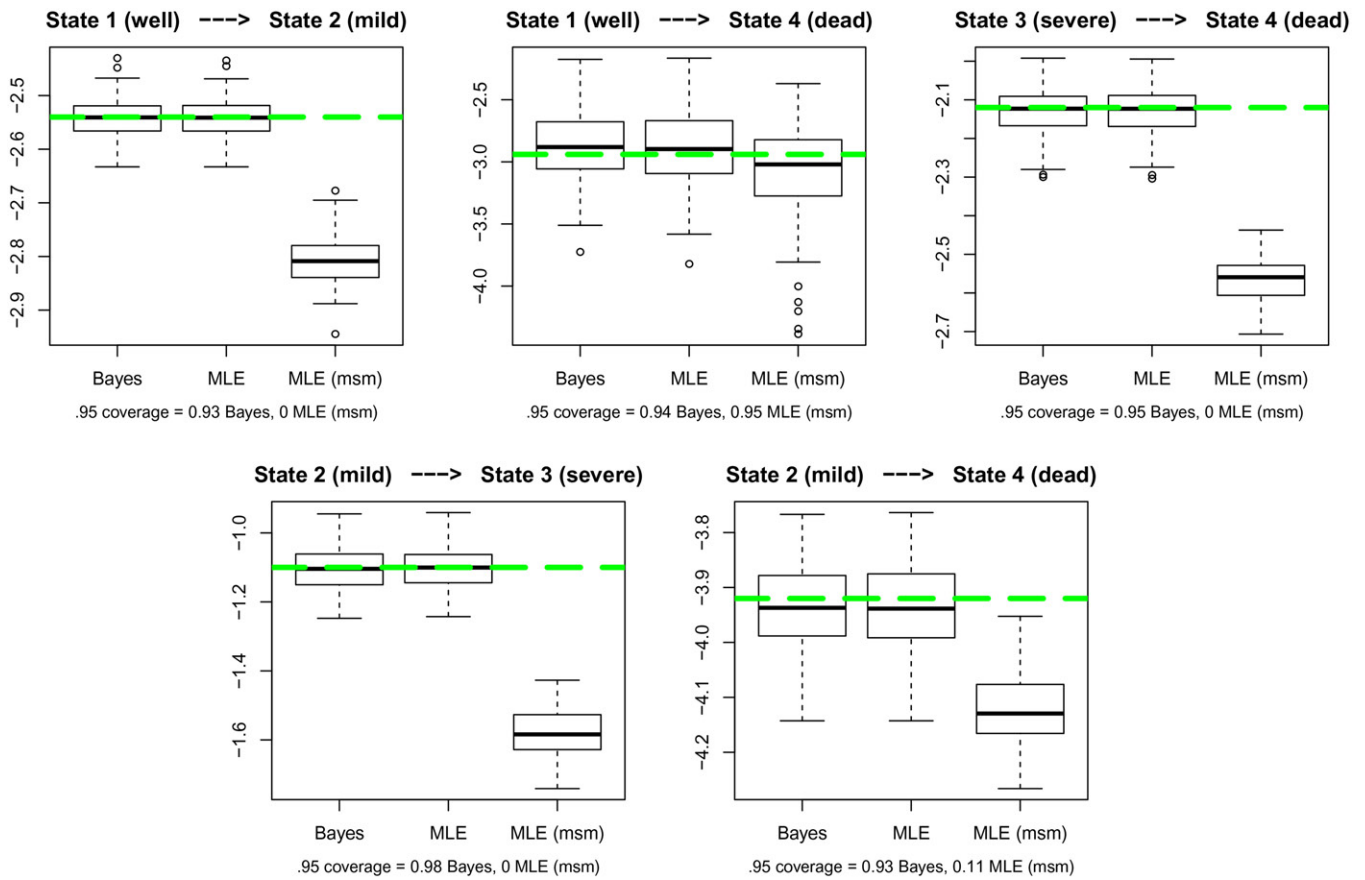
One of the advantages of the Bayesian approach of estimation is that imposing this long list of important constraints in the model can be easily handled through specification of the priors, and the priors have a natural ability to accommodate other known information about the model parameters. In particular, the MCSA sampled subjects are from the greater Rochester, MN

area, and Minnesota death rates for women and men of all ages (from 1970 to 2004) are made available in the US Decennial Census, which are captured in the "survexp.mn" dataset in the `survival` package in R (Therneau 2015). These gender-specific rates are used to set the prior means placed on the baseline and gender coefficients for the log death transition rate (Equation (12)), see Table 2. It is assumed that all non-dementia to death rates are the same (and hence share the same coefficients in Equation (12)). The fact that the A and N states do not have obvious external manifestations makes this assumption reasonable.

A multivariate normal prior is placed on the vector of HMM parameters, and multivariate normal proposals are used. The full table of prior distribution specifications for each of the 81 parameters in the model is provided in Table 2. For parameters which are nonnegative, such as the variance parameters for the response functions, the Gaussian priors are placed (and Gaussian proposals are made) on the natural logarithm of the parameter. For parameters which are constrained to be between 0 and 1, such as the dementia misclassification parameters and the baseline state probability parameters, the logit transformation of the parameters is used. For example, the baseline state probability vector is re-expressed as

$$\boldsymbol{\pi}_0 = \left[1, e^{\xi_1}, e^{\xi_2}, e^{\xi_3}, 0, 0, 0\right]' \frac{1}{1 + e^{\xi_1} + e^{\xi_2} + e^{\xi_3}}, \quad (15)$$

where $\xi_1$, $\xi_2$, and $\xi_3$ are assigned Gaussian priors and proposals.

**Figure 6.** Intercept coefficient estimates for a *delayed enrollment* study. "Bayes" corresponds to the Bayesian posterior mean estimates, "MLE" corresponds to the MLE obtained via optimizing the likelihood function from Section 2.4 using the "optim" function in R, and "MLE (msm)" corresponds to maximum likelihood estimates computed from the "msm" function. Green dashed lines represent the true values. Coverage is the proportion of 0.95 probability credible intervals (confidence intervals for the MLE) which contain the true parameter value.

Although commonly done in the literature (mostly due to conjugacy), we hesitate to use inverse-gamma priors on variances for reasons discussed in Gelman (2006); namely, the inverse-gamma($\varepsilon$,$\varepsilon$) family of priors is very sensitive to the choice of $\varepsilon$, which does not naturally lend itself as weakly informative nor uninformative. Moreover, we view our priors in general as weakly informative in the sense of Gelman (2006). That is, they are set intentionally weaker than what we believe the expert domain knowledge warrants, with the exception of the priors for parameters associated with the non-dementia to death rate and the *death rate bias*. Further, without having more knowledge about the shape of the prior distributions, the symmetry, exponential tails, and mathematical simplicity of Gaussian priors make them a natural choice. They allow us to be as diffuse as we believe appropriate without affording a questionably large amount of mass at the extreme values of the distributions. We investigate/discuss sensitivities of our posterior estimates to specifications of our priors in Section 5.1.

The simulation study consisted of simulating 50 synthetic datasets resembling the MCSA data. The synthetic datasets were simulated to contain similar amounts of information to the real dataset. That is, 4742 subjects were simulated starting from random ages and assigned covariate values randomly from the empirical distribution of the MCSA data. The simulated subjects are "observed" at times which are determined by sampling from the actual inter-observation times in the MCSA data.

The maximum length of time in the study for subjects in the MCSA is not much longer than 12 years. Thus, for reasonable comparison, subjects in the synthetic datasets are observed for a maximum of 12 years or until time of death, whichever comes first. To simulate attrition in the study, for each generated clinical visit there is a 0.9125 probability that the given subject remains in the study. The probability of 0.9125 was chosen because it yields a comparable distribution of number of clinical visits and observed deaths in the synthetic data to that of the real MCSA data, as discussed shortly. Exactly 2718 (approximately 57%) of subjects in the MCSA dataset have at least one observed biomarker measurement, and both biomarkers were observed for 1740 subjects. Of the 2718 subjects, the proportions of observed biomarkers over all visits is presented in the following table,
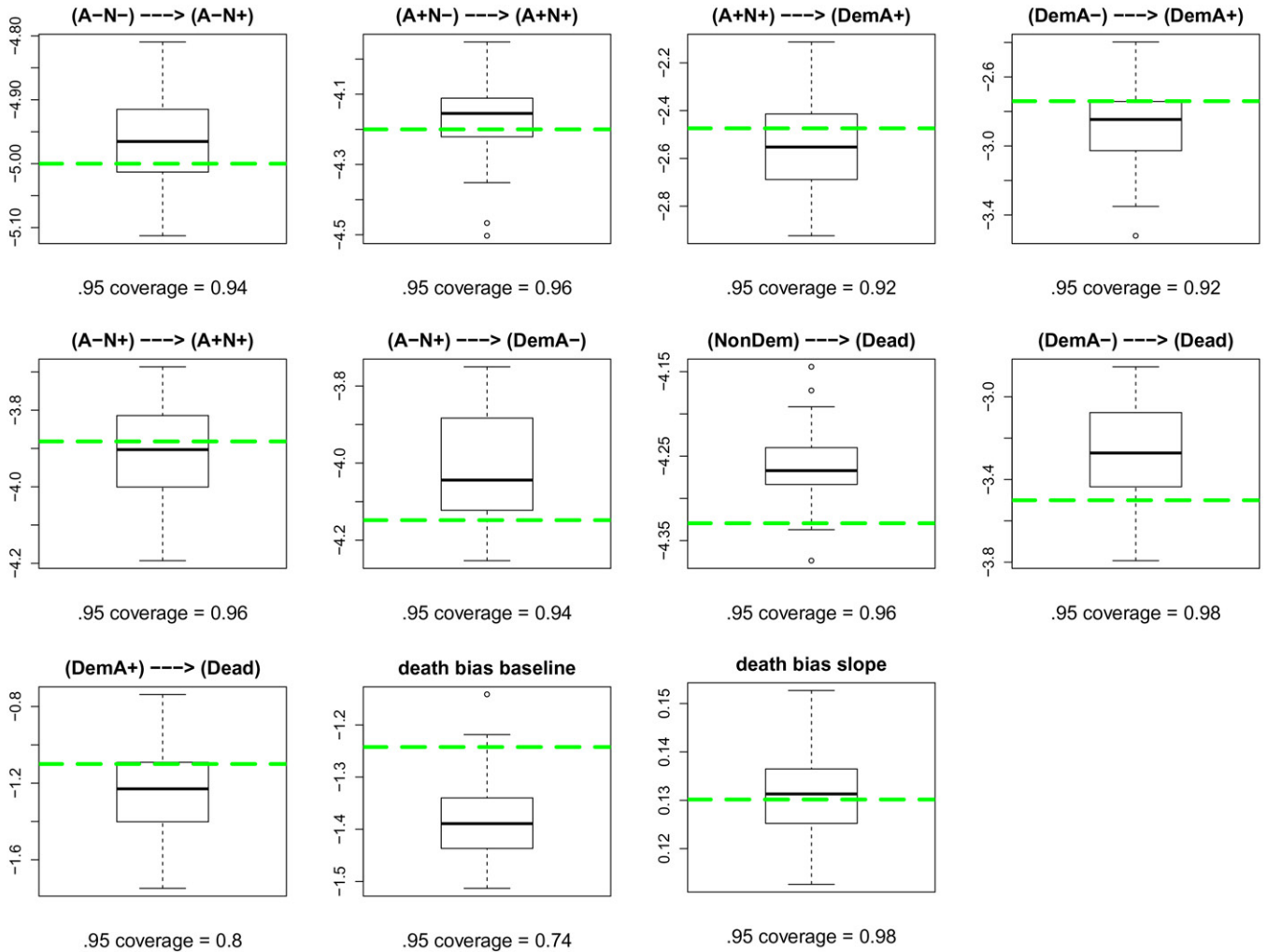
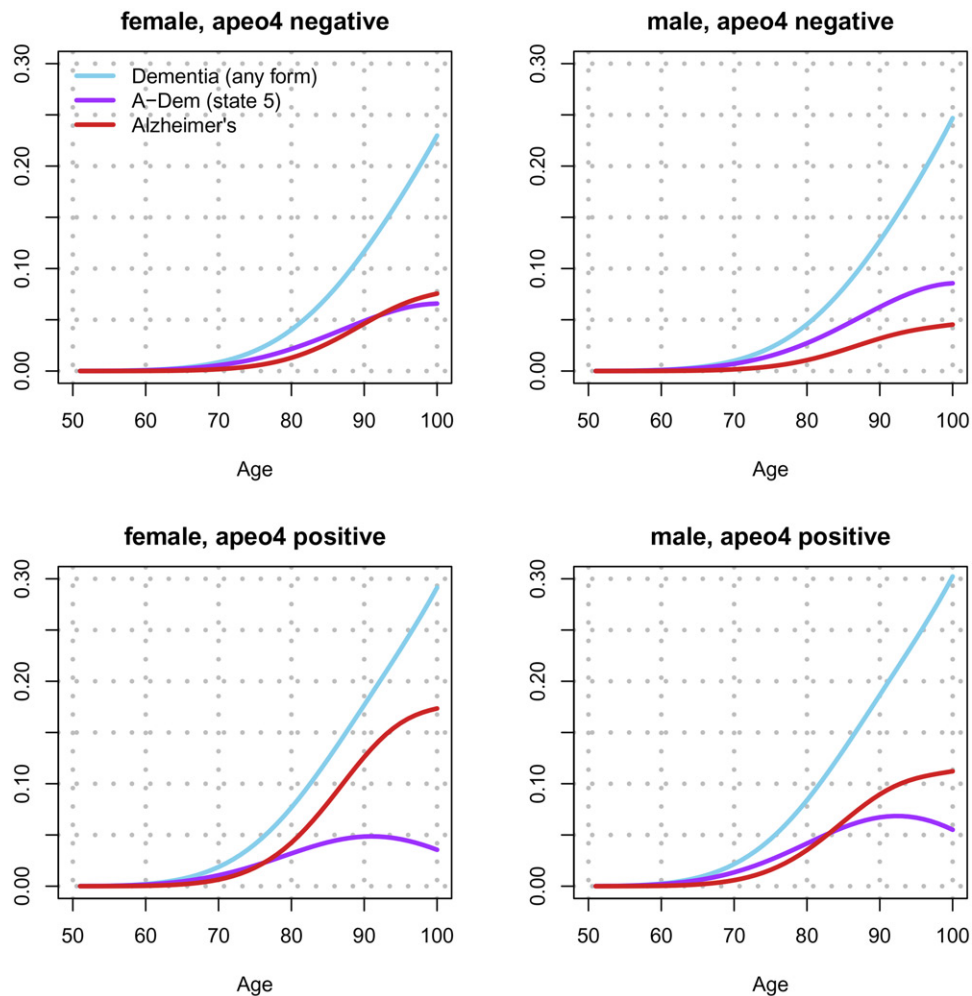|  |  | PIB | |
|---|---|---|---|
|  |  | measured | not measured |
| Thickness | measured | 0.227 | 0.231 |
|  | not measured | 0.002 | 0.540 |

To keep consistent with this feature, ∼43% of the synthetically generated subjects were given no biomarker data for any of their visits, while the remaining ∼57% were randomly given observed PIB or thickness measurements, at each clinical visit, according to the above distribution.

**Table 2.** Displayed are the means and SDs of the (independent) normal priors placed on the 81 model parameters.

| | | | Prior means and SDs | | | | |
|---|---|---|---|---|---|---|---|
| **State transition parameters (see (5) and (12))** | | | | | | | |
| Transition | $\beta_0^{(l)}$ | $\beta_1^{(l)}$ | $\beta_2^{(l)}$ | $\beta_3^{(l)}$ | $\beta_4^{(l)}$ | $c$ | $c + d$ |
| non-dem→7 | −4.41 (0.1) | 0.094 (0.01) | 0.47 (0.05) | 0 (0.1) | 0 (1) | −0.75 (0.375) | −0.60 (0.3) |
| 5→7 and 6→7 | −4 (1) | 0.1 (0.05) | 0 (1) | 0 (0.1) | 0 (1) | | |
| all others | −3 (1) | 0.1 (0.05) | 0 (1) | 0 (0.1) | 0 (1) | | |
| **Cubic spline parameters for state 1 to state 2 as a function of age** | | | | | | | |
| $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | $c_6$ | $c_7$ | $c_8$ |
| −5 (1) | −4 (2) | −3 (2) | −2 (2) | −1 (2) | 0 (3) | 1 (3) | 2 (3) |

| **log(PIB − 1) response (see (7))** | | | | **Thickness response (see (8))** | | |
|---|---|---|---|---|---|---|
| $\mu_{A-}$ | $\mu_{A+}$ | $\log(\sigma_{pib}^2)$ | | $\mu_{N-}$ | $\mu_{N+}$ | $\log(\sigma_{thick}^2)$ |
| −1.3 (0.2) | −0.5 (0.2) | $\log\left((0.4/3)^2\right)$ (2) | | 3.14 (0.2) | 2.34 (0.2) | $\log\left((0.4/3)^2\right)$ (2) |

| **MMSE response (see (9))** | | | | | | | |
|---|---|---|---|---|---|---|---|
| $\alpha_1$–$\alpha_4$ | $\alpha_5$–$\alpha_6$ | $\alpha_7$ | $\alpha_8$ | $\alpha_9$ | $\alpha_{10}$ | $\alpha_{11}$ | $\log(\sigma_{mmse}^2)$ |
| −0.28 (0.75) | −7.3 (3) | 0 (1) | 0 (1) | 0 (1) | 0 (1) | 0 (1) | −0.7 (2) |

| **Dem misclass (see (10))** | | | | **Baseline probabilities (see (15))** | | |
|---|---|---|---|---|---|---|
| logit($p_0$) | logit($p_1$) | | | $\xi_1$ | $\xi_2$ | $\xi_3$ |
| −3 (1) | −3 (1) | | | −3.5 (0.25) | −6 (1) | −6 (1) |

NOTE: SDs are in parentheses, and these priors assume that the data have been centered. The parameters $c_1, \ldots, c_8$ are the control points used to estimate the cubic spline for the state 1 to state 2 transition (for baseline and age), as discussed at the end of Section 2.3.



**Figure 7.** Intercept and *death rate bias* coefficient estimates for the synthetic MCSA data, see (5) and (12). Note that the covariates in the data are centered. Presented are boxplots of posterior means of the labeled parameters, from 50 synthetic MCSA datasets. Green dashed lines represent the true values. Coverage is the proportion of 0.95 probability credible intervals which contain the true parameter value.

**Figure 8.** Evolution of transition probabilities. The curves represent the probability of transitioning to the respective state for each given age, computed using the posterior mean estimates of the HMM parameters. The probabilities are conditional on not transitioning to state 7 (Dead), and correspond to an individual in state 1 (A−N−) at the baseline age of 50. The label "apoe4 negative" corresponds to an individual with no APOE-$\varepsilon$4 alleles, and "apoe4 positive" corresponds to an individual with at least one APOE-$\varepsilon$4 allele. The curve labeled "Alzheimer's" depicts the probability of making the transition from state 4 (A+N+) to state 6 (A+Dem), given not dead.

Death was observed for just over 28% of the actual MCSA study subjects, and the number of clinical visits for study subjects varied between 1 and 10 visits, with a median of 4. The synthetic datasets observe death for ~31% of subjects on average, and the number of clinical visits for synthetic study subjects varies between 1 and 11 visits, with a median of around 6.

Figure 7 provides a summary of the results in the form of boxplots of the posterior mean estimates, and coverage for 95% credible intervals, for 11 of the more interesting model parameters. The supplementary materials contains a more detailed summary of the results, including boxplots, coverages, histograms, and MCMC trace plots for all 81 model parameters. Overall, this simulation exercise lends some confidence to the results of the actual MCSA analysis presented next.
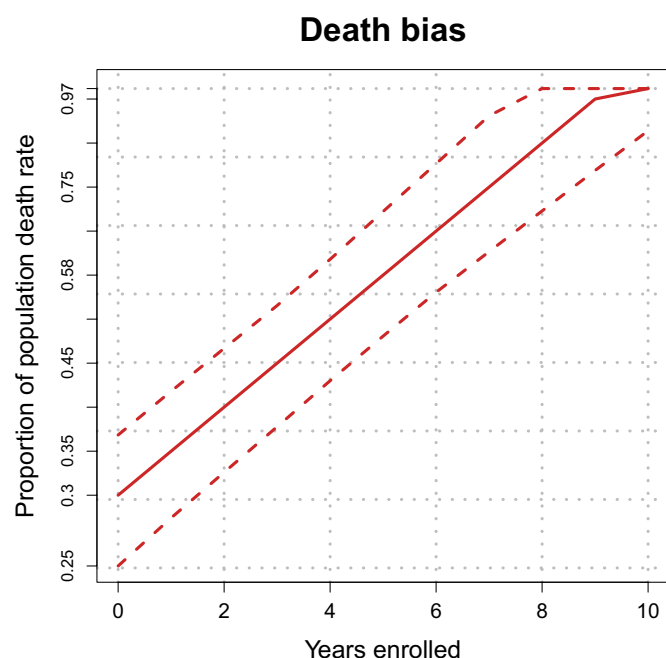
## 5. Analysis of the Mayo Clinic Study of Aging Data

This section presents analysis and interpretation of the HMM parameter estimates of the actual MCSA data. Since there are 81 HMM parameters, some playing very different roles, presenting the output in a concise manner is challenging and depends on

the question being asked. We present a targeted summary of a few important questions here. See the supplementary materials for estimates of all 81 HMM parameters, including the MCMC trace plots, histograms, and 95% credible intervals. We parallelize our likelihood computation (within each MCMC step) over 30 threads on a computing cluster, and for 10 unique random number generator seeds we run the MCMC algorithm (i.e., using a total of 300 threads) for about 3 days on the real MCSA dataset.

The state space as it is defined in Figure 1 allows for the computation of transition probabilities broken down for particular types of dementia. Most notably, the development of Alzheimer's disease as defined here corresponds to a transition from state 4 (A+N+) to state 6 (A+Dem), that is, there was Amyloid build up prior to the neurodegeneration leading to Dementia. See Figure 8 for the estimates of how these transition probabilities evolve over time. While Alzheimer's disease is slightly more prominent among females of a given age versus males of the same age, it is interesting that the likelihood of dementia (of any kind) is nearly the same, that is, males are more likely to develop non-Alzheimer's related dementia.

## Death bias



**Figure 9.** Posterior mean estimate of the *death rate bias* (12). Values are interpreted as the proportion of the population death rate that is experienced by subjects in the MCSA, for each integer year a subject is enrolled in the study. For example, subjects just enrolled in the study experience a death rate which is 31% of the population death rate. The dotted lines are 95% credible bands.

**Table 3.** Posterior mean estimates of the dementia diagnosis response parameters.

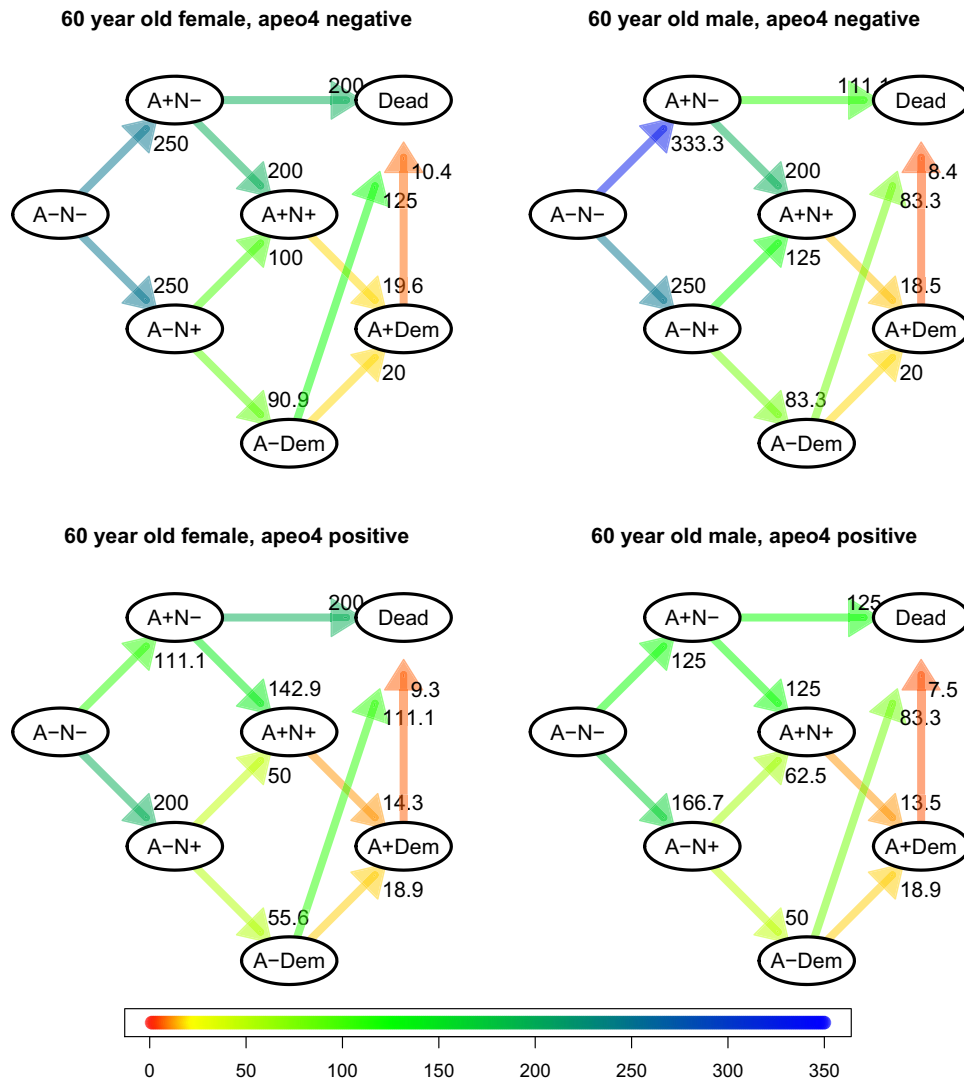| | | Observed status | |
|---|---|---|---|
| | | Diagnosed not demented | Diagnosed demented |
| True status | Not demented | 0.992 | 0.008 [0.007, 0.009] |
| | Demented | 0.107 [0.072, 0.152] | 0.893 |

NOTE: The components are probabilities. Note that each row corresponds to only one parameter, but both columns have been filled in for ease of interpretability (rows must sum to one). The brackets represent 95% credible intervals (for the components directly corresponding to the parameters which were estimated).

There are many variables whose effect is known medically that could be used to validate some of the results. For example, it is observed that the rates are relatively dormant for the first two decades starting from 50 years old. After that, transition rates among the first four states seem to increase slower than the transition rates to more advanced states. Transitions particularly to A+Dem from A+N+ or from A−Dem are the most intense over almost all ages. APOE-$\varepsilon 4$ is known to increase risk of A+ by a factor of 2–3. Based on this, the model should yield a higher rate estimate of A−N− (state 1) to A+N− (state 2) and A−N+ (state 3) to A+N+ (state 4) among APOE-$\varepsilon 4$ carriers than noncarriers. Examination of Figures 10 and 11 illustrates that at both ages and for both men and women, the relationship between both of these transition rates and APOE-$\varepsilon 4$ in the model output is exactly as would be expected. Additionally, the rate of A+N+ to A+Dem should be greater in APOE-$\varepsilon 4$ carriers than noncarriers. This is the case in the model for both ages shown and for both men and women as would be predicted from the known biology.

### 5.1. Sensitivity to Prior Densities

As with any Bayesian analysis one must consider the sensitivity of the resulting posterior estimates with respect to the prior specifications. Since the non-dementia to death rate is known to not deviate far from the overall Minnesota death rate, tight priors serve to untangle ambiguity between the death rate and other parameters in a broad sense, or "on average." It should be noted that the non-demented to death rate and *death rate bias* parameters are quite sensitive to the variances placed on their respective priors. We remark once more that the methodology we propose in Section 3.1 for estimating/correcting the *death rate bias* relies on strong prior information available for the overall population death rate. In investigating the sensitivity of the HMM parameter estimates to the priors for the non-dementia to death rate we find that the only unstable parameter estimates are those associated with the rate itself and the *death rate bias*, due to a lack of identifiability.

To study sensitivities to the specification of all priors other than those related to the non-dementia to death rate, we increased all other prior SDs by a factor of 10 and recomputed the HMM parameter estimates. See the supplementary materials to compare the posterior estimates between the more and less diffuse prior specifications. It was observed that the estimated posterior distributions of each parameter were largely unchanged, with the exception of the log-cubic spline parameters for the transition from state 1 to state 2, and the baseline state probabilities. This can most likely be attributed

The estimated *death rate bias* for this study is also of interest, particularly due to the novel approach taken to account for it. It is observed that individuals just enrolled in the study experience a death rate which is ∼31% (posterior mean) of the population death rate and it remains lower than the rest of the population for several years after enrollment; see Figure 9. This suggests that the *death rate bias* cannot be ignored.

Another feature of the analysis is that it is cut-point agnostic, by design. Instead of hardwiring cut-points, the suggested cut-points in the medical literature have been used as prior information for the response distributions from high/low amyloid burden and high/low cortical thickness loss burden in the log(PIB − 1) and thickness measurements, respectively. Recall, Figure 2 displayed the estimated distribution of these response biomarkers for high/low burden states.

Table 3 shows the estimated dementia misclassification probabilities. These estimates indicate that physicians tend to be conservative in diagnosing dementia. That is, they very seldom diagnose an individual without dementia as demented, but about 1 out of 10 individuals that truly have dementia is not diagnosed as such.

Finally, Figures 10 and 11 provide heat maps of the state space corresponding to individual specific estimated transition intensities (posterior mean estimates). These depictions of the state space in relation to the estimated parameters capture a holistic picture of the model as a physical system. With these plots, it is easy to identify which transition rates are most intense at different ages, and to get a general sense of when rates begin to "heat up." In fact, these plots can be created for every integer age greater than or equal to 50, and they are presented as a movie in the supplementary materials.

**Figure 10.** Posterior mean estimates of the *reciprocal* transition rate components of *Q* at age 60. The numerical values can be interpreted as the estimated mean times (in years) to transition, conditional on age. These plots correspond to an individual with a college degree. Recall that transitions to dead from states 1–4 are constrained to be equal, and so for ease of presentation only one transition arrow is shown.
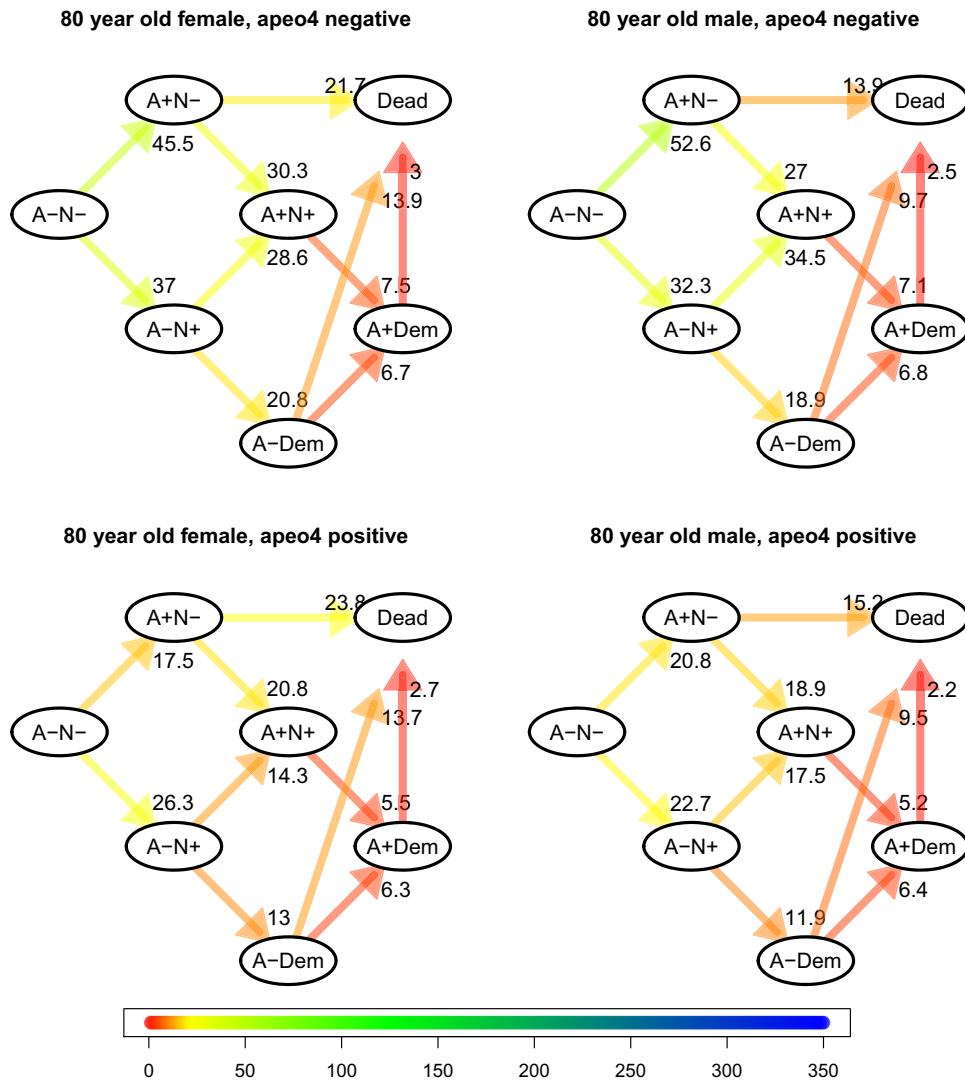
to the large amount of flexibility afforded by a cubic spline, and a degree of un-identifiability between the cubic spline and baseline state parameters because all other parameter estimates are unaffected. Moreover, the original tighter priors set in Table 2 for the baseline state probabilities are well-informed and reasonable, according to population data.

## 6. Conclusions and Future Work

A continuous-time HMM was developed for the analysis of the MCSA data. Much care was taken to make this model as realistic of an approximation to the actual data generating process as possible, including the treatment of important features such as *delayed enrollment* and *death rate bias*. A Bayesian computational framework was developed to facilitate computation and quantification of uncertainty, as well as allow for essential prior information on many of the parameters. The model and its estimation performance were validated via simulation studies,

prior to presenting the results of its application to the MCSA data. Several important findings were that (i) the *delayed enrollment* and *death rate bias* play a significant role in this study, (ii) females of a given age are more prone to Alzheimer's related dementia than male counterparts, but they are less prone to non-Alzheimer's related dementia, and (iii) individuals with at least one APOE-$\varepsilon$4 allele are more than twice as likely to develop Alzheimer's than those with no alleles.

Our work builds on the simpler Markov model of the MCSA from Jack et al. (2016), and could be viewed as a competing model. We include the additional covariates for gender, education, and presence of an APOE-$\varepsilon$4 allele on the state transition rates, and introduce the emitted response variables associated with amyloid and cortical thickness to allow for agnostic biomarker cut-points. Additionally, we consider the emitted response variables for MMSE scores and physician diagnosis misclassification of dementia, as well as provide methodology for estimating/correcting the *death rate bias*. The model from Jack et al. (2016) used fixed death rates from Minnesota

**Figure 11.** Posterior mean estimates of the *reciprocal* transition rate components of **Q** at age 80.

population data. These added features of our model allow for greater flexibility in the theorized data generating model, and deeper understanding of the biology. Admittedly, though, if the proposed model became much more complex with additional covariates (or other features) it would require some decisions about which covariates to include in which equations. Stochastic search variable selection is a viable option as it has been used successfully on event rate models (Storlie et al. 2013; George and McCulloch 1993). Other options include spike and slab priors (Ishwaran and Rao 2005), or computing competing models and comparing them with some criterion such as the widely applicable information criterion (WAIC, see Watanabe 2010).

One limitation of the proposed approach is that it does not allow for rates to vary depending on the extent of Amyloid (or cortical thickness) burden. It is known that once there is sufficient Amyloid build up, then the rate of neurodegeneration is elevated, however, this effect may not be constant across all levels of Amyloid build up. This could be tested by allowing for multiple discrete Amyloid states (e.g., low, medium, high). However, this feature would be best addressed with a continuous-state space for both Amyloid and cortical thickness burden. This is a subject of future work.

An additional remark is that the various constraints placed within our model such as on the transition rates and on the response function parameters are rather stringent, but they reflect expert domain knowledge. Moreover, constraints such as those on the PIB and thickness mean parameters simply make the parameters identifiable. Nonetheless, imposition of such constraints should not be decided ad hoc. And when concerns arise as to the reasonability of various constraints, sensitivity tests should be performed to more fully understand the implications. A similar remark applies to prior specifications, and as we noted in Section 5.1 we find that our estimated model, particularly for parameters associated with the death rate and *death rate bias*, are heavily reliant on the availability of strong prior information for population death rates.

Finally, it is likely the case that our computations could be made more efficient by implementing Hamiltonian Monte Carlo methods. This article lends credibility to and provides verification for our theorized model and features of the MCSA data. Accordingly, a natural next step is to sharpen our computational strategies which would facilitate deeper explorations of the features of the data.

## Supplementary Materials

The online supplementary materials contain details of the computations, the code for reproducing all the results of the article, and further analysis of the presented simulation studies and MCSA data results.

## References

Bureau, A., Shiboski, S., and Hughes, J. P. (2003), "Applications of Continuous Time Hidden Markov Models to the Study of Misclassified Disease Outcomes," *Statistics in Medicine*, 22, 441–462. [17]

Gelman, A. (2006), "Prior Distributions for Variance Parameters in Hierarchical Models (Comment on Article by Browne and Draper)," *Bayesian Analysis*, 1, 515–534. [25]

George, E. I., and McCulloch, R. E. (1993), "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889. [30]

Ishwaran, H., and Rao, J. S. (2005), "Spike and Slab Variable Selection: Frequentist and Bayesian Strategies," *The Annals of Statistics*, 33, 730–773. [30]

Jack, C. R., Therneau, T. M., Wiste, H. J., Weigand, S. D., Knopman, D. S., Lowe, V. J., Mielke, M. M., Vemuri, P., Roberts, R. O., Machulda, M. M., Senjem, M. L., Gunter, J. L., Rocca, W. A., and Petersen, R. C. (2016), "Transition Rates Between Amyloid and Neurodegeneration Biomarker States and to Dementia: A Population-Based, Longitudinal Cohort Study," *The Lancet Neurology*, 15, 56–64. [16,17,18,20,29]

Jackson, C. H. (2011), "Multi-state Models for Panel Data: The **msm** Package for **r**," *Journal of Statistical Software*, 38, 1–29. [22,23]

Jackson, C. H., Sharples, L. D., Thompson, S. G., Duffy, S. W., and Couto, E. (2003), "Multistate Markov Models for Disease Progression With Classification Error," *The Statistician*, 52, 193–209. [17]

Karlin, S., and Taylor, H. M. (1981), *A Second Course in Stochastic Processes*, New York, NY: Academic Press. [19]

Lange, J. M., and Minin, V. N. (2013), "Fitting and Interpreting Continuous-Time Latent Markov Models for Panel Data," *Statistics in Medicine*, 32, 4581–4595. [17]

Satten, G. A., and Longini, I. M. (1996), "Markov Chains With Measurement Error: Estimating the 'True' Course of a Marker of the Progression of Human Immunodeficiency Virus Disease," *Journal of the Royal Statistical Society*, Series C, 45, 275–309. [17,21]

Sharples, L. D., Jackson, C. H., Parameshwar, J., Wallwork, J., and Large, S. R. (2003), "Diagnostic Accuracy of Coronary Angiography and Risk Factors for Post-heart-transplant Cardiac Allograft Vasculopathy," *Transplantation*, 76, 679–682. [22]

Storlie, C. B., Michalak, S. E., Quinn, H. M., Dubois, A. J., Wender, S. A., and Dubois, D. H. (2013), "A Bayesian Reliability Analysis of Neutron-Induced Errors in High Performance Computing Hardware," *Journal of the American Statistical Association*, 108, 429–440. [30]

Therneau, T. M. (2015), "A Package for Survival Analysis in S," Version 2.38, available at *https://CRAN.R-project.org/package=survival*. [24]

Titman, A. C., and Sharples, L. D. (2008), "A General Goodness-of-Fit Test for Markov and Hidden Markov Models," *Statistics in Medicine*, 27, 2177–2195. [17]

Watanabe, S. (2010), "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory," *Journal of Machine Learning Research*, 11, 3571–3594. [30]

Xu, X., Chong, E., Hilal, S., Ikram, M. K., Venketasubramanian, N., and Chen, C. (2015), "Beyond Screening: Can the Mini-Mental State Examination be Used as an Exclusion Tool in a Memory Clinic?," *Diagnostics*, 5, 475–486. [18]

Yu, L., Griffith, W. S., Tyas, S. L., Snowdon, D. A., and Kryscio, R. J. (2010), "A Nonstationary Markov Transition Model for Computing the Relative Risk of Dementia Before Death," *Statistics in Medicine*, 29, 639–648. [18]

Zhao, T., Wang, Z., Cumberworth, A., Gsponer, J., de Freitas, N., and Bouchard-Côté, A. (2016), "Bayesian Analysis of Continuous Time Markov Chains With Application to Phylogenetic Modeling," *Bayesian Analysis*, 11, 1203–1237. [17]