Published in final edited form as:

Stat Med. 2013 September 10; 32(20): 3569–3589. doi:10.1002/sim.5782.

Using Group-Based Latent Class Transition Models to Analyze Chronic Disability Data from the National Long Term Care Survey 1984–2004

Toby A. White,

Drake University, Actuarial Science & Finance

Elena A. Erosheva

University of Washington, Statistics & Social Work

Abstract

Latent class transition models track how individuals move among latent classes through time, traditionally assuming a complete set of observations for each individual. In this paper, we develop group-based latent class transition models that allow for staggered entry and exit, common in surveys with rolling enrollment designs. Such models are conceptually similar to, but structurally distinct from, pattern mixture models of the missing data literature. We employ group-based latent class transition modeling to conduct an in-depth data analysis of recent trends in chronic disability among the U.S. elderly population. Using Activities of Daily Living data from the National Long Term Care Survey (NLTCs), 1982–2004, we estimate model parameters using the EM algorithm, implemented in SAS PROC IML. Our findings indicate that declines in chronic disability prevalence, observed in the 1980s and 1990s, did not continue in the early 2000s as previous NLTCs cross-sectional analyses have indicated.

1 Introduction

Panel data sets, for which the number of waves is small but the assessment interval is often large, are adept at handling complex dynamic behaviors, but they often suffer from missing data problems such as those due to death and drop-out [1]. Many longitudinal analyses of survey data assume a balanced panel design whereby either all subjects are observed across all waves, or if there are subjects with any missing data, they are removed from further consideration [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. However, if a survey has late entrants or early drop-outs, these individuals may have different measurement profiles than those who are present across all waves, which may lead to biased findings if the probability of drop-out depends on the measurement process itself [13].

An unbalanced panel design attempts to replace those respondents who drop-out by an equal proportion of respondents who are new to the survey [1]. Some longitudinal analyses do allow for early drop-out [14, 15, 16, 17], but do not account for late entry after the initial

wave, especially if individuals entering late were eligible and could have potentially been observed in earlier waves. However, surveys with rolling enrollment, such as the National Long-Term Care Survey (NLTCs) [18], National Youth Survey [19], and Denver Youth Survey [20] have groups of individuals that by design may enter a survey late (e.g., after first age eligible) or leave early (e.g., before the last wave), which creates a missing data problem.

The issue then becomes how to account for the missing data among respondents who either drop-out early or who enter a survey late. First, though, it is helpful to distinguish between the three types of missing data mechanisms. If the missingness process is not related to either the observed data or the missing data, the data are said to be missing completely at random (MCAR). If the missingness process is not related to the missing data, but does depend on the observed data, the data are said to be missing at random (MAR). Finally, if the missingness process also depends on the missing data itself, the data are said to be missing not at random (MNAR or informative missingness) [21].

For an unbalanced panel survey design, the MNAR assumption is the most appropriate. In other words, the missing data mechanism is non-ignorable, for if it were ignored, the estimation of parameters would be biased and not consistent. There have been several approaches suggested to handle this problem, including jointly modeling drop-out and the response variable [22], imposing a missing data indicator [23], creating a logit covariate-based model for non-response [24], classifying partial and complete responses into a matrix or an array [25], and incorporating missing data to model the repeated outcomes data and missingness process jointly [26].

Three modeling approaches are available for this purpose. First, selection models factor the joint density of the full data into the marginal outcome density and the response (missing or not) density conditional on the outcomes [13]. Second, pattern mixture models factor the joint density of the full data into the marginal response density and the outcome density conditional on the missingness pattern [27, 28]. Third, shared parameter models incorporate random effects, which drive both the outcome and response processes [29]. For a more complete theoretical background on these three modeling types, see Hogan et.al. [30].

In this paper, we present a framework to handle partial non-response with a group-based extension of latent class transition models (LCTM), and we apply our models to chronic disability binary response data from the National Long-Term Care Survey (NLTCs). We use the term ‘group’ to denote a collection of individuals who enter the survey at the same time, leave the survey at the same time, and may share some other observed characteristics (e.g. - they may be within a certain common age interval). Our models are most similar to pattern mixture models, as we consider each possible pathway of observations as one group, and then combine such observations within a traditional likelihood-based estimation approach. However, unlike pattern mixture models, we avoid estimating separate parameter sets for each pattern of observations, which helps with interpretation. Furthermore, we incorporate both early drop-out and late entry, and distinguish between the case where late entry is due to age restrictions and when it occurs for other reasons (e.g. - non-response, sample selection, etc.).

The NLTCs is the largest and most representative longitudinal survey of chronic disability in the U.S. elderly population, 1982–2004. Previous substantive analyses based on the NLTCs have included comparisons of cross-sectional slices of the population over time [31, 32], estimation of individual longitudinal pathways [33, 34], as well as longitudinal analyses based on the Grade of Membership (GoM) model [35, 36, 37]. The cross-sectional comparisons [31, 32] were aggregate-based, and consider the populations as a whole at particular snapshots in time, but they did not account for individual transitions across waves. The approaches that estimated individual longitudinal pathways [33, 34] across multiple time periods assumed that responses across time, conditional on latent trajectory, were independent; the likelihood function for such approaches was simply a product of probabilities over time. These pathways formed continuous trajectories, and assumed an individual remained in the same growth trajectory throughout. One analysis based on the Grade of Membership (GoM) Model incorporated longitudinal settings by pooling the data across waves but ignoring within-person variability [35], while another analysis employed transition probability matrices for each age but ignored other risk factors affecting transition behavior [36]. Most recently, a GoM Trajectory model combined features of both the cross-sectional GoM model and longitudinal trajectory growth models to study the evolution of certain classes of individuals over time [37].

Among all the previous analyses of trends in chronic disability that rely on Activities of Daily Living (ADL) data from the NLTCs, the dominant opinion [31] is that disability declined throughout the period from 1984–2004, and that the rate of this decline has continually accelerated. However, other studies based on the same NLTCs data reached conflicting conclusions with respect to either the timing or the rate of disability decline during this period [38, 39, 40]. Using a different, longitudinal modeling framework that focuses on individual transitions across waves, we conduct an independent assessment of these conclusions, and also explore a key link between the presence of arthritis and various ADLs.

In Section 2, we develop group-based LCTM, and distinguish between initial latent class weights and period class weights, which are used to interpret longitudinal shifts among classes. In Section 3, we conduct a brief simulation study of group-based LCTM, so as to verify that the true model parameters are recovered by estimation procedures. In Section 4, we incorporate covariate extensions to the three components of LCTM and discuss the diagnostics of assumptions pertaining to LCTM with covariates. In Section 5, we conduct a longitudinal analysis of the NLTCs ADL data and address issues relating to model diagnostics, parameter estimation, computing, and the calculation of standard errors. In Section 6, we discuss results of our analyses with respect to both the progression of recent U.S. disability decline and the effect of arthritis on functional limitations. Finally, in section 7, we summarize the methodology and its application to chronic disability data.

2 Group-Based Latent Class Transition Models

2.1 Latent Class Models

Let M be the number of observed binary responses, $y_1, \dots, y_M$. At each point in time, assume a K -class latent class model, where K is fixed and unknown [41]. Let $I = 1, \dots, K$ denote

latent class membership. Then, let $\gamma_{ij} = Pr(L_i = j)$ represent the K latent class weights. Here, L_i is the latent class indicator for individual i . Note that the model parameters don't directly depend on individual but rather on the corresponding latent class to which an individual belongs. Let $\rho_{im|l} = Pr(Y_{im} = 1 | L_i = l)$ represent the conditional response probability of individual i to item m , $m = 1, \dots, M$, given individual i is in class l .

Conditional independence [42] is the core assumption of latent class models, which states that:

$$Pr(Y_{i1} = y_1, \dots, Y_{iM} = y_M | L_i = l) = \prod_{m=1}^M Pr(Y_{im} = y_m | L_i = l). \quad (1)$$

Incorporating conditional independence, and defining:

$$\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iM}), \mathbf{y} = (y_1, y_2, \dots, y_M), \quad \boldsymbol{\gamma} = (\gamma_1, \gamma_2, \dots, \gamma_K), \text{ and}$$

$$\boldsymbol{\rho} = ((\rho_{1|1}, \rho_{2|1}, \dots, \rho_{M|1}), (\rho_{1|2}, \rho_{2|2}, \dots, \rho_{M|2}), \dots, (\rho_{1|K}, \rho_{2|K}, \dots, \rho_{M|K}))$$

we can write the marginal distribution for an individual's observed data (at a single point in time, and dropping the subscript i for model parameters) as:

$$Pr(\mathbf{Y}_i = \mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\rho}) = \sum_{l=1}^K \gamma_l \prod_{m=1}^M \rho_{m|l}^{y_{im}} (1 - \rho_{m|l})^{1 - y_{im}}. \quad (2)$$

For now, we also assume that no covariates are present, although we relax this assumption later.

2.2 Latent Class Transition Models

Latent Class Transition Models (LCTMs) extend latent class models to include multiple waves of observations. Standard LCTMs assume that all individuals have been observed across all waves. Considering a three-wave example, for illustrative purposes, let I_t denote latent class membership for the t th wave, where $t=1,2,3$. Then, $\gamma_{iI_t} = Pr(L_{it} = I_t)$ gives the latent class weights at wave t for $I_t = 1, \dots, K$. Also, $\rho_{im|I_t} = Pr(Y_{imt} = 1 | L_{it} = I_t)$ are the conditional response probabilities for indicator m given class I_t at time t for $m = 1, \dots, M$ and $t = 1, 2, 3$. Note that individual i 's complete set of observations over time is comprised of $t = 3$ vectors of length M . Finally, denote latent class transition probabilities by $\pi_{iI_t|I_{t-1}} = Pr(L_{it} = I_t | L_{i(t-1)} = I_{t-1})$ for moves from wave $t-1$ to an adjacent wave t . Here, there are two possible transition opportunities, so that the pair $(t-1, t)$ is either (1,2) or (2,3).

A core assumption of LCTMs is that transition probabilities follow a first-order Markovian process [6]; in the three-wave case, this implies:

$$Pr(L_{i3} = l_3 | L_{i1} = l_1, L_{i2} = l_2) = Pr(L_{i3} = l_3 | L_{i2} = l_2) = \pi_{il_3 l_2}. \quad (3)$$

An additional assumption of most LCTMs is that the response probabilities (given by ρ) are time-homogeneous [43], meaning that one does not have to estimate a separate set of such probabilities each wave. This is a necessary restriction to keep the nature of latent classes consistent over time. Two slightly less common assumptions, that are also employed throughout this paper, are that transition probabilities are time-homogeneous [43], and waves are evenly-spaced [9].

We can now define:

$$\mathbf{Y}_i = ((Y_{i11}, Y_{i21}, \dots, Y_{iM1}), (Y_{i12}, Y_{i22}, \dots, Y_{iM2}), (Y_{i13}, Y_{i23}, \dots, Y_{iM3})),$$

$$\mathbf{y} = ((y_{11}, y_{21}, \dots, y_{M1}), (y_{12}, y_{22}, \dots, y_{M2}), (y_{13}, y_{23}, \dots, y_{M3})), \text{ and}$$

$$\boldsymbol{\pi} = ((\pi_{1|1}, \pi_{2|1}, \dots, \pi_{K|1}), (\pi_{1|2}, \pi_{2|2}, \dots, \pi_{K|2}), \dots, (\pi_{1|K}, \pi_{2|K}, \dots, \pi_{K|K}))$$

and then write the marginal distribution for an individual's observed data as:

$$\begin{aligned} Pr(\mathbf{Y}_i = \mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{\pi}) &= \sum_{l_1=1}^K \left[\gamma_{l_1} \prod_{m=1}^M \rho_{ml_1}^{y_{im1}} (1 - \rho_{ml_1})^{1-y_{im1}} \right] \\ &\times \sum_{l_2=1}^K \left[\pi_{l_2|l_1} \prod_{m=1}^M \rho_{ml_2}^{y_{im2}} (1 - \rho_{ml_2})^{1-y_{im2}} \right] \times \sum_{l_3=1}^K \left[\pi_{l_3|l_2} \prod_{m=1}^M \rho_{ml_3}^{y_{im3}} (1 - \rho_{ml_3})^{1-y_{im3}} \right]. \end{aligned} \quad (4)$$

The $K \times K$ set of time-homogeneous transition probabilities for individual i can be represented succinctly in a matrix Π_i :

$$\Pi_i = \begin{pmatrix} \pi_{il_t=1|l_{t-1}=1} & \pi_{il_t=2|l_{t-1}=1} & \cdots & \pi_{il_t=K|l_{t-1}=1} \\ \pi_{il_t=1|l_{t-1}=2} & \pi_{il_t=2|l_{t-1}=2} & \cdots & \pi_{il_t=K|l_{t-1}=2} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{il_t=1|l_{t-1}=K} & \pi_{il_t=2|l_{t-1}=K} & \cdots & \pi_{il_t=K|l_{t-1}=K} \end{pmatrix}.$$

The rows of Π_i represent latent class assignments for the previous wave ($t-1$), and the columns represent latent class assignments for the present wave (t). If any logical ordering of

the latent classes exists, both the rows and columns should be ordered correspondingly.

Although $\sum_{t=1}^K \pi_{l_t|l_{t-1}} = 1$, it is not generally true that $\sum_{(t-1)=1}^K \pi_{l_t|l_{t-1}} = 1$.

2.3 Group-Based Extension and Initial Latent Class Weights

Define a group to be a set of individuals who both share the same times of observation and a common age interval. For simplicity, assume initially there are only two waves (1 and 2) and one transition opportunity. Then, there will be four groups of individuals that appear in the two-wave data set, specifically:

- group A = individuals with observed records in waves 1 and 2.
- group B = individuals with observed records in wave 1, but not wave 2.
- group C = individuals with observed records in wave 2, but not wave 1, due to not being age-eligible in wave 1.
- group D = individuals with observed records in wave 2, but not wave 1, who were age-eligible in wave 1, but were not in the survey sample at that time.

Now, the full marginal distribution can be composed as a product of the marginal distributions for each of the four groups, connected via a group indicator function as follows:

$$\begin{aligned}
 & Pr(\mathbf{Y}_i = \mathbf{y}; \gamma_{l_1}, \gamma_{l_2}, \rho, \pi) \\
 &= \left[\sum_{l_1=1}^K \gamma_{l_1} \prod_{m=1}^M \rho_m^{y_{im1}} (1 - \rho_m|_{l_1})^{1-y_{im1}} \sum_{l_2=1}^K \pi_{l_2|l_1} \prod_{m=1}^M \rho_m^{y_{im2}} (1 - \rho_m|_{l_2})^{1-y_{im2}} \right]^{I(G_i=A)} \\
 &\times \left[\sum_{l_1=1}^K \gamma_{l_1} \prod_{m=1}^M \rho_m^{y_{im1}} (1 - \rho_m|_{l_1})^{1-y_{im1}} \right]^{I(G_i=B)} \\
 &\times \left[\sum_{l_2=1}^K \gamma_{l_2} \prod_{m=1}^M \rho_m^{y_{im2}} (1 - \rho_m|_{l_2})^{1-y_{im2}} \right]^{I(G_i=C)} \\
 &\times \left[\sum_{l_1=1}^K \gamma_{l_1} \sum_{l_2=1}^K \pi_{l_2|l_1} \prod_{m=1}^M \rho_m^{y_{im2}} (1 - \rho_m|_{l_2})^{1-y_{im2}} \right]^{I(G_i=D)}.
 \end{aligned} \tag{5}$$

Equation (5) contains new parameters that do not appear in traditional LCTM. Specifically, γ_{l_1} and γ_{l_2} are two sets of latent class weights, which we will refer to as *initial* latent class weights hereafter. More formally, the initial latent class weights correspond to the wave during which an individual was first age-eligible to participate in the survey, regardless of whether they actually participated or not. Thus, all individuals from Groups A, B, and D have initial latent class weights $\gamma_{il_1} = Pr(L_{i1} = l_1)$ because they were all age-eligible to

participate in the first of the two waves. However, Group C individuals were not age-eligible to participate in the first wave, and hence they have the initial latent class weights $\gamma_{il_2} = \Pr(L_{i2} = l_2)$.

The likelihood for Group A looks exactly like the traditional LCTM likelihood, albeit without missing data. The likelihoods for Groups B and C contain data only for wave 1 and wave 2, respectively, with wave-specific initial latent class weights and conditional response probabilities. Finally, the likelihood for Group D pertains to individuals first observed in wave 2, but eligible to have been observed in wave 1. We assume that the latent class specification at wave 1 would have been appropriate for these individuals. Hence, we use $\gamma_{l_1} \pi_{l_2|l_1}$ to establish initial wave 2 latent class weights, rather than just using γ_{l_2} , as for group C. We treat group D wave 1 data as missing at random, and only incorporate the wave 2 data in the likelihood.

2.4 Period Class Weights

One of the primary objectives of our longitudinal model is to observe how initial latent class weights change over time. Time here refers to a period effect, rather than an age effect. However, in the above example, with only two waves, it is awkward to simply compare γ_{l_1} and γ_{l_2} , since many individuals incorporated in the estimation of γ_{l_1} would have been age-eligible to have been observed in earlier waves (had there been earlier waves), whereas all the individuals incorporated in the estimation of γ_{l_2} were only initially age-eligible in wave

2. Furthermore, such a multi-wave comparison focuses only on individuals who are initially age-eligible in a wave (after the first wave), and thus has limited scope relative to the entire range of ages represented. Thus, it is necessary to introduce the idea of period class weights, which are no longer age-specific, but pertain to the calendar time at which observations are made.

Period class weights are a function of estimated initial latent class weights, and are defined as the proportions of observed individuals from a given period that the model assigns to a particular class. Note that a period could equivalently be interpreted as the fixed calendar time for which a given wave is administered. Consider a simple 2-wave example, where $N_1(l_2)$ and $N_2(l_2)$ are counts of individuals observed in period 2 who are subject to initial class weights from periods 1 and 2, respectively (i.e., γ_{l_1} and γ_{l_2}). Thus, $N_1(l_2)$ is the

combined size of Groups A and D, as defined in the previous section, and $N_2(l_2)$ is the size of Group C. Also, let $K+1$ denote an extra, non-latent class for individuals dropping out between periods 1 and 2. Then, we can express period class weights for waves 1 and 2 as follows:

$$\text{Period 1 Class Weight} = \Pr(L_1 = l_1) = \gamma_{l_1}. \quad (6)$$

$$\text{Period 2 Class Weight} = Pr(L_2 = l_2) = \left(\frac{\sum_{l_1=1}^K \gamma_{l_1} \pi_{l_2|l_1}}{1 - Pr(L_2 = K+1)} \times \frac{N_1(l_2)}{N_1(l_2) + N_2(l_2)} \right) + \left(\gamma_{l_2} \times \frac{N_2(l_2)}{N_1(l_2) + N_2(l_2)} \right)$$

$$\text{where } Pr(L_2 = K+1) = \sum_{l_1=1}^K \gamma_{l_1} \pi_{l_2=K+1|l_1}. \quad (7)$$

Period 2 class weights are not only a function of γ_{l_2} , but also of both γ_{l_1} and $\pi_{l_2|l_1}$. For further explanation and mathematical derivation concerning equations (6)–(7), along with a corresponding extension for cases with $T > 2$ waves, see Appendix A.

2.5 Restrictions on Parameter Estimation

For a two-wave LCTM without covariates, restrictions on the parameter set include:

$$\sum_{l_1=1}^K \gamma_{l_1} = 1, \quad \sum_{l_2=1}^K \gamma_{l_2} = 1. \quad (8)$$

$$\sum_{l_2=1}^K \pi_{l_2|l_1=1} = 1, \quad \sum_{l_2=1}^K \pi_{l_2|l_1=2} = 1, \dots, \quad \sum_{l_2=1}^K \pi_{l_2|l_1=K} = 1. \quad (9)$$

Thus, to ensure identifiability, we estimate $\gamma(l_1 = K)$, $\gamma(l_2 = K)$, and $\pi(l_2 = K|l_1 = K)$ as functions of other parameters (for $l_1 = 1, \dots, K$ and $l_2 = 1, \dots, K$).

3 Simulation Study

The purpose of this study is to show that our group-based LCTM are able to consistently recover a set of given (true) model parameters for a simulated model when the estimation process is complete. Then, if successful, this will lend credibility to the performance of the estimators for our actual data example (see Section 6). To emphasize the clarity of the group-based LCTM approach, we keep the structure of the simulated example relatively simple. More specifically, our simulation has 5,000 individuals, 2 waves (and thus, 4 groups), 5 binary response outcome variables, and 3 latent classes, with an additional class for drop-out. Table 1 provides the full set of model parameters used to generate the complete data. The groups are defined as they were in Section 2.3, and the 3 latent classes are ordered from least healthy to most healthy. We initially assume complete data so that all 5,000 individuals are observed both waves.

In generating the initial complete data for wave 1, we had to ensure that conditional independence was maintained (see Section 4.2 for more detail). We then used the given transition probability parameters in Table 1 to allocate a proportion of the total count of each (response pattern, latent class) combination into a particular latent class for the second wave. Once the total sizes of the latent classes were determined for the second wave, we then allocated individuals in each assigned class into 1 of the 32 response patterns using the same conditional response probabilities that were used for wave 1. Finally, although all of the wave 1 data were ordered by response pattern and latent class, and held constant across simulations, the wave 2 data were randomly sorted within each given latent class for each simulation. The complete simulated data set, when run as a traditional LCTM, yielded a recovery of all model parameters to within a very close margin of error (not shown). This was expected, and was done as a prerequisite check to ensure we were ready to expand into the group-based setting.

It was now necessary to arbitrarily determine the proportions of the 5,000 individuals that would be allocated into each group. 2,000 (40%) individuals would retain their complete data across both waves, and thus be placed into Group A, while 1,250 (25%) would have their wave 2 observations removed, and thus be placed into Group B. The remaining 1,750 would have their wave 1 observations removed, with 1,000 (20%) going into Group C, and 750 (15%) going into Group D. We also assumed that the probabilities of individuals being placed into each group were associated with the original latent class assignment from the data. For example, individuals in class 1 (the least healthy class) during wave 1 had the greatest probability of having their wave 2 observations removed (reflecting drop-out), and thus going to Group B, whereas individuals in class 3 during wave 1 had the lowest such probability. Also, individuals who were in class 3 (the most healthy class) during both waves were more likely to have their wave 1 observations removed and to be placed into Group C, which by definition is the youngest such group. After Groups B and C were formed for each simulation, the remaining individuals were randomly assigned to Groups A and D, according to the overall proportions of those groups defined above. This allocation reflects that Group D individuals could have been observed wave 1 but were not present in the sample and assumed to be missing at random during wave 1.

Table 2 provides the revised full set of model parameters, after individuals have been allocated into groups. The latent class weights and transition probabilities in Table 2 are different than they are in Table 1 because some individuals who formerly had complete data are now only observed in one wave (and the first such wave may be wave 2 so that a separate set of initial latent class weights is necessary), and there is now an additional class in the transition probability matrix to reflect drop-out during wave 2. Note that these changes necessarily had the effect of altering both the wave 1 initial latent class weights and the transition probabilities among all latent classes used in the complete data generation.

After the removal process was complete, there were a fixed number of individuals in each group for each latent class combinations across the two waves. These counts would be retained across all simulations, but the specific individuals that were placed into each group were randomized for each simulation. In all, 30 simulations were performed, and a separate set of complete parameter estimates was generated for each simulation. The sample size of

30 simulations was sufficient to show successful parameter recovery. For the first few simulations, multiple sets of starting values were used, all of which led to the same maximum likelihood solution. After the 30 simulations were complete, the average and standard deviation of each individual parameter estimate was recorded to verify parameter recovery. The averages were compared to the actual parameters that had been used to generate the data, and the sample-based standard deviations were compared to the theoretically correct standard deviations. The averages for all 33 estimated parameters were within .015 of their true values (see Table 2 for these differences), with 28 of the 33 averages being within .010 of their true value. The standard deviations also matched up quite well, with 28 of 33 within .008 of the theoretical value. Thus, the simulation study did indeed enhance the credibility of our group-based LCTM estimators so that the conclusions of our upcoming real data example will be enhanced.

4 Covariate Extensions to LCTMs

4.1 Covariates by Model Component

In this section, we incorporate covariates in all three components of an LCTM, using a 2-wave 1-group case for simplicity of exposition. Extensions to more waves and groups are straightforward. Let x_{it} be a covariate for individual i that impacts both initial latent class weights (at time $t=1$) and transition probabilities (at time $t=2$). Let z_{it} be a covariate for individual i for the response probabilities (at both times $t=1$ and 2). When including covariates that affect response probabilities, the interpretation of both the latent class weights and transition probabilities becomes more challenging. This is because the characterization of the latent classes may no longer be constant over the population [44]. Furthermore, due to the increased model complexity from having covariate sets in multiple model components, greater diligence is needed to guard against identification problems. However, according to Huang and Bandeen-Roche [45], the incorporation of such covariates can “adjust for characteristics that determine responses other than underlying classes, hence hopefully improving the accuracy of classifying individuals.”

For illustrative purposes, note that x_{it} and z_{it} represent only single covariates, but that multi-dimensional covariate sets can be included as well. Note also that these covariates may or may not vary with time. If a covariate is time-varying (e.g., an indicator for whether some health condition is present or not), additional checks are necessary to ensure that the latent class structure remains comparable across time; if this fails to be the case, then the interpretation of the transition probabilities becomes problematic.

Due to identifiability constraints and the ability to readily interpret parameter effects, it is prudent to keep the covariates x and z (or covariate sets $\mathbf{x}$ and $\mathbf{z}$) mutually exclusive [8, 46]. We suggest using the same covariate set ($\mathbf{x}$) for both the initial latent class weights and transition probabilities model components because these two components both govern the assignment of latent classes to individuals across time. The marginal distribution of longitudinal outcomes for individual i in the 2-wave 1-group example is:

$$Pr(\mathbf{Y}_i = \mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{\pi}, \mathbf{x}, \mathbf{z}) = \sum_{l_1=1}^K \left[\gamma(x_{1i})_{l_1} \prod_{m=1}^M \rho(z_{1i})_{m|l_1}^{y_{im1}} \left(1 - \rho(z_{1i})_{m|l_1} \right)^{1-y_{im1}} \right] \quad (10)$$

$$\times \sum_{l_2=1}^K \left[\pi(x_{2i})_{l_2|l_1} \prod_{m=1}^M \rho(z_{2i})_{m|l_2}^{y_{im2}} \left(1 - \rho(z_{2i})_{m|l_2} \right)^{1-y_{im2}} \right].$$

We use the canonical (logit) link to model covariates for all three model components. Let $(\boldsymbol{\alpha}, \boldsymbol{\nu}, \boldsymbol{\tau})$ be the intercept and let $(\boldsymbol{\beta}, \boldsymbol{\omega}, \boldsymbol{\kappa})$ be the slope parameters for the $(\boldsymbol{\gamma}, \boldsymbol{\pi}, \boldsymbol{\rho})$ model components, respectively. Then, the equations that link the regression coefficients to the core $(\boldsymbol{\gamma}, \boldsymbol{\pi}, \boldsymbol{\rho})$ model components can be written as:

$$Pr(L_{i1} = l_1 | x_{1i}) = \gamma(x_{1i})_{l_1} = \frac{e^{\alpha_{l_1} + \beta_{l_1} x_{1i}}}{\sum_{k=1}^K e^{\alpha_k + \beta_k x_{1i}}} \quad (11)$$

$$Pr(L_{i2} = l_2 | L_{i1} = l_1, x_{2i}) = \pi(x_{2i})_{l_2|l_1} = \frac{e^{\nu_{l_2|l_1} + \omega_{l_2|l_1} x_{2i}}}{\sum_{k=1}^K e^{\nu_{k|l_1} + \omega_{k|l_1} x_{2i}}} \quad (12)$$

$$Pr(Y_{im} = 1 | L_{it} = l_t, z_{ti}) = \rho(z_{ti})_{m|l_t} = \frac{e^{\tau_{m|l_t} + \kappa_m z_{ti}}}{1 + e^{\tau_{m|l_t} + \kappa_m z_{ti}}}, \text{ where} \quad (13)$$

$$\tau_{m|l_t} = \text{logit}[Pr(Y_m = 1 | L_t = l_t)] \quad (14)$$

$$= \ln \left[\frac{Pr(Y_m = 1 | L_t = l_t)}{1 - Pr(Y_m = 1 | L_t = l_t)} \right].$$

Note that we use the same intercept ($\boldsymbol{\tau}$) and slope ($\boldsymbol{\kappa}$) parameters for both waves ($t = 1$ and 2), which is consistent with the assumption that response probabilities are homogeneous across time.

When covariates are involved, one often makes the additional assumption of no differential item functioning [42]; that is, the value of a covariate does not affect the estimation of conditional response probabilities. In other words, given one's latent class, we assume that the responses and covariates are independent:

$$Pr(Y_{im} = y_m | L_i = l, \mathbf{x}) = Pr(Y_{im} = y_m | L_i = l). \quad (15)$$

For a two-wave LCTM with covariates, we follow the methodology of Bandeen-Roche et.al. [42] in specifying identifiability constraints. Again, we do not have to estimate:

$$\gamma(x_{1i})_K = \frac{e^{\alpha_K + \beta_K x_{1i}}}{\sum_{k=1}^K e^{\alpha_k + \beta_k x_{1i}}}. \quad (16)$$

Thus, to maintain model identifiability, we let $\alpha_K = \beta_K = 0$. We also estimate $\pi(x_{2i})_{K|I_1}$ as functions of other parameters, and let $\nu_{K|I_1} = \omega_{K|I_1} = 0$ for $I_1 = 1, \dots, K$.

Before incorporating a potential covariate into the appropriate covariate set, $\mathbf{x}$ or $\mathbf{z}$, one must understand the difference in interpretation between two types of covariates. Huang and Bandeen-Roche [45] state that covariates affecting the underlying variable seek to “estimate the effects of the independent variables on the conceptual outcome”, whereas covariates affecting the responses attempt to “adjust for characteristics associated with measurement, hence preventing possible misclassification of underlying variable categories.” To date, there has not been a consistent covariate classification scheme employed. Perhaps this is because such classification depends on expert judgment about the anticipated effects [47]. For more supporting detail on covariate-related estimation and classification, see Section 4.5, and also White (2008) [48].

4.2 Diagnostic Checks of Model Assumptions

Model assumptions, as discussed in Sections 2 and 4.1, can be broken into two categories: those that apply to a single point in time, and those that are longitudinal in nature. The two main cross-sectional assumptions are conditional independence and no differential item functioning, for which diagnostics already exist [42]. There is less guidance in the literature about how to verify that the longitudinal assumptions relied upon here are reasonable. These include response probabilities being time-homogeneous, transition probabilities being time-homogeneous, and the number of latent classes remaining constant across waves.

Let’s begin with an examination of conditional independence and no differential item functioning, assuming the simpler latent class modeling framework. We’ll also assume that covariates are only allowed to affect latent classes, as in Dayton and MacReady [49]. Then, the procedure to check for both of these assumptions begins as follows [42]:

1. Fit a latent class regression model to the data, where covariates are only allowed to affect latent classes.
2. Use the fitted parameter estimates to calculate the probabilities of class membership given an individual’s responses and covariates.

3. Randomly assign each individual i into a class $L_i = 1 \dots K$ according to the probabilities from the previous step.

Now, all N individuals in the sample should belong completely to one latent class, and the overall latent class allocations should be approximately the same as the model's estimated latent class weights. At this point, Bandeen-Roche et.al [42] suggest performing tests of mutual independence among each pair of responses within each latent class (i.e., testing for 'conditional independence) or between each (response, covariate) combination within each latent class (i.e., testing for 'no differential item functioning).

However, we suggest the following alternative methodology to check for conditional independence:

1. Within each class of assigned individuals, find $Pr(Y_1 = y_1, \dots, Y_M = y_M | L = l)$.
2. Within each class of assigned individuals, also find $\prod_{m=1}^M Pr(Y_{im} = y_m | L = l)$.
3. Take the difference between the probabilities, as calculated in the previous two steps. The smaller these differences, the greater the likelihood that the assumption of conditional independence is reasonable; if sizeable differences occur, investigate further.

The same idea can be applied when examining the assumption of 'no differential item functioning,' except now one is comparing $Pr(Y_m = y_m | L = l, \mathbf{x})$ and $Pr(Y_m = y_m | L = l)$ for each (response m , covariate x) combination.

For most latent class transition model applications, it is both reasonable and wise to assume that conditional responses probabilities remain time homogeneous. Otherwise, the qualitative nature of the latent classes would change through time, which would make it challenging (or even meaningless) to track shifts in class composition across waves. Furthermore, if one had to estimate separate response probabilities for each wave, especially if covariates were allowed to affect response probabilities, the number of estimated parameters would quickly become unwieldy, and identification problems would become threatening.

Also, it is reasonable to conceive of the transition probabilities varying across waves. In fact, allowing each transition opportunity to have its own, distinct transition probability matrix is a better reflection of reality. However, the increase in model complexity may not be worth the added benefit, due to the increased number of parameters and the potential lack of identifiability. Still, because many applications of latent class transition models do assume time-varying transition probabilities, we suggest that the assumption of time-homogeneous transition probabilities be checked by decomposing a T wave model into $T-1$ 2-wave models, each of which can be run separately so that transition probabilities can be compared. If there are only a few significant mismatches, these transitions could be allowed to vary, subject to the constraint that all other transition probabilities remain the same. One can also compromise between the two extremes of using one constant transition probability matrix and using separate matrices for each transition opportunity by using a small number of separate matrices (e.g., 2 or 3), where transition probabilities change every few waves.

Finally, in a latent class transition model, should the optimal number of latent classes vary across waves? The answer to this question depends on both the study design and the nature of the latent variable. For applications with a rolling enrollment design, it makes sense for the number of latent classes to remain fixed across waves. To check whether this assumption is reasonable given the data, one can temporarily assume a cross-sectional perspective, and examine each wave of data separately, and verify that the optimal number of latent classes remains fixed through time.

5 Data

5.1 NLTCS Group-Based LCTM Analysis

Elderly (age 65+) Americans often require long-term care. To better allocate resources to those subgroups that need, or may potentially need assistance, it is prudent to understand the dynamics of chronic disability across time and age. For example, if one could identify common patterns of disability progression, then health care and its associated cost could be directed to at-risk groups early on. This would induce a cost saving relative to what might be spent later when such individuals are more seriously disabled and/or close to death.

We use the NLTCS (National Long Term Care Survey) to examine chronic disability in the U.S. aged 65-and-over population. Here, individuals are sampled from Medicare enrollment files, which cover over 95% of the U.S. elderly (65+) population [50]. One strength of the NLTCS is its extremely low loss-to-follow-up rates [51]. The NLTCS began in 1982, and was repeated every 5 years from 1984 to 2004. The National Institute on Aging, along with support from Duke University and the Census Bureau, have sponsored and helped administer the NLTCS.

Activities of Daily Living (ADLs) and Instrumental Activities of Daily Living (IADLs) are typically used to assess chronic disability. ADLs represent essential everyday physical functioning, and IADLs represent various household and lifestyle-oriented tasks that tend to be more cognitive in nature. Here, we focus on ADLs only, for which the NLTCS measures 6 ADLs each wave. Binary outcomes (1=presence or 0=absence of disability) for ADL measures are determined using a sequence of triggering questions [51]. In Table 3, we provide the complete set of ADLs from least to most common, based on unweighted relative frequencies, pooled over the community detailed portion of the NLTCS 1982–2004.

There was also a screener survey, for which 49,274 unique individuals in the NLTCS sample received at least one of the 6 survey waves. After 1982, about 5,000 additional seniors were given the screener survey each wave, both in order to keep the age composition of the survey relatively consistent across waves and to replace those that had left the survey. Among the 49,274 individuals that received the initial screener survey, 18,224 unique individuals were (at some point) classified as ‘chronically disabled,’ which meant that they had a problem performing at least one basic or instrumental ADL that was expected to last at least 90 days. These 18,224 individuals were then also administered a detailed survey, not only in the initial wave for which they were deemed ‘chronically disabled,’ but also in each subsequent wave for which they were alive (even if no longer meeting the definition for ‘chronic disability’). We focus on these individuals who are/were ‘chronically disabled’ for our

subsequent analyses. We use the detailed survey data to determine both the ADL history and covariates for these individuals, rather than the screener survey, which was not always administered to returning survey participants and thus lacks important covariates. Thus, our substantive conclusions only apply to the U.S. elderly population of individuals who (at some point) have been ‘chronically disabled,’ rather than the U.S. elderly population at large.

With the exception of the healthy and oldest-old supplements that were introduced in the later waves of the survey (see Section 5.2), only those individuals who were classified as chronically disabled by the screen interview were eligible to participate in the more detailed interview. For more information on the relationships between the screen and detailed interview, and the various definitions of chronic disability in these two surveys, see Erosheva and White (2010). [51]

5.2 Group Selection

Of the 18,224 individuals who participated in the detailed NLTCs, observation patterns for 268 individuals fall across non-consecutive waves, for which group-based LCTMs are not appropriate. We remove these 268 individuals, leaving 17,956 individuals in the data. In addition, because the gap between the first (1982) and the second (1984) wave was only 2 years instead of 5, we only retain the final 5 waves (1984, 1989, 1994, 1999, and 2004) in order to maintain evenly spaced intervals across both groups and age brackets. Eliminating 1982 data, we have 16,201 individuals left.

Finally, to make valid conclusions about the chronically disabled elderly population, we remove individuals who were in both the healthy and oldest-old supplements, as they are not representative of the population of interest. We are left with data on 13,710 individuals across 5 waves. We assign these individuals to 35 groups, where each group is defined as a collection of individuals that are born within a specified period, enter the survey at the same wave, and leave the survey at the same wave (see Table 4).

5.3 Latent Class Characterization

To choose the optimal number of latent classes to model, one approach is to begin with goodness-of-fit procedures for models without covariates, and then rely on the ‘marginalization procedure’ [42]. This states that the number of latent classes to include for models without covariates can be fixed at the level determined for corresponding models without covariates. Furthermore, it is advisable to begin with a single wave of data, and later perform checks for consistency across other waves of the survey.

After comparing observed and expected ADL pattern counts, using a combination of goodness-of-fit criteria including AIC [52], BIC [53], and relative entropy [54], we determined that the appropriate number of latent classes (K) was between 4 and 6, plus an additional class for drop-out. Note that a bootstrapped likelihood ratio test [55] has recently been shown to also be helpful in determining the optimal number of latent classes to model.

Table 5 shows the AIC and BIC values for six fitted models, one for each of $K = 2, \dots, 7$.

Table 5 shows that the minimum (best) AIC occurs when $K=6$, and the minimum (best) BIC

occurs when $K=4$. Entropy is the degree to which latent classes are clearly distinguishable by the data (higher=better). Interestingly, the best fitting models by AIC/BIC ($K=4, 5, 6$) actually have the greatest amount of fuzziness with respect to the entropy measure. This indicates that the best fitting model will have some classes that are somewhat similar to each other.

Following the principle of parsimony, we narrowed down our final choice for K to be either 4 or 5 for our Group-Based LCTM analysis. However, the qualitative characterization of each latent class was more clearly defined for $K=4$ than for $K=5$. In the end, we used 4 stochastic classes for living disability states and 1 extra, deterministic class for drop-out. Based on the fit of this 5-class model, the four latent classes can be characterized as follows: Class 1 is the most disabled, Class 4 is the least disabled, and the remaining Classes (2 and 3) differ mainly on the types of associated functional limitations. More specifically, individuals with impaired-mobility ADLs (i.e., getting in and out of bed and inside mobility) are much more likely to be in Class 2 than in Class 3.

To ensure that K does not vary from wave to wave, we used the same methodology as above for all other waves. Not only did AIC and BIC identify the same number of latent classes to model each wave, but the structure of the latent classes, as measured by the conditional response probabilities, remained consistent across waves. However, we also found that K can vary depending on covariate structure, especially if the core assumptions of LCTMs (i.e., conditional independence, no differential item functioning) are not satisfied.

5.4 Diagnostic Checks of Assumptions for NLTCs Data

We diagnosed the model assumptions discussed in Sections 2 and 4.1 using the 1999 NLTCs data, and found that the assumption of conditional independence was quite reasonable. For all but one class, all 64 ($= 2^6$) possible response sequences had differences in calculated probabilities of less than .01, and for the remaining class, 60 of the 64 possible responses sequences had differences in calculated probabilities of less than .02 (with the remaining 4 differences being between .03 and .06). However, there was considerable violation of the assumption of no differential item functioning, especially with respect to the covariate that indicates the presence of arthritis. In other words, even after controlling for latent class, the presence of arthritis may affect the response probabilities. This is evidence in favor of allowing covariates like arthritis to affect the response probabilities, as will be shown with an example in Section 6.2. Similar conclusions with respect to the diagnostics were reached when looking at other waves (1984, 1989, 1994, 2004) in isolation. Due to space constraints, we do not show further numerical results here.

Next, let's briefly consider the longitudinal assumptions. Since the NLTCs has a rolling enrollment design, and attempts to study trends in chronic disability among the entire elderly populations over time, it makes sense for the number of latent classes to remain fixed across waves. If instead, the NLTCs had followed a single group of commonly-aged individuals as they progressed through time (e.g., from age 65 in 1984 to age 85 in 2004), it would have been more appropriate to have the number of latent classes change across waves as to reflect changes in population heterogeneity. For example, at age 65, there would be less heterogeneity in chronic disability patterns than at age 75 (or especially at age 85). After

comparing the response probabilities, transition probabilities, and optimal number of latent classes across all waves of the NLTCs data, it was determined that the assumptions of time-homogeneous response probabilities, time-homogeneous transition probabilities, and a fixed number of latent classes across waves were all reasonable.

5.5 Computing

All models were estimated using an EM Algorithm, implemented in SAS with PROC IML (Interactive Matrix Language), which allows the conversion of multi-dimensional arrays into two-dimensional matrices. When covariates were incorporated into LCTM models, regression parameters were also estimated using the EM algorithm, but in conjunction with a Newton-Raphson step. Most models converged within about 200 iterations.

It is well known that the estimation of latent class models may be subject to multimodality [63]. We used multiple sets of starting values for the entire estimated parameter set for each model run. For most model runs, the final sets of estimated parameters were quite similar across distinct sets of starting values, and produced log-likelihoods in close proximity.

To estimate standard errors for each parameter using the LCTM framework, we assumed the number of latent classes was fixed and known, and calculated the Hessian matrix of second partial derivatives with respect to the expected log-likelihood [56]. We then inverted the observed information matrix [56] (the negative of the Hessian matrix), to form the conditional variance-covariance matrix of parameter estimates, evaluated at the parameter estimates [56]. The standard error of each estimated parameter is then found by taking the square root of the appropriate cell in the main diagonal of the conditional variance-covariance matrix.

To obtain standard errors for constrained parameters like γ_{l_K} , we compute:

$$Var(\gamma_{l_K}) = Var\left(1 - \sum_{l_1=1}^{K-1} \gamma_{l_1}\right) = \sum_{l_1=1}^{K-1} Var(\gamma_{l_1}) - \sum_{l_{1a}=1}^{K-1} \sum_{l_{1b}=1}^{K-1} Cov(\gamma_{l_{1a}}, \gamma_{l_{1b}}), a \neq b \quad (17)$$

$$s.e.(\gamma_{l_K}) = \sqrt{Var(\gamma_{l_K})} \quad (18)$$

Here, the covariance terms come from the off-diagonal elements of the estimated variance-covariance matrix. A similar computation is used to derive standard errors for the transition probabilities (π).

In the context of group-based LCTMs, there is a more fundamental issue involved in estimating standard errors. When forming the expected log-likelihood upon execution of the EM algorithm, it is necessary to rely on a set of individual-specific parameters, calculated

over all K^T possible latent class combinations across T waves. For example, if $T=1$, then these individual-specific parameters can be written as:

$$\delta_{il} = Pr(L = l | Y_i) = \frac{Pr(L = l, Y_i)}{Pr(Y_i)} = \frac{\gamma_l R}{\sum_{j=1}^K \gamma_j R}, \text{ where} \quad (19)$$

$$R = \prod_{m=1}^M \rho_{m|l}^{y_{im}} (1 - \rho_{m|l})^{1 - y_{im}}. \quad (20)$$

Note that δ_{il} is a function of both γ and ρ , the latter contained within R . Furthermore, if $T=2$, then this parameter set can be written as:

$$\begin{aligned} \delta_{i(l_1, l_2)} &= Pr(L_1 = l_1, L_2 = l_2 | Y_{1i}, Y_{2i}) = \frac{Pr(L_1 = l_1, L_2 = l_2, Y_{1i}, Y_{2i})}{Pr(Y_{1i}, Y_{2i})} \quad (21) \\ &= \frac{\gamma_{l_1} R_1 \pi_{l_2 | l_1}}{\sum_{j_1=1}^K \gamma_{j_1} R_1 \sum_{j_2=1}^K \pi_{j_2 | j_1} R_2}, \text{ where} \end{aligned}$$

$$R_t = \prod_{m=1}^M \rho_{m|l_t}^{y_{imt}} (1 - \rho_{m|l_t})^{1 - y_{imt}}, \text{ for } t = 1, 2. \quad (22)$$

Now, $\delta_{i(l_1, l_2)}$ is a function of γ , π , and ρ , the latter contained within both R_1 and R_2 . For more discussion on the interpretation and treatment of δ_{il} and $\delta_{i(l_1, l_2)}$ when estimating standard errors in group-based LCTM, see Appendix B.

6 Results

6.1 The Progression of Disability Decline: 1984–2004

Researchers disagree on whether or not chronic disability prevalence in the U.S. has continuously declined from 1984–2004, and on the rate of such decline. It is often stated that chronic disability prevalence declined throughout the 1980s and 1990s [31, 38, 39, 40, 57, 58, 59]. However, some researchers did not expect this trend to continue into the 21st century [60, 61], while by 2004, other researchers have indicated that this trend did indeed subside [38, 39]. We apply group-based LCTMs to examine dynamics of disability throughout the period 1984–2004 to further explore whether the rate of decline has been accelerating or leveling off in recent waves.

The 5-wave, LCTM marginal distribution for Group A individuals can be written as:

$$Pr(\mathbf{Y}_i = \mathbf{y}; \boldsymbol{\gamma}, \boldsymbol{\rho}, \boldsymbol{\pi}) = \sum_{l_1=1}^K \gamma_{l_1} R_1 \sum_{l_2=1}^K \pi_{l_2|l_1} R_2 \times \sum_{l_3=1}^K \pi_{l_3|l_2} R_3 \sum_{l_4=1}^K \pi_{l_4|l_3} R_4 \sum_{l_5=1}^K \pi_{l_5|l_4} R_5, \quad (23)$$

where

$$R_t = \prod_{m=1}^M \rho_{m|l_t}^{y_{imt}} \left(1 - \rho_{m|l_t}\right)^{1-y_{imt}}, \text{ for } t = 1, 2, 3, 4, 5. \quad (24)$$

The corresponding marginal distribution across all 35 groups can be written down following our example in Section 2.3.

Table 6 provides parameter estimates for the full (35-group) Group-Based LCTM without covariates. It is interesting to compare the differential transition rates into the drop-out class ($\pi_{l_t=5|l_{t-1}}$ for $l_{t-1} = 1, \dots, 4$, and $(t-1, t) = (1, 2), \dots, (4, 5)$) from each of the 4 living disability classes. For example, someone in class 1, the most disabled class, is more than 3 times more likely to drop out within the next 5 years than someone in class 4, the most healthy class ($\pi_{5|1}=.661$ versus $\pi_{5|4}=.191$).

We convert initial latent class weights ($\gamma_{l_1}, \dots, \gamma_{l_5}$) from Table 6 into period class weights, following the process discussed in Section 2.4. The converted period weights are given in Table 7, which provide a succinct representation of disability trends over the past 20 years. We observe that from 1984–1999, there was a steady decline in disability prevalence, as reflected in declining proportions of the 3 most disabled classes (1, 2, and 3), and in increasing proportions of the healthiest class (4). Interestingly, the changes in period class weights were almost exactly linear during this time frame.

However, these trends did not continue as such from 1999 to 2004; the declines in disability observed in the 1980s and 1990s appear to have at least stabilized in the early 2000s. In fact, there even appears to have been a slight reversal. This is most evident when observing the (healthy) class 4 weight declining from .686 in 1999 to .676 in 2004. This is contrary to what some NLTCs researchers have often claimed [31, 35, 57].

6.2 The Effect of Arthritis on Functional Limitations

For our next NLTCs-based example, we consider the group-based LCTM with covariates. Here, we use sex as a covariate that affects initial latent class weights and arthritis as a covariate that affects conditional response probabilities. We choose arthritis as a covariate of interest in our study of the dynamics of disability since trends in arthritis quite closely mirror trends in ADL limitations [17, 59, 62]. Thus, if we were to ignore arthritis as a covariate in

our model, this would likely cause a violation of the no differential item functioning assumption [42].

We apply group-based LCTMs to examine if arthritics have a different disability class profile than non-arthritics, and also explore the types of disability with which arthritis is associated the most. Then, the 5-wave, LCTM marginal distribution for Group-A individuals is:

$$Pr(Y_i = y; \alpha, \beta, \tau, \kappa, \pi, \mathbf{x}, \mathbf{z}) = \sum_{l_1=1}^K \frac{e^{\alpha_{l_1} + \beta_{l_1} x_{1i}}}{\sum_{k=1}^K e^{\alpha_k + \beta_k x_{1i}}} R_1(\tau, \kappa) \times \sum_{l_2=1}^K \pi_{l_2|l_1} R_2(\tau, \kappa) \quad (25)$$

$$\times \sum_{l_3=1}^K \pi_{l_3|l_2} R_3(\tau, \kappa) \times \sum_{l_4=1}^K \pi_{l_4|l_3} R_4(\tau, \kappa) \sum_{l_5=1}^K \pi_{l_5|l_4} R_5(\tau, \kappa), \text{ where}$$

$$R_t(\tau, \kappa) = \prod_{m=1}^M \left[\frac{e^{\tau_{m|l_t} + \kappa_{m|z_{ti}}}}{1 + e^{\tau_{m|l_t} + \kappa_{m|z_{ti}}}} \right]^{y_{imt}} \left[1 - \frac{e^{\tau_{m|l_t} + \kappa_{m|z_{ti}}}}{1 + e^{\tau_{m|l_t} + \kappa_{m|z_{ti}}}} \right]^{1 - y_{imt}}, \text{ for } t = 1, 2, 3, 4, 5. \quad (26)$$

Table 8 represents regression coefficients pertaining to arthritis (τ, κ). However, since it is difficult to make conclusions from the regression coefficients in Table 8 directly, we convert them back into conditional response probabilities (ρ) to distinguish between individuals with and without arthritis. For example, given that non-arthritics are coded as 0 and arthritics are coded as 1, we have for the bathing ADL ($m=6$):

$$Pr(Y_{m=6}|l=4, arthritis=no) = .099 = \frac{e^{-2.20}}{1 + e^{-2.20}}$$

$$Pr(Y_{m=6}|l=4, arthritis=yes) = .137 = \frac{e^{-2.20 + .360}}{1 + e^{-2.20 + .360}}$$

We provide a full set of conditional response probabilities in Table 9. Note that there is a considerable amount of differential item functioning with respect to arthritis. Differences are highest for the inside mobility ADL ($m=5$), but are also high for toileting ($m=3$), getting in and out of bed ($m=4$) and bathing ($m=6$). The two most difficult ADLs of eating and dressing are less affected by arthritis. Interestingly, those without arthritis have slightly higher ADL prevalence on eating than those with arthritis. This indicates that arthritis may be more associated with mild-to-moderate disability rather than severe disability. Note that including covariates did not change conclusions with respect to disability declines; that is,

initial and period class weights (not shown) were similar to those presented for models without covariates (see Tables 6 and 7).

6.3 Summary of NLTCs group-based LCTM Analyses

To summarize, we have employed group-based LCTM to analyze NLTCs ADL data from 1984–2004. These analyses are longitudinal, as they follow individuals across several waves while incorporating transitions among classes. This is in contrast to the aggregate, population-based, cross-sectional analyses completed by Manton et.al. [31, 32] for which the disability and gerontology communities have often relied. Furthermore, our models are able to accommodate any subset of consecutive observation times among all five survey waves observed.

First, we used group-based LCTM to show that there was a steady decline in chronic disability prevalence from 1984–1999, but that this trend reversed from 1999–2004. This conflicts with the mainstream opinion [31], which states that disability declined throughout the whole period, 1984–2004, and did so at an accelerating rate.

Second, we applied group-based LCTM including arthritis and gender as covariates, and found that arthritics have a much different chronic disability profile than non-arthritics. Specifically, arthritics have higher incidence of ADLs that are associated with mild-to-moderate disability like inside mobility, toileting, getting in and out of bed, and bathing. However, arthritics have lower incidence of ADLs that are associated with severe disability like eating and dressing.

7 Discussion

Latent Class Transition Models (LCTM) offer a way to model change in a latent construct over time based on panel survey data. LCTMs have been around since the early 1990s, most typically applied in settings where individuals are observed through all waves (or when all start at the same time, but some drop-out before the last possible observation time) [6, 7]. However, for surveys with rolling enrollment, standard LCTM analyses are not readily applicable and require adjustments for the various observation patterns that can occur in a survey.

Group-Based LCTM, developed in this paper, enable researchers to take into account any subset of consecutive, evenly spaced observation times. Note that pattern mixture models accomplish the same goal, but with group-based LCTM, we do not need to estimate unique sets of parameters for each observation pattern. Furthermore, unlike the traditional LCTM, our group-based extension incorporates ‘period weights’ that allow for the comparison of calendar-based class weights across each observation time in the survey. These models can be especially useful when applied to longitudinal survey data with long assessment intervals. There is some evidence that, in the context of morbidity or disability, the progression occurs in stages rather than being continually degenerative [15].

One limitation of our NLTCs analysis, when compared to other analyses of NLTCs data, is that we do not consider Instrumental Activities of Daily Living (IADLs), which tend to be

more cognitive in nature compared to ADLs. Our attempt to incorporate IADLs resulted in a very high number of latent classes. This is consistent with earlier work on maximum likelihood estimation in latent class models with the NLTCs data, which indicated that more than 20 latent classes are necessary to adequately describe ADLs and IADLs together [63]. As the number of latent classes increases beyond reasonable levels, the qualitative nature of each latent class became blurred and non-sensical. Note that more complex models, such as those employed by Erosheva, Fienberg, and Joutard [64], are needed to incorporate both ADLs and IADLs simultaneously.

Another limitation of our group-based LCTM models is the requirement that observation waves be equally spaced. Thus, we did not incorporate the 1982 wave, which was only two years before the next wave in 1984. Finally, as previously mentioned, when calculating standard errors for our parameters, we treat the posterior probabilities of class membership as given, without accounting for uncertainty in those memberships. Thus, our estimated standard errors are slightly lower than they would have been otherwise. Note that a close companion to LCTM is Hidden Markov Models [65], and LCTM estimation procedures might borrow from the relatively more efficient algorithms used for Hidden Markov Models [66, 67]. Likewise, Hidden Markov Models could be extended to encompass the group-based LCTM methodology introduced here.

Now that the basic groundwork has been established for Group-Based LCTM, a natural methodological extension would be to relax some of the restrictive assumptions such as equal spacing of the waves and time homogeneity for both conditional response probabilities and transition probabilities. Diagnostics could also be formed to verify that the first-order Markov property, on which the LCTM framework is based, is appropriate. Furthermore, with respect to Group-Based LCTM with covariates, model identifiability conditions have yet to be developed. The Group-Based LCTM framework could also be extended so that the number of latent classes is treated as random, which would make the results of any application less dependent on choosing a single, optimal number of classes.

We have also assumed that the pattern of missing data was monotone, which means that if a subject has missing data on a response or covariate, it continues to have missing data on that same variable in all future waves [68]. Most applications using joint models for the outcome data and the missingness process assume monotone missing data, so that the drop-out process is irreversible [69]. If re-entry was permissible, the number of observation patterns in a survey wave would quickly become unwieldy (which would result in many patterns having an insufficient number of observations), and that could lead to several model identifiability problems [68, 70]. See Tsonaka et.al (2010) for an approach that addresses non-monotonic, or intermittently missing, data [71].

Analyzing NLTCs data up to 2004, our results indicate that the declines in chronic disability among the U.S. elderly population in the 1980s and 1990s may not have been sustained after 1999. This finding has important ramifications in both the health care and insurance industries, especially with respect to long-term care. As the baby-boom generation is set to retire, and longevity continues to expand, the number of individuals qualifying for long-term care will increase substantially over the next few decades. If the severity of disability among

these individuals is also increasing, then additional pressure will be placed on our already constrained health care system.

Appendices

8

8 8.1 Appendix A:: More on Period Weights Derivations

For individuals subject to initial latent class weights (given by γ_{l_1}), the first ratio in the

period 2 class weight given in equation (7) (i.e. $\frac{\sum_{l_1=1}^K \gamma_{l_1} \pi_{l_2|l_1}}{1 - \Pr(L_2 = K+1)}$) is derived as follows:

$$\begin{aligned} \Pr(L_2 = l_2 | L_2 < K+1) &= \frac{\Pr(L_2 = l_2, L_2 < K+1)}{\Pr(L_2 < K+1)} = \frac{\sum_{l_1=1}^K \Pr(L_1 = l_1, L_2 = l_2)}{1 - \Pr(L_2 = K+1)} \\ &= \frac{\sum_{l_1=1}^K \Pr(L_1 = l_1) \Pr(L_2 = l_2 | L_1 = l_1)}{1 - \Pr(L_2 = K+1)} = \frac{\sum_{l_1=1}^K \gamma_{l_1} \pi_{l_2|l_1}}{1 - \Pr(L_2 = K+1)}. \end{aligned}$$

For cases with $T > 2$ waves, the numerator in equation (7) can be extended by incorporating a chain of transition probabilities to sum over the previous T-1 class indices. For example, for 3 waves, we can express the period 1, 2, and 3 weights as follows:

$$\text{Period 1 Weight} = \Pr(L_1 = l_1) = \gamma_{l_1}. \quad (27)$$

$$\text{Period 2 Weight} = \Pr(L_2 = l_2) = \frac{\sum_{l_1=1}^K \gamma_{l_1} \pi_{l_2|l_1}}{1 - \Pr(L_2 = K+1)} \times \frac{N_1(l_2)}{N_1(l_2) + N_2(l_2)} + \gamma_{l_2} \times \frac{N_2(l_2)}{N_1(l_2) + N_2(l_2)},$$

$$\text{where } \Pr(L_2 = K+1) = \sum_{l_1=1}^K \gamma_{l_1} \pi_{l_2=K+1|l_1}. \quad (28)$$

$$\begin{aligned} \text{Period 3 Weight} = \Pr(L_3 = l_3) &= \frac{\sum_{l_1=1}^K \sum_{l_2=1}^K \gamma_{l_1} \pi_{l_2|l_1} \pi_{l_3|l_2}}{1 - \Pr(L_3 = K+1)} \times \frac{N_1(l_3)}{N_1(l_3) + N_2(l_3) + N_3(l_3)} \\ &+ \frac{\sum_{l_2=1}^K \gamma_{l_2} \pi_{l_3|l_2}}{1 - \Pr(L_3 = K+1)} \times \frac{N_2(l_3)}{N_1(l_3) + N_2(l_3) + N_3(l_3)} + \gamma_{l_3} \times \frac{N_3(l_3)}{N_1(l_3) + N_2(l_3) + N_3(l_3)}, \end{aligned}$$

$$\text{where } Pr^*(L_3 = K + 1) = \sum_{l_1=1}^K \sum_{l_2=1}^{K+1} \gamma_{l_1} \pi_{l_2|l_1} \pi_{l_3=K+1|l_2},$$

$$\text{and } Pr(L_3 = K + 1) = \sum_{l_2=1}^K \gamma_{l_2} \pi_{l_3=K+1|l_2}. \quad (29)$$

Note that when calculating $Pr^*(L_3 = K + 1)$ for the period 3 weight, the probability $Pr^*(L_3 = K + 1)$ must encompass both individuals who drop-out between periods 2 and 3, as well as individuals who drop-out between periods 1 and 2. Thus, for $Pr^*(L_3 = K+1)$ above, the summation for l_2 goes from 1 to $K + 1$ rather than from 1 to K .

8 8.2 Appendix B:: More on Standard Error Estimation

Qualitatively, parameter δ_{il} is the conditional probability of membership in latent class l for individual i , given the estimated parameter values [44]. Thus, if these estimated parameter values are taken as given, then δ_{il} is a constant. However, if these estimates are considered with error, then δ_{il} is also subject to error, and is in effect, treated like an estimated parameter, or a function of estimated parameters [42]. The main benefit of the approach that treats δ_{il} as a constant is the relative ease in computation, both with respect to derivative calculations and coding. However, strictly speaking, such an approach is improper as it systematically underestimates standard errors.

In 2002, Lystig and Hughes [72] provided methodology to compute asymptotic standard errors, using exact computation of the observed information matrix. Although their approach was given in the context of HMMs, an extension to LCTMs is possible since the likelihoods are equivalent. More importantly, the HMM estimation approach of Lystig and Hughes [72] is more accurate, as it accounts for the extra variability inherent within the δ_{il} parameters. When the LCTM method is adjusted to treat δ_{il} as a function of LCTM estimated parameters rather than as a constant, the two methods are numerically equivalent. However, such an approach is numerically cumbersome and awkward, especially in a 5-wave, 35-group LCTM context.

Implementing the proper calculation for some lower-dimensional cases, we found that differences in estimated standard errors between the two methods of incorporating δ_{il} did not appear to be important for substantive conclusions in our analysis. Thus, when estimating standard errors for more complex multi-wave group-based LCTMs, we used the method that treats the δ_{il} quantities as constants rather than as functions of all the core estimated parameters, since this was much more computationally manageable, and only slightly less accurate.

References

- [1]. Baltagi BH, Song SH. Unbalanced panel data: A survey. *Statistical Papers* 2006; 47:493–523.

- [2]. Bartolucci F Likelihood Inference for a Class of Latent Markov Models under Linear Hypotheses on the Transition Probabilities. *Journal of the Royal Statistics Society: Series B* 2006; 68-2:155–178.
- [3]. Bartolucci F, Nigro V. Maximum Likelihood Estimation of an Extended Latent Markov Model for Clustered Binary Panel Data. *Computational Statistics & Data Analysis* 2007; 51:3470–3483.
- [4]. Bockenholt U, Langeheine R. Latent Change in Recurrent Choice Data. *Psychometrika* 1996; 61(2):285–301.
- [5]. Chung H, Park YS, Lanza ST. Latent Transition Analysis with Covariates: Pubertal Timing and Substance Use Behaviours in Adolescent Females. *Statistics in Medicine* 2005; 24:2895–2910. [PubMed: 16134129]
- [6]. Collins LM, Wugalter SE. Latent Class Models for Stage-Sequential Dynamic Latent Variables. *Multivariate Behavioral Research* 1992; 27:131–157.
- [7]. Graham JW, Collins LM, Wugalter SE, Chung NK, Hansen WB. Modeling Transitions in Latent Stage-sequential Processes: A Substance Use Prevention Example. *Journal of Consulting and Clinical Psychology* 1991; 59:48–57. [PubMed: 2002142]
- [8]. Huang GH. Selecting the Number of Classes under Latent Class Regression: A Factor Analytic Analogue. *Psychometrika* 2005; 70:325–345.
- [9]. Kim JS, Bockenholt U. Modeling Stage-sequential Change in Ordered Categorical Responses. *Psychological Methods* 2000; 5:380–400. [PubMed: 11004875]
- [10]. Miglioretti DL. Latent Transition Regression for Mixed Outcomes. *Biometrics* 2003; 59:710–720. [PubMed: 14601773]
- [11]. Pfeiffermann D, Skinner C, Humphreys K. The Estimation of Gross Flows in the Presence of Measurement Error Using Auxiliary Variables. *Journal of the Royal Statistics Society: Series A* 1998; 161:13–32.
- [12]. Reboussin BA, Reboussin DM, Liang KY, Anthony JC. Latent Transition Modeling of Progression of Health-Risk Behavior. *Multivariate Behavioral Research* 1998; 33(4):457–478. [PubMed: 26753825]
- [13]. Diggle P, Kenward MG. Informative Drop-out in Longitudinal Data Analysis. *Applied Statistics* 1994; 43:49–93.
- [14]. Anderson RT, James MK, Miller ME, Worley AS, Longino CF. The Timing of Change: Patterns in Transitions in Functional Status among Elderly Persons. *The Journals of Gerontology: Social Sciences* 1998; 53B:S17–S27.
- [15]. Ferruci L, Guralnik JM, Simonsick E, Salive ME, Corti C, Langlois J. Progressive versus Catastrophic Disability: A Longitudinal View of the Disablement Process. *The Journals of Gerontology: Medical Sciences* 1996; 51A:M123–M130.
- [16]. Hardy SE, Allore HG, Guo Z, Dubin JA, Gill TM. The Effect of Prior Disability History on Subsequent Functional Transitions. *The Journals of Gerontology: Medical Sciences* 2006; 61A: 272–277.
- [17]. Jagger C, Matthews R, Matthews F, Robinson T, Robine J, Brayne C. The Burden of Diseases on Disability-Free Life Expectancy in Later Life. *The Journals of Gerontology: Medical Sciences* 2007; 62A:408–414.
- [18]. National Long-Term Care Survey. National Institute on Aging, Duke University 1982–2004.
- [19]. National Youth Survey - Family Study. Institute of Behavioral Science, University of Colorado 1976–2004.
- [20]. Denver Youth Survey. Office of Juvenile Justice and Delinquency Prevention 1987–1999.
- [21]. Rubin DB. Inference and Missing Data. *Biometrika* 1976; 63: 581–592.
- [22]. Wu MC, Carroll RJ. Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modelling the Censoring Process. *Biometrics* 1988; 44:175–188.
- [23]. Little RJA. Modelling the Drop-out Mechanism in Repeated-Measure Studies. *Journal of the American Statistical Association: Review Paper* 1995; 90:1112–1121.
- [24]. Conaway M Non-ignorable Non-response Models for Time-ordered Categorical Variables. *Applied Statistics* 1993; 42:105–115.

- [25]. Stasny EA. Some Markov-chain Models for Non-response in Estimating Gross Labor Force Flows. *Journal of Official Statistics* 1987; 3:359–373.
- [26]. Akacha M, Hutton JL. Modelling the rate of change in a longitudinal study with missing data, adjusting for contact attempts. *Statistics in Medicine* 2011; 30:1072–1089. [PubMed: 21284012]
- [27]. Little RJA, Rubin DB. *Statistical Analysis with Missing Data* (2nd edn). Wiley: New York, 2002.
- [28]. Wu MC, Bailey KR. Analysing Changes in the Presence of Informative Right Censoring Caused by Death and Withdrawal, *Statistics in Medicine* 1988; 7:337–346. [PubMed: 3281208]
- [29]. Follmann D, Wu M. An Approximate Generalized Linear Model with Random Effects for Informative Missing Data. *Biometrics* 1995; 51:151–168. [PubMed: 7766771]
- [30]. Hogan JW, Roy J, Korkontzelou C. Tutorial in Biostatistics: Handling drop-out in longitudinal studies. *Statistics in Medicine* 2004; 23:1455–1497. [PubMed: 15116353]
- [31]. Manton KG, Gu X, Lamb VL. Change in Chronic Disability from 1982 to 2004/2005 as Measured by Long-term Changes in Function and Health in the U.S. Elderly Population. *Proceedings of the National Academy of Sciences* 2006; 103:18374–18379.
- [32]. Manton KG, Stallard E, Corder LS. The Dynamics of Dimensions of Age-Related Disability 1982 to 1994 in the U.S. Elderly Population. *Journal of Gerontology: Biological Sciences* 1998; 53A:B59–B70.
- [33]. Connor J Multivariate Mixture Models to Describe Longitudinal Patterns of Disability in American Seniors. Ph.D. Thesis, Carnegie Mellon University 2006.
- [34]. Nagin DS. Analyzing Developmental Trajectories: A Semi-parametric, Group-Based Approach. *Psychological Methods* 1999; 4:139–157.
- [35]. Manton KG, Lamb VL, Gu X. Medicare Cost Effects of Recent U.S. Disability Trends in the Elderly. *Journal of Aging and Health* 2007; 19:359–381. [PubMed: 17496239]
- [36]. Stallard E Trajectories of Disability and Mortality among the U.S. Elderly Population: Evidence from the 1984–1999 NLTCs Presented at the Living to 100 and Beyond Symposium, Center for Demographic Studies 2005.
- [37]. Manrique D Longitudinal Mixed Membership Models With Applications to Disability Survey Data. Ph.D. Thesis, Carnegie Mellon University 2010.
- [38]. Spillman BC. Changes in Elderly Disability Rates and the Implications for Health Care Utilization and Cost. *Millbank Quarterly* 2004; 82:157–194.
- [39]. Freedman VA, Crimmins E, Schoeni RF, Spillman BC, Aykan H, Kramarow E. Resolving Inconsistencies in Trends in Old-Age Disability: Report from a Technical Working Group. *Demography* 2004; 41:417–441. [PubMed: 15461008]
- [40]. Wolf DA, Hunt K, Knickman J. Perspectives on the Recent Decline in Disability at Older Ages. *The Millbank Quarterly* 2005; 83:365–395.
- [41]. Lazarsfeld PF, Henry NW. *Latent Structure Analysis*, Houghton Mifflin: Boston 1968.
- [42]. Bandeen-Roche K, Miglioretti DL, Zeger SL, Rathouz PJ. Latent Variable Regression for Multiple Discrete Outcomes. *Journal of the American Statistical Association: Theory and Methods* 1997; 92:1375–1386.
- [43]. Bartolucci F, Pennoni F, Francis B. A Latent Markov Model for Detecting Patterns of Criminal Activity. *Journal of the Royal Statistical Society: Series A* 2007; 170:115–132.
- [44]. Clogg CC, Goodman LA. Latent Structure Analysis of a Set of Multidimensional Contingency Tables. *Journal of the American Statistical Association: Application Section* 1984; 79:762–771.
- [45]. Huang GH, Bandeen-Roche K. Building an Identifiable Latent Class Model with Covariate Effects on Underlying and Measured Variables. *Psychometrika* 2004; 69:5–32.
- [46]. Formann AK. Linear Logistic Latent Class Analysis for Polytomous Data. *Journal of the American Statistical Association: Theory and Methods* 1992; 87:476–486.
- [47]. Legler JM, Ryan LM. Latent Variable Models for Teratogenesis using Multiple Binary Outcomes. *Journal of the American Statistical Association: Application and Case Studies* 1997; 92:13–20.
- [48]. White T Extensions of Latent Class Transition Models with Application to Chronic Disability Survey Data. Ph.D. Thesis, University of Washington 2008.
- [49]. Dayton CM, MacReady GB. Concomitant-Variable Latent-Class Models. *Journal of the American Statistical Association: Theory and Methods* 1988; 83:401: 173–178.

- [50]. Clark RF. An Introduction to the National Long-Term Care Survey. Office of Disability, Aging, and Long-Term Care Policy within the U.S. Department of Health and Human Services. 1998.
- [51]. Erosheva E, White TA. Issues in Survey Measurement of Chronic Disability: An Example from the National Long Term Care Survey. *Journal of Official Statistics* 2010; 26:317–339.
- [52]. Akaike H A New Look at the Statistical Model Identification. *IEEE: Transactions on Automatic Control* 1974; 19:716–723.
- [53]. Schwarz GE. Estimating the Dimension of a Model. *Annals of Statistics* 1978; 6:461–464.
- [54]. Kullback S, Liebler RA. On Information and Sufficiency. *Annals of Mathematical Statistics* 1951; 22:79–86.
- [55]. Nylund KL, Asparouhov T, Muthen BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. *Structural Equation Modeling* 2007; 14(4):535–569.
- [56]. Lehman EL, Casella G. *Theory of Point Estimation*. Second Edition Springer: New York 1998.
- [57]. Manton KG, Gu X, Lamb VL. Changes in the Prevalence of Chronic Disability in the United States Black and Nonblack Population above age 65 from 1982 to 1999. *Proceedings of the National Academy of Sciences* 2001; 98:6354–6359.
- [58]. Wolf DA, de Leon CFM, Glass TA. Trends in Rates of Onset of and Recovery from Disability at Older Ages: 1982–1994. *The Journals of Gerontology: Social Sciences* 2007; 62B:S3–S10.
- [59]. Cutler DM. The Reduction in Disability among the Elderly. *PNAS* 2001; 98-12: 6546–6547. [PubMed: 11390991]
- [60]. Freedman VA, Martin LG, Schoeni RE. Recent Trends in Disability and Functioning among Older Adults in the United States. *Journal of the American Medical Association* 2002; 288:3137–3146. [PubMed: 12495394]
- [61]. Wolf DA. Population Change: Friend or Foe of the Chronic Care System? *Health Affairs* 2001; 20:28–42.
- [62]. Manton KG, Stallard E, Corder LS. Changes in Morbidity and Chronic Disability in the U.S. Elderly Population: Evidence from the 1982, 1984, and 1989 National Long Term Care Surveys. *Journals of Gerontology: Social Sciences* 1995; 50B:S194–S204.
- [63]. Fienberg SE, Hersh P, Rinaldo A, Zhou Y. *Maximum Likelihood Estimation in Latent Class Models for Contingency Table Data Algebraic and Geometric Methods in Statistics*, Cambridge University Press: Cambridge 27–62. 2010.
- [64]. Erosheva EA, Fienberg SE, Joutard C. Describing Disability through Individual-level Mixture Models for Multivariate Binary Data. *Annals of Applied Statistics* 2007; 1-2:502–537.
- [65]. MacDonald IL, Zucchini W. *Hidden Markov and Other Models for Discrete-valued Time Series*, Chapman and Hall: London 1997.
- [66]. Baum LE. An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes Inequalities, III (Proceedings of the Third Symposium on Inequalities, Academic Press: New York 1–8. 1972.
- [67]. Lystig TC. *Evaluation of Hidden Markov Models*. Ph.D. Thesis, University of Washington 2001.
- [68]. Xu L, Shao J. Estimation in Longitudinal or Panel Data Models with Random-Effect-Based Missing Responses. *Biometrics* 2009; 65:1175–1183. [PubMed: 19432782]
- [69]. Alfo M, Maruotti A. A selection model for longitudinal binary responses subject to non-ignorable attrition. *Statistics in Medicine* 2009; 28:2435–2450 [PubMed: 19424960]
- [70]. Kaciroti NA, Raghunathan TE, Schork MA, Clark NM. A Bayesian model for longitudinal count data with non-ignorable dropout. *Applied Statistics* 2008; 57-5:521–534. [PubMed: 21072316]
- [71]. Tsonaka R, Rizopoulos D, Verbeke G, Lesaffre E. Nonignorable Models for Intermittent Missing Categorical Longitudinal Responses. *Biometrics* 2010; 66:834–844. [PubMed: 19995352]
- [72]. Lystig TC, Hughes JP. Exact Computation of the Observed Information Matrix for Hidden Markov Models. *Journal of Computational and Graphical Statistics* 2002; 11:678–689.

Table 1:

True Parameters for Simulation Study - Complete Data Across 2 Waves

l	1	2	3
γ_l	.20	.30	.50
$\rho_{1 l}$	.60	.05	.01
$\rho_{2 l}$	.70	.10	.02
$\rho_{3 l}$	.80	.15	.03
$\rho_{4 l}$	.90	.20	.04
$\rho_{5 l}$	.99	.90	.05
$\pi_{l_2 l_1=1}$	.70	.25	.05
$\pi_{l_2 l_1=2}$	.35	.45	.20
$\pi_{l_2 l_1=3}$	.10	.30	.60

Table 2:

True v. Estimated Average Parameters (and Differences) for Simulation Study - Grouped Data with Missingness

l	True/Simulated	1	2	3	4
γ_{l_1}	True	.220	.312	.468	N/A
γ_{l_2}	True	.140	.260	.600	N/A
$\rho_{1 l}$	True	.600	.050	.010	N/A
$\rho_{2 l}$	True	.700	.100	.020	N/A
$\rho_{3 l}$	True	.800	.150	.030	N/A
$\rho_{4 l}$	True	.900	.200	.040	N/A
$\rho_{5 l}$	True	.990	.900	.050	N/A
$\pi_{l_2 l_1=1}$	True	.288	.096	.015	.602
$\pi_{l_2 l_1=2}$	True	.247	.296	.105	.352
$\pi_{l_2 l_1=3}$	True	.103	.287	.460	.151
$\pi_{l_2 l_1=4}$	True	.000	.000	.000	1.000
γ_{l_1}	Simulated	.217 (−.003)	.300 (−.012)	.483 (.015)	N/A
γ_{l_2}	Simulated	.140 (.000)	.251 (−.009)	.610 (.010)	N/A
$\rho_{1 l}$	Simulated	.603 (.003)	.051 (.001)	.011 (.001)	N/A
$\rho_{2 l}$	Simulated	.701 (.001)	.106 (.006)	.021 (.001)	N/A
$\rho_{3 l}$	Simulated	.807 (.007)	.154 (.004)	.031 (.001)	N/A
$\rho_{4 l}$	Simulated	.903 (.003)	.208 (.008)	.041 (.001)	N/A
$\rho_{5 l}$	Simulated	.990 (.000)	.912 (.012)	.061 (.011)	N/A
$\pi_{l_2 l_1=1}$	Simulated	.287 (−.001)	.092 (−.004)	.019 (.004)	.602 (.000)
$\pi_{l_2 l_1=2}$	Simulated	.248 (.001)	.299 (.003)	.094 (−.011)	.359 (.007)
$\pi_{l_2 l_1=3}$	Simulated	.102 (−.001)	.279 (−.008)	.466 (.006)	.154 (.003)
$\pi_{l_2 l_1=4}$	Simulated	.000 (.000)	.000 (.000)	.000 (.000)	1.000 (.000)

Table 3:

List of NLTCS ADLs and Corresponding Frequencies from Pooled Data 1982–2004

ADL	Frequency from Pooled Data 1982–2004
eating	.10
dressing	.19
toileting	.25
getting in and out of bed	.27
inside mobility	.39
bathing	.43

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

Age Decomposition of Individuals in each NLTC Period Observed, 1984–2004

Age	First / Last year	1984	1989	1994	1999	2004	Total
65+	1984	3,133	1,602	689	312	112	5,848
65–69	1989		186	101	94	79	460
70+			802	237	110	41	1,190
65–69	1994			164	119	165	448
70–74				123	95	89	307
75+				864	406	179	1,449
65–69	1999				243	380	623
70–74					82	94	176
75–79					289	222	511
80+					668	245	913
65–69	2004					377	377
70–74						232	232
75–79						288	288
80–84						350	350
85+						538	538
Total		3,133	2,590	2,178	2,418	3,391	13,710

Table 5:

AIC, BIC, and Entropy for Six Fitted Latent Class Models ($K = 2-7$), Based on 1999 ADL Data; p is the number of estimated parameters

K	p	AIC	BIC	Entropy
2	13	23,776	23,862	.899
3	20	22,511	22,642	.864
4	27	22,366	22,543	.847
5	34	22,325	22,547	.795
6	41	22,299	22,567	.798
7	48	22,310	22,624	.817

Table 6:

Estimated Parameters (and Standard Errors) for Full (35-group) Group-Based LCTM 1984–2004

<i>l</i>	1	2	3	4	5
γ_{l_1}	.113 (.003)	.167 (.004)	.109 (.003)	.610 (.007)	N/A
γ_{l_2}	.053 (.006)	.090 (.007)	.035 (.005)	.822 (.011)	N/A
γ_{l_3}	.063 (.008)	.129 (.011)	.056 (.008)	.752 (.017)	N/A
γ_{l_4}	.100 (.010)	.132 (.012)	.068 (.009)	.699 (.020)	N/A
γ_{l_5}	.149 (.018)	.206 (.021)	.204 (.021)	.441 (.042)	N/A
$\rho_{eat l}$	.624 (.008)	.019 (.002)	.052 (.004)	.002 (.000)	N/A
$\rho_{dress l}$	.966 (.003)	.111 (.005)	.185 (.008)	.009 (.001)	N/A
$\rho_{toilet l}$	.904 (.005)	.407 (.007)	.302 (.009)	.017 (.001)	N/A
$\rho_{bed l}$	.965 (.003)	.556 (.007)	.079 (.005)	.016 (.001)	N/A
$\rho_{mobility l}$	.975 (.003)	.945 (.003)	.157 (.007)	.067 (.002)	N/A
$\rho_{bath l}$	.998 (.001)	.693 (.007)	.826 (.007)	.098 (.003)	N/A
$\pi_{l_2 l_1=1}$	.206 (.007)	.048 (.004)	.036 (.003)	.049 (.004)	.661 (.010)
$\pi_{l_2 l_1=2}$	.108 (.004)	.363 (.006)	.015 (.002)	.068 (.003)	.447 (.010)
$\pi_{l_2 l_1=3}$	.109 (.005)	.105 (.005)	.274 (.008)	.066 (.004)	.445 (.014)
$\pi_{l_2 l_1=4}$	.045 (.002)	.101 (.002)	.070 (.002)	.593 (.004)	.191 (.006)
$\pi_{l_2 l_1=5}$	.000 (N/A)	.000 (N/A)	.000 (N/A)	.000 (N/A)	1.000 (N/A)

Table 7:

Estimated Period Class Weights (and Standard Errors) for Full Group-Based LCTM 1984–2004

Year	l=1	l=2	l=3	l=4
1984	.113 (.003)	.167 (.004)	.109 (.003)	.610 (.007)
1989	.107 (.003)	.159 (.003)	.101 (.003)	.632 (.006)
1994	.099 (.003)	.151 (.003)	.092 (.003)	.657 (.005)
1999	.092 (.003)	.141 (.003)	.081 (.002)	.686 (.005)
2004	.093 (.003)	.142 (.004)	.087 (.003)	.676 (.007)

Table 8:

Estimated Conditional Response Probability Arthritis Regression Coefficients (and Standard Errors) for Group-Based LCTM in Equation (8.12), NLTCs 1984–2004

m	1(eat)	2(dress)	3(toilet)	4(bed)	5(mobility)	6(bath)
κ_m	-.198 (.071)	.125 (.070)	.406 (.051)	.526 (.060)	1.182 (.046)	.360 (.048)
$\tau_{m l=1}$	.24 (.06)	3.51 (.11)	2.77 (.08)	3.91 (.12)	5.24 (.16)	6.45 (.39)
$\tau_{m l=2}$	-4.16 (.13)	-1.92 (.07)	.14 (.05)	.93 (.06)	4.77 (.11)	1.29 (.05)
$\tau_{m l=3}$	-3.56 (.10)	-1.74 (.07)	-.64 (.05)	-1.83 (.07)	-.20 (.05)	1.23 (.05)
$\tau_{m l=4}$	-6.84 (.29)	-4.77 (.12)	-4.39 (.11)	-3.56 (.08)	-1.20 (.04)	-2.20 (.05)

Table 9:

Estimated Conditional Response Probabilities (and Standard Errors) for Group-Based LCTM 1984–2004, by Arthritis Status

	ADL(m)	Arthritis	l=1	l=2	l=3	l=4
$\rho_{m=1 l}$	eat	No	.559 (.015)	.015 (.002)	.028 (.003)	.001 (.000)
$\rho_{m=2 l}$	dress	No	.971 (.003)	.128 (.008)	.149 (.009)	.008 (.001)
$\rho_{m=3 l}$	toilet	No	.941 (.004)	.536 (.013)	.346 (.011)	.012 (.001)
$\rho_{m=4 l}$	bed	No	.980 (.002)	.718 (.012)	.139 (.008)	.028 (.002)
$\rho_{m=5 l}$	mobility	No	.995 (.001)	.992 (.001)	.450 (.012)	.231 (.006)
$\rho_{m=6 l}$	bath	No	.998 (.001)	.783 (.009)	.774 (.009)	.099 (.004)
$\rho_{m=1 l}$	eat	Yes	.510 (.010)	.013 (.001)	.023 (.002)	.001 (.000)
$\rho_{m=2 l}$	dress	Yes	.974 (.003)	.142 (.006)	.165 (.007)	.010 (.001)
$\rho_{m=3 l}$	toilet	Yes	.960 (.003)	.634 (.008)	.443 (.009)	.018 (.002)
$\rho_{m=4 l}$	bed	Yes	.988 (.001)	.812 (.005)	.214 (.009)	.046 (.003)
$\rho_{m=5 l}$	mobility	Yes	.998 (.000)	.997 (.000)	.728 (.008)	.495 (.010)
$\rho_{m=6 l}$	bath	Yes	.999 (.000)	.838 (.005)	.831 (.005)	.137 (.005)
$\rho_{m=1 l}$	eat	Difference	-.049 (.018)	-.003 (.002)	-.005 (.003)	.000 (.000)
$\rho_{m=2 l}$	dress	Difference	.003 (.004)	.015 (.010)	.017 (.011)	.001 (.001)
$\rho_{m=3 l}$	toilet	Difference	.019 (.005)	.098 (.015)	.097 (.015)	.006 (.002)
$\rho_{m=4 l}$	bed	Difference	.008 (.003)	.094 (.013)	.075 (.012)	.018 (.004)
$\rho_{m=5 l}$	mobility	Difference	.004 (.001)	.006 (.001)	.277 (.014)	.264 (.012)
$\rho_{m=6 l}$	bath	Difference	.000 (.001)	.055 (.010)	.057 (.010)	.037 (.006)