Published in final edited form as:

J Am Stat Assoc. 2013 June 1; 108(502): 370–380. doi:10.1080/01621459.2013.770307.

Partially ordered mixed hidden Markov model for the disablement process of older adults

Edward H. Ip, Qiang Zhang, W. Jack Rejeski, Tamara B. Harris*, and Stephen Kritchevsky*

Abstract

At both the individual and societal levels, the health and economic burden of disability in older adults is enormous in developed countries, including the U.S. Recent studies have revealed that the disablement process in older adults often comprises episodic periods of impaired functioning and periods that are relatively free of disability, amid a secular and natural trend of decline in functioning. Rather than an irreversible, progressive event that is analogous to a chronic disease, disability is better conceptualized and mathematically modeled as states that do not necessarily follow a strict linear order of good-to-bad. Statistical tools, including Markov models, which allow bidirectional transition between states, and random effects models, which allow individual-specific rate of secular decline, are pertinent. In this paper, we propose a mixed effects, multivariate, hidden Markov model to handle partially ordered disability states. The model generalizes the continuation ratio model for ordinal data in the generalized linear model literature and provides a formal framework for testing the effects of risk factors and/or an intervention on the transitions between different disability states. Under a generalization of the proportional odds ratio assumption, the proposed model circumvents the problem of a potentially large number of parameters when the number of states and the number of covariates are substantial. We describe a maximum likelihood method for estimating the partially ordered, mixed effects model and show how the model can be applied to a longitudinal data set that consists of $N = 2,903$ older adults followed for 10 years in the Health Aging and Body Composition Study. We further statistically test the effects of various risk factors upon the probabilities of transition into various severe disability states. The result can be used to inform geriatric and public health science researchers who study the disablement process.

Keywords

Latent Markov model; continuation ratio model; EM algorithm; generalized linear model; Health ABC study

1 Introduction

Because of the rapid growth of the aging population in developed countries, the prevention of late-life disability is quickly becoming a major theme in public health policy. The personal, social, and economic costs of late-life disability are staggering, and within the U.S. alone, more than one-third of older adults are affected by disability (for a recent review, see Freedman et al. 2002). Until recently, disability was conceptualized much like other chronic diseases, as an irreversible, progressive event or trait of the person. Modeling disability as a trait meant that it could be measured on a continuum from low to high. Unfortunately, this approach ignores the fact that disability is a dynamic condition of the person — i.e., a bidirectional state rather than a trait — that is inherently multidimensional — e.g., Gill and

*for the Health ABC Study

Kurland (2003), Hardy and Gill (2004), Gill et al. (2006), and Rejeski et al. (2008b). Older adults, especially relatively healthy individuals, may experience abrupt and unpredictable episodes of disability and recover from these states (Hardy et al. 2005). For example, Gill et al. (2006) examined the prospective transition to physical frailty, a condition that is strongly related to disability and that was defined as meeting at least three of the following criteria: weight loss, exhaustion, muscle weakness, and slow walking speed. During an 18-month observational period, 43.3% of the sample transitioned to more severe states of frailty, whereas 23.0% became less frail. In a longitudinal study of older adults with knee pain, Rejeski et al. (2008b) showed that there were six states of disability. The states were characterized by multiple dimensions of physical functioning that were not linearly ordered. They found that patients often cycled through many functional states over time, exhibiting patterns of episodic disabled status. Using a stratified analysis, they also demonstrated that obesity was an important risk factor in acceleration down the “slippery slope” of disability. These studies provide support for the position that disability is neither immutable nor an invariant trait-like condition; rather, it is a dynamic process that is more appropriately conceptualized as a state. This conceptual perspective on disability has been succinctly summarized by Leveille et al. (2000, p. S41) in noting that: “disability status is highly variable across individuals, and persons may move back and forth between nondisability and disability, or from disability or nondisability to death.” Moreover, the classification or distinction between nondisability and disability is likely to be more subtle than coarse.

The shift from conceptualizing disability as a unidimensional trait-based concept to one that is multidimensional and state-based has important methodological implications. First, to operationalize disability from the latter perspective requires a portfolio of variables that describe functioning along the major domains of disability, including mobility, activities of daily living (ADL), and instrumental activities of daily living (IADL) (Katz 1983; Rejeski et al. 2008a). This necessitates multi-variable and multi-state statistical models. Most previous studies have relied on expert judgment to generate protocols for classifying individuals into multiple categories of disability — see, e.g., Hardy et al. (2005), Yang and George (2005), and Guilley et al. (2008). One limitation of this approach is the rather “coarse” quality of the resulting state classifications. For example, using this strategy, Hardy and colleagues created three categories: not disabled, mildly disabled, and severely disabled. In addition to being subjective, this type of model could obscure subtle features of, and changes in, the disablement process. Perhaps more importantly, with predetermined disability states, the validity of these states is rarely evaluated against data. Without data, it is a challenging task, even for gerontologists, to decide whether or not a specific profile of measurements would constitute a distinct, clinically meaningful state. In our view, a data-driven approach that statistically determines both the number of, and the profile of various disability states, is more theoretically sound; it could uncover otherwise overlooked but clinically meaningful transitions in disability.

Second, the role that various risk and protective factors play in modifying the dynamics of transition between various disability states poses significant modeling challenges. In general, for a model that contains S states, there are S^2 transitions. For example, Altman (2007) proposed a generalized mixed effects linear model for each individual transition within a hidden Markov model (HMM):

$$\text{logit } p(z_{it}=s|z_{it-1}=r) = x_i' \gamma_{rs} + w_i' \theta_i, \quad (1)$$

where z_{it} denotes the latent state for individual i at time t , γ_{rs} is a vector of regression coefficients for covariate x_i for transitioning from state r at time $t-1$ (state of origination) to state s at time t (state of destination), and w_i and θ_i are the random-effects design vector and random-effects vector, respectively. The number of parameters in the multinomial transition

model described by (1) increases rapidly with the number of covariates and states. For a modest 5-state model of disability with 15 predictors, the multinomial model will involve a total of $5 \times 4 \times 16 = 320$ parameters, including an intercept term. For many data sets collected in aging studies, the sample size will not be sufficiently large to support such complex models. A plausible solution is to resort to only examining stability; that is, examining only the diagonal elements of the transition matrix or only focusing on transitions to a specific state such as death. However, a more comprehensive and conceptually advanced approach that examines the overall transition pattern would be more informative.

A third methodological challenge is the need to account for individual differences in age-related decline to disability within a general population. Clearly, there is heterogeneity in the general downward trend in terms of their physical functioning. From a modeling perspective, this secular, long-term decline — likely due to the gradual loss of one's capacity to maintain homeostasis with age — is best captured by both random and fixed effects. The random effects provide a model of individual differences in the rate of decline, whereas the fixed effects capture the linear or nonlinear relationship in decline due to factors such as age and gender. Previous research has shown that the decline in physical functioning is more rapid in older adults than in younger subjects (Peres et al. 2005) and that women have a steeper rate of decline compared with men — see, e.g., Beckett et al. (1996) and Crimmins et al. (1997). In the current paper, we adopt mixed effects for modeling transition probabilities, implying a non-homogeneous transition model.

It is our view that the disablement process can be conceptualized as both a series of short-term transitions between various disability states and the secular long-term deterioration of health. To overcome the methodological challenges described above for operationalizing disability, we propose a data-driven statistical approach for (1) delineating the different states of disability, (2) conducting inference using a parsimonious partial-order model for describing the predictive relationship between risk factors and transition between disability states, and (3) accounting for the natural long-term decline in functioning within possible trajectories of short-term cycling back and forth between states. The basic model is a multivariate discrete HMM (e.g., MacDonald and Zucchini 1997; Vermunt et al. 1999; Bartolucci and Farcomeni 2009; Ip et al. 2010; Zhang et al. 2010).

As a solution to the inherently complex problem of disability state modeling, our model presentation is similar to that described in Scott et al. (2005, continuous outcomes, no covariate) and Altman (2007, single outcome, multinomial transition model). The methodological focus in this paper is on modeling transitions between hidden or latent states produced by the HMM, and for this purpose we developed a novel, partially ordered mixed model HMM (POM-HMM). The latent states are treated as partially ordered sets (posets) such that a parsimonious mixed model can be developed to directly estimate effects for risk factors and interventions and to test the significance of specific factors after adjusting for relevant covariates. The approach can also estimate the individual-specific rate of decline. We report the statistical analysis of a longitudinal data set from the Health Aging and Body Composition Study (Health ABC, Harris et al. 2006) for studying disability states and for examining the role of selected risk factors on the disablement process such as hypertension, obesity, and alcohol intake (Verbrugge and Jette 1994).

The remainder of the article proceeds as follows. Section 2 describes the motivating application for the study — the Health ABC data set. Section 3 presents the basic HMM and its various extensions, including the POM-HMM. Section 4 reports the results of the analysis, and Section 5 provides a discussion and concluding remarks.

2 The Health ABC Data Set

The Health Aging and Body Composition Study (HealthABC) is a cohort study focused on understanding the relationship between body composition, weight-related health conditions, and incident functional limitation. This rich data set now contains more than 10 years of follow-up since its inception in 1997/98, when baseline data were collected. The study sample consisted of 3,075 well-functioning black and white men and women aged 68 to 80 years. Potential participants were identified from a random sample of white Medicare beneficiaries and all age-eligible, community-dwelling black residents in designated zip code areas surrounding the Pittsburgh, PA, and Memphis, TN field centers. Entry criteria were: no self-reported difficulty walking 1/4 mile or walking up 10 steps without resting, and no difficulty with any of the three ADLs (transferring, bathing, and dressing). Exclusion criteria were: active treatment for cancer, plans to move out of the study area, problems in communication with study staff, or participation in a life-style intervention study. Consistent with its focus on disability, Health ABC's primary outcome is persistent lower-extremity limitation as assessed by self-reported difficulty in walking 1/4 mile or climbing 10 steps on two consecutive semi-annual questionnaires. In this study, we supplement the primary outcomes with several other self-reported measures, all of which were measured annually. Only annually collected data were used in the HMM analysis. Participants were asked, "Because of a health or physical problem, do you have any difficulty ...?" Besides the walking 1/4 mile or climbing 10 steps items, observed outcomes included binary responses regarding getting in and out of bed, bathing, dressing, and lifting a weight of 10 pounds. Baseline and nine yearly follow-ups are included in the current study, with participants with only baseline data ($n = 134$) and participants never observed consecutively two years in a row ($n = 34$) being excluded. For example, a participant with observations at time points 1,2,3, and 5 would be included, but one with observations at times 1,3,5, and 7 would not. This left 2,903 participants in the analysis. Of note, loss-to-follow-up participants could be present in the final sample as long as they had consecutive measurements. Attrition due to loss to follow-up, which totaled $n=250$, averaged at 4.2% (range 2.5%, 7.0%) across time points. Among the 2,903 subjects, a significant portion ($n = 801$, or 27.6%) became deceased within the study period. Instead of treating missed outcomes due to death as missing completely at random (MCAR) or missing at random (MAR), we include Death as an observed absorbing state, as one of the mix of latent states identified through HMM. Figure 1 shows both the percentage of deceased adults and the percentage of older adults reporting difficulty in performing the tasks over time. The item "lifting 10 lb." was inadvertently left out of the HealthABC survey at time 7, so this specific outcome was treated as missing.

Other variables that were used for studying the disablement process included anthropometric measurement of body mass index (BMI), mental health measurement (Center of Epidemiologic Study — Depression Scale CESD), co-morbidity (previous incidence of cardiovascular disease, diabetes, and hypertension), lifestyle variables (smoking, alcohol consumption), and demographics (gender, race, education, marital status, and baseline age). Table 1 shows the characteristics of the data set. The BMI range for this sample is 14.6-52.0.

3 The Partially Ordered Mixed Hidden Markov Model

3.1 The Basic Multivariate Discrete Hidden Markov Model

Let y_{ijt} denote the discrete response of subject i at occasion t on outcome j , $i = 1, \dots, N$; $j = 1, \dots, J$; $t = 1, \dots, T$. Here, an outcome could be a response to an item such as a survey question, or a physiological measure, such as whether systolic blood pressure ≥ 140 . Hereinafter, the terms response and outcome are used interchangeably. The variable y_{ijt}

takes on discrete value $k = 1, \dots, K_j$. Specifically, $y_{it} = (y_{it1}, \dots, y_{itJ})'$ denotes the response vector of subject i at occasion t , and $y_i = (y'_{i1}, \dots, y'_{iT})'$ represents the complete response pattern of subject i . The categorical (unobserved) latent state of subject i at occasion t is denoted by $z_{it} = 1, \dots, S$. Thus, $z_i = (z_{i1}, \dots, z_{iT})'$ is the state sequence of subject i through the latent space over time. A first-order Markov chain for the latent variable assumes that the latent state at occasion $t + 1$ depends only upon the latent state at occasion t —i.e., for any i , $p(z_{it} | z_{i1}, \dots, z_{iT}) = p(z_{it} | z_{it-1})$. Accordingly, the basic HMM contains the following three sets of parameters:

1. $\tau = (\tau_{rs})$, an $S \times S$ matrix of time-homogeneous transition probabilities between latent states, where for any i and t , $\tau_{rs} = p(z_{it} = s | z_{it-1} = r)$, $t = 2, \dots, T$. In the extended HMM for poset, the transition probabilities would be subject-specific and time-heterogeneous, and we use $\tau_{it, z_{it} z_{it-1}}$ to denote the transition probability from state z_{it-1} at time $(t-1)$ to state z_{it} at time t for individual i ; i.e.,

$$\tau_{itrs} = p(z_{it} = s | z_{it-1} = r), \quad (2)$$

for given i and t .

2. $\alpha_1 = (\alpha_{11}, \dots, \alpha_{1S})$, an $S \times 1$ vector of marginal probabilities of the states at occasion 1, $\alpha_{1s} = p(z_{i1} = s)$. The vector $\alpha_t = (\alpha_{t1}, \dots, \alpha_{tS})$, $t = 2, \dots, T$, can be recursively derived by the formula $\alpha'_t = \alpha'_{t-1} \tau$.
3. $\pi = (\pi_{jsk})$, an $J \times S \times K_j$ array of state-conditional response probabilities, where k is the index of outcome category, $k = 1, \dots, K_j$ and for any i , t , $\pi_{jsk} = p(y_{itj} = k | z_{it} = s)$, where $\sum_{k=1}^{K_j} \pi_{jsk} = 1$.

The notation for POM-HMM will involve a more complex set of indexes. To improve readability, we do not boldface vectors and use the shorthand form to represent vectorized entries, e.g., $z_i = \text{vec}(z_{it})$.

Assuming conditional independence between responses given the latent state, the marginal probability of a response pattern y_i is given by

$$p(y_i) = \sum_{\mathcal{Z}} p(y_i | z_i) p(z_i) = \sum_{\mathcal{Z}} p(z_i) \prod_{t=1}^T p(y_{it} | z_{it}), \quad (3)$$

with the summation over the entire latent state space $\mathcal{Z}$ of size S^T ,

$$p(z_i) = \alpha_1 z_{i1} \prod_{t=2}^T \tau_{z_{it-1} z_{it}}, \quad (4)$$

and

$$p(y_{it} | z_{it} = s) = \prod_{j=1}^J \prod_{k=1}^{K_j} \pi_{jsk}^{\delta(y_{itj}, k)}, \quad (5)$$

where $\delta(y_{itj}, k) = 1$ if $y_{itj} = k$, and 0 otherwise.

An observed State of Death is included into the set of $\{z_{it}\}$ and was designated, conveniently, as the last state S , in which the entire probability mass was assigned to the conditional probability for the “worst” response category (assumed to be K_j here); i.e.,

$$p(y_{itj}=K_j|z_{it}=S)=1. \quad (6)$$

The Death State is absorbing, and we have:

$$\tau_{ss}=0 \quad \text{for } s=1, \dots, S-1, \quad \text{and} \quad \tau_{SS}=1. \quad (7)$$

When estimating HMM, the latent states of the HMM of deceased subjects were assigned the value S from the year of death to the last year of the study period; otherwise data prior to death were treated no differently from other subjects.

The overall likelihood of the HMM is given by

$$p(y) = \prod_{i=1}^N \sum_{z_i \in \mathcal{Z}} p(z_i) \prod_{t=1}^T p(y_{it}|z_{it}). \quad (8)$$

3.2 Partially Ordered Mixed Hidden Markov Model (POM-HMM)

States of physical functioning as identified by HMM often do not necessarily follow a rank order. Figure 2 shows one proposed dominance relationship of the latent states in the Health ABC data set (to be discussed in Section 4). For binary outcomes, dominance here is coordinate-wise determined — i.e., State 1 (S_1) dominates State 2 (S_2) if the conditional probabilities for positive response (assumed to be category 1 here) of each individual variable in State 1 is higher than the corresponding conditional probability in State 2 (i.e., for all j , $p_{js1} > p_{js2}$), and we write $S_1 \geq S_2$, or $S_1 > S_2$ if dominance is strict. For outcome with more than 2 categories, one could resort to using a weighted sum of the conditional probabilities for determining the dominance structure. An alternative option is to use expert judgment in determining whether one state is “better” or “worse” than another state. This is especially useful for nominal outcomes. As a general principle, the statistical procedure for POM-HMM requires a fixed, user-defined poset structure, regardless of whether the choice relies on numerical computation.

In Section 1, we point out that treating latent states as nominal variables without rank ordering could result in unnecessarily complex transition models in which every transition requires its own set of parameters for all possible states of origination and destination. To circumvent this problem, one approach is to take full advantage of the partially ordered set (POS) structure of the latent states.

As an example, Figure 2 shows a bounded POS in which one state — the healthiest state (labeled 1) — dominates all the other states, and another state — the most disabled state (labeled 9) — is dominated by the remaining states. The states in between the two extreme states may or may not dominate one another. The partially ordered mixed HMM (POM-HMM) is designed for data that exhibit the kind of structure shown in Fig. 2, and it extends the continuation ratio model (Agresti 2002) for ordinal responses to partially ordered responses.

First recall the continuation ratio logit model for ordinal responses in variable Y with common fixed effects coefficients and random coefficients across categories:

$$\log \left(\frac{p(Y_i=k|x_i, w_i)}{p(Y_i>k|x_i, w_i)} \right) = \phi_k + x_i' \gamma + w_i' \theta_i, \quad (9)$$

where ϕ_k is the category-specific parameter, $k = 1, \dots, K$, x_i and w_i are vectors of covariates for individual i , and $\epsilon_i \sim N(0, \sigma^2)$. This specification is equivalent to a sequential model (Fahrmeir and Tutz (2001), Section 3.3.4) of the form

$$p(Y_i = k | Y_i \geq k, x_i, w_i) = h(-\phi_k - x_i' \gamma - w_i' \theta_i), \quad (10)$$

where $h(\cdot)$ is the logistic distribution function. Tutz (1990) argued that the sequential model can be considered as a collection of simple process models in which an observed response $Y_j = k$ implies successive conditional transitions from a lower category to the next higher category until at the k -th category. The sequential model thus is more appropriate for discrete responses that have a substantive ordering, as opposed to a coarse grouping of a finer scale — e.g., trichotomization of an original continuous response; see also McCullagh (1980).

The POM-HMM extends the sequential model by replacing the $\geq$ operator in (10) with the dominance operator $\succeq$, which is reflexive, non-symmetric, and transitive. One apparent difference between the two operators is that $p(Y_i \geq k) + p(Y_i < k)$ is not necessarily equal to 1. In Figure 2, for example, $P(s \geq 3) = P(\{1, 3\})$, $P(s < 3) = P(\{8, 9, D\})$, and clearly the two probabilities do not sum to one. Hence, the calculus of cumulative odds ratio for ordinal data can not be immediately applied to the partially ordered data.

To put the partially ordered mixed model into the current context, the POM-HMM specifies the ratio of the conditional log odds in transitioning from a state r at time $(t-1)$ to a state s at time t versus to a better state t as a function of fixed effects with common regression coefficients and random effects with common prior distributions across states of origination and destination:

$$\log \left(\frac{p(z_{it} > s | z_{it-1} = r, x_{it}, w_{it})}{p(z_{it} = s | z_{it-1} = r, x_{it}, w_{it})} \right) = \phi_{rs} + x_{it}' \gamma + w_{it}' \theta_i, \quad (11)$$

where $\epsilon_i \sim N(0, \sigma^2)$. The parameterization in (11) is intended to facilitate interpretation of transitioning to a better state and as such is slightly different from the usual parameterization in continuation-ratio modeling. Both the vectors of covariates x_{it} and w_{it} can be time varying. In our subsequent derivation, we focus on the case in which ϵ_i is a scalar and $\epsilon_i \sim N(0, \sigma^2)$. Here, we also assume that there exists a unique maximal state that dominates all other states so that the set $\{z_{it} > s\}$ would always be non-empty for s other than the maximal and hence there would be exactly $S-1$ equations from (11) for the model to be identifiable.

Compared with a multinomial logit transition model of type (1), the specification in (11) only requires $S(S-1)$ transition-specific intercept terms and a set of common regression coefficients and variance components instead of a distinct set of regression parameters for each transition. The interpretation of the regression parameters is similar to the proportional odds interpretation in cumulative logit models (Agresti 2002). For example, if x is an indicator of treatment, then the associated exponentiated coefficient represents the ratio of the conditional odds between the treatment and the control group in transitioning to a better state, given that the remaining variables in the model are held constant. We shall continue to call this assumption the proportional odds assumption (for POM-HMM). Conditional odds here are defined as $p(z_{it} > s | z_{it-1} = r, x_{it}, w_{it}) / p(z_{it} = s | z_{it-1} = r, x_{it}, w_{it})$.

To formalize the partial order, a dominance matrix is first created to represent the dominance relationship amongst the states (Meulders et al. 2005). The dominance matrix $A = (a_{ij})$ for a bounded POS of order S is defined by $a_{ij} = 1$ if state $i \succeq j$ and 0 otherwise. For example, the dominance matrix for the POS in Figure 2 is given by:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

To directly model the dynamics in transitioning, we use the subject-specific and time-heterogeneous transition probabilities τ_{itrs} defined in (2). The cumulative probability of states dominating a specific state s can be obtained by left multiplying the transition probabilities by the dominance matrix. Equation (11) can be expressed, in a shorthand vector form, as:

$$\log \left(\frac{1 - \widehat{A} \tau_{itr}^-}{\tau_{itr}^-} \right) = \phi_r + X_{it} \gamma + W_{it} \theta_i, \quad (12)$$

where $\widehat{A} \in \mathbb{R}^{(S-1) \times (S-1)}$ and $\widehat{A} = (A^T - \mathbf{1}_{2:S, 2:S})$. Here, the subscripts, $(2 : S, 2 : S)$, represent taking only rows from the second to the last, and columns from the second to the last, $\mathbf{1}$ is a constant matrix of ones, $\mathbf{1}$ is a constant vector of ones, and $\tau_{itr}^- \in \mathbb{R}^{(S-1) \times 1}$ is the transition probability vector except without the first element — i.e., $\tau_{itr}^- = (\tau_{itr2}, \dots, \tau_{itrS})'$. On the right-hand side, vector $\phi_r \in \mathbb{R}^{(S-1) \times 1}$ is the intercept term, vector $\gamma \in \mathbb{R}^{p \times 1}$ and θ_i are assumed to be constant across different r and s , and p represents the number of covariates. To illustrate this poset model, we use the POS depicted in Fig. 2. The corresponding model (12) contains a set of nine regression equations, which are explicitly written out in an appendix.

In the subsequent estimation procedure, which uses the EM algorithm, the E-step requires the “unwinding” of the probabilities in (12). In other words, we need to compute τ_{itrs} given the poset cumulative probabilities. To obtain the individual transition probabilities from the model, we first denote the linear predictor $\exp(\phi_r + X_{it} \gamma + W_{it} \theta_i)$ by η_{itr} and then we rewrite (12), again in shorthand form, as:

$$\eta_{itr} = \frac{1 - \widehat{A} \tau_{itr}^-}{\tau_{itr}^-}, \quad (13)$$

where division is coordinate-wise. As a result, τ_{itr}^- can be solved using the following equation:

$$\tau_{itr}^- = [\widehat{A} + \text{diag}(\eta_{itr})]^{-1} \mathbf{1}. \quad (14)$$

Note that even though $\widehat{A}$ can be rank deficient and thus non-invertible, τ_{itr} is always positive, and $\widehat{A} + \text{diag}(\eta_{itr})$ is always invertible. Therefore, the model is identifiable. The following lemma ensures that (14) leads to proper probability values in $(0, 1)$. The proof is provided in an appendix.

Lemma 1. Given a finite poset structure with a unique maximal and its dominance matrix $A \in \mathbb{Z}_2^{K \times K}$, the vector $\tau \in \mathbb{R}^{K \times 1}$ computed through (14) on a positive vector $\eta \in \mathbb{R}_+^{K \times 1}$ has entries $\tau_i \in (0, 1)$, $i = 1, \dots, K$.

The matrix formulation described above allows the representation of the “winding” and “unwinding” of transition probabilities, respectively, as processes involving an integral operator ($\hat{A}$) that computes the cumulative probabilities, and a differential operator (inverse of matrices derived from A and the linear predictor) that computes the individual probabilities.

Besides the sequential interpretation of the continuation ratio model, which is more appropriate for the POS structure, there are also technical reasons that we prefer using the continuation ratio model instead of the cumulative model. Similar to (11), the cumulative ratio model can be expressed as

$$\log \left(\frac{p(z_{it} > s | z_{it-1}=r, x_i, w_i)}{p(z_{it} \leq s | z_{it-1}=r, x_i, w_i)} \right) = \phi_{rs} + x'_{it} \gamma + w'_{it} \theta_i, \quad (15)$$

or in vector form,

$$\log \left(\frac{1 - \hat{A} \tau_{itr}^-}{\hat{A} \tau_{itr}^-} \right) = \phi_r + X_{it} \gamma + W_{it} \theta_i. \quad (16)$$

However, this model might not be identifiable when unwinding for the individual transition probabilities:

$$\tau_{itr}^- = \hat{A}^{-1} \frac{1}{1 + \eta_{itr}}, \quad (17)$$

where in the completely ordered case, $\hat{A}$ would be an upper diagonal matrix with all entries having value 1, and with $\hat{A}$ thus being invertible. In the poset case shown above, $\hat{A}$ is rank deficient, and thus the model is unidentifiable. Another possible model can be expressed as

$$\log \left(\frac{p(z_{it} > s | z_{it-1}=r, x_i, w_i)}{p(z_{it} \leq s | z_{it-1}=r, x_i, w_i)} \right) = \phi_{rs} + x'_{it} \gamma + w'_{it} \theta_i, \quad (18)$$

or

$$\log \left(\frac{1 - \hat{A} \tau_{itr}^-}{\hat{A}^- \tau_{itr}^-} \right) = \phi_r + X_{it} \gamma + W_{it} \theta_i, \quad (19)$$

where $\hat{A}^- = A_{2:S,2:S}$. Note that the left-hand side is no longer a logit function because the numerator and the denominator do not sum to one. The disadvantage of this model is the possible negative probabilities in the unwinding step, where

$$\tau_{itr}^- = \left[\hat{A} + \text{diag}(\eta_{itr}) \hat{A}^- \right]^{-1} 1. \quad (20)$$

3.3 Maximum Likelihood Estimation of POM-HMM

The POM-HMM is estimated by the maximum likelihood method. The EM algorithm (Dempster et al. 1977) is particularly useful to handle the presence of latent states and

random effects within the POM-HMM. For historical reasons, a version of the EM algorithm used to estimate the basic HMM is also known as the Baum-Welch algorithm (Baum and Petrie 1966). Here, we extend the Baum-Welch algorithm for inference for the mixed effects model for transition probabilities in (12). Briefly, in the EM terminology, the latent state and random effects are treated as “missing data.” The objective here is to maximize the conditional expected log-likelihood of the complete data, which includes the observed data y and the latent states z , given y and some provisional estimates for the parameters, where y and z , respectively, denote the vectors of the collection of data (y_i) and (z_i).

Denote the entire parameter set defined in (3)-(5) and (12) by β and its estimate at the n^{th} iteration of the EM algorithm by $\beta^{(n)}$. Furthermore, let

$$\beta^{(n+1)} = \underset{\beta}{\operatorname{argmax}} Q(\beta, \beta^{(n)}),$$

where the conditional expected log-likelihood of the complete data $Q(\beta, \beta^{(n)})$ is given by :

$$Q(\beta, \beta^{(n)}) = E \left[\log p(y, z, \theta | \beta) | y, \beta^{(n)} \right]. \quad (21)$$

Here $\theta = (\theta_1, \dots, \theta_N)$ is the random effect. Assuming that θ and z are independent, we have

$$p(z_i, \theta_i | y_i, \beta^{(n)}) = p(z_i | y_i, \beta^{(n)}) f_{\beta^{(n)}}(\theta_i | y_i),$$

where $f^{(n)}(z_i | y_i)$ is the posterior density of z_i given y_i . Thus, the objective function is

$$Q(\beta, \beta^{(n)}) = \sum_i \sum_{\mathcal{Z}} \int p(z_i | y_i, \beta^{(n)}) \log p(y_i, z_i, \theta_i | \beta) f_{\beta^{(n)}}(\theta_i | y_i) d\theta_i, \quad (22)$$

in which the joint distribution of the complete data of subject i within (22) is given by:

$$p(y_i, z_i, \theta_i | \beta) = \alpha_{1z_{i1}} \left(\prod_{t=2}^T \tau_{z_{it-1}z_{it}} \right) \prod_{t=1}^T \prod_{j=1}^J \prod_{k=1}^{K_j} \pi_{jsk}^{\delta(y_{itj}, k)} f_{\beta}(\theta_i), \quad (23)$$

where $\tau_{z_{it-1}z_{it}}$ is given by the set of equations in (12)-(14).

By taking the logarithm on (23) and substituting (22), we rewrite the objective function as

$$\begin{aligned} Q(\beta, \beta^{(n)}) &= \sum_i \sum_{\mathcal{Z}} p(z_i | y_i, \beta^{(n)}) \log \alpha_{1z_{i1}} \\ &+ \sum_i \sum_{\mathcal{Z}} p(z_i | y_i, \beta^{(n)}) \int \sum_{t=2}^T \log h(\phi_{z_{it}z_{it-1}} + x'_{it}\gamma + w'_i\theta_i) f_{\beta^{(n)}}(\theta_i | y_i) d\theta_i \\ &+ \sum_i \sum_{\mathcal{Z}} p(z_i | y_i, \beta^{(n)}) \sum_t \sum_j \sum_k \delta(y_{itj}, k) \pi_{jsk} \\ &+ \sum_i \sum_{\mathcal{Z}} p(z_i | y_i, \beta^{(n)}) \int \log [f_{\beta}(\theta_i)] f_{\beta^{(n)}}(\theta_i | y_i) d\theta_i, \end{aligned} \quad (24)$$

where h is the inverse link function specified in (11). One appealing feature about (24) is the separability of the three sets of parameters, namely α_{1s} , π_{jsk} , and (ϕ, σ^2) , which can be optimized individually.

Before we look into the second and fourth terms of the RHS of (24), we see that the first and third terms do not involve the random effects and can be computed separately. Specifically, both terms can be written in a multinomial log-likelihood of the form :

$$\sum_i \sum_I \mathcal{Y}_{iI} \log(\psi_{iI}), \quad (25)$$

where I is a vector of summation indices; $\mathcal{Y}_{iI}$ are the “pseudo-counts,” which may take non-integer values, and ψ_{iI} are functions of model parameters. In the first term, $I = \{s/s = 1, \dots, S\}$, $\mathcal{Y}_{iI} = \tilde{\alpha}_{i1s}^{(n)} = p(z_{i1} = s | y_i, \beta^{(n)})$, and $\psi_{iI} = \alpha_{1s}$ whereas in the third term, $I = \{(j, s, k) | j = 1, \dots, J, s = 1, \dots, S, k = 1, \dots, K_j\}$, $\mathcal{Y}_{iI} = \sum_t \tilde{\alpha}_{its}^{(n)} \delta(y_{itj}, k)$, and $\psi_{iI} = \pi_{jsk}$. The updating equation for maximum likelihood estimates (MLEs) of α_{1s} and π_{jsk} are the solutions of the re-estimation formulas of the basic HMM (Rabiner 1990):

$$\begin{aligned} \alpha_{1s} &= \frac{\sum_{i=1}^N p(z_{i1} = s | y_i, \beta^{(n)})}{N} = \frac{\sum_{i=1}^N \tilde{\alpha}_{i1s}^{(n)}}{N}, \\ \pi_{jsk} &= \frac{\sum_{i=1}^N \sum_{t=1}^T p(y_{itj} = k, z_{it} = s | y_i, \beta^{(n)})}{\sum_{i=1}^N \sum_{t=1}^T p(z_{it} = s | y_i, \beta^{(n)})} = \frac{\sum_{i=1}^N \sum_{t=1}^T \tilde{\alpha}_{its}^{(n)} \delta(y_{itj}, k)}{\sum_{i=1}^N \sum_{t=1}^T \tilde{\alpha}_{its}^{(n)}}. \end{aligned} \quad (26)$$

The above description for the maximization of the first and third terms highlights the E- and the M-steps of the EM algorithm: the E-step computes the posterior probabilities and pseudo-counts, whereas the M-step updates the estimates using pseudo-counts via closed-form equations (26). Details of EM without mixed effects are provided in a recent article (Zhang et al. 2010); here we expand the estimation procedure to include the poset mixed effects model.

For the second and fourth terms of the RHS of (24), which contains the fixed- and random-effects parameters in the linear poset model, the E-step is similar. However, the M-step does not have a closed-form solution and requires an iterative method. The estimation of fixed effects will be described further below. For the random effects, first we use Gauss-Hermite quadrature for integration. Denote the vector of q quadrature points by $v = (v_p)$. The

corresponding collection of weights $w^{(n)} = (w_i^{(n)})$, which depend upon the current estimate of the variance structure, $\Sigma^{(n)}$, are used to approximate $f^{(n)}(\gamma)$. At a quadrature point v_p , denote the transition probability by $\pi_{itrs|l} = [\pi_{itrs}(v_p)]_{j=v_p}$. The integration notwithstanding, computationally the conditional expected log-likelihood still retains the multinomial form. In other words, the computation of random effects can be conducted using codes for fixed effects HMM by adding a dimension for discretized γ . When there are more than one random effects and they are independent, this method applies with straightforward adaptation. When the random effects are not independent, the Gauss-Hermite quadrature can still be used, but the computational burden would increase. In this application, γ_i is a scalar. We reparameterize γ_i by $\theta_i = \sigma \tilde{\theta}_i$, and thus $\tilde{\theta}_i \sim N(0, 1)$. The quadrature points and weights can both be taken from the standard normal distribution and remain constant throughout the iterations, while σ is treated as one of the covariate coefficients, β .

The M-step for updating the poset parameters in (11) — the second term in the RHS of (24) — involves an iterative procedure within each EM cycle: denote the parameters of the poset

model (ϕ, γ, θ) by ξ . Maximize the following function for the estimate of ξ , given $\xi^{(n)}$ at the n^{th} iteration:

$$\xi^{(n+1)} = \underset{\xi}{\operatorname{argmax}} \sum_i \sum_l \mathcal{Y}_{il}(\xi^{(n)}) \log(\psi_{il}(\xi)), \quad (27)$$

where $\mathcal{Y}_{il}$ is the pseudo-counts vector $\tilde{\tau}_{itrl}^{(n)} = p(z_{it}=s, z_{it-1}=r | y_i, \xi^{(n)}, w_i^{(n)})$ and $\psi_{il}(\xi) = h(\phi_{rs} + x'_{it}\gamma + w'_{it}\theta_i)$. Updating parameter values in (27) follows the Fisher scoring iteration (Fahrmeir and Tutz 2001, p. 42):

$$\xi^{(n+1)} = \xi^{(n*)} + \sum_{i,t,r,l} (X'_{it} D_{itrl} V_{itrl}^{-1} D'_{itrl} X_{it})^{-1} X_{it} D_{itrl} V_{itrl}^{-1} (\tilde{\tau}_{itrl}^{(n)} - \tau_{itrl}), \quad (28)$$

where X_{it} is the design matrix for subject i at time t , D_{itrl} is the derivative matrix of the inverse link function; i.e.,

$$D_{itrl} = \partial h(\eta_{itrl}) / \partial \eta_{itrl} = - [\widehat{A} + \operatorname{diag}(\eta_{itrl})]^{-1} \operatorname{diag}(\eta_{itrl} \cdot \tau_{itrl}), \quad (29)$$

where the linear predictor vector $\eta_{itrl} = \exp(\phi_{rs}^{(n*)} + x'_{it}\gamma^{(n*)} + w'_{it}\theta_i^{(n*)})$, $\tau_{itrl} = [\widehat{A} + \operatorname{diag}(\eta_{itrl})]^{-1} 1$, and the operation $\cdot$ indicates the coordinate-wise product of two vectors. The variance function V_{itrl} is given by,

$$V_{itrl} = \operatorname{diag}(\tau_{itrl}) - \tau_{itrl} \tau_{itrl}'. \quad (30)$$

Note that n^* is the inner Fisher scoring iteration index and n is the outer EM loop index.

Throughout the inner Fisher scoring iteration, the pseudo-counts $\tilde{\tau}_{itrl}^{(n)}$ remain unchanged, while D_{itrl} , V_{itrl} , and τ_{itrl} are updated after each iteration.

We also note that in the E-step, the posterior probabilities, or the pseudo-counts, $\tilde{\tau}_{itrl}^{(n)}$, can be computed by first computing $p(z_{it}=s | z_{it-1}=r, y_i, \xi^{(n)}, \theta_i) = h(\phi_{rs}^{(n)} + x'_{it}\gamma^{(n)} + w'_{it}\theta_i^{(n)})$, and then using a junction tree algorithm (Lauritzen and Spiegelhalter 1988), which is a generalized Baum-Welch algorithm for general Bayesian network (in which the HMM is a special case; see Smyth 1997). The junction tree is especially efficient for computing the joint distribution when the dimension of the HMM is high (Zhang et al. 2010).

3.4 Computation of Standard Errors

The exact standard errors can be computed through an augmented junction tree algorithm, which is described together with simulation studies elsewhere (Blinded 2010, provided in Supplementary Materials). Briefly, the likelihood for the POM-HMM can be viewed as a Bayesian network with both observable and hidden nodes. The augmented junction tree algorithm builds upon the existing junction tree algorithm by augmenting the message-passing of the likelihood function with both the first and the second derivatives. By exploiting the highly efficient junction tree algorithm for computing the joint and marginal distribution of a Bayesian network, the augmented junction tree computes the exact information matrix and the exact standard error for the parameter estimates.

4 Data Analysis and Results

We used a two-stage procedure for the analysis of the Health ABC data set. In stage 1, we fit plain HMMs and selected the best-fitting model of $\mathcal{S}$ -states. The poset relationship between

the S states was then determined. Once the states and poset structure were both determined, we layered the transition model on top of the S -state model and jointly fitted the transition model and the state conditional probabilities. Both S and the poset structure were kept fixed in the joint estimation.

4.1 Stage 1 analysis

The Bayesian Information Criterion (BIC), one of the most commonly used goodness-of-fit indexes, was used for model selection in stage 1. While the BIC may not be an ideal choice (Scott 2002), it has the advantage of being easy to compute for maximum likelihood procedures. The appropriate number of hidden states was determined by selecting the model with the lowest BIC value.

For participants who missed on disability variables, we treated missing values as MAR and used the EM algorithm for fitting the model. In order to assess the impact of missing values on the HMM-derived states, a multiple imputation procedure was used. The SAS program PROC MI, which implemented the Markov Chain Monte Carlo (MCMC)-based method (Schafer 1997), was used to generate 5 sets of imputed values for the missing data. Specifically, the procedure implemented in our analysis estimated the multivariate covariance matrix using a non-informative prior and directly drew a sample from the estimated matrix with the MCMC procedure. We only used a single MCMC chain, and the convergence of the chain was assessed by inspection of the time-series plots of parameters of interest. The effects of missing values on state conditional probabilities were subsequently assessed through visualization. Furthermore, to solve a possible label-switching problem across imputation — e.g., State 1 in imputation data set 1 does not correspond to State 1 in imputation data set 2 — we fixed the model estimated from the original data set as an anchor model, and for each imputed data set we determined a state label by matching the state to the closest state in the anchor model. The label-switching regime is a complex issue (Stephens 2000). Admittedly, the rather ad hoc procedure used here may not be optimal. However, for the current application, in which the proportion of missing value was relatively small, the proposed anchoring procedure was found to perform satisfactorily.

4.2 Stage 2 analysis

Subsequent to the stage 1 procedure, POM-HMMs were fitted under a joint model for both the transitional and the HMM components. Both fixed-effects-only and mixed effects models were fitted for the transitional component. The fixed effects predictor included in the POM-HMM were listed in Table 1, among which only BMI, CESD, and time (t), are time-varying, whereas the others are baseline measurements. The categories within the variables education and hypertension were coded as equal intervalled (0,1,2) and subsequently treated as continuous. In the transition model, we included the quadratic term for BMI. The rationale for its inclusion is that recent studies suggested that mortality and other health indicators often exhibit a U-shaped relationship with BMI (e.g., Thinggaard et al. 2010). On the other hand, the random effects model includes a random slope for time t for capturing the individual heterogeneity in age-related decline of functioning. For the set of covariates, missing values for non-time-varying variables were imputed using a single imputation procedure. The imputation model used all variables present in the set of predictors. On the other hand, last-value-carry-forward was used for time-varying variables. For example, as part of the study protocol, the time-varying covariate CESD was only measured during alternate years. Consistent with prior HABC analysis procedures, last-value-carry-forward was used on CESD for handling such missing-by-design values. Unlike the Stage 1 analysis, in Stage 2 we assumed MAR for missing outcome variables and did not use multiple imputation.

4.3 Results

Applying BIC to the competing models, the 11-state model appears to have the lowest BIC (Fig. 3). However, in this model, the two “best” states appear to have very similar state conditional probabilities. Moreover, one of the two states has only a 0.01 probability of staying in that state over time, and a 0.48 probability of transitioning into the other one. As a result, it was decided that these two states be merged into a single state to form a more parsimonious and interpretable 10-state model. While the manual procedure created a model that has slightly inferior BIC, the difference in BIC value seems small enough to trade-off the statistically stronger 11-state model for a more stable and clinically interpretable 10-state model. Figure 4 shows the conditional probabilities of nine states, excluding the Death State, sorted more or less according to their partial order as graphically shown in Fig. 2. The effect of partially missing values of the indicator measures of the HMM is visualized in Fig. 4 by two vertical bookends at the end of each horizontal bar that indicate the minimum and maximum probabilities computed among the five imputed data sets. The interpretation of the 10-state model does not appear to be materially changed by the missing values. In addition, the standard errors for the conditional probabilities are generally small (mean 0.006, min/max 0.0003/0.036) and are not depicted in the figure.

State 1 is the nondisabled state; participants within this state generally show no difficulty with any of the six different tasks. State 2 to 5 all have some difficulty of one form or another, and States 6 to 8 show difficulties in two or more tasks. Generally, States 2 to 5 can be characterized as having subclinical conditions that involve one task, or mild disability, while States 6 to 8 describe participants who are severely disabled. In particular, the worst state — State 9 — shows difficulties in all six tasks, including the basic ADLs, dressing and bathing oneself. This delineation of disability states is consistent with the literature, in which older adults are classified as having no disability, mild disability, or severe disability, and death (Hardy et al. 2005). Beyond this general classification, the identified states reveal subtle subclinical conditions not captured when disability is conceptualized as a trait. For example, State 2, which is driven by problems with getting in and out bed, suggests some musculoskeletal decline that precedes the loss of mobility. State 3 suggests dysfunction in either the joints and/or strength of the arms and shoulder joints; State 4 suggests horizontal mobility problems (walking 1/4 mile), and State 5 indicates problems with moving the center of mass vertically, possibly due to a decline in muscle strength/power (walking up 10 steps).

Table 2 provides the prior and transition probabilities of the 10 states. The diagonal of the transition probability matrix, reveals that all the states are relatively stable — i.e., with the highest probability to transit into itself — and that the transitions are not unidirectional. Older adults do not always transit into a worse state even as they age. For example, there exists a 12% probability of transitioning from State 3 to 1 and a 13% probability of transitioning from State 5 to 1. Interestingly, the probability of transitioning into the Death State increases as one moves across the poset from State 1 to State 9.

The unrestricted transition matrix is also telling when we consider the dynamic for the subclinical states of disability. State 5 (vertical mobility problem) has a good chance (13%) of returning to normal (State 1), while State 4 (horizontal mobility problem) tends to stay within the same state and only has a 3% chance of returning to normal. Those with upper-body issues (State 3) also tend to have a substantial chance of returning to normal (12%).

Fixed and random effects transition models were included in stage 2 analysis, as reported above. Table 3 compares the result of a fixed-effects-only model to the mixed effects model of the POM-HMM. The mixed effects model includes time as a random effect, implying an individual-specific rate of decline. The inclusion of random effects does not appear to have a

large impact on the estimates of the fixed effects. The variance of the random effect has a p-value of 0.347, while the likelihood ratio test has a p-value of 0.267, providing very little evidence that the addition of random effects marginally improves model fit. This could be due to the fact that the set of fixed covariates effectively explained away a significant proportion of the variance ¹.

For both fixed and mixed effects POM-HMM models, BMI shows some interesting trends as both the linear and quadratic effects are significant. When the entire sample was used, the maximum occurred at an BMI value of 11.0. As suggested by a reviewer, this could be an artifact due to a few individuals with extremely high BMI values. The result shown in Table 3 is based on a subsequent analysis after removing individuals ($n = 82$) with BMI values > 40 . Figure 5 shows a plot of the conditional odds ratio for BMI, which exhibits peak at the BMI value of 20.5, implying that lower or higher BMI than 20.5 results in lower transitional probability in transitioning to a positive state. Co-morbidities of a history of cardiovascular disease (CVD), depression (defined as having a score greater than or equal to 16), and a history of diabetes are all significant factors negatively affecting transitioning into better states. Hypertension, however, is not statistically significant. Demographic factors, including education level, baseline age, gender, and race, are significant. The lifestyle variable smoking has a negative effect, but alcohol consumption is positive. Similar to the interpretation of coefficients in logistic regression, the estimate of a specific variable in Table 3 can be interpreted as log odds ratios conditioned on the other variables. For example, individuals with CVD are less likely to transition to a better state (odds = $0.79 = \exp(-0.2323)$) than individuals without CVD. Everything else being equal, individuals with BMI values at 30 are also less likely to transition to a better state with odds ratio $0.74 (\hat{f}(30)/\hat{f}(20.5))$ where $\hat{f}(x) = \exp(0.1493 \times x + 0.0036 \times x^2)$ than individuals at the BMI value of 20.5.

Figure 6 shows the prevalences of states across time, stratified by BMI status. The prevalences of poorer states generally increase with BMI for BMI ≥ 25 . The quadratic trend is especially marked for BMI ≥ 35 , where the prevalences of poorer states increase dramatically. Figure 7 shows the individual specific decline in the posterior mean log conditional odds of transition to any better state for a random sample of 50 participants. The graph shows a general downward secular trend with individual variations.

We also noticed that there was a rather abrupt change in the prevalence of State 1 from the first to the second time point; Fig. 6 shows this change across all BMI statuses. At the first time point, the probability of staying in State 1 is 0.59, while the mean probability of staying in State 1 is around 0.9 from the second transition on. One possible explanation is that the study was designed to recruit the better functioning half of the age range. Because there was a high concentration of older adults who had varying small degrees of mobility problems but were otherwise healthy, a substantial number of participants recruited only marginally met the inclusion criteria and were at the borderline of a poor disability state. As a result, regression to the mean was observed at a later time. To account for this artifact, we included an indicator variable (labeled ($t = 1$) in Table 3) to the models for capturing this effect. The recruitment effect is found to be highly significant.

4.4 Goodness-of-fit, missing data and residual analysis

Table 4 compares goodness-of-fit statistics across several models that contain different numbers of states, and with and without the POM transitional model component.

¹To provide further evidence of this conjecture, we fit a random-effect-only model without fixed effects to the data. The estimated standard deviation of the random effect was 0.43 ($p < 0.01$), which suggested that indeed a large part of the heterogeneity is attributable to observed covariates.

Particularly, a 4-state ordered model, which was suggested by a reviewer, was included. This model groups the 10 states into 4 ordered states — not disabled ($\{1\}$), mildly disabled ($\{2,3,4,5\}$), severely disabled ($\{6,7,8,9\}$), and death ($\{D\}$). Besides the BIC, cross-validation loglikelihood (Smyth 2000) was also reported. The cross-validation procedure was based on a 50-50 split for randomly selecting participants for the in- and out-of-samples. All models only include fixed-effects. The 10-state POM model shows the best fit in terms of both BIC and loglikelihood for the test set.

The percentage of missing values among the disability variables in this sample per inclusion criterion described in Section 2 is 14.8% (range is (5.0%, 21.0%)), with smaller variation among the variables (within 5%) and higher variation across time points. As explained earlier, the impact of the missing values of disability variables on the HMM-derived state is minimal. The perturbation to the marginal prevalences (not shown) also appears to be small.

The main threat to the MAR assumption, as we see it, arises from poorly functioning individuals who were too frail to respond, resulting in missing values in the measurements. The missing data mechanism in such a case implies that the missingness of an outcome variable depends its value, invalidating MAR. Although it is not possible to fully assess the validity of MAR, we took several steps in handling the MAR assumption. First, we excluded individuals without two consecutive observations. The step is likely to exclude individuals who were too weak to be able to respond. All in all, we have excluded $n = 38$ such individuals, as opposed to $n = 2,903$ who were included in the analysis. The summed scores on the disability responses, which indicated general disability (higher score is better), were respectively 5.23 (0.93) and 5.02 (1.05) for the included and excluded groups ($p = 0.16$ using a t-test). Thus, the impact on the overall conclusion of excluding individuals without two consecutive measurements was likely to be minimal. For partially missing values (e.g., missing on one disability item but not on another), the MAR assumption seems reasonable. The rationale is that if an older adult is too frail to respond to one item, then it is more likely that the individual will miss the entire questionnaire. In the multiple imputation procedure, we have used all available disability responses in the imputation model. Finally, the creation of the Death State in the model allows us to not have to make the assumption of MAR due to death.

We investigated the validity of the proportional odds assumption in (11). As far as we know, there is no existing procedure for formally assessing the proportional odds assumption for poset data. We approached the problem using two different methods: (a) applying the existing score test to directly evaluate the proportional odds assumption in a simplified 4-state ordered model, which has been previously described; and (b) adapting an existing residual checking method for ordinal data (Harrell 2001) and applying the adapted method to our poset data. The rationale for (a) is to provide checking for an ordinal model that is structurally similar to the full poset model.

We fit a stratified model for the 4-state ordered HMM. The score test (Allison 1999) for the stratified model evaluated the proportional odds assumption for each of all the originating states except the Death state. The p-values for States 1, 2, and 3 were respectively 0.06, 0.83, and 0.002. It is difficult to tell if the proportional odds assumption has been violated in the case when State 3 is the originating state; there were very few transitions from this state to other states (e.g., there were only 5 instances in transitioning to State 1, out of a total of 2,673 transitions).

For the residual checking procedure, the POS residual for each category s and the m^{th} predictor is given by:

$$r_{itrs} = \hat{\gamma}_m x_{im} + \frac{\tilde{\tau}_{itrs} - \tau_{itrs}}{\tau_{itrs}(1 - \tau_{itrs})}, \quad (31)$$

where $m = 1, \dots, p$ is the covariate index, $(\tilde{\tau}_{itrs}, s=1, \dots, S) = \tilde{\tau}_{itr}1 - \hat{A}\tilde{\tau}_{itr}^-$ is a vector of the pseudo-count probabilities of observing better than state s given state r and τ_{itrs} is the model estimate of the probability of observing state s given state r . Due to space limitation, in this paper we only show residual plots at $r = 5$ and $t = 1$ for four covariates (Fig. 8). We excluded the states ($s = 2, 3, 4, 9$), because their transition probabilities were too close to the boundary (0.00 to 2 decimal places). To highlight the general trend, we applied the LOWESS smoothing algorithm (Cleveland 1979) to the residuals with a smoothing factor of 0.5. Non-parallelism of the lines would indicate possible violation of the proportional odds assumption. The graphs for other variables and states look similar to the graphs in Fig. 8 and are not reported here. Generally, the plots suggest that except for a few categories, there is no serious violation of the POS proportional odds assumption, as exemplified by Fig. 8. See further discussion on the proportional odds assumption in the following section.

5 Discussion

Motivated by the study of the disablement process in aging, we propose a partially ordered response model within a hidden Markov framework to directly address several methodological challenges. We use a data-driven approach to delineate granular disability states as an alternative to a rules-based approach, a partial-order proportional odds model to circumvent the complexity of a full-fledged multinomial transition model, and a random-effects model to capture the secular long-term decline of functioning that is above and beyond episodic disability states.

The application of HMM to the HealthABC data uncovered subtle subclinical states that were not discernable using a generic rules-based definition of disability. These subclinical states might be recognized by physicians post hoc, but they are difficult to identify without reliance on data. With even a modest number of binary measurements, it is challenging for a physician to determine if a state is “real” because of the large number of different combinations to consider. Individual physicians’ experience with patients also varies, adding to the difficulty in identifying subclinical states that can be agreed upon.² We thus argue that a data-driven approach is more appropriate for identifying patterns and different states of disabilities within a large sample of patients. The characterization of subclinical states and the close examination of their dynamics could be used by gerontologists to inform prevention, treatment, and care decisions.

Because the POM-HMM modeling framework can incorporate generalized linear mixed models (GLMM), it retains the flexibility that GLMM offers and therefore can be used to statistically evaluate a broad range of scientific hypotheses. For example, the effects of risk factors on transitions between states can be statistically tested using the POM-HMM. Although the example used in this article comes from an observational study, there is no reason to stop at epidemiological predictors. For example, the POM-HMM can be used to evaluate the effects of an exercise intervention program on transition between disability states (Rejeski et al. 2012).

²It is hard for us to see that the 10-state model that was estimated from this data set can be entirely created by a clinician and agreed upon by other experts without reference to any data. For example, should weak upper extremity (as indicated by the lifting 10-lb item), strong lower extremity (quarter-mile walk item), and low flexibility (in/out of bed item) form a “state” of its own? There are just too many combinations for a clinician to consider.

The POM-HMM requires several important assumptions to function well in practice. First, the poset would require a unique maximal element for (11) to be estimable. Of course, if the partial order is changed in direction (from $\succ$ to $\prec$) within the poset model, then a unique minimal element is required. The assumption of a single maximum (minimum) may not be universally true for every empirical study; however, our experience with a dozen or so clinical applications has revealed that without exception, a unique maximal and/or a unique minimal element always existed. If circumstances exist in which multiple states appear as maximal/minimal elements, we may need to merge the states to circumvent the modeling constraint. Other models, such as one that imposes a total order on partitions of partially ordered sets and then builds unordered models within each partition, are possible (e.g., Zhang and Ip 2011). Work is currently ongoing along such lines of research. Second, we use a simple rule of coordinate-wise dominance for determining the poset relationship. This works quite naturally for binary responses. For ordinal responses, the issue becomes more complicated. Some researchers would prefer the poset structure to be always specified by users and to be kept fixed in the transition analysis.³ We have adopted this approach in the analysis. The dominance matrix could also be entirely specified by users, and this is an especially viable solution for situations in which there are categorical responses with more than two categories, or when the dominance relation between states can be clinically determined.⁴ A somewhat similar issue concerns the determination of traversal between states. It is possible for users to specify structural zero probabilities for certain transitions. The constrained model could be of interest in special situations in which transitions between certain states are clinically not plausible. We did not impose such constraint in the current application of POM-HMM.

Third, the evaluation of the poset proportional odds assumption – the constraint that the regression coefficient is constant over different transitions – requires further investigation. Admittedly, proportional odds is a rather strong assumption. For the HealthABC data set, this assumption implies that a risk factor such as CVD has more or less uniform impact on transitions from a specific health status to better (defined by the partial order) health statuses. Many, if not all, of the factors included in our model are either general health indicators — depression, comorbidity, lifestyle, and so on — or demographic variables. Thus, the “uniform impact” assumption is reasonable. We paid special attention to the age variable because age is expected to differentially affect the odds ratios in transition. However, we did not find evidence that this is the case (e.g., see Fig. 8). In other applications, the uniform impact assumption may be difficult to defend. For example, in a study of domestic violence (Ip et al. 2010), a batterer intervention program may have a stronger effect on light and first-time offenders than on persistent offenders. To evaluate the proportional odds assumption, we have adopted an exploratory approach by testing the assumption under a simplified, ordered model, and also inspecting the residuals across the different categories of transitions. We recognize that the two reported evaluation methods have limitations. There may exist a subset of predictors that does not satisfy the proportional odds assumption. The visual inspection method is also subject to bias and divergent interpretations. A comprehensive analysis is beyond the scope of this article, but it seems that a hypothesis-testing approach such as score functions would be a welcomed development for checking the poset proportional odds assumption. It is also worth pointing out here that an alternative solution to making the proportional odds assumption is applying a hierarchical Bayesian HMM approach to the data, such as the one described in Shirley et al. (2010).

³We thank a reviewer for this suggestion.

⁴We thank another reviewer for this suggestion.

Like many recent investigations into changes in physical functioning (e.g., Hubbard et al. 2009), we assume that the disablement process follows the first-order Markov assumption. However, as pointed out by Hubbard et al. (2009), if the history of disability status were more relevant to the transition between states, then semi-Markov process models would be more appropriate. While such conceptualization can be modeled through a dual process in which one is a Markov chain and the other a process that is a function of consecutive states, the modeling and estimation become more complex. Evaluation of the first-order Markov assumption and development of potential models for deviation from such an assumption would be an important area of research for the disablement process.

Appendix

A Regression equations for the Health ABC study

$$\begin{aligned}
 \log\left(\frac{\tau_{itr1}}{\tau_{itr2}}\right) &= \phi_{r2} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{\tau_{itr1}}{\tau_{itr3}}\right) &= \phi_{r3} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{\tau_{itr1}}{\tau_{itr4}}\right) &= \phi_{r4} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{\tau_{itr1}}{\tau_{itr5}}\right) &= \phi_{r5} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{\tau_{itr1} + \tau_{itr2} + \tau_{itr4} + \tau_{itr5}}{\tau_{itr6}}\right) &= \phi_{r6} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{\tau_{itr1} + \tau_{itr2}}{\tau_{itr7}}\right) &= \phi_{r7} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{\tau_{itr1} + \tau_{itr2} + \tau_{itr3} + \tau_{itr4} + \tau_{itr5}}{\tau_{itr8}}\right) &= \phi_{r8} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{\tau_{itr1} + \tau_{itr2} + \tau_{itr3} + \tau_{itr4} + \tau_{itr5} + \tau_{itr6} + \tau_{itr7} + \tau_{itr8}}{\tau_{itr9}}\right) &= \phi_{r9} + x'_{it}\gamma + w'_{it}\theta_i, \\
 \log\left(\frac{1 - \tau_{itr10}}{\tau_{itr10}}\right) &= \phi_{r10} + x'_{it}\gamma + w'_{it}\theta_i.
 \end{aligned}$$

B Proof of Lemma 1

Proof. Consider a positive linear predictor vector $\tau = (\tau_1, \dots, \tau_{K-1})$ and the probability vector $\eta = (\eta_1, \dots, \eta_{K-1})$ that are related by the following equation:

$$[\hat{A} + \text{diag}(\eta)]\tau = \mathbf{1}. \quad (32)$$

Assume we have proved $\tau_i > 0$. Because all entries of $\hat{A} + \text{diag}(\eta)$ are positive with diagonal entries greater than 1, whereas the right-hand side of (32) is $\mathbf{1}$, we have, coordinate-wise, $\tau_i < 1$.

For proving that (32) has a positive solution, we rely on Farkas' Lemma (Fang and Puthenpura 1993, p. 60), which states that there exists either a positive solution τ , or a $\hat{\tau}$

such that $[\hat{A} + \text{diag}(\eta)]^T \hat{\tau} \geq 0$ and $\mathbf{1}^T \hat{\tau} < 0$. We now prove that indeed $\hat{\tau}$ would not exist and hence (32) can only have a positive solution.

If $\hat{\tau}$ does exist, then from $\mathbf{1}^T \hat{\tau} < 0$ we know at least one $\hat{\tau}_i < 0$. From the assumption of the presence of a unique maximal in the poset structure, there exists a column, say column j of $\hat{A}$, being all ones. By the first condition on $\hat{\tau}$, we have for the j^{th} coordinate,

$$\sum_{i=1}^{K-1} \hat{\tau}_i + \eta_j \hat{\tau}_j = \mathbf{1}^T \hat{\tau} + \eta_j \hat{\tau}_j \geq 0. \quad (33)$$

Thus $\eta_j \widehat{\tau}_j \geq -1^T \widehat{\tau} > 0$, implying that $\widehat{\tau}_j > 0$. If all states other than the maximal state are only dominated by the maximal state, then the proof is complete because, by the above argument, all columns of $\hat{A}$ are $\mathbf{1}$ and hence all $\widehat{\tau}_j > 0$, contradicting the condition $1^T \widehat{\tau} = \sum_{i=1}^{K-1} \tau_i < 0$.

If there is a state, say k , that is dominated by states directly dominated by the maximal, then denote the set of states dominating k by S_k . It is easy to check that all entries in column k in $\hat{A}$ are 1 except at positions in the set S_k , which are zero. For example, if $1 > 2 > 3$, then $\hat{A}$ contains two columns, $[1, 1]^T$ and $[0, 1]^T$. Note that from the above argument, $\forall j \in S_k, \widehat{\tau}_j > 0$, because states within S_k are directly dominated by the maximal. Now, we have

$$\sum_{i=1, i \notin S_k}^{K-1} \widehat{\tau}_i + \eta_k \widehat{\tau}_k = \sum_{i=1}^{K-1} \widehat{\tau}_i + \eta_k - \sum_{i \in S_k} \widehat{\tau}_i \geq 0, \quad (34)$$

which leads to

$$\sum_{i=1}^{K-1} \widehat{\tau}_i + \eta_k \widehat{\tau}_k \geq \sum_{i \in S_k} \widehat{\tau}_i > 0, \quad (35)$$

and hence we can conclude that $\widehat{\tau}_k > 0$. By recursively considering states directly dominated by other states, we can prove that $\widehat{\tau}_i > 0$ for all i . The process is guaranteed to complete given a finite poset. Thus $\widehat{\tau}$ cannot exist and, by Farkas' Lemma, (32) can only have a positive solution.

For the probability of the last category, τ_K , from (32) we know $\forall j \in 1, \dots, K-1$,

$$\sum_{i=1}^{K-1} \tau_i + \eta_j \tau_j = 1, \quad (36)$$

and because $\tau_j > 0$ and $\eta_j > 0$, we know $0 < \sum_{i=1}^{K-1} \tau_i = 1 - \eta_j \tau_j < 1$. Hence $0 < \tau_K = 1 - \sum_{i=1}^{K-1} \tau_i < 1$.

References

- Agresti, A. Categorical data analysis. Wiley-Interscience; New York: 2002.
- Allison, P. Logistic regression using the SAS system: theory and application. SAS Publishing; 1999.
- Altman R. Mixed hidden Markov models: An extension of the hidden Markov model to the longitudinal data setting. Journal of the American Statistical Association. 2007; 102:201–210.
- Bartolucci F, Farcomeni A. A multivariate extension of the dynamic logit model for longitudinal data based on a latent Markov heterogeneity structure. Journal of American Statistical Association. 2009; 104:816–831.
- Baum L, Petrie T. Statistical inference for probabilistic functions of finite state Markov chains. The Annals of Mathematical Statistics. 1966; 37:1554–1563.
- Beckett L, Brock D, Lemke J, de Leon C, Guralnik J, Fillenbaum G, Branch L, Wetle T, Evans D. Analysis of change in self-reported physical function among older persons in four population studies. American Journal of Epidemiology. 1996; 143:766–778. [PubMed: 8610686]
- Blinded. Technical Report. 2010.
- Cleveland W. Robust locally weighted regression and smoothing scatterplots. Journal of the American Statistical Association. 1979:829–836.

- Crimmins E, Saito Y, Reynolds S. Further evidence on recent trends in the prevalence and incidence of disability among older Americans from two sources: the LSOA and the NHIS. *Journals of Gerontology Series B*. 1997; 52:59–71.
- Dempster A, Laird N, Rubin D, et al. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1977; 39:1–38.
- Fahrmeir, L.; Tutz, G. *Multivariate statistical modelling based on generalized linear models*. Springer Verlag; New York: 2001.
- Fang, S.; Puthenpura, S. *Linear optimization and extensions: theory and algorithms*. Prentice-Hall, Inc.; Upper Saddle River, NJ, USA: 1993.
- Freedman V, Martin L, Schoeni R. Recent trends in disability and functioning among older adults in the United States: a systematic review. *Journal of the American Medical Association*. 2002; 288:3137–3146. [PubMed: 12495394]
- Gill T, Gahbauer E, Allore H, Han L. Transitions between frailty states among community-living older persons. *Archives of Internal Medicine*. 2006; 166:418–423. [PubMed: 16505261]
- Gill T, Kurland B. The burden and patterns of disability in activities of daily living among community-living older persons. *Journals of Gerontology Series A: Biological and Medical Sciences*. 2003; 58:70–75.
- Guilley E, Ghisletta P, Armi F, Berchtold A, d'Epinay C, Michel J, de Ribaupierre A. Dynamics of frailty and ADL dependence in a five-year longitudinal study of octogenarians. *Research on Aging*. 2008; 30:299–317.
- Hardy S, Dubin J, Holford T, Gill T. Transitions between states of disability and independence among older persons. *American Journal of Epidemiology*. 2005; 161:575–584. [PubMed: 15746474]
- Hardy S, Gill T. Recovery from disability among community-dwelling older persons. *Journal of the American Medical Association*. 2004; 291:1596–1602. [PubMed: 15069047]
- Harrell, F. *Regression modeling strategies*. Springer; New York: 2001.
- Harris T, Visser M, Everhart J, Cauley J, Tylavsky F, Fyfe T, Zamboni M, Taaffe D, Resnick H, Scherzinger A, et al. Waist circumference and sagittal diameter reflect total body fat better than visceral fat in older men and women: the Health, Aging and Body Composition Study. *Annals of the New York Academy of Sciences*. 2006; 904:462–473. [PubMed: 10865790]
- Hubbard R, Inoue L, Diehr P. Joint modeling of self-rated health and changes in physical functioning. *Journal of the American Statistical Association*. 2009; 104:912–928. [PubMed: 20151036]
- Ip E, Snow-Jones A, Heckert D, Zhang Q, Gondolf E. Latent Markov model for analyzing temporal configuration for violence profiles and trajectories in a sample of batterers. *Sociological Methods and Research*. 2010; 39:222–255.
- Katz S. Assessing self-maintenance: activities of daily living, mobility, and instrumental activities of daily living. *J Am Geriatr Soc*. 1983; 31:721–27. [PubMed: 6418786]
- Lauritzen S, Spiegelhalter D. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B*. 1988; 50:157–224.
- Leveille S, Penninx B, Melzer D, Izmirlian G, Guralnik J. Sex differences in the prevalence of mobility disability in old age: the dynamics of incidence, recovery, and mortality. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2000; 55:S41–S50.
- MacDonald, I.; Zucchini, W. *Hidden Markov and other models for discrete-valued time series*. CRC Press; 1997.
- McCullagh P. Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1980; 42:109–142.
- Meulders M, Ip E, De Boeck P. Latent variable models for partially ordered responses and trajectory analysis of anger-related feelings. *British Journal of Mathematical and Statistical Psychology*. 2005; 58:117–143. [PubMed: 15969843]
- Peres K, Verret C, Alioum A, Barberger-Gateau P. The disablement process: factors associated with progression of disability and recovery in French elderly people. *Disability & Rehabilitation*. 2005; 27:263–276. [PubMed: 16025753]
- Rabiner L. A tutorial on hidden Markov models and selected applications in speech recognition. *Readings in speech recognition*. 1990; 53:267–296.

- Rejeski J, Ip E, Bertoni A, Bray G, Evans G, Gregg E, Zhang Q, the Look AHEAD Research Group. Lifestyle change and mobility in obese adults with Type 2 diabetes. *New England Journal of Medicine*. 2012 in press.
- Rejeski W, Ip E, Marsh A, Miller M, Farmer D. Measuring disability in older adults: The international classification system of functioning, disability and health (ICF) framework. *Geriatrics and Gerontology International*. 2008a; 8:48–54. [PubMed: 18713189]
- Rejeski W, Ip E, Marsh A, Zhang Q, Miller M. Obesity influences transitional states of disability in older adults with knee pain. *Archives of Physical Medicine and Rehabilitation*. 2008b; 89:2102–2107. [PubMed: 18996238]
- Schafer, J. Analysis of incomplete multivariate data. Chapman & Hall/CRC; 1997.
- Scott S. Bayesian methods for hidden Markov models. *Journal of the American Statistical Association*. 2002; 97:337–351.
- Scott S, James G, Sugar C. Hidden Markov models for longitudinal comparisons. *Journal of the American Statistical Association*. 2005; 100:359–370.
- Shirley K, Small D, Lynch K, Maisto S, Oslin D. Hidden Markov models for alcoholism treatment trial data. *The Annals of Applied Statistics*. 2010; 4:366–395.
- Smyth P. Belief networks, hidden Markov models, and Markov random fields: a unifying view. *Pattern Recognition Letters*. 1997; 18:1261–1268.
- Smyth P. Model selection for probabilistic clustering using cross-validated likelihood. *Statistics and Computing*. 2000; 10:63–72.
- Stephens M. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2000; 62:795–809.
- Thinggaard M, Jacobsen R, Jeune B, Martinussen T, Christensen K. Is the relationship between BMI and mortality increasingly U-shaped with advancing age? A 10-year follow-up of persons aged 70–95 years. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*. 2010; 65:526.
- Tutz G. Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*. 1990; 43:39–55.
- Verbrugge L, Jette A. The disablement process. *Social Science & Medicine*. 1994; 38:1–14. [PubMed: 8146699]
- Vermunt J, Langeheine R, Bockenholt U. Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*. 1999; 24:179.
- Yang Y, George L. Functional disability, disability transitions, and depressive symptoms in late life. *Journal of Aging and Health*. 2005; 17:263–292. [PubMed: 15857959]
- Zhang Q, Ip EH. Generalized linear model for partially ordered data. *Statistics in Medicine*. 2011; 31:56–68. [PubMed: 22086670]
- Zhang Q, Snow Jones A, Rijmen F, Ip E. Multivariate discrete hidden markov models for domain-based measurements and assessment of risk factors in child development. *Journal of Computational and Graphical Statistics*. 2010; 19:746–765.

Figure 1. The percentage of deceased adults and the percentage of older adults reporting difficulty in performing the tasks over time.

Figure 2.

The dominance structure of the 10-state model for the Health ABC data. An arrow that goes from State X to State Y implies $X \geq Y$. Transition between states is unrestricted between states — e.g., state 3 can transition into state 4 and vice versa.

Figure 3.
BICs of computed HMMs.

Figure 4.

State conditional probabilities for the “No difficulty” response, excluding the Death State.

The leftmost State 1 has no difficulty in any one of the six physical functions; the rightmost State 9 has difficulties in all six. The bookend marks indicate the minimum and maximum of values for five imputed data sets for missing values. The color indicates different domains: horizontal and vertical mobilities, ADLs, and upper-body strength.

Figure 5.
Estimated conditional odds as a function of BMI.

Figure 6.
State prevalences across time, stratified by BMI status.

Figure 7.
Smoothed mean log conditional odds vs. time for a random sample of 50 participants.

Figure 8. Residual plots for the transition probabilities from different originating states to state 5 assessing proportional odds assumption. A random sample of 10% of the individual residuals is shown. States 2,3,4, and 9 are not included because their transition probabilities from state 5 are too close to zero.

Table 1Baseline Characteristics of Health ABC Participants ($N = 2,903$)

Characteristics		Mean $\pm$ SD / Percentage	Missing Percentage
Age (year)		73.6 $\pm$ 2.9	
BMI (Kg/m^2)		27.6 $\pm$ 4.8	0
Education			0.3
	Less than high school	24	
	High school graduate	32	
	Postsecondary	44	
CESD			1
	< 16	95	
	≥ 16	5	
Previous vascular disease			2.7
	No	71	
	Cerebro, or cardio, or both	29	
Previous diabetes			0.5
	No	85	
	Yes	15	
Previous hypertension			0
	No	41	
	Low ^a	21	
	High ^b	38	
Race			0
	White	60	
	Black	40	
Gender			0
	Female	52	
	Male	48	
Marital status			8.8
	Married	52	
	Unmarried	48	
Smoking status			0.2
	Never	45	
	Before or Now	55	
Alcohol consumption			0.5
	< 1 drink/week	71	
	1 drink/week	29	

^a 130 SBP<140 and DBP<90 or SBP<140 and 85 DBP<90, where SBP stands for the systolic blood pressure and DBP stands for the diastolic blood pressure.

^b SBP 40 or DBP 90.

Table 2

Estimated Prior State Probabilities () and Transition Probabilities (^a)

	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9	Death
:	0.95	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
:	0.83	0.01	0.04	0.04	0.03	0.00	0.01	0.01	0.00	0.02
	0.02	0.74	0.00	0.02	0.00	0.15	0.01	0.00	0.03	0.03
	0.12	0.00	0.75	0.01	0.00	0.00	0.07	0.00	0.00	0.04
	0.03	0.00	0.00	0.83	0.00	0.02	0.00	0.07	0.00	0.05
	0.13	0.01	0.01	0.00	0.71	0.03	0.00	0.06	0.00	0.04
	0.00	0.04	0.00	0.00	0.01	0.81	0.00	0.00	0.06	0.07
	0.01	0.00	0.10	0.00	0.01	0.00	0.69	0.04	0.08	0.06
	0.01	0.00	0.00	0.08	0.01	0.03	0.00	0.78	0.03	0.07
	0.00	0.01	0.00	0.00	0.00	0.06	0.02	0.02	0.79	0.11
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

^a _{rs} value indicates the transition probability of ^rth row to ^sth column.

Table 3

Parameters for POM-HMM Estimated from the Health ABC Study

	Fixed Effects Model				Mixed Effects Model ^a			
	Estimates	SE	t	Pr > t	Estimates	SE	t	Pr > t
BMI	.1493	.0500	2.99	.003	.1511	.0511	2.96	.003
BMI ²	-.0036	.0009	4.08	< .001	-.0037	.0009	4.02	< .001
Education	.1417	.0296	4.78	< .001	.1453	.0308	4.72	< .001
CESD	-.0585	.0049	12.06	< .001	-.0589	.0051	11.46	< .001
CVD	-.2323	.0416	5.58	< .001	-.2400	.0446	5.38	< .001
Diabetes	-.3987	.0609	6.55	< .001	-.4057	.0628	6.46	< .001
Hypertension	-.0061	.0246	.25	.805	-.0060	.0249	.24	.809
Black vs. White	-.1215	.0498	2.44	.015	-.1235	.0510	2.42	.015
Female vs. Male	-.2589	.0501	5.16	< .001	-.2675	.0529	5.06	< .001
Baseline Age	-.0524	.0077	6.80	< .001	-.0528	.0080	6.64	< .001
Marital Status	-.0135	.0297	.46	.648	-.0125	.0301	.42	.678
Smoking	-.0433	.0241	1.80	.072	-.0452	.0246	1.84	.066
Alcohol Consumption	.0604	.0241	2.50	.012	.0622	.0246	2.52	.012
t	-.1232	.0153	8.05	< .001	-.1217	.0157	7.74	< .001
t=1	2.5139	.1159	21.68	< .001	.2509	.0117	21.44	< .001
					.0394	.0419	.94	.347
Log Likelihood	-30,385.82				-30,385.20			
Likelihood Ratio Test					1.23 .267			

^aThe mixed model is based on Eq (11) such that the log conditional odds= $\phi_{rs} + x'_{it}\gamma + \theta_i t_{it}$, where the random slope $\theta_i \sim N(0, \sigma^2)$, and ϕ_{rs} and x_{it} are respectively the intercept and the fixed effects. Mean imputation was used for missing values in the predictor variables.

Table 4

Goodness-of-Fit Statistics for HealthABC Data

Model	Loglikelihood for Training Set	Loglikelihood for Test Set	BIC
2-state	-23908	-24037	95946
4-state (ordered)	-17643	-17776	71070
4-state (ordered) + POM	-16870	-16961	67893
10-state	-16389	-16666	67249
10-state + POM	-15822	-16042	64868