Statistics in Biosciences (2014), **6**(1): 127–161
DOI: 10.1007/s12561-013-9087-8

Statistical Issues in Modeling Chronic Disease in Cohort Studies

RICHARD J. COOK

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: rjcook@uwaterloo.ca

JERALD F. LAWLESS

Department of Statistics and Actuarial Science,
University of Waterloo, Waterloo, ON, N2L 3G1, Canada
E-mail: jlawless@uwaterloo.ca

Summary

Observational cohort studies of individuals with chronic disease provide information on rates of disease progression, the effect of fixed and time-varying risk factors, and the extent of heterogeneity in the course of disease. Analysis of this information is often facilitated by the use of multistate models with intensity functions governing transition between disease states. We discuss modeling and analysis issues for such models when individuals are observed intermittently. Frameworks for dealing with heterogeneity and measurement error are discussed including random effect models, finite mixture models, and hidden Markov models. Cohorts are often defined by convenience and ways of addressing outcome-dependent sampling or observation of individuals are also discussed. Data on progression of joint damage in psoriatic arthritis and retinopathy in diabetes are analysed to illustrate these issues and related methodology.

Keywords: heterogeneity, intermittent observation, Markov processes, multistate models, life history studies

The final publication is available at Springer via http://dx.doi.org/10.1007/s12561-013-9087-8.

1 INTRODUCTION

Chronic disease processes are often complex, and feature periods of acute disease activity and gradual deterioration in function of one or more organ systems. Cohort studies can offer important information about these sorts of processes through careful follow-up of affected individuals over time. Such studies have played a central role in the identification of demographic, genetic and environmental factors associated with disease onset and progression, the assessment of therapeutic interventions, and the specification of models for prediction at the individual or population level.

Multistate stochastic models offer an appealing framework for the analysis of disease activity and progression, of related events, and of concomitant time-varying risk factors. Our purpose is to discuss the utility of multistate models in studies of chronic disease, to highlight some of the modeling issues arising from such studies, and to consider ways in which these challenges may be addressed within the multistate framework.

When disease status is measured on an ordinal scale in progressive conditions, intensity-based methods offer a convenient and informative way to model dynamic aspects of disease and to identify prognostic variables. This strategy has been widely used, in studies of psoriatic arthritis (O'Keeffe et al., 2011), hepatitis (Sweeting et al., 2006), complications from diabetes (Al-Kateb et al., 2008), HIV infection (Gentleman et al., 1994), bone marrow transplantation (Keiding et al., 2001) and age-related cognitive decline and dementia (Tyas et al., 2007). Many challenges arise in the use of such models, however. Often there is considerable variability between individuals in the nature and rate of progression, making it difficult to specify plausible models. This is particularly true when states are based on categorization of continuous measures of disease severity. Imperfect measurement of disease status can lead to apparent improvement in what is believed to be a progressive condition across consecutive observation times. Fitting even simple models can be challenging when data are only collected at random and variably spaced assessment times, and this is particularly true when modeling the effects of concomitant variables which are only observed intermittently. Further challenges arise with other types of missing information, when modeling relationships between different aspects of disease status, and when dealing with the effects of disease-related selection or follow-up.

We now introduce two studies that illustrate and motivate subsequent development.

EXAMPLE 1.1 THE UNIVERSITY OF TORONTO PSORIATIC ARTHRITIS COHORT

Psoriatic arthritis (PsA) is an immunological disease associated with considerable joint pain, inflammation and destruction which can ultimately lead to serious disability and poor quality of life (Chandran et al., 2010). The Centre for Prognosis Studies in Rheumatic Disease is a tertiary referral center at the Toronto Western Hospital which treats patients with a variety of rheumatic diseases and maintains a clinic registry of patients with psoriatic arthritis. This cohort was established in 1976 and has been recruiting and following patients since its formation; today it is one of the largest cohorts of patients with PsA in the world.

Upon entry to the clinic patients undergo a detailed clinical and radiological examination and provide serum samples which are subsequently stored. Follow-up clinical and radiological assessments are scheduled annually and biannually, respectively, in order to track changes in joint damage, functional ability, and quality of life. Additional serum samples are taken at follow-up clinic visits to measure dynamic markers of inflammation and to store for future analysis of genetic data. To date, 1191 patients have been recruited and there is a median of 4.84 years of follow-up and a median of 6 clinical follow-up assessments. Given this data, disease progression can be modeled in a number of ways including the development of newly damaged joints (Gladman et al., 1995; Sutradhar and Cook, 2008), the involvement of particular types (e.g. spinal) of joints (Tolusso and Cook, 2009), or progression to a state of clinically important damage (Chandran et al., 2012).

EXAMPLE 1.2 DCCT/EDIC STUDY OF TYPE 1 DIABETES

The Diabetes Control and Complications Trial (DCCT) was a randomized study that involved Type 1 diabetics, and ran from 1984 to 1993. Subjects in the trial were randomized to an intensive therapy treatment arm or to a conventional therapy control arm. The intensive therapy treatment was designed to achieve and maintain near-normal glucose levels, whereas the conventional therapy was designed to prevent hyperglycemic symptoms (The Diabetes Control and Complications Trial Research Group, 1993). The DCCT showed that the intensive therapy was associated with a significant reduction in the onset of diabetic retinopathy and neuropathy. After it was terminated in 1993, most subjects (1375 of 1441) joined the observational Epidemiology of Diabetes Interventions and Complications (EDIC) study, which began in 1994. Subjects in the DCCT were seen approximately every 3 months, and eye and kidney measurements were taken every 6 months; during EDIC these measurements are taken roughly every two years. The degree of retinopathy was expressed on the ordinal ETDRS scale (Early Treatment Diabetic Retinopathy Study Research Group, 1991). Measurements of renal function in-

cluded urinary albumin excretion rates (AER) and two states, persistent microalbuminaria and severe nephropathy, were based on this (Al-Kateb et al., 2008). Many fixed covariates and time-varying biomarkers were also measured on each individual; the most important biomarker is glycosylated hemoglobin (HbA1c), which measures average blood glucose over a prolonged period up to the measurement date. In addition to establishing the value of intensive therapy for the control of HbA1c in Type 1 diabetics, the DCCT and EDIC studies have produced information on complications from diabetes and on biological and genetic factors associated with progression.

In the remainder of the paper we discuss several issues in the context of multistate models. They include challenges in modeling heterogeneous disease processes and in dealing with missing data arising from intermittent observation of individuals; alternative models for individual or joint processes; and the possibility that selection or observation of individuals is outcome-dependent. Section 2 reviews likelihood function construction for multistate models, emphasizing the case of intermittent observation. Section 3 discusses several aspects of modeling progressive disease processes, including families of multistate models, heterogeneity versus process dynamics, models with latent (unobserved) components, and ways of handling time-varying covariates. Section 4 considers models for multiple processes. Section 5 illustrates modeling and analysis in the psoriatic arthritis and Type 1 diabetes cohorts introduced in Examples 1.1 and 1.2. Section 6 gives brief discussions of outcome-dependent selection or observation of processes, and Section 7 contains some concluding remarks and notes areas where further development is needed.

2 Likelihood for Multistate Analyses

2.1 Likelihood Construction with Right-Censored Data

Consider a multistate model with K states where $Z(t)$ indicates the state occupied at time t and $\{Z(s), 0 < s\}$ is the associated stochastic process. States are ideally formed based on well-defined stages of the disease process. Sometimes it is appropriate to include one or more absorbing states to reflect death, different causes of death, or other reasons for termination of the disease process. Let $X(t)$ denote a $p \times 1$ vector of left-continuous, time-varying covariates. The history for the two processes is denoted $\mathcal{H}(t) = \{(Z(s), X(s)), 0 < s < t\}$. We may specify intensity functions for the process $\{Z(s), 0 < s\}$ as

$$\lambda_{k\ell}(t|\mathcal{H}(t)) = \lim_{\Delta t \downarrow 0} \frac{P(Z(t + \Delta t^-) = \ell | Z(t^-) = k, \mathcal{H}(t))}{\Delta t} \tag{1}$$

where $\ell = k$ (Andersen et al., 1993).

It is often helpful to express models equivalently in terms of counting processes, particularly when data are right-censored. Let $dN_{kl}(t) = 1$ if a transition from state k to l occurs at time t and let $dN_{k\ell}(t) = 0$ otherwise, $N_{kl}(t) = \int_0^t dN_{kl}(s)$ count the number of $k \to l$ transitions over $(0, t]$ and let $\{N_{kl}(s), 0 < s\}$ be the corresponding right-continuous counting process. Let $N_k(t) = (N_{k1}(t), \ldots, N_{kK}(t))'$ denote the multivariate counting process of transitions out of state k and $N(t) = (N_1'(t), \ldots, N_K'(t))'$. With counting process notation the intensity can be alternatively expressed as

$$\lambda_{k\ell}(t|\mathcal{H}(t)) = \lim_{\Delta t \downarrow 0} \frac{P(\Delta N_{k\ell}(t) = 1 | Y_k(t) = 1, \mathcal{H}(t))}{\Delta t} ,$$

where $\Delta N_{k\ell}(t) = N_{k\ell}(t + \Delta t^-) - N_{k\ell}(t^-)$ counts the number of $k \to \ell$ transition over $[t, t + \Delta t)$ and $Y_k(t) = I(Z(t^-) = k)$ indicates whether an individual is at risk of a transition out of state k at t. To accommodate right-censoring times $C > 0$ we denote $\bar{Y}_k(t) = I(t \le C)Y_k(t)$.

The partial likelihood for the parameters specifying the intensities is

$$L \propto \prod_{k=1}^{K} \prod_{\ell \in \mathcal{S}_k} L_{k\ell}$$

for a given individual, where S_k is the set of states ℓ for which $\lambda_{k\ell}(t|H(t)) > 0$,

$$L_{k\ell} = \prod_{s \in D_{k\ell}} \lambda_{k\ell}(s|X(s), \mathcal{H}(s)) \exp\left(- \int_0^\infty \bar{Y}_k(u) \lambda_{k\ell}(u|X(u), \mathcal{H}(u)) du \right) ,$$

and $D_{k\ell}$ is the set of transition times observed from state k to state ℓ, $\ell \in \mathcal{S}_k, k = 1, 2, \ldots, K$ (Andersen et al. 1993; Cook and Lawless, 2007). This likelihood is similar to those used in survival analysis and this connection enables us to fit multistate models using software for survival data.

2.2 Likelihood with Intermittent Observation

Now we consider the case of intermittent inspection and consider the likelihood contribution from a single individual. Let $v_0 < v_1 < \cdots < v_n$ denote the times of n clinic visits where assessments are carried out and the health state and covariates are measured. Let $H(v_r) = \{(z(v_s), x(v_s), v_s), s = 0, 1, 2, \ldots, r-1\}$ denote the history of the observed data under this inspection process. This is sometimes referred to as panel data in the literature (e.g. Kalbfleisch and Lawless, 1985). In this case the full likelihood is

$$L \propto P(Z(v_0), X(v_0), V_0 = v_0) \prod_{r=1}^{n} P(Z(v_r), X(v_r), V_r = v_r | H(v_r)) .$$

It is common to condition on the state and covariate values at the initial assessment, in which case we omit the term $P(Z(v_0), X(v_0), V_0 = v_0)$; when $v_0 = 0$ this corresponds to the onset of the process so if $P(Z(0) = 1|X(0)) = 1$, this omission is innocuous. More generally, there can be some loss in efficiency, particularly when the number of follow-up assessments is low, but this information can only be extracted with strong assumptions about the initial conditions.

Premature loss to followup (LTF) is also common, in which case the final observation time v_n may precede an intended final assessment time v. We assume the "sequential missingness at random" (SMAR) condition (Hogan et al., 2004) which requires that if an individual is observed up to v_{r-1}, then conditional on the event history at that time, the probability they are LTF and not observed at v_r cannot depend on events in $[v_{r-1}, v_r)$. In studies with widely spaced visit times this condition is often violated to some extent. We discuss this further in Section 6.1; see also Lawless (2013).

Under the assumption $V_r \perp Z(v_r), X(v_r)|H(v_r)$,

$$P(Z(v_r), X(v_r), V_r = v_r|H(v_r)) = P(Z(v_r), X(v_r)|v_r, H(v_r))P(V_r = v_r|H(v_r)) ,$$

and if the inspection process is non-informative, we disregard the second term on the right (Grüger et al., 1991). The factorization

$$P(Z(v_r), X(v_r)|v_r, H(v_r)) = P(Z(v_r)|v_r, H(v_r))P(X(v_r)|v_r, Z(v_r), H(v_r))$$

is typically adopted for modeling, since it is natural to use covariate information at v_{r-1} to model the state occupied at v_r. In this case we may focus on the partial likelihood (Cox, 1975) based on the first components of this factorization,

$$\prod_{r=1}^{n} P(Z(v_r)|H(v_r)) . \tag{2}$$

In (2) and henceforth we omit v_r in the conditioning event for notational convenience, but it is implicitly present.

When assessment times are evenly spaced and common to all individuals, discrete time models offer a convenient framework. When there is considerable variation in the spacing of assessment times, models for continuous-time processes are usually necessary, and in this case dealing with time-varying covariates under intermittent inspection is challenging. If the covariates are discrete, one approach is to model the covariate process jointly with the outcome (Tom and Farewell, 2011). If for example, $X(t)$ is an indicator of whether a concomitant comorbidity has developed, joint models with an expanded state space encompassing both the covariate and outcome data are useful. Continuous time-varying covariates are much more difficult to deal with and these will typically require stronger assumptions to specify tractable joint models. We return to this in Section 3.6.

3 Considerations in Modeling Progressive Processes

3.1 Models for Progressive Disease

We consider multi-state models with $K-1$ transient states and a single absorbing state representing death or another event that terminates the process. The emphasis here is on chronic conditions for which progression tends to occur over time, in which case we often assume that transitions from a transient state can only be made to the next highest state or the absorbing state; see Figure 1. In some settings, entries to state K may only come from state $K-1$.

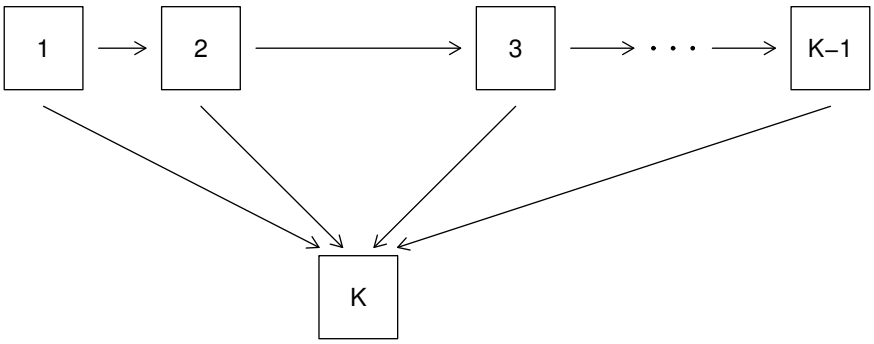

Figure 1: Diagram of a progressive multistate model with an absorbing state

It is common in longitudinal studies for an individual to show improvement from one visit to the next, so that $Z(v_r) < Z(v_{r-1})$; this may be due to random fluctuations in a person's status, to misclassification of the discrete disease states, or to errors in the measurement of an underlying continuous score. A common *ad hoc* convention involves defining states operationally based on sustained observation above a threshold for a prescribed period of time or number of visits. Progression of diabetic retinopathy, for example, has sometimes been defined as occuring when a person has ETDRS scores on two consecutive visits that are 3 or more grades higher than the score at the baseline visit (The Diabetes Control and Complications Trial Research Group, 1993). This approach ties the state definitions to the sequence of visit times $\{v_r, r = 0, 1, 2, \ldots\}$, which can be undesirable when times between visits are highly variable. It might also be undesirable for processes in which significant real improvement may occur. In such cases we may prefer to allow transitions in both directions between adjacent transient states; this complicates model fitting and analysis but provides a more realistic representation of the observed disease process. Another option is to consider the "true" disease process $\{Z(s), 0 < s\}$

to be as in Figure 1 but unobserved, with the observed process, denoted $\{Z^*(s), 0 < s\}$, related to the true process. Hidden Markov models (HMMs), in which $\{Z(s), 0 < s\}$ is a Markov process, have been used by several authors (e.g. Satten and Longini, 1996; Bureau et al., 2003; Jackson and Sharples, 2002) and can be fit with the `msm` package in R (Jackson, 2011). We discuss these in Section 3.5.

A thorough comparison of competing models and of estimability issues with intermittent observation would be valuable. We compare different approaches for analysing diabetic retinopathy in Section 5.2.

3.2 Times scales, families of models and model fitting

The types of models used for disease processes may depend on the main objectives of analysis. If we aim to understand individual process dynamics and related factors, then models whose intensities reflect such dynamics are desirable. However, if the main objective is to assess a treatment or intervention then models which consider marginal process characteristics such as time of entry to a particular state are preferred. The reason for this is that intensity-based models which condition on previous event history confound treatment covariates with previous history, thus making interpretation of treatment effects problematic. For more discussion see Aalen et al. (2008, Ch. 8) and for examples of marginal analysis, Cook et al. (2009).

In modeling transition intensities (2), an initial consideration is whether to emphasize age of a process ("global" time) or the duration of time spent in each state visited. Markov models are basic for the former situation and semi-Markov models for the latter. With Markov models the transition intensities take the form $\lambda_{k\ell}(t|H(t), X) = q_{k\ell}(t; X)$, where for now we restrict attention to fixed covariates X. This requires specification of a time origin ($t = 0$) that is comparable across individuals; it particularly affects relative risk modeling of covariate effects. For example, in the DCCT/EDIC study of Example 1.2 we could let t represent (a) age, (b) time since onset of diabetes, or (c) time since entry to the study. Because individuals were randomized to treatment at study entry, (c) is preferable when analyses are directed at treatment comparisons, but this may not provide the most appropriate or parsimonious model of process dynamics. Pencina et al. (2007) provide some general discussion of time origins and relative risk models.

Markov models with fixed covariates can be fitted fairly easily with intermittently observed data. Estimation of model parameters by maximum likelihood is based on the likelihood function (2), which for Markov models depends only on transition probabilities $P(Z(v_r) = \ell | Z(v_{r-1}) = k, X)$. Time-homogeneous models for which $q_{k\ell}(t; X) = q_{k\ell}(X)$ are especially straightforward. In this case the $K \times K$ transition probability matrix $\mathbb{P}(t; X)$ with entries $p_{k\ell}(t; X) = P(Z(s + t) = \ell | Z(s) = k; X)$ is given by the matrix exponential function,

$$\mathbb{P}(t; X) = \exp\{t\mathbb{Q}(X)\} , \tag{3}$$

where $\mathbb{Q}(X)$ is the $K \times K$ matrix with (k, ℓ) entries $q_{k\ell}(X)$ for $k \neq \ell$, and $q_{kk}(X) = -\Sigma_{\ell \neq k} q_{k\ell}(X)$ as the diagonal entries (Cox and Miller, 1965). Kalbfleisch and Lawless (1985) discuss computation of (3) and algorithms for maximizing the associated likelihood function. This has been implemented in software packages, including the `msm` package in R (Jackson, 2011) which we use later.

Non-homogeneous Markov models with intensities $q_{k\ell}(t; X)$ are somewhat harder to fit with intermittent observation and general software is not available. However, by using the product-integral representation for a transition probability matrix $P(s, t; X)$, (e.g. Aalen et al., 2008, Section A.2.4),

$$P(s, t; X) = \prod_{(s,t]} \{\mathbb{I} + \mathbb{Q}(u; X)\, du\} , \tag{4}$$

where $\mathbb{I}$ is a $K \times K$ identity matrix, we can approximate transition probabilities closely by discrete approximation. See Titman (2011) for a similar approach in terms of the Kolmogorov differential

equations for the process (e.g. Aalen et al., 2008, Section A.2.3). A simple and convenient alternative is to assume transition intensities are piecewise constant; the `msm` package can deal with this. Models with time-varying covariates are more complicated to handle when the covariates are observed only intermittently. It is generally necessary to make simplifying assumptions about their behavior or to model them, and we defer discussion to Section 3.6.

Models in which intensities (2) depend on time since entry to a state are also harder to handle with intermittent observation because the time of entry is usually unobserved. For progressive models as in Figure 1, however, it is possible to obtain transition probabilities needed for the likelihood (2); see Titman and Sharples (2010a) and Yang and Nair (2011) for recent discussions of semi-Markov models in this context.

Realistically, the transition intensities for a process may depend on other aspects of the history besides t, covariates and time since entry to the current state. Aalen et al. (2008, Chapters 8 and 9) provide an informative discussion of process dynamics and how they can be modeled. With intermittently observed continuous time processes it is difficult or impossible to fit complex models, although progress is sometimes possible for progressive models. In cases where the times between successive visits are more or less constant across individuals, a viable option is to use discrete time models for $Z(v_r), r = 0, 1, \ldots, n$; Bacchetti et al. (2010) consider logistic regression models with dependence on process history in this setting.

Finally, we note the importance of initial conditions in settings where individuals are not all followed from the origins of their processes. Initial conditions include baseline covariates and relevant process history such as the duration of the disease process at the start of followup. Failure to adjust for important factors can bias conclusions, especially in observational studies. For example, Prentice et al. (2005) discuss time-dependent adverse health effects of hormone therapy (HT) for post-menopausal women that were seen in a Women's Health Initiative study where women were randomized to receive HT or a placebo at the start of followup. Previous observational studies in which there was insufficient adjustment for the prior duration of HT conversely indicated beneficial effects, in contradiction to the randomized trial. Relevant initial information should be collected in observational studies; if it is not, then we must rely on modeling assumptions.

3.3 Heterogeneity versus Process Dynamics

Fixed covariates (e.g. genetic variables, family history) that help explain variation in the course of disease across individuals are of considerable interest. The effects of such covariates are necessarily confounded to some extent with the dependence on the process history when process intensities are considered. Random effects are sometimes used in multistate models to account for unexplained heterogeneity (e.g. Lin et al., 2008; Mandel and Betensky, 2008; O'Keeffe et al., 2013) and their effects can also be hard to distinguish from the effects of process history. As an illustration, consider a 3-state progressive model and suppose that conditional on an unobservable random effect α, the conditional transition intensities are $\lambda_{12}(t|\mathcal{H}(t), \alpha) = \alpha\lambda_{12}$ and $\lambda_{23}(t|\mathcal{H}(t), \alpha) = \alpha\lambda_{23}$. If α has a gamma distribution with mean 1 and variance ϕ, after marginalizing over the random effect the "observable" intensity $\lambda_{23}(t|\mathcal{H}(t))$ is,

$$\begin{aligned} \lambda_{23}(t|\mathcal{H}(t)) &= P\left(dN_{23}(t) = 1 | Y_2(t) = 1, t_1\right)/dt \\ &= \frac{(1+\phi)\lambda_{23}}{1+\phi\{\lambda_{12}t_1 + \lambda_{23}(t-t_1)\}} \end{aligned} \tag{5}$$

where t_1 is the transition time from state 1 to state 2. The random effect model is therefore equivalent to a model with a very specific type of dependence on $\mathcal{H}(t)$ given by (5). This model has also been considered by Aalen et al. (2008, pp. 262–263).

More generally, covariates, their effects, and random effects may all vary with time. A deconstruction of process dynamics that clearly distinguishes covariate effects, random effects and history-dependence is unachievable without strong assumptions about the mathematical form of the model. In settings where process dynamics and prediction are of interest and fairly complete data are available on process histories, we might prefer to develop intensity-based models and to add random effects only if they improve significantly the predictive power of the model. Given the incompleteness of process history information in some studies, particularly those with intermittent observation, random effects can offer a tractable and convenient approach. For example, the intensity function (5) is unusable if we do not observe (or know to a reasonable approximation) the transition time t_{i1}. If we use the random effects formulation on the other hand, we can take the likelihood function for an individual conditional on their random effect α, and then integrate it with respect to α; this is tractable.

Aalen et al. (2008, Ch. 7 and 8) provide a lucid discussion of the issues in this section. On a final important point not discussed by them, we note that random effects modeling can be problematic when an individual is followed from a time v_0 at which their disease process has been underway for some time. In that case the distribution of a fixed random effect α will in general depend on process history $\mathcal{H}(v_0)$ prior to v_0; this is frequently ignored, with a single distribution assumed for the random effects of all individuals. For additional discussion, see Lawless and Fong (1999).

3.4 Finite Mixture Models

Continuous mixtures as in the preceding section involve very particular specifications on the nature of the heterogeneity and sometimes discrete mixtures are preferable. In many contexts, the proportion of individuals remaining progression-free is much higher than would be expected from a given model. This type of heterogeneity is naturally modeled by allowing a fraction of the population to be at zero risk of progression. A mixture model can be specified in this context with a Bernoulli random variable α, where $P(\alpha = 1) = \pi = 1 - P(\alpha = 0)$. Such models are called mover-stayer models (Goodman, 1961; Frydman, 1984) since individuals will either be "movers" who transition through the process, or "stayers" who do not.

A natural generalization of the mover-stayer model is to accommodate G classes labeled $1, 2, \ldots, G$ where individuals in the same class experience a disease course governed by a common process. We let W be a latent random variable indicating the sub-population to which a particular individual belongs; that is, $W = g$ if the individual belongs to class g, and $P(W = g|X;\gamma) = \pi_g(X)$, $g = 1, 2, \ldots, G$, with $\sum_{g=1}^{G} \pi_g(X) = 1$. The disease process in the population is then governed by a mixture of the class-specific processes. Estimating the number of such classes is challenging and we focus on the case where this is specified in advance. Accommodation of distinct classes is often motivated by apparent clusters of life history paths that share similar features, evident from examination of raw data. In many cohort studies, for example, some individuals are seen to experience rapid progression, some progress at much slower rates and some may not progress even after extensive follow-up. It is most appropriate to consider mixture models when there is evidence of distinct clusters after adjusting for important prognostic variables.

To allow the intensity for transitions from state k to ℓ to differ across classes we let

$$\lambda_{k\ell}(t|\mathcal{H}(t), W = g) = \lambda_{k\ell}^{(g)}(t|\mathcal{H}(t);\theta_g)$$

denote the intensity for class g where θ_g denotes the set of parameters governing the disease process, $g = 1, \ldots, G$, with $\theta = (\theta_1', \ldots, \theta_G')'$, and $\psi = (\theta', \gamma')'$. The resultant complete data likelihood (assuming W is observed) is

$$L_{comp}(\psi) \propto \prod_{g=1}^{G} \Big\{ L_g(\theta_g)\pi_g(\gamma) \Big\}^{I(W=g)}$$

where $L_g(\theta_g)$ is of the form (2) but with the probabilities evaluated according to the class g model. The structure of this likelihood suggests the use of an EM algorithm for estimation. If ψ^r denotes the estimate of ψ obtained at the rth iteration,

$$\omega_g^r = P(W = g|H(\infty); \psi^r) = \frac{L_g(\theta_g^r)\pi_g(X|\gamma^r)}{\sum_{g=1}^G L_g(\theta_g^r)\pi_g(X|\gamma^r)} ,$$

and $Q_{\psi;\psi^r} = E(\log L_{comp}(\psi)|H(\infty); \psi^r)$, we obtain $Q(\psi; \psi^r) = Q_1(\theta; \psi^r) + Q_2(\gamma; \psi^r)$ where

$$\begin{aligned} Q_1(\theta; \psi^r) &= \sum_{g=1}^G \omega_g^r \log L_g(\theta_g) , \\ Q_2(\gamma; \psi^r) &= \sum_{g=1}^G \omega_g^r \log \pi_g(X; \gamma) . \end{aligned}$$

Under a Markov model for each class maximization of $Q_1(\theta; \psi^r)$ is straightforward. let $p_{k\ell}(s, t|X, W = g; \theta_g) = P(Z(t) = \ell|Z(s) = k, X, W = g; \theta_g)$. At the rth iteration of the EM algorithm we can then write (for progressive models)

$$Q_1(\theta; \psi^r) = \sum_{g=1}^G \sum_{r=1}^n \sum_{k=1}^{K-1} \sum_{\ell=k}^K [\omega_g^r \cdot n_{k\ell} \cdot \log P_{k\ell}(v_{r-1}, v_r|X; \theta_g) ,$$

which can be maximized separately for each θ_g using pseudo-transition counts $\omega_g^r \cdot n_{k\ell}$ and the Fisher-scoring algorithm of Kalbfleisch and Lawless (1985).

Such models have considerable appeal in their flexibility. They can, for example, involve stratification so that in the multiplicative framework, baseline intensities can differ across classes but regression coefficients may be common. Tests for homogeneity of covariate effects can likewise be carried out. In some classes, some transition intensities may be estimated to be zero, suggesting that individuals are at negligible risk of certain events. In fact the models for the different classes may be very different in nature; some classes may involve Markov assumptions and others could involve semi-Markov assumptions. Such hybrid models do not seem to have received much attention in the literature.

Care is needed in fitting these models, with restrictions on parameter values typically needed to avoid multimodal likelihoods. Considerable thought must also go into how best to incorporate covariate effects in finite mixture models. While in theory covariates may be used to model both class membership and the transition intensities, parameter estimation can be difficult when one or more covariates appear in both parts of the model. Many questions are naturally addressed by modeling only class membership as a function of covariates.

3.5 Hidden Markov Models

When it is difficult to accurately classify individuals, disease status may be imperfectly recorded. In disease processes observed through radiological examination, for example, quality of radiographs, ambiguous definitions of states, and inter- or intra-observer variation may lead to misclassification of states. In these settings it is common to adopt a model based on an underlying process and a misclassification model. As before we let $Z(v_r)$ represent the true state occupied at time v_r, but now we let $Z^*(v_r)$ denote the state recorded at v_r. With a perfect classification procedure $P(Z(v_r) = Z^*(v_r)|x) = 1$, but more generally we let $\zeta_{kh} = P(Z^*(s) = k|Z(s) = h, x)$ denote the conditional probability that an individual is recorded to be in state k given they are in state h at time s; $\sum_{k=1}^K \zeta_{kh} =$

1, for $h = 1, 2, \ldots, K$. In this case the process $\{Z(s), 0 < s\}$ is latent (unobserved) or "hidden" and the data relate to $\{Z^*(s), 0 < s\}$.

With fixed covariates we let $H(v_s) = \{(Z(v_s), v_s), s = 1, 2, \ldots, r-1, X\}$ denote the history of the latent states, covariates, and observed inspection times. Let $H^*(v_r) = \{(Z^*(v_s), Z(v_s)), v_s, s = 1, 2, \ldots, r-1, X\}$ denote the history of the observed and latent processes. We make the following assumptions:

Assumption A.1 $\quad P((Z(v_r), v_r)|H^*(v_r)) = P((Z(v_r), v_r)|H(v_r))$,

Assumption A.2 $\quad P(Z^*(v_r)|Z(v_r), v_r, H^*(v_r)) = P(Z^*(v_r)|Z(v_r))$.

Assumption A.1 states that the inspection time and state occupied at the next assessment are conditionally independent of the recorded states given the history of the true states, covariates, and past inspection times. This is most reasonable in settings where individuals are passively observed and no potentially influential actions are taken as a consequence of the assessments. It may be less plausible in cohorts of patients in a clinic that provides routine medical care. Assumption A.2 states that the classification of the state occupied at each inspection time depends only on the current true state and possibly covariates but is conditionally independent of the observation process and prior recorded and true disease states.

Then, conditioning on the fixed covariates and the first assessment at $v_0 = 0$, a complete data likelihood can be written as

$$\begin{aligned} L_c &\propto P(Z^*(v_0), Z(v_0)|X, v_0) \prod_{r=1}^{n} P(Z^*(v_r), Z(v_r), v_r|H^*(v_r)) \\ &= P(Z^*(v_0)|Z(v_0), X, v_0) P(Z(v_0)|X, v_0) \prod_{r=1}^{n} P(Z^*(v_r)|Z(v_r), v_r, H^*(v_r)) P(Z(v_r), v_r|H^*(v_r)) \end{aligned}$$

which by Assumptions A.1 and A.2 we write as

$$\begin{aligned} L_c \propto & \left[P(Z^*(v_r)|Z(v_r), X) \prod_{r=1}^{n} P(Z^*(v_r)|Z(v_r), X) \right] \\ & \left[P(Z(v_0)|X, v_0) \prod_{r=1}^{n} P(Z(v_r), v_r|H(v_r)) \right] \end{aligned} \tag{6}$$

The observed data likelihood can be obtained by summing (6) over all possible realization of $\{Z(v_r), r = 0, 1, \ldots, n\}$, and this is done in Jackson and Sharples (2002), Jackson et al. (2003), and the `msm` package. A EM algorithm can alternatively be used since the complete data likelihood factors into two relatively simple parts; the first term in (6) denoted by L_c^* contains the parameters of the misclassification distribution, and the second term denoted by L is indexed by the parameters of the latent process of interest. It is important to note, however, that at the E-step the expectation of $\log L_c$ is with respect to the entire path $\{Z(v_r), r = 0, 1, \ldots, n\}$ since none of these values are observed. The Kalman filter offers an alternative computationally convenient framework for estimation in this setting (e.g. Fahrmeir and Tutz, 2001, Ch. 8).

With limited sample sizes, estimation of the misclassification probabilities and the transition probabilities are confounded to some extent, and maximum likelihood estimation can be challenging. This is especially true when the times between visits are sufficiently large that multiple transitions might occur. An additional caveat is that model checking can be difficult. We examine these issues in the application in Section 5.2.

3.6 COVARIATES AND REGRESSION MODELS

With continuously observed but right-censored data, regression with fixed covariates is in principle easy. Multiplicative Markov models having

$$\lambda_{k\ell}(t|H(t)) = q_{k\ell 0}(t)\exp(X'\beta_{k\ell}) \tag{7}$$

are widely used, and recent developments in the theory and related software for additive models have also increased their popularity (e.g. Aalen et al., 2008, Ch. 4).

With intermittently observed processes the general situation is more difficult, because the likelihood function (2) does not factor into separate pieces for each type (k, ℓ) of transition. Estimation when there are several covariates becomes a high-dimensional optimization problem and maximization of the likelihood function can be challenging, especially when multiple transitions between observation times may occur. Progressive models with relatively few states are much less problematic since there are few transition types, and the transition probabilities needed when individuals progress at most one or two states forward are straightforward to compute.

As mentioned in Section 2.2, dealing with time-varying covariates is considerably more challenging. If covariates are observed only at visit times then there are basically two options: (i) adopt some convention whereby covariates are assumed fixed between visit times or for some maximum acceptable period of time following assessment, and (ii) model the covariate process jointly with the multistate process in such a way that likelihood contributions can be calculated. An example of (i) is the "last value carried forward" (LVCF) approach, where it is assumed that $X(t) = X(v_{r-1})$ for $t \in (v_{r-1}, v_r]$. This and similar approaches can be satisfactory when visit times are not too far apart but generally they bias estimation (e.g. Raboud et al., 1993).

There is a large literature on joint models for covariate processes and single survival times or repeated events; Tsiatis and Davidian (2004) provide an excellent review of work up to 2004. It is generally assumed in this work that $X(t)$ is observed only intermittently and in that case, rather strong simplifying assumptions are needed concerning the dependency of transition intensities on the covariate values. We outline here two approaches that are reasonably tractable.

The first approach assumes the existence of a vector α of random effects such that, given α, the processes $\{Z(s), 0 < s\}$ and $\{X(s), 0 < s\}$ are independent. The transition intensities for $Z(t)$ are of conditional Markov form $q_{k\ell}(\alpha)$, where for simplicity we do not accommodate the presence of fixed covariates. In addition, we assume some model for the $X(t)$ process, conditional on α. Letting $\mathcal{Z}(v_r) = (Z(v_0), \ldots, Z(v_{r-1}))$ and $\mathcal{X}(v_r) = (X(v_0), \ldots, X(v_{r-1}))$, the likelihood function requires terms

$$P\{Z(v_r), X(v_r)|\mathcal{Z}(v_r), \mathcal{X}(v_r)\} \;=\; \int P\{Z(v_r)|\mathcal{Z}(v_r), \alpha\}\, P\{X(v_r)|\mathcal{X}(v_r), \alpha\}\, dG(\alpha) \tag{8}$$

where $G(\cdot)$ is the distribution function for α. If the second term in the integrand has a tractable form then (8) can be obtained by numerical integration.

The shared random effects model makes a strong conditional independence assumption. In addition, it implicitly assumes that given $H(t)$, $Z(t)$ depends on values $X(s)$ for $s > t$, and that $X(t)$ exists after a terminal event like death has occurred. These assumptions are often undesirable, especially in the case of internal covariates (biomarkers), and we now consider an idealized but simple modeling strategy that avoids such assumptions. The key idea is to replace $X(t)$ with a categorical variable $\tilde{X}(t)$ obtained by grouping the values of $X(t)$ into distinct levels $1, 2, \ldots, G$. The expanded state space for $(Z(t), \tilde{X}(t))$ consists of states $\{(z, x), z = 1, \ldots, K; \tilde{x} = 1, \ldots, G\}$. If we now assume that $\{(Z(t), \tilde{X}(t)), t > 0\}$ is a Markov process, then likelihood contributions

$$P\{Z(v_r), \tilde{X}(v_r)|Z(v_{r-1}), \tilde{X}(v_{r-1})\} \tag{9}$$

are readily obtained. This approach has recently been discussed by Tom and Farewell (2011) and by earlier authors such as Kay (1986), and Gentleman et al. (1994). In contexts where the internal covariate process is a marker of disease severity and broadly reflects the general trend in the disease process, Markov assumptions for such joint models may be reasonable. When the covariate reflects more transient acute disease activity the former approach may be preferred. We discuss such an approach further in the context of models for multiple disease processes in Section 4.

4 MULTIPLE PROGRESSIVE DISEASE PROCESSES

In studies involving paired or multiple organ systems, two or more processes may be of interest. In patients with psoriatic arthritis, onset of damage in the sacroiliac (SI) joints signals the onset of a spondyloarthropy, an arthritic condition of the vertebral joints. Damage may occur in the left SI joint, the right SI joint, or both. In the DCCT/EDIC study, retinopathy was measured in the left and right eyes and interest may lie in characterizing progression in both. Interest may also lie in joint modeling of retinopathy and nephropathy, both of which are of a vascular nature.

Let $\{Z_1(t), 0 < t\}$ and $\{Z_2(t), 0 < t\}$ denote two processes of interest; when they arise from paired organs the parameters governing the marginal processes may be the same, but for different but correlated processes they are not. We assume for convenience the processes have the same number of states, as in Figure 2. Interest lies in jointly modeling the processes to estimate the association between them, to consider the extent to which one process can be used to improve efficiency in estimation of another, and to enhance prediction.

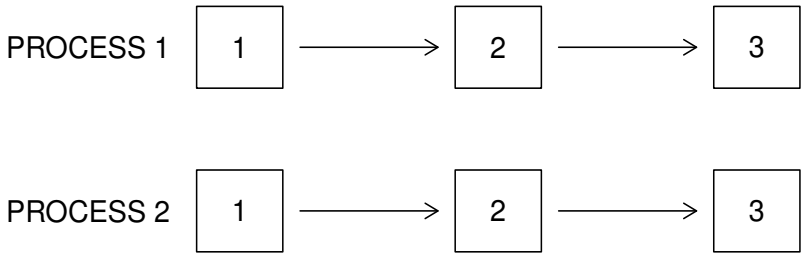

Figure 2: State space diagram for two parallel processes

Random effects modeling is often used. Let $\mathcal{H}^r(t) = \{Z_r(s), 0 < s < t, X\}$ denote the history for process r in the case where there are fixed covariates. Let $Y_j^r(s) = I(Z_r(s^-) = j)$, $r = 1, 2$, and

$$\lambda_{jk}^{(r)}(t|\mathcal{H}^r(t), \alpha_r) = \lim_{\Delta t \downarrow 0} \frac{P(Z_r(t + \Delta t^-) = k | Y_j^r(t) = 1, \mathcal{H}^r(t), \alpha_r)}{\Delta t} \tag{10}$$

denote the conditional intensity of a $j \to k$ transition for process r given the random effect α_r; we assume $E(\alpha_r) = 1$ and $\mathrm{var}(\alpha_r) = \phi_r$. A conditional Markov model for process r has intensity of the form

$$\lambda_{jk}^{(r)}(t|\mathcal{H}^r(t), \alpha_r) = \alpha_r Y_j^r(t)\, q_{jk}(t; X)\ .$$

If $\alpha_1 \perp \alpha_2 | x$, then the two processes are independent given X and if $\alpha_1 = \alpha_2$ the random effect is shared. If $\alpha = (\alpha_1, \alpha_2)'$ follows a bivariate distribution with $\mathrm{cov}(\alpha_1, \alpha_2) = \phi_{12} \neq 0$, then an

association is accommodated for the processes. It is important to note that this does not typically lead to easily interpreted measures of association, and the resulting marginal processes do not have simple forms (e.g. it is not possible to have Markov models for the marginal processes after marginalizing over the random effect), although as noted in Section 3.3, random effects models can be helpful with intermittently observed data. However, the measures of association they imply are really only valid with correct specification of the conditional intensities. The shared random effect model is more easily fit (Satten, 1999), but here, however, the parameter ϕ is a measure of history dependence as in (5) and also reflective of the association between the two processes, and making this model even more problematic when interest lies primarily in association.

Another framework for joint models arises by specification of an expanded state-space defined by combinations of the two processes; this was discussed in the context of time-varying covariates in Section 3.6. To this end we may define transition intensities in terms of

$$\lim_{\Delta t \downarrow 0} \frac{P(Z(t+\Delta t^-) = (j',k') | Z(t^-) = (j,k), \mathcal{H}(t))}{\Delta t} = Y_j^1(t) Y_k^2(t) \lambda_{jk,j'k'}(t|\mathcal{H}(t))$$

where $\mathcal{H}(t) = \{Z(s), 0 < s < t, X\}$. While this retains the simple form and interpretability of multiplicative intensity-based models, we pay the price of losing direct interpretation of covariate effects on features of the marginal processes. The same is true for more general models in which $\mathcal{H}^r(t)$ in (10) is replaced by $\mathcal{H}(t)$. We illustrate this point in the applications of Section 5.1.

Copula functions offer an appealing framework for constructing multivariate survival models (Nelsen, 2006), since the marginal distributions retain their simple interpretation. There has been little work to date on the use of copula functions with multistate models, but here we outline one such approach. Let T_{rk} be the time of entry into state k for process r, and for illustration, consider two 3-state processes depicted in Figure 2; we assume that each process is marginally Markov. Let $\mathcal{F}^r(t|X) = P(T_{r3} > t|X)$ denote the survival distribution for the entry time to state 3 for process r, $r = 1, 2$. In terms of the transition intensities,

$$\begin{aligned} \mathcal{F}^r(t|X) &= P_{11}^r(t|X) + P_{12}^r(t|X) \\ &= \exp\left(-Q_{12}^r(t|X)\right) + \int_0^t q_{12}^r(s|X) \exp\left(-Q_{12}^r(s|X)\right) \exp\left(-Q_{23}^r(t-s|X)\right) ds \end{aligned}$$

where $P_{jk}^r(t|X) = P(Z_r(t) = k | Z_r(0) = j, X)$ and $Q_{k\ell}^r(s|X) = \int_0^s q_{k\ell}^r(u|z)du$. A copula function $C(u_1, u_2; \tau)$ may be used to accommodate an association between T_{13} and T_{23} while retaining the Markov properties of the marginal processes. More generally we can do this for the time of entry to any state in a process. However, it is important to note that the different models for different pairs of states will not be compatible with any single model for the full bivariate process. This approach is most appealing when entry to a specific state is of special interest.

In some contexts interest may lie in simultaneous inferences regarding two or more processes, but not in the association between processes. In the DCCT study, for example, the effectiveness of the intensive glucose control program in delaying progression of nephropathy and retinopathy, was of interest. In this case a working independence assumption can furnish parameter estimates of effects on the marginal processes and robust sandwich-type variances estimates can be used to ensure valid simultaneous inferences regarding two or more processes. Lee and Kim (1998) describe such an approach for interval-censored data when the marginal processes are Markov; this approach is similar in spirit to the Wei et al. (1989) approach for multivariate failure time data.

5 Analyses and Empirical Studies

5.1 Axial Involvement in Psoriatic Arthritis

Here we consider a joint analysis of the onset and progression of damage in the sacroiliac (SI) joints among patients in the University of Toronto Psoriatic Arthritis Cohort with normal SI joints at clinic entry. Damage was assessed radiologically at visits scheduled biannually using the modified Steinbrocker method (Rahman et al., 1998). A joint was classified in state 1 if it was normal or if there was soft tissue swelling. Joints showing early signs of damage through surface erosions were classified in state 2, and joints with clinically important damage including joint space narrowing or "disorganization" were classified in state 3. The left and right SI joints were assessed at each radiological examination. The analysis here is for 538 patients who had two or more examinations.

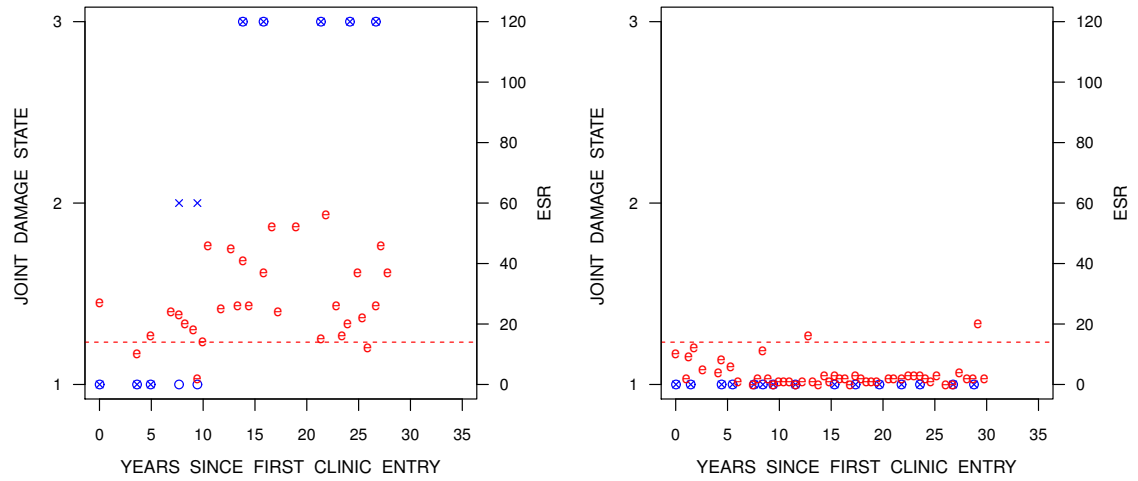

Figure 3: Plots of damage state for left (cross) and right (circle) sacroiliac joints with erythro sedimentation rates (ESR; mm/hour) for two individuals for the University of Toronto Psoriatic Arthritis Cohort

Figure 3 contains a plot of data from two individuals with roughly thirty years of follow-up and frequent measurement of the inflammatory biomarker erythro sedimentation rate (ESR). Individual A (on the left) experienced a long period of time with an elevated ESR value and both SI joints were observed to enter state 3; individual B (on the right) had consistently low values of ESR and no damage developed in the left or right SI joints over a period of 30 years.

Table 1 contains estimates from separate analysis of damage in the left ($r = 1$) and right ($r = 2$) SI joints. The set-up is as in Figure 2 and we fitted separate time-homogeneous Markov models as well as non-homogeneous models for which the transition intensities were piecewise-constant. The naive standard errors are relevant for inferences regarding the left or right SI joints themselves, but the robust covariance matrix would be required if interest lay in simultaneous inferences regarding the two processes. This would also be the case if regression models were specified and interest lay in joint inferences across the left and right sides.

The Markov model with piecewise constant intensities had cut-points at 5 and 10 years; the results are given in the bottom of Table 1. Based on this model we can test the hypothesis of time homogeneity. For the left SI joint we obtain a likelihood ratio statistic of 7.78 which gives $p = 0.1$, so there is insufficient evidence to reject the time homogeneous model. For the right SI joint we obtain a likelihood ratio statistic of 20.64 giving $p < 0.001$; hence there is evidence of a need to accommodate

time non-homogeneity in the transition intensities for the right SI joints. The p-values reported in Table 1 are for 2 degree of freedom tests directed at the specific intensities and reveal that the evidence against time homogeneity comes from the $1 \rightarrow 2$ transitions for the right SI joints.

Table 1: Parameter estimates, standard errors for transition intensities of three state model for damage in left and right sacroiliac joints in patients with psoriatic arthritis (n = 538); cut points for the piecewise constant time non-homogeneous model are 5 and 10 years

	LEFT $(r=1)$		RIGHT $(r=2)$	
Parameter	EST.	S.E.	EST.	S.E.
	TIME HOMOGENEOUS MODEL			
$\lambda_{12}^{(r)}$	0.033	0.003	0.038	0.003
$\lambda_{23}^{(r)}$	0.057	0.009	0.046	0.007
log L	-540.931		-581.810	
	PIECEWISE-CONSTANT MODEL			
$\lambda_{12,1}$	0.039	0.005	0.054	0.006
$\lambda_{12,2}$	0.031	0.007	0.021	0.006
$\lambda_{12,3}$	0.022	0.005	0.023	0.006
p-value*		0.0871		0.0001
$\lambda_{23,1}$	0.092	0.028	0.054	0.018
$\lambda_{23,2}$	0.043	0.017	0.037	0.014
$\lambda_{23,3}$	0.050	0.013	0.047	0.011
p-value*		0.2384		0.7883
log L	-537.039		-571.492	

p-value* is for a 2 d.f. test of the null hypotheses of a time-homogeneous intensity for the respective transition and joint

Figure 4 shows the state-space diagram for a joint Markov model for the left and right SI joints. From this we can estimate the 9×9 transition intensity matrix and extract features of the marginal processes. If we let $Z(t) = (Z_1(t), Z_2(t))$ where $\{Z_r(s), 0 < s\}$ is the process for the left $(r = 1)$ and right $(r = 2)$ SI joints, then for example

$$P(Z_1(t) = j | Z(0) = (1,1)) = \sum_{k=1}^{3} P(Z(t) = (j,k) | Z(0) = (1,1)) \; .$$

Figure 5 shows estimates of the marginal probabilities of damage for the left and right SI joints from a non-parametric estimate of the time to entry into state 3 (Turnbull, 1976), the probability $P(Z_r(t) = 3 | Z_r(0) = 1)$ from the marginal analysis, and the corresponding probability from the joint analysis based on the 9-state model. There is good agreement between the three estimates, particularly for the first 10 years of follow-up. The joint model is more reflective of the history of the two processes, and enables one to characterize the dependence between the processes in a convenient

way. For example, if as before $N_{jr}(t) = I(T_{jr} < t)$, we may write

$$\lambda_{jk,j+1,k}(t|\mathcal{H}(t)) = \lambda^{(1)}_{j,j+1}(t) \exp\left(\delta_{j2}N_{12}(t^-) + \delta_{j3}N_{23}(t^-)\right)$$

and

$$\lambda_{jk,j,k+1}(t|\mathcal{H}(t)) = \lambda^{(2)}_{k,k+1}(t) \exp\left(\gamma_{k2}N_{12}(t^-) + \gamma_{k3}N_{13}(t^-)\right) \ ,$$

where $\lambda^{(1)}_{j,j+1}(t)$ is the baseline intensity for transitions into state $j+1$ for left SI joints, δ_{j2} is the log relative intensity reflecting the effect of mild damage in the right SI joint (versus no damage) on $j \rightarrow j+1$ progression in the left SI joint; δ_{j3} reflects the further change in the intensity for progression in the left SI joint when the right SI joint has clinically important damage. If we let $\delta = (\delta_{12}, \delta_{13}, \delta_{22}, \delta_{23})'$ and $\gamma = (\gamma_{12}, \gamma_{13}, \gamma_{22}, \gamma_{23})'$, then testing $H_0 : \delta = \gamma = 0$ is a test of independence of the two processes; under H_0, the so-called naive analyses in Table 1 are valid. Aalen et al. (2008) and Aalen (2012) give a nice discussion about the utility of this model for conducting causal inference in the Granger school (Granger, 1969) through studying the nature of local-dependence.

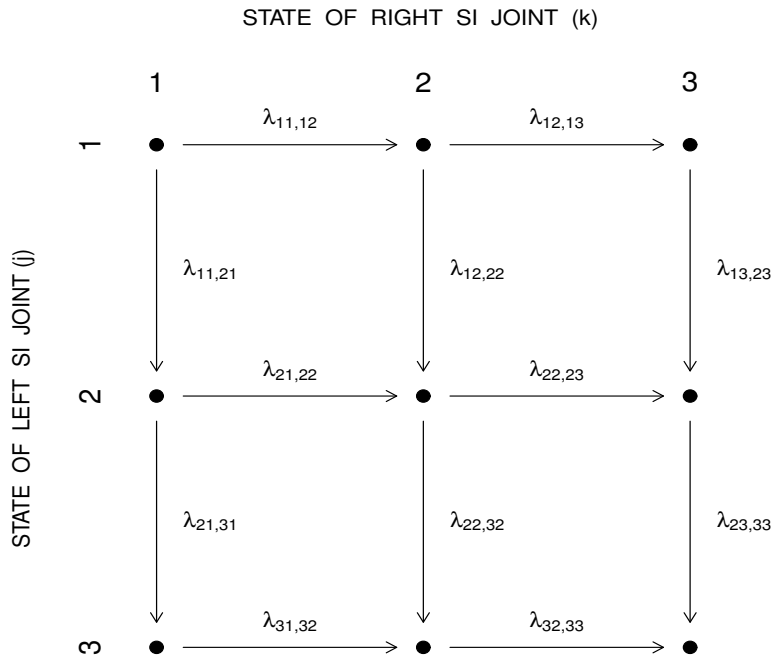

Figure 4: State space diagram for joint process for damage in left and right sacroiliac joints among patients with psoriatic arthritis

Relative risks are given in Table 2 based on δ and γ, where it is evident that progression of damage in one side has a highly significant effect on progression in the other side. For example, individuals with moderate damage (state 2) in the right SI joint have a highly significant 16-fold higher risk of moderate damage developing in the left SI joint. There is an insignificant further increase in risk upon occurrence of clinically important damage on the right, but there is very little information in the data concerning this. Those with clinically important damage on the right have a significant 9-fold increased risk of developing clinically important damage in the left over persons with moderate damage on the right. Interestingly there is evidence that any damage on the left side leads to higher

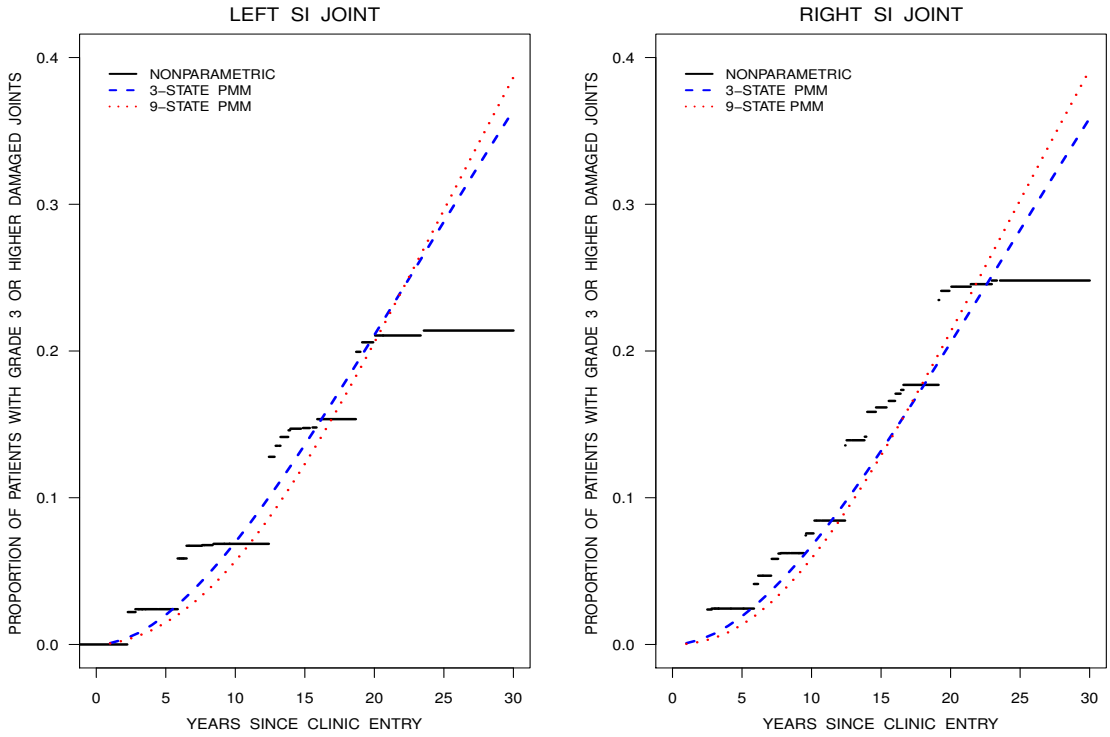

Figure 5: Plot of the cumulative probability of state 3 damage for the left and right SI joints from separate non-parametric analyses of time to State 3, separate marginal Markov models, and a joint 9-state Markov model

transition rates on the right side for all transitions with the exception of state 3 damage for $2 \rightarrow 3$ transitions.

An appealing feature of this formulation over a random effect formulation is that it accommodates an asymmetric dependence. If $\gamma = 0$, then the status of process 2 does not alter risks of progression in process 1, but if $\delta \neq 0$, changes in process 1 alter intensities for process 2. Such a structure is not accommodated in standard random effect models where associations are presumed to be symmetric and arise because of latent traits that have common effects; modification are possible which weaken the assumption of symmetry, but such models are not as appealing when interest lies in a more detailed understanding of dependencies.

Gamma random effects models as described in Section 3.3 were also fitted, separately for the left and right SI joints, and then a joint model as described in Section 4. In the separate fits, the variance parameter quantifies the dependence on t_{i1} on the intensity of $2 \rightarrow 3$ transitions as shown in (5). In both the left and right SI joints there is significant evidence of dependence; the null value in a test of $H_0 : \phi = 0$ is on the boundary of the parameter space and so the likelihood ratio statistic is a $50 : 50$ mixture of a point mass of zero and a χ^2_1 distribution; tests of $H_0 : \phi = 0$ yield $p = 0.0011$ and $p < 0.0001$ for the left and right SI joints, respectively. The (baseline) intensities in the time-homogeneous model and the corresponding parameters in gamma random effects models are not strictly comparable, however it is noteworthy that the rates of $2 \rightarrow 3$ transitions are higher than the $1 \rightarrow 2$ transitions in Table 1, but not in Table 2; this is a consequence of the form of (5). In the joint model the estimated transition intensities are broadly similar to those of the separate fits, but here the estimate of ϕ is considerably larger. This is in part a consequence of the dual role of ϕ in this model; it serves both as a measure of dependence as in (5) but also as a measure of association between the two processes. As discussed earlier, this makes interpretation of ϕ more challenging in this setting.

Techniques for modeling the development of axial involvement in psoriatic arthritis were consid-

Table 2: Multiplicative measures of local dependence between left and right SI joints from the 9-state Markov model

Side	Effect of Transition Type	Complementary Side State	RR	95% CI	p−value
Left	$1 \to 2$	2 vs. 1	16.418	(9.598 28.082)	< 0.0001
		3 vs. 2	0.029	(0, ∞)	0.8351
	$2 \to 3$	2 vs. 1	0.530	(0.186, 1.511)	0.2347
		3 vs. 2	9.391	(4.400, 20.045)	< 0.0001
Right	$1 \to 2$	2 vs. 1	4.707	(2.422 9.146)	< 0.0001
		3 vs. 2	10.777	(4.421 26.271)	< 0.0001
	$2 \to 3$	2 vs. 1	18.927	(1.429 250.754)	0.0257
		3 vs. 2	0.281	(0, ∞)	0.9328

Table 3: Estimates from fitting separate and joint gamma random effects models to left and right SI joints

		LEFT		RIGHT		JOINT	
	Parameter	EST.	S.E.	EST.	S.E.	EST.	S.E.
Left	$\lambda_{12}^{(r)}$	0.046	0.007			0.070	0.011
	$\lambda_{23}^{(r)}$	0.035	0.008			0.025	0.005
Right	$\lambda_{12}^{(r)}$			0.064	0.010	0.094	0.014
	$\lambda_{23}^{(r)}$			0.026	0.005	0.022	0.004
Variance	ϕ	1.629	0.649	2.175	0.658	4.555	0.621
log L		-535.602		-572.103		-995.8464	

ered by Chandran et al. (2010), who modeled a composite outcome based on left and right SI joints; they found a broadly comparable cumulative risk of spondylitis but did not explore the relation in progression in the left and right sides. Finally, we recommend the reader to O'Keeffe et al. (2011) who give an extended discussion on the progression of joint damage for persons in the PsA cohort considered here, focusing on joints in the hands. They use multistate models and, among other things, consider the possibility of causal relationships.

5.2 Models for Diabetic Retinopathy

We examine here some features associated with the progression of retinopathy in persons randomized to treatment in the DCCT trial introduced in Example 1.2. There were two cohorts in the DCCT: a Primary Intervention prevention (PI) cohort consisting of 726 individuals with no retinopathy and diabetes disease duration of 1–5 years at recruitment, and a Secondary Intervention (SI) cohort consisting of 715 individuals with mild retinopathy and disease duration of 1–15 years at recruitment.

Within each cohort individuals were randomized to Intensive Therapy (IT) or Conventional Therapy (CT) treatments; the experimental IT involved the administration of insulin three or more times daily and was intended to maintain blood glucose concentrations at close to normal levels. Within the primary prevention cohort, 348 (48%) individuals were randomized to the intensive therapy and within the secondary intervention cohort 363 (51%) were randomized to the intensive therapy. Recruitment occurred over 1983–1989 and subjects were followed until the termination of the trial in 1993. The Diabetes Control and Complications Trial Research Group (1993, 1995) describes findings from the trial concerning the progression of retinopathy and other complications, and concludes that while intensive therapy does not prevent retinopathy, it slows its progression.

The ETDRS scale used to measure retinopathy is the "final" scale in the DCCT (1995). It has 23 levels but this can be reduced to 12 levels for our purposes, with the final level 12 interpreted as 12 or higher. Broadly, level 1 is no retinopathy, levels 2 and 3 represent mild retinopathy, levels 4–9 represent increasing levels of moderate retinopathy, and levels 10 and over represent severe retinopathy. The measures here represent composite scores based on the assessment of both eyes, and as we discuss, there is considerable variation in ETDRS scores both within and between individuals. In DCCT (1993) an outcome termed "progression" was defined as a sustained change in ETDRS over consecutive semi-annual visits of three or more levels from the baseline measurement (at recruitment), and this was used for treatment comparisons. All individuals in the PI cohort had ETDRS = 1 (no retinopathy) at baseline and so for them "progression" represents a move to level 4 or higher. Later, in the DCCT (1995), the primary outcome for the PI cohort was defined as the first occurrence of an ETDRS score of 2 or higher on two consecutive 6-monthly visits. Our discussion here will focus on the issue of defining progression and, more specifically, primary outcomes for the comparison of treatment groups. For simplicity we will focus on the PI cohort.

We consider data from the DCCT followup period for the 651 white subjects among the 726 who were randomized to treatment in the PI cohort. Box plots of the ETDRS scores at 6-monthly follow-up visits for each treatment group (see Figure 6) show that median scores slowly move upwards over time, starting at level 1 and reaching level 2 in the IT group and level 4 in the CT group at 9 years from the start of the study. Variability at any given time is substantial.

Figure 7 shows plots of successive ETDRS scores for four individuals from each of the IT and CT groups. The scores for an individual are expected to increase over time, but there is considerable variability. In particular, an individual may show improvement (a lower ETDRS) from one visit to the next; this can be due to variation in the assignment of a discrete score on the basis of photos of the eyes, and to natural fluctuations in the condition of each eye.

We will explore progression of retinopathy through multistate models based on ETDRS scores. Multistate models have not been used in previous papers on retinopathy in this study. Such models provide a clearer picture of the longitudinal patterns of ETDRS scores and the trade-offs in defining

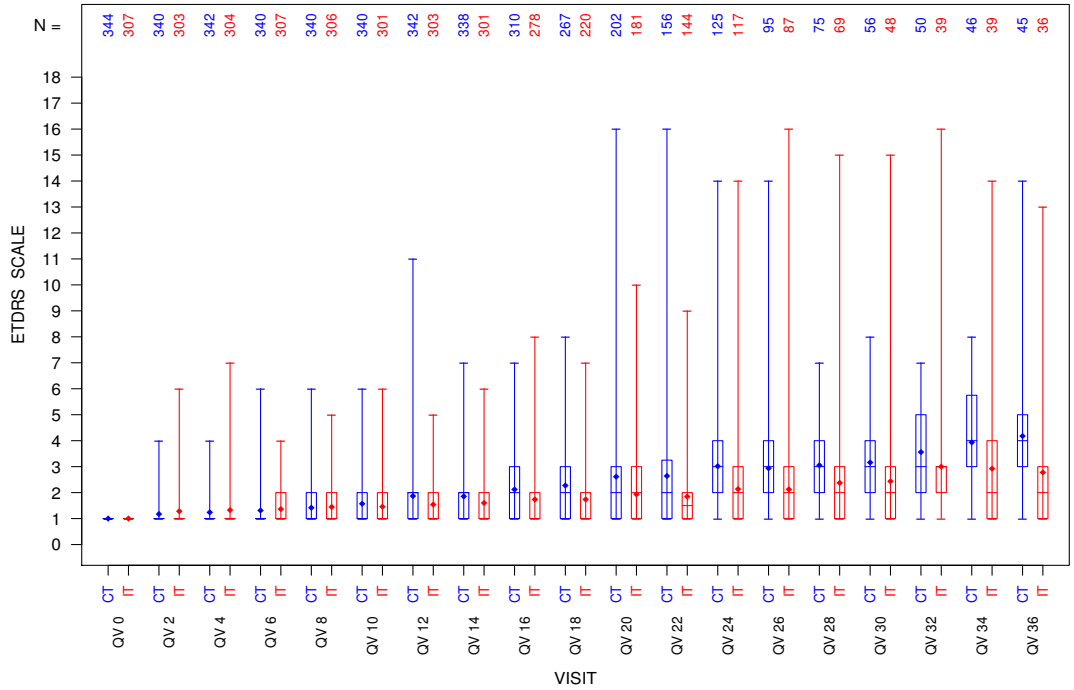

Figure 6: Box plots of the ETDRS scores for the CT (Blue Box) and IT (Red Box) treatment group. Lines show minimum, Q_{25}, median, Q_{75} and maximum scores; Diamond denotes the average value of ETDRS; Numbers denotes the total number of individuals in the indicated visit

an event termed "progression". We begin with three-state models in which state 1 = ETDRS 1–3, state 2 = ETDRS 4–9 and state 3 = ETDRS 10 or higher; the states correspond roughly to no or mild, moderate and severe retinopathy, respectively. One advantage of this categorization is that everyone in the PI cohort starts in state 1, and a move to state 2 corresponds to an increase in ETDRS of 3 or more from baseline, consistent with the early definition of progression (DCCT, 1993). However, diagnostic checks suggested that models with more states would be preferable, and we report here on 5-state models with states as follows: 1 =ETDRS 1, 2 =ETDRS 2–3, 3 =ETDRS 4–6, 4 =ETDRS 7–9, and 5 =ETDRS ≥ 10. Given the variability in longitudinal ETDRS patterns exemplified in Figure 7, there is no clearly superior way to specify a single outcome that can be used for treatment comparisons even in the PI cohort. We consider here a number of Markov models and then discuss their utility as follows: Model M1 – transitions from states to any adjacent states are possible (reversible Markov model, RMM); Model M2 – only transitions from a state to the next higher state are possible, and once an individual progresses to the next higher state they stay there (progressive Markov model, PMM); Model M3 – the same as M2 but a person is considered to have progressed to the next state only when they have been in the state for two consecutive 6-monthly visits (sustained progressive Markov model, SPMM); Model M4 – a hidden Markov model in which the true retinopathy state follows the progressive 5-state model but where the observed state at any time is considered an imperfect measure of the true state (Jackson and Sharples, 2002; Jackson et al., 2003).

For the 651 white subjects who were randomized to treatment in the PI cohort, we now consider separate models for the two treatment groups (Conventional - CT, Experimental - IT). It was apparent from fits of time-homogeneous Markov models that some time-dependence should be incorporated, and we discuss results here for models in which the intensities were constants over 0-4 years from randomization and (different) constants beyond 4 years. Table 4 shows parameter estimates from

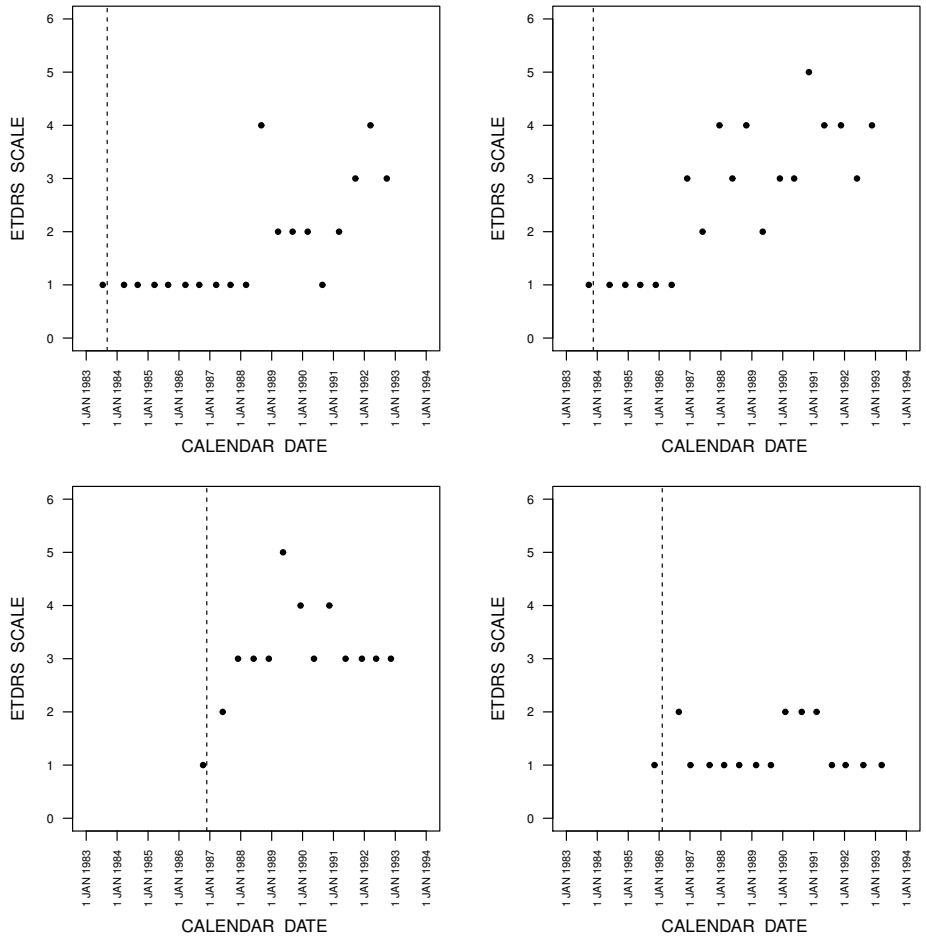

Figure 7: Profile plots of two individuals in the primary intervention cohort: IT (Left Panel) and CT (Right Panel) group

models M1 to M3. Plots of observed transition counts (not shown) show that the reversible Markov model M1 mimics the observed data quite well and it reflects the fact that many subjects in each group move both up and down between states over time.

The downward transition intensities are larger than the upward intensities from the same state in most cases, indicating that a higher level of retinopathy is less likely than some degree of recovery. An exception is that the 2 to 1 intensities are smaller than the 2 to 3 intensities after 4 years in the CT group; this reflects the fact that eventually, many subjects (especially in the CT group) experience some degree of retinopathy. Some transition rates involving states 4 and 5 are essentially inestimable because almost no direct transitions were observed and to achieve estimability we restricted some intensities to be equal (see Table 4). Models M2 and M3 were fitted to derived transition data in which a subject is assumed to stay in a higher state from the time they move to it. The results from the two models are similar, and there is relatively little difference whether we consider progression to have occurred when it is first observed (M2) or when it is first observed on two consecutive visits. Models M2 and M3 agree well with the (derived) data in terms of transition counts (results not shown). We note that models M1, M2 and M3 use three different data sets, so it is not possible to compare maximized log likelihoods or AIC values. Since all three provide reasonable fits to the corresponding observed data, we now consider them as a basis for comparison of the treatment groups.

The DCCT (1993) compared the two groups by considering the time of ordinary or sustained 3-step progression of ETDRS, treating the endpoint as a survival time. This can also be done using models M2 or M3, by considering time to entry to state 3; in either model the probability of entry to state 3 by time t is given by $P_{13}(t) + P_{14}(t) + P_{15}(t)$. Figure 8 shows plots based on model M2 and a nonparametric estimate that treats time of first entry to state 3 as interval-censored (Turnbull, 1976).

Plots for sustained progression (Model M3) in the bottom panels of Figure 8 look very similar. There is a substantial difference between the treatment groups, with the intensive treatment associated with much slower progression. We also show in the top two panels of Figure 8 the estimated distribution of time to first entry into state 3 based on the reversible model M1. These estimates are well above the other two. This is at least in part due to the intermittent (6-monthly) observation of retinopathy. Model M2 assumes that the first entry to state 3 is when it is first observed; however, it is possible that first entry occurred earlier but was unobserved because the individual returned to state 2 (or 1) before the next observation time. Model M1 allows for this and the estimates in Figure 8 reflect the fact that first entry to state 3 may precede the first observed entry. An "intermediate" reversible model in which transitions from state 2 to state 1 are allowed but not transitions from state 3 to state 2, gives estimates close to those from model M2.

An examination of transition intensities in models M1 to M3 provides additional comparisons and if desired, a formal test of no treatment effect can be based on the hypothesis of no difference between the two groups. For M2 and M3 the IT group has intensities over the first 4 years that are comparable to those for the CT group; this is consistent with the observation of some initial worsening in the IT group (The Diabetes Control and Complications Trial Research Group, 1998). After 4 years all IT group intensities are substantially lower than the CT intensities, indicating a slowing of the rate of progression at all stages of retinopathy. The reversible model M1 better represents the observed patterns in the ETDRS scores, but does not provide simple treatment comparisons. However, as would be expected, the tendency to move down a state decreases with time, and tends to be higher in the IT group, but the tendency to move up is, after the first few years, lower in the IT group.

The hidden Markov model M4 was also fitted to the raw data. It gives substantially lower rates of progression to states 2 and 3 for the IT group than for the CT group; the results are qualitatively similar, but not directly comparable, to those for models M2 and M3. Model M4 is not very appealing here because it compares treatments in terms of an unobservable process. In addition, some of the misclassification probabilities are rather large, reflecting not just measurement error but also the substantial up and down fluctuations in actual ETDRS scores. A further complication is that the

misclassification probabilities could be expected to vary with time, because as time passes many individuals are expected to "progress" within a state in the sense that they are more likely to have an ETDRS score at the top end of the range for the state. It is considerably more difficult to fit models with this feature. The interpretation of parameter estimates in fitted models M4 is therefore awkward and we do not show them in Table 4.

In the PI cohort considered here, everyone began with an ETDRS score of 1 whereas in the SI cohort subjects had higher scores at baseline. The specification of a treatment effect is more complex in this case. In addition, there were few persons in the PI cohort who progressed to ETDRS scores of 10 or higher during DCCT followup; more information about progression to higher scores is present in the SI cohort and in the EDIC study that follows the DCCT. The analysis just described did not consider covariates because we want to focus on progression and on the comparison of treatment groups. Analyses directed at understanding the disease process would incorporate covariates and in particular, the time-varying biomarker glycosylated hemoglobin percentage (HbA1c). It is predictive of retinopathic progression, and the intensive therapy was designed to maintain glucose levels and this biomarker at near normal levels. An assessment of treatment within models that contain HbA1c is complicated because treatment affects the biomarker. One option is to define direct versus indirect treatment effects in the sense described by Aalen et al. (2008, Section 9.3.2). However, the net effect of treatment on retinopathy, whether acting through HbA1c or otherwise, is clear from the marginal models considered above.

6 Outcome-Dependent Observation and Selection Processes

In many settings an individual's selection for a study, or the chance of premature loss to followup (LTF) or missing data, may depend on their disease process. A thorough discussion of these issues is beyond our present scope but because of their importance we discuss a few key points.

6.1 Outcome-Dependent Inspection Times or Loss to Followup

We first consider the case where observation times, $v_1 < \cdots < v_n$ for an individual are prespecified. Using the notation of earlier sections, we suppose that models for $P(Z(v_r)|H(v_r))$ are of interest, and that if there is no missing data the likelihood function (2) provides consistent estimation of model parameters θ_0. To allow for missed visits we let $R(v_r)$ equal 1 if the individual is seen at time v_r and 0 otherwise. To discuss LTF, suppose that an individual is treated as LTF (not seen again) at the first time v_r with $R(v_r) = 0$. Provided that $R(v_r)$ is conditionally independent of $Z(v_s)$ and $X(v_s)$ for $s \geq r$, given $H(v_r)$, we can use the censored partial likelihood analogous to (2):

$$L(\theta) = \prod_{r=1}^{C} P(Z(V_r)|H(v_r);\theta) \ , \tag{11}$$

where $C = \max(v_r : R(v_r) = 1)$. If the preceding SMAR conditional does not hold but there exists a vector $X^c(t)$ of observed explanatory variables such that $R(v_r)$ is conditionally independent of $Z(v_s)$ and $X(v_s)$ for $s \geq r$, given $X^c(v_r)$ and $H(v_r)$, then inverse probability of censoring weights (IPCW) can be used to adjust the log-likelihood or score components from (11). This requires specification of a model for $R(v_r)$ given $H(v_r)$ and $X^c(v_r)$ (Robins et al., 1995; Hajducek and Lawless, 2012).

If an individual misses occasional visits, considerable information might be lost if they are treated as LTF at the first missed visit. A SMAR assumption, that $Z(v_r)$ and $R(v_r)$ are conditionally independent given $H(v_r)$, in principle allows "non-monotone" missing data patterns to be handled by maximum likelihood (e.g. Fitzmaurice et al., 2009, Chapters 17, 22) but this is often intractable. In recent work on multistate models (e.g. Chen et al., 2010; Sweeting et al., 2010) likelihood functions

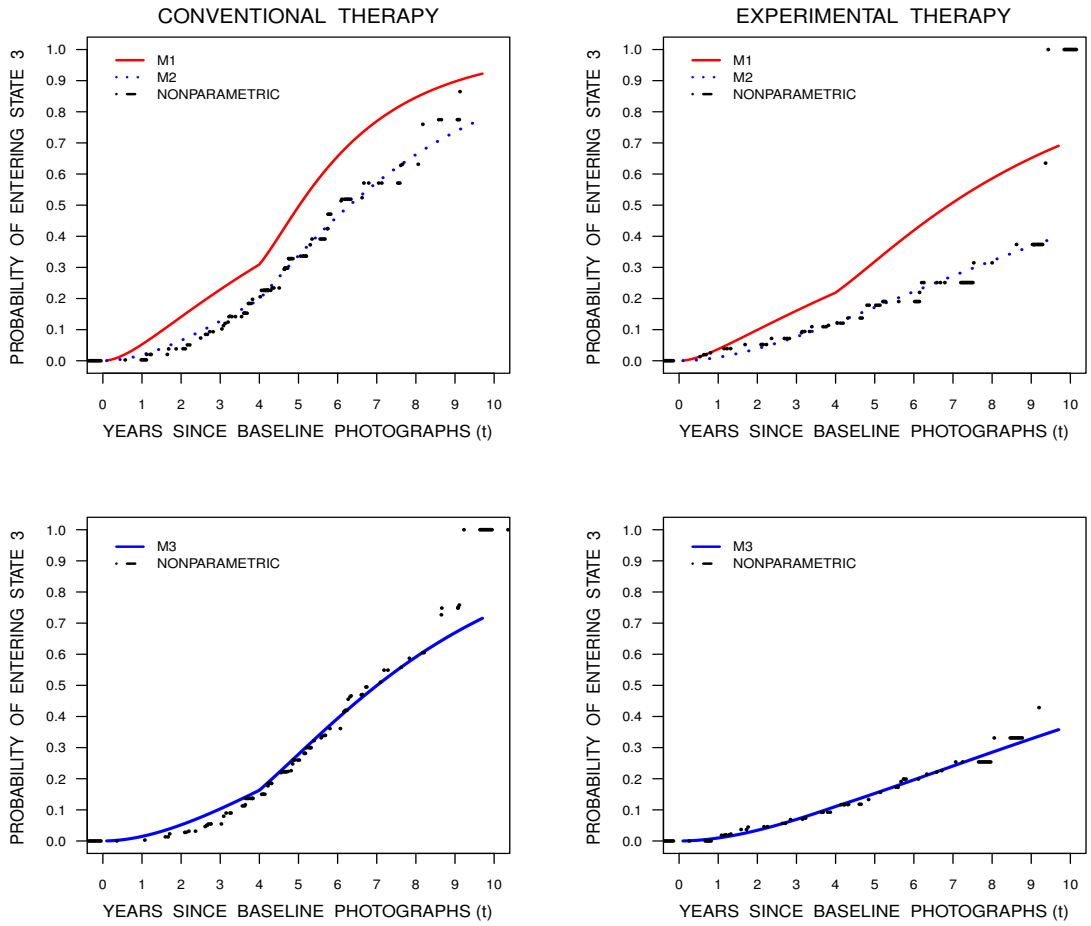

Figure 8: Plots of the probability of entering state 3 (ETDRS $4 - -6$) over time for conventional therapy group (left panels) and experimental therapy group (right panels) using the estimates obtained from fitting time non-homogeneous M1 (reversible Markov model), M2 (progressive Markov model), and M3 (sustained progressive Markov model) with corresponding nonparametric estimates

Table 4: Estimates of transition intensities and 95% confidence intervals for reversible Markov model, progressive Markov model, and sustained progressive Markov model fitted to data from the primary prevention cohort in the DCCT period (n = 651); cut point for the piecewise constant time non-homogeneous model are 4 years

			CONVENTIONAL			EXPERIMENTAL		
Model†	Period	Parameter	EST.	95% CI	p*	EST.	95% CI	p*
M1	$[0,4)$	λ_{12}	0.496	(0.442,0.557)	< 0.0001	0.557	(0.492,0.630)	0.4934
		λ_{21}	0.825	(0.705,0.967)	0.0008	1.082	(0.931,1.259)	< 0.0001
		λ_{23}	0.348	(0.262,0.463)	< 0.0001	0.233	(0.158,0.343)	0.0635
		λ_{32}	2.032	(1.458,2.833)	0.0002	2.637	(1.752,3.969)	0.2425
		$\lambda_{34} = \lambda_{45}^{\ddagger}$	0.144	(0.033,0.622)	0.7700	0.118	(0.020,0.699)	0.9836
		$\lambda_{43} = \lambda_{54}^{\ddagger}$	3.069	(0.481,19.573)	0.5413	0.749	(0.001,669.887)	0.8718
	$[4,\infty)$	λ_{12}	1.084	(0.907,1.295)		0.512	(0.417,0.627)	
		λ_{21}	0.511	(0.407,0.640)		0.398	(0.314,0.504)	
		λ_{23}	0.706	(0.586,0.850)		0.376	(0.275,0.514)	
		λ_{32}	0.925	(0.727,1.175)		1.905	(1.352,2.686)	
		$\lambda_{34} = \lambda_{45}^{\dagger}$	0.182	(0.109,0.305)		0.115	(0.044,0.300)	
		$\lambda_{43} = \lambda_{54}^{\ddagger}$	1.648	(0.844,3.218)		0.424	(0.114,1.570)	
log L				-2762.282			-2173.535	
M2	$[0,4)$	λ_{12}	0.326	(0.288,0.368)	0.0443	0.344	(0.302,0.392)	0.6939
		λ_{23}	0.135	(0.105,0.174)	< 0.0001	0.074	(0.053,0.103)	0.8075
		$\lambda_{34} = \lambda_{45}^{\dagger}$	0.029	(0.008,0.112)	0.3181	0.037	(0.009,0.142)	0.7689
	$[4,\infty)$	λ_{12}	0.478	(0.338,0.675)		0.317	(0.215,0.467)	
		λ_{23}	0.273	(0.218,0.341)		0.079	(0.054,0.116)	
		$\lambda_{34} = \lambda_{45}^{\dagger}$	0.062	(0.037,0.101)		0.028	(0.009,0.084)	
log L				-1388.908			-992.704	
M3	$[0,4)$	λ_{12}	0.251	(0.220,0.285)	< 0.0001	0.271	(0.237,0.311)	0.9435
		λ_{23}	0.133	(0.099,0.177)	0.0030	0.080	(0.055,0.114)	0.7655
		$\lambda_{34} = \lambda_{45}^{\dagger}$	0.038	(0.010,0.153)	0 .6992	0.020	(0.003,0.141)	0.7724
	$[4,\infty)$	λ_{12}	0.513	(0.386,0.681)		0.267	(0.180,0.396)	
		λ_{23}	0.235	(0.186,0.296)		0.073	(0.049,0.109)	
		$\lambda_{34} = \lambda_{45}^{\dagger}$	0.051	(0.028,0.093)		0.028	(0.009,0.086)	
log L				-1348.579			-985.507	

† M1 - Reversible Markov model; M2 - Progressive Markov model; M3 - Sustained Progressive Markov model. ‡ Parameters constrained to be the same. * p−value for test of time homogeneity in corresponding transition intensity.

involve multiple sums over states $Z(v_r)$ at times v_r with $R(v_r) = 0$; for complex models and general patterns of missingness these are complicated to compute unless the total number of visits n is small.

When observation times are widely spaced it is likely that SMAR assumptions are at least to some extent violated, with the probability of a missed visit at v_r depending on disease history or covariate values over the interval $(v_{r-1}, v_r]$. Many authors (including Chen et al., 2010 and Sweeting et al., 2010) have proposed "not missing at random" (NMAR) models, typically through parametric assumptions about the distribution of $R(v_r)$ given $Z(v_r)$ and $H(v_r)$. However, such models cannot be checked using the observed data (Fitzmaurice et al., 2009, Ch. 22). NMAR models are sometime useful for sensitivity analysis concerning LTF (e.g. Barrett et al., 2011; Lawless, 2012) but in settings with complex patterns of missingness the models are generally so complex that transparent sensitivity analysis is impossible. It is best to minimize the occurrence of missing data, and to identify and record covariates that make SMAR assumptions plausible.

In observational studies the visit times v_{ir} may vary widely within and between individuals. If it can be assumed as in Section 2.2 that v_{ir} is scheduled on the basis of data observed up to time $v_{i,r-1}$ then the likelihood function (2) holds in the absence of LTF. If there are wide departures from scheduled visit times then NMAR concerns arise. For example, Lawless et al. (2001) consider a cohort of children with hydrocephalus, with individuals scheduled to return for clinic visits every 6 or 12 months. However, at the administrative end of follow-up for a particular study it was found that many individuals had not been seen for over two or three years. In such cases it is useful to budget for tracing of some such individuals; Farewell et al. (2003) describe this in connection with a psoriatic arthritis cohort.

6.2 Outcome-Dependent Selection

Chronic disease studies often involve conditions for recruitment. Suppose that an individual experiences onset of a chronic disease process at age a_0 and that they are selected for a study at age a_1. If $a_1 \leq a_0$ we have full prospective information on disease history (subject to right censoring) and inference can be based on $P(H(t), t > a_0 | H(a_0))$, where $H(a_0)$ includes covariates related to selection conditions. If $a_1 > a_0$ the prospective likelihood based on terms $P(H(t), t > a_1 | H(a_1))$ can be used, subject to the availability of relevant information in $H(a_1)$, as discussed in Section 3.2. Matters are much more complicated if we wish to use data or the disease process prior to selection as a response; this can occur when there are rich retrospective data and the duration of followup after selection is limited. For example, Kvist et al. (2010) describe studies on hospitalization of persons with psychiatric affective disorders. In one study (Kessing et al., 2004) individuals were included if they were admitted to a specific set of hospitals during the years 1959 to 1963. They were followed forward until 1989, but data on disease history and admissions prior to 1959 were also available from medical records. If we wish to treat all of the data for an individual as a response instead of just the data following their (initial) admission during 1959–1963, we need to consider $P(H(a_3)|H(a_0)$, at least one admission in $1959-1963)$, where a_3 is the individual's age at the end of 1989. Except in special cases involving simple progressive disease models (e.g. Copas and Farewell, 2001), there has been little study of effective ways to deal with such likelihoods.

6.3 Outcome-Based Subsampling

Outcome-based subsampling is frequently used for the measurement of expensive covariates of selected members of a cohort. For studies involving a single event or survival time, case-cohort and case-control designs are familiar and methods of analysis based on Cox models have been developed (e.g. Samuelsen et al., 2007; Breslow et al., 2009). For example, some studies of the genetics of kidney disease in Type 1 diabetics have selected cohort members with short times to severe nephropathy

("cases") and members without severe nephropathy ("control") for genotyping (e.g. Mueller et al., 2006).

The use of similar designs in connection with more complicated disease history models has received little attention. For example, consider a simple plan where at a common visit time v_r a subset of cohort members is selected and variables W are measured on them. Let $R = 1$ indicate that individual i is selected and $R = 0$ that they are not, and for convenience let Z and X refer to an individual's full history $\{Z(v_s), X(v_s), s = 0, 1, \ldots, n\}$ before and after v_r. Assuming that R depends only on observed information in (Z, X) up to time v_r, the likelihood function

$$L = \{P(Z, X|W)g(W)\}^R \{P(Z, X)\}^{1-R} \tag{12}$$

can be used (e.g. Lawless et al., 1999), where $g(W)$ is the probability mass function for W in the cohort, assuming for simplicity that W is discrete. There are challenges in using (12), however. In particular, we need to specify and estimate $g(w)$, since

$$P(Z, X) = \sum_w P(Z, X|w)g(w) \ .$$

In addition we need to consider how $P(Z, X|w)$ is modeled. If we are primarily interested in $P(Z|X, w)$ then it is necessary to specify a model for $P(X|w)$ in order to compute (12). In cases where X involves time-varying factors this may be of some interest. If X represents only baseline factors, then an alternative to modeling $P(X|w)$ is to model $g(w|X) = P(W = w|X = x)$ and to replace (12) with a likelihood conditional on X. Alternatives that involve weighted and/or augmented estimating function methods provide other options that have not been investigated much in the current framework. Breslow et al. (2009) and Lumley et al. (2011) provide examples involving single event times.

7 Discussion

Multistate models offer a convenient framework for the analysis of chronic disease processes and there are many excellent review papers on this theme (Commenges, 1999; Hougaard, 1999; Andersen, 2002). Intensity-based models can provide very useful information on the dynamic aspects of a disease process including the nature of the dependence on previous events and the effect of internal and external time-varying covariates. Undertaking such an analysis requires careful thought about how best to formulate models, including how to define states, which time scale to adopt, how to formulate the dependence on covariates, etc. Fitting semiparametric models is particularly challenging with interval-censored data but convenient flexible alternatives include methods based on piecewise-constant baseline intensities; Joly et al. (2002) develop intensity-based methods based on splines. Heterogeneity poses another modeling challenge and when random effects models are adopted care must be taken to ensure inferences about heterogeneity are not unduly influenced by misspecification of the model conditional on the random effects. When multiple aspects of a disease process are of interest, as in the survival setting, simultaneous inference can be based on (correlated or shared) random effect models, copula models or models based on a "working independence" assumption with robust variance estimates.

Model assessment has not been discussed in this paper. In general, we recommend this be done through the usual device of model expansion, particularly when there are covariates. In settings where processes are observed continuously, residual plots and other techniques for model assessment in survival analysis can be used, since sojourns in a state are equivalent to survival times in a competing risk model. When Markov models are adopted for processes under intermittent observation, it is possible to compare expected (model-based) transition counts and corresponding observed counts in

some cases (Titman and Sharples, 2010b), and some such checks are provided by the `msm` package (Jackson, 2011). Further exploration of model checking methods would be valuable.

Many progressive chronic diseases involve a gradual deterioration in the function of one or more organ systems (e.g. eyes, kidneys) over a long time horizon. When assessments are carried out frequently, data suggesting a short-term improvement may be obtained and decisions are required about how to handle such information. Reasons for the apparent improvement include a legitimate reversal of the underlying condition or an imperfect measurement process leading to spurious evidence of improvement. In the former case, models accommodating transitions to better health states may be appealing, but models based on latent processes (e.g. hidden Markov models) may be preferred in the latter. When serum or urinary markers of disease severity are to be used short-term variation may be considerable. In studies of HIV/AIDS, for example, markers of immune function (e.g. CD4 cell counts) have been used effectively to characterize the disease course, but these are, in some sense, surrogate data to the more difficult to observe clinically important events. The considerable variation in these measurements motivates the use of methods which accommodate measurement error (Satten and Longini, 1996).

Aalen et al. (2001) and Datta and Satten (2001) showed that with right-censored multistate data the usual Aalen-Johansen estimator of the state occupancy probabilities, motivated by Markov assumptions, are robust and valid for non-Markov models when censoring is completely independent. When states are intermittently observed nonparametric estimation based on Markov assumptions is challenging (Frydman, 1992, 1995; Frydman and Szarek, 2009); exploration of robustness in this context is warranted.

Regression models involving multiplicative covariate effects are routinely specified but there has been increased interest in the use of additive models (Aalen, 1989; Martinussen and Scheike, 2006), or hybrid models (Scheike and Zhang, 2002). There is scope for development of methods based on these alternative specification in the context of multistate analysis. When interest lies in formally comparing two or more treatment groups, intensity-based models are less natural since they inherently involve conditioning on the process history through internal and responsive covariates (Kalbfleisch and Prentice, 2002). Here alternative frameworks for treatment comparisons are of interest. A series of papers beginning with Andersen et al. (2003) explored the quite different approach of using of resampling methods and model fitting based on pseudo-observations to model covariate effects on state occupancy probabilities at particular time points with right-censored data. There is some appeal to this in that analysis is carried out as in generalized linear models and the dependencies in estimators from fits at different time points are accommodated by use of robust variance estimates. Little work has been done to adapt these methods to interval-censored data and this is an area warranting development.

ACKNOWLEDGEMENTS

We thank the Editor and Jianwen Cai for inviting this article in connection with the Symposium on Emerging Methodological Issues in Population-Based Chronic Disease Research, in honor of Ross Prentice (October 20–21, 2011). This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Canadian Institutes for Health Research. Richard Cook is a Canada Research Chair in Statistical Methods for Health Research. The authors thank Dr. Dafna Gladman and Dr. Vinod Chandran for collaboration and helpful discussions regarding the research at the Centre for Prognosis Studies in Rheumatic Disease at the University of Toronto. The Diabetes Control and Complications Trial was sponsored by the Division of Diabetes, Endocrinology, and Metabolic Diseases of the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institutes of Health, through cooperative agreements and a research grant. Additional support was provided by the National Heart, Lung, and Blood Institute, the National Eye Institute, and the National Center for Research Resources.

REFERENCES

Aalen OO (1989) A linear regression model for the analysis of life times. Statistics in Medicine 8:907–925

Aalen OO, Borgan O, Fekjaer H (2001) Covariate adjustment of event histories estimated with Markov chains: The additive approach. Biometrics 57: 993–1001

Aalen OO, Borgan O, Gjessing HK (2008) Survival and Event History Analysis: A Process Point of View. Springer Science + Business Media, LLC, New York, NY

Aalen OO (2012) Armitage lecture 2010: Understanding treatment effects: The value of integrating longitudinal data and survival analysis. Statistics in Medicine (in press)

Al-Kateb H, Boright AP, Mirea L, Xie X, Sutradhar R, Mowjoodi A, Bharaj B, Liu M, Bucksa JM, Arends VL, Steffes MW, Cleary PA, Sun W, Lachin JM, Thorner PS, Ho M, McKnight AJ, Maxwell AP, Savage DA, Kidd KK, Kidd JR, Speed WC, Orchard TJ, Miller RG, Sun L, Bull SB, Paterson AD and the Diabetes Control and Complications Trial/Epidemiology of Diabetes Interventions and Complications Research Group (2008) Multiple superoxide dismutase 1/splicing factor serine alanine 15 variants are associated with the development and progression of diabetic nephropathy: The diabetes control and complications trial/epidemiology of diabetes interventions and complications genetics study. Diabetes 57: 218–228

Andersen PK, Borgan O, Gill RD, Keiding N (1993) Statistical Models based on Counting Processes. Springer-Verlag, New York

Andersen PK (2002) Multi-state models for event history analysis. Statistical Methods in Medical Research 11: 91–115

Andersen PK, Klein JP and Rosthoj S (2003) Generalised linear models for correlated pseudo-observations, with applications to multi-state models. Biometrika 90: 15–27

Bacchetti P, Boylan RD, Terrault NA, Monto A, Berenguer M (2010) Non-Markov multistate modeling using time-varying covariates, with application to progression of liver fibrosis due to Hepatitis C following liver transplant. The International Journal of Biostatistics 6 (1): Article 7

Barrett JK, Siannis F, Farewell VT (2011) A semi-competing risks model for data with interval-censoring and informative observation: An application to the MRC cognitive function and ageing study. Statistics in Medicine 30: 1–10

Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M (2009) Using the whole cohort in the analysis of case-cohort data. American Journal of Epidemiology 169 (11): 1398–1405

Bureau A, Shiboski S, Hughes JP (2003) Applications of continuous time hidden Markov models to the study of misclassified disease outcomes. Statistics in Medicine 22 (3): 441–462

Chandran V, Cook RJ, Edwin J, Shen H, Pellett FJ, Shanmugarajah S, Rosen CF, Gladman DD (2010) Soluble biomarkers differentiate patients with psoriatic arthritis from those with psoriasis without arthritis. Rheumatology 49 (7): 1399–1405

Chandran V, Tolusso DC, Cook RJ and Gladman DD. (2010). Risk factors for axial inflammatory arthritis in patients with psoriatic arthritis. Journal of Rheumatology 37: 809–815.

Chandran V, Cook RJ, Thavaneswaran A, Lee KA, Pellett FJ, Gladman DD (2012) Parametric survival analysis as well as multi-state analysis confirms the association between human leukocyte antigen alleles and the development of arthritis mutilans in patients with psoriatic arthritis. Journal of Rheumatology (in press)

Chen B, Yi GY, Cook RJ (2010) Analysis of interval-censored disease progression data via multi-state models under a nonignorable inspection process. Statistics in Medicine 29 (11): 1175–1189

Commenges D (1999) Multi-state models in epidemiology. Lifetime Data Analysis 5: 315–327

Cook RJ, Lawless JF (2007) The Statistical Analysis of Recurrent Events. Springer Science + Business Media, LLC, New York, NY

Cook RJ, Lawless JF, Lakhal-Chaieb L, Lee K-A (2009) Robust estimation of mean functions and treatment effects for recurrent events under event-dependent censoring and termination: application to skeletal complications in cancer metastatic to bone. Journal of the American Statistical Association 104 (485): 60–75

Copas AJ, Farewell VT (2001) Incorporating retrospective data into an analysis of time to illness. Biostatistics 2 (1): 1–12

Cox DR, Miller HD (1965) The Theory of Stochastic Processes. Chapman and Hall, London

Cox DR (1975). Partial likelihood. Biometrika 62: 269–276.

Datta S, Satten GA (2001) Validity of the Aalen-Johansen Estimators of Stage Occupation Probabilities and Nelson-Aalen Estimators of Integrated Transition Hazards for non-Markov Models, Statistics and Probability Letters 55: 403–411

Early Treatment Diabetic Retinopathy Study Research Group (1991) Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12. Ophthalmology 98 (Suppl): 823–833

Fahrmeir L, Tutz G (2001) Multivariate Statistical Modeling Based on Generalized Linear Models. Springer, New York

Farewell VT, Lawless JF, Gladman DD, Urowitz MB (2003) Tracing studies and analysis of the effect of loss to follow-up on mortality estimation from patient registry data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 52 (4): 445–456

Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (2009) Longitudinal Data Analysis. Handbooks of Modern Statistical Methods, Chapman and Hall/CRC, Boca Raton.

Frydman H (1984) Maximum likelihood estimation in the mover-stayer model. Journal of the American Statistical Association 79 (387): 632–638

Frydman H (1992) A nonparametric estimation procedure for a periodically observed three-state Markov process, with applications to AIDS. Journal of the Royal Statistical Society: Series B 54: 853–866

Frydman H (1995) Semiparametric estimation in a three-state duration-dependent Markov model from interval-censored observations with application to AIDS data. Biometrics 51: 502–511

Frydman H, Szarek M (2009) Nonparametric estimation in a Markov “illness-death” process from interval censored observations with missing intermediate transition status. Biometrics 65: 142–151

Gentleman RC, Lawless JF, Lindsey JC, Yan P (1994) Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. Statistics in Medicine 13 (8): 805–821

Gladman DD, Farewell VT, Nadeau C (1995) Clinical indicators of progression in psoriatic arthritis : multivariate relative risk model. Journal of Rheumatology 22 (4): 675–679

Goodman LA (1961) Statistical methods for the mover-stayer model. Journal of the American Statistical Association 56: 841–868

Granger CWJ (1969) Investigating causal relationships by econometric models and cross-spectral methods. Econometrica 37: 424–438

Grüger J, Kay R, Schumacher M (1991) The validity of inferences based on incomplete observations in disease state models. Biometrics 47: 595–605

Hajducek DM, Lawless JF (2012) Duration analysis in longitudinal studies with intermittent observation times and losses to followup. Canadian Journal of Statistics 40 (1): 1–21

Hogan JW, Roy J and Korkontzelou C (2004) Handling dropouts in longitudinal studies. Statistics in Medicine 23: 1455–1497.

Hougaard P (1999) Multi-state models. A review. Lifetime Data Analysis 5: 239–264

Jackson CH, Sharples LD (2002) Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. Statistics in Medicine 21: 113–128

Jackson CH, Sharples LD, Thompson SG, Duffy SW, Couto E (2003) Multistate Markov models for disease progression with classification error. Journal of the Royal Statistical Society: Series C (The Statistician) 52: 193–209

Jackson CH (2011) Multi-state models for panel data: the msm package for R. Journal of Statistical Software 38 (8): 1–28

Joly P, Commenges D, Helmer C and Letenneur L (2002) A penalized likelihood approach for an illness-death model with interval-censored data : application to age-specific incidence of dementia, Biostatistics 3: 433–443

Kalbfleisch J, Lawless J (1985) The analysis of panel data under a Markov assumption. Journal of the American Statistical Association 80 (392): 863–871

Kalbfleisch JD, Prentice RL (2002) The Statistical Analysis of Failure Time Data. John Wiley and Sons, New York

Kay R (1986) A Markov model for analysing cancer markers and disease states in survival studies. Biometrics 42: 855–865

Keiding N, Klein JP, Horowitz MM (2001). Multi-state models and outcome prediction in bone marrow transplantation. Statistics in Medicine 20: 1871–1885.

Kessing LV, Hansen MG, Andersen PK, Angst J (2004) The predictive effect of episodes on the risk of recurrence in depressive and bipolar disorders - a life-long perspective. Acta Psychiatrica Scandinavica 109 (5): 339–344

Kvist K, Andersen PK, Angst J, Kessing LV (2010) Event dependent sampling of recurrent events. Lifetime Data Analysis 16: 580–598

Lawless JF, Fong DYT (1999) State duration models in clinical and observational studies. Statistics in Medicine 18: 2365–2376

Lawless JF, Kalbfleisch JD, Wild CJ (1999) Semiparametric methods for response-selective and missing data problems in regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61 (2): 413–438

Lawless JF, Wigg MB, Tuli S, Drake J, Lamberti-Pasculli M (2001) Analysis of repeated failures of durations, with application to shunt failures for patients with paediatric hydrocephalus. Journal of the Royal Statistical Society: Series C (Applied Statistics) 50: 449 – 465

Lawless JF (2013) The design and analysis of life history studies. Statistics in Medicine (in press).

Lee EW, Kim MY (1998) The analysis of correlated panel data using a continuous-time Markov model. Biometrics 54 (4): 1638–1644

Lin H, Guo Z, Peduzzi PN, Gill TM, Allore HG (2008) A semiparametric transition model with latent traits for longitudinal multistate data. Biometrics 64 (4): 1032–1042

Lumley T, Shaw PA and Dai JY (2011) Connections between survey calibration estimators and semiparametric models for incomplete data (with discussion). International Statistical Review 79: 200–220

Mandel M, Betensky RA (2008) Estimating time-to-event from longitudinal ordinal data using random-effects Markov models: application to multiple sclerosis progression. Biostatistics 9 (4): 750–764

Martinussen T, Scheike TH (2006) Dynamic Regression Models for Survival Data Springer, New York

Mueller PW, Rogus JJ, Cleary PA, Zhao Y, Smiles AM, Steffes MW, Bucksa J, Gibson TB, Cordovado SK, Krolewski AS, Nierras CR, Warram JH (2006) Genetics of Kidneys in Diabetes (GoKinD) study: a genetics collection available for identifying genetic susceptibility factors in diabetic nephropathy in type I diabetes. Journal of the American Society of Nephrology 17 (7): 1782–1790

Nelsen RB (2006) An Introduction to Copulas, 2nd Edition. Springer Science + Business Media, LLC, New York, NY

O'Keeffe AG, Tom BDM, Farewell VT (2011) A case-study in the clinical epidemiology of psoriatic arthritis: multistate models and causal arguments. Journal of the Royal Statistical Society: Series C (Applied Statistics) 60 (5): 675–699

O'Keeffe AG, Tom BDM and Farewell VT (2013) Mixture distributions in multi-state modelling: Some considerations in a study of psoriatic arthritis. Statistics in Medicine 32: 600–619.

Pencina MJ, Larson MG, D'Agostino RB (2007) Choice of time scale and its effect on significance of predictors in longitudinal studies. Statistics in Medicine 26 (6): 1343–1359

Prentice RL, Langer R, Stefanick ML, Howard BV, Pettinger M, Anderson G, Barad D, Curb JD, Kotchen J, Kuller L, Limacher M, Wactawski-Wende J and for the Women's Health Initiative Investigators (2005) Combined postmenopausal hormone therapy and cardiovascular disease: toward

resolving the discrepancy between observational studies and the Women's Health Initiative Clinical Trial. American Journal of Epidemiology 162 (5): 404–414

Raboud J, Reid N, Coates RA, Farewell VT (1993) Estimating risks of progressing to AIDS when covariates are measured with error. Journal of the Royal Statistical Society: Series A 156 (3): 393–406

Rahman P, Gladman P, Cook RJ, Zhou Y, Young G, Salonen D (1998) Radiological assessment in psoriatic arthritis. British Journal of Rheumatology 37: 760–765

Robins JM, Rotnitzky A, Zhao LP (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association 90 (429): 106–121

Samuelsen SO, Anestad H, Skrondal A (2007) Stratified case-cohort analysis of general cohort sampling designs. Scandinavian Journal of Statistics 34: 103–119

Satten GA, Longini IM (1996) Markov chains with measurement error: estimating the "true" course of a marker of the progression of human immunodeficiency virus disease. Applied Statistics 45(3): 275–309

Satten GA (1999) Estimating the extent of tracking in interval-censored chain-of-events data. Biometrics 55: 1228–1231

Scheike TH and M-J Zhang (2002) An additive-multiplicative Cox-Aalen regression model. Scandinavian Journal of Statistics 29: 75–88

Sweeting MJ, De Angelis D, Neal KR, Ramsay ME, Irving WL, Wright M, Brant L, Harris HE; Trent HCV Study Group; HCV National Register Steering Group (2006) Estimated progression rates in three United Kingdom hepatitis C cohorts differed according to method of recruitment. Journal of Clinical Epidermiology 59: 144–152

Sweeting MJ, Farewell VT, De Angelis D (2010) Multi-state Markov models for disease progression in the presence of informative examination times: an application to hepatitis C. Statistics in Medicine 29 (11): 1161–1174

Sutradhar R, Cook RJ (2008) Analysis of interval-censored data from clustered multistate processes: application to joint damage in psoriatic arthritis. Journal of the Royal Statistical Society: Series C (Applied Statistics) 57 (5): 553–566

The Diabetes Control and Complications Trial Research Group (1993) The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. The New England Journal of Medicine 329 (14): 977–986

The Diabetes Control and Complications Trial Research Group (1995) Progression of retinopathy with intensive versus conventional treatment in the Diabetes Control and Complications Trial. Opthalmology 102: 647–661

The Diabetes Control and Complications Trial Research Group (1998) Early worsening of diabetic retinopathy in the Diabetes Control and Complications Trial. Archives of Opthalmology 116: 874–887

Titman AC, Sharples LD (2010a) Semi-Markov models with phase-type sojourn distributions. Biometrics 66 (3): 742–752

Titman AC, Sharples LD (2010b) Model diagnostics for multi-state models. Statistical Methods in Medical Research 19: 621–651.

Titman AC (2011) Flexible nonhomogeneous Markov models for panel observed data. Biometrics 67 (3): 780–787

Tolusso D, Cook RJ (2009) Robust estimation of state occupancy probabilities for interval-censored multistate data: an application involving spondylitis in psoriatic arthritis. Communications in Statistics - Theory and Methods 38 (18): 3307–3325

Tom BDM, Farewell VT (2011) Intermittent observation of time-dependent explanatory variables: a multistate modelling approach. Statistics in Medicine 30 (30): 3520–3531

Tsiatis AA, Davidian M (2004) Joint modeling of longitudinal and time-to-event data: an overview. Statistica Sinica 14: 809–834

Turnbull B (1976) The empirical distribution function with arbitrarily grouped, censored and truncated data. Journal of the Royal Statistical Society: Series B 38: 290–295

Tyas SL, Salazar JC, Snowdon DA, Desrosiers MF, Riley KP, Mendiondo MS, Kryscio RJ (2007) Transitions to mild cognitive impairments, dementia, and death: findings from the nun study. American Journal of Epidemiology 165 (11): 1231–1238

Wei LJ, Lin DY, Weissfeld L (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association 84 (408): 1065–1073

Yang Y, Nair VN (2011) Parametric inference for time-to-failure in multi-state semi-Markov models: a comparison of marginal and process approaches. Canadian Journal of Statistics 39 (3): 537–555