



A case-study in the clinical epidemiology of psoriatic arthritis: multistate models and causal arguments

Aidan G. O’Keeffe, Brian D. M. Tom and Vernon T. Farewell

Medical Research Council Biostatistics Unit, Cambridge, UK

[Received December 2009. Final revision March 2011]

Summary. In psoriatic arthritis, permanent joint damage characterizes disease progression and represents a major debilitating aspect of the disease. Understanding the process of joint damage will assist in the treatment and disease management of patients. Multistate models provide a means to examine patterns of disease, such as symmetric joint damage. Additionally, the link between damage and the dynamic course of disease activity (represented by joint swelling and stress pain) at both the individual joint level and otherwise can be represented within a correlated multistate model framework. Correlation is reflected through the use of random effects for progressive models and robust variance estimation for non-progressive models. Such analyses, undertaken with data from a large psoriatic arthritis cohort, are discussed and the extent to which they permit causal reasoning is considered. For this, emphasis is given to the use of the Bradford Hill criteria for causation in observational studies and the concept of local (in)dependence to capture the dynamic nature of the relationships.

Keywords: Bradford Hill criteria; Causality; Composable Markov process; Damage; Disease activity; Dynamic modelling; Granger causality and non-causality; Interval censoring; Local dependence and independence; Multistate models; Psoriatic arthritis; Random effect; Robust information sandwich estimator; Temporality

1. Introduction

From a clinical perspective, it is generally held that better understanding of a disease process will lead to more appropriate treatment and disease management of patients. Often multistate models (Hougaard, 1999; Commenges, 1999; Andersen and Keiding, 2002; Meira-Machado *et al.*, 2009) are particularly useful for this. In this paper we illustrate their use for the study of disease progression in psoriatic arthritis.

In his 1970 seminal paper, Schweder (1970) introduced the concept of local (in)dependence between components of a composable finite Markov process. He felt that many studied phenomena can be realistically described by time continuous finite Markov processes. If, in addition, the Markov process representing the phenomenon under study could be defined to be composable (i.e. represented as a vector of distinct subprocesses, whereby no two subprocesses or components can change state ‘simultaneously’), then (in)dependences between these subprocesses can be explicitly expressed through the transition intensities of the original Markov process.

More explicitly, let \mathbf{Y} be a composable finite Markov process with components Y_1, \dots, Y_p , which we denote as $\mathbf{Y} \sim (Y_1, \dots, Y_p)$, and let $V = \{1, \dots, p\}$. Here the dependence on time in the notation is left out for convenience but is implicit. A component Y_k is said to be locally

Address for correspondence: Aidan G. O’Keeffe, Medical Research Council Biostatistics Unit, University Forvie Site, Robinson Way, Cambridge, CB2 0SR, UK.
E-mail: aidan.okeeffe@mrc-bsu.cam.ac.uk

Reuse of this article is permitted in accordance with the terms and conditions set out at <http://wileyonlinelibrary.com/onlineopen#OnlineOpenTerms>.

independent of Y_j , $j \neq k$, given the remaining components $\mathbf{Y}_{V \setminus \{j,k\}}$, if and only if the transition intensities of \mathbf{Y} corresponding to transitions only between states in Y_k are constant functions for any state in Y_j , within a specified infinitesimal time interval. If Y_k is not locally independent of Y_j , it is said to be locally dependent on Y_j .

The above relationship is a ‘local’ property as it holds in an infinitesimal time interval. Furthermore, recall that, because $\mathbf{Y} \sim (Y_1, \dots, Y_p)$ is composable, no two components of \mathbf{Y} can change state at the same time over this infinitesimal interval and therefore the transition intensities for such simultaneous transitions in more than one component are zero. Moreover, local independence is an asymmetric relationship, i.e. it has a direction, and so, Y_k being locally independent of Y_j does not necessarily imply that Y_j is locally independent of Y_k over the same infinitesimal time interval.

This important concept of local independence was extended in Aalen (1987) to apply to more general stochastic processes that admit a Doob–Meyer decomposition, with unrelated innovations. Aalen stressed the usefulness of ‘dynamic’ models where dynamic refers to how the future relates to the past, and it is this dynamic nature of Schweder’s work, rather than the Markov assumption *per se*, which he suggested is important for statistical analysis. Earlier works by Cox (1972) and Aalen *et al.* (1980) also reflect this dynamic viewpoint. More recently, Didelez (2007, 2008) introduced dynamic graphical models to describe these local dependences, which further allowed local independences to be read off. In our discussion of the progression of psoriatic arthritis disease that follows, we use Schweder’s local (in)dependence concept as a means of characterizing the findings from dynamic analyses based on multistate models.

Because we focus on arthritic disease progression at the individual joint level, our analysis is based on the use of correlated multistate models. Cook *et al.* (2004) considered such an analysis for progressive processes but with a discrete multivariate random-effects distribution used to account for correlation. For a four-state progressive model, we extend this to allow the use of gamma-distributed random effects. In addition, for a three-state model with some reversible transitions, we outline an approach that is based on generalized estimating equations to account for correlation between processes.

Our reported investigations are based on observational clinical data and it is well recognized that causal relationships can never be proved with such data. However, as Weiss (1986) argued, ‘it is necessary to attempt to draw inferences of cause and effect, even from inevitably incomplete data, for the alternative is to make no inference at all, which would preclude taking preventive or therapeutic action’.

Therefore, after presenting results of our analyses, we consider the extent to which they might allow inference concerning causal relationships. In doing so, we take up the implied challenge of Aalen (Aalen *et al.* (2008), page 348) who wrote:

‘One major danger of avoiding the subject of causality in statistical education and statistical literature, is that one never gets any insight into this fascinating concept, which has such an old history in philosophy and science. The fact is that statistics plays a major role in looking for causal connections in many fields and statisticians who know next to nothing about causality as a larger concept will be far less useful than they could have been.’

After 30 years of data collection and 20 years of previous analyses, it seems particularly appropriate to take up this challenge with the psoriatic arthritis cohort data that are discussed in the next section.

2. Psoriatic arthritis

2.1. Background

Psoriatic arthritis is an inflammatory arthritis associated with psoriasis, which is usually sero-

negative for rheumatoid factor (Wright and Moll, 1976; Gladman, 2004). It is distinguished as an entity from the prototypal arthritis, rheumatoid arthritis (RA), because of its unique clinical features: the association with the skin disease psoriasis; equal gender frequency, as opposed to the preponderance in females in RA; the asymmetric presentation involving large joints and distal interphalangeal joints which are not commonly affected in RA; the majority (more than 85%) of psoriatic arthritis patients lack rheumatoid factor; roughly half of the patients with psoriatic arthritis have spinal involvement which is distinctly rare in RA. Moreover, in the fingers, patients with psoriatic arthritis demonstrate dactylitis, or inflammation of the whole digit, as well as enthesitis and other extra-articular features that are typical to the seronegative spondyloarthritides. Recent studies have demonstrated that psoriatic arthritis is a progressive disease, leading to joint destruction, disability and reduced quality of life (Husted *et al.*, 2001; Kane *et al.*, 2003; McHugh *et al.*, 2003; McKenna *et al.*, 2004; Queiro-Silva *et al.*, 2003; Sokoll and Helliwell, 2001).

Disease progression in psoriatic arthritis, as with RA, is often taken to be reflected in the accumulation and severity of damaged joints, evaluated either clinically or through radiographic imaging (e.g. X-rays). Clinical damage is determined by the presence of a limitation in range of movement of more than 20% of the range not related to the presence of joint effusion, the presence of joint deformities, subluxation, flail joints or ankylosis (Siannis *et al.*, 2006). The damage process is generally irreversible; therefore once a joint is damaged it will remain so; thus efforts by clinicians to prevent or slow this process are crucial in the care of patients. Disease activity, in contrast, is a reversible process, and is reflected in part by joints being described as either tender (the presence of stress pain and/or joint line tenderness) only or effused (joint swelling with or without tenderness), with the latter representing a more severe level of activity than the former. Various types of medication are available to treat the disease activity. The use of non-steroidal anti-inflammatory drugs and disease modifying anti-rheumatic drugs as first- and second-line treatments, and more recently biologics that bind to tumour necrosis factor α , preventing activation of its receptors, are all available in the clinician's arsenal of medications for controlling the activity of the disease. In addition, if these front-line therapies are not effective at reducing inflammation, the use of intra-articular steroids injected directly into the specific active joint(s) may also be considered.

In 1978, the Toronto Psoriatic Arthritis Clinic was established by Professor Dafna Gladman at the University of Toronto after recognizing that there was a paucity of knowledge regarding psoriatic arthritis at the time. Since then, much has been learned about the disease through research done at the Toronto Psoriatic Arthritis Clinic, which has the largest and most comprehensively studied cohort of psoriatic arthritis patients in the world. At present, the clinic includes over 1000 patients who have been closely followed up prospectively over the years. Visits to the clinic are aimed to be scheduled 6–12 months apart. At these visits, patients are evaluated in a standard way according to a defined protocol that includes a complete history, physical examination (including rheumatological assessment) and routine blood and urine tests, with X-rays also being performed biennially. The rheumatological examination includes validated and reliable determination of active and damaged joints. This longitudinal information provides a valuable resource for research.

Up to the end of 2006, which, in the main, was the period predating the introduction of biologics, longitudinal data on 790 patients were collected from the Toronto Psoriatic Arthritis Clinic. A subset of 517 patients, which is defined in Section 3, of the 790 will be used in this study to examine patterns of disease and to investigate the link between clinical damage and the dynamic course of disease activity at both the individual joint level and otherwise. Of these 517 patients, 289 (55.9%) were male and 228 (44.1%) female. The mean age at entry was 41 years

and 7 months, with a standard deviation of 12 years and 7 months. The number of visits to the clinic recorded up to 2006 ranged from 2 to 47, with a median of 7. The mean gap time (i.e. the time between successive clinic visits) was 10.8 months (standard deviation 15.2 months) and the median gap time was 6.3 months.

2.2. Previous investigations

For both psoriatic arthritis and RA there is a strong belief among clinicians that active inflammation (i.e. persistent inflammatory synovitis) results in or causes joint damage. A number of research groups have repeatedly shown an association between disease activity and progression to damage in both psoriatic arthritis and RA, with this association consistently seen irrespective of the measure of activity used (i.e. whether it be the number of active (tender or effused) joints, biochemical markers or some form of composite measure of disease activity) or whether damage is determined radiologically or clinically (Aletaha *et al.*, 2009; Bond *et al.*, 2007; Gladman *et al.*, 1995; Gladman and Farewell, 1999; Molenaar *et al.*, 2004; Mulherin *et al.*, 1996; Scott, 2004; Smolen *et al.*, 2009; Welsing *et al.*, 2004).

For psoriatic arthritis, it has been found that radiological damage precedes clinical damage in the majority of patients (Siannis *et al.*, 2006), and that the same predictors of disease progression are seen regardless of how damage is detected (Bond *et al.*, 2007). In a recent investigation of the link between activity and damage in psoriatic arthritis based on the Toronto Psoriatic Arthritis data (Bond *et al.*, 2007; Bond and Farewell, 2009), negative binomial regression models for the increase in the total damaged joint count between visits were fitted, with previous damage incorporated as a dynamic explanatory variable in the models to account for the within-patient correlation. Disease activity was included initially in these models both in terms of total active joint counts (effused and tender) at entry to the clinic and as time-dependent explanatory variables, with the total joint counts updated at visits to the clinic. These models found that time varying activity (both effused and tender total joint counts) was associated with the progression of radiological and clinical damage, but that activity variables at entry to the clinic were not, when in the presence of their time varying counterparts.

These results, coupled with similar results in RA, argue for a close link between activity and damage, which, additionally, is supported by the time ordering. This argument is further strengthened by the dose–response nature of the relationship between activity and damage. In our recent psoriatic arthritis investigation, differential effects of tender-only joint and effused (usually also tender) joint counts were found to some extent, with the relative rate estimate for the more severe effused joint count larger than the relative rate estimate for tender joint count. Similar findings were reported in a recent study in RA, in which a positive association was found between the level of disease activity (defined by using the simplified disease activity index) and the level of radiographic progression (defined by using the modified total Sharp score) (Aletaha *et al.*, 2009).

However, because total joint counts or other ‘global’ measures of disease activity and damage are used, the link that was found between activity and damage is essentially between the systemic activity process and the systemic damage process. If we can consider a finer level of detail and examine the link at the individual joint level then similar results would further our understanding of the relationships between activity and damage. Additionally, issues that are related to disease patterns can be investigated.

3. Four-state model for damage process

Clinical damage and disease activity information at the individual joint level for the hands

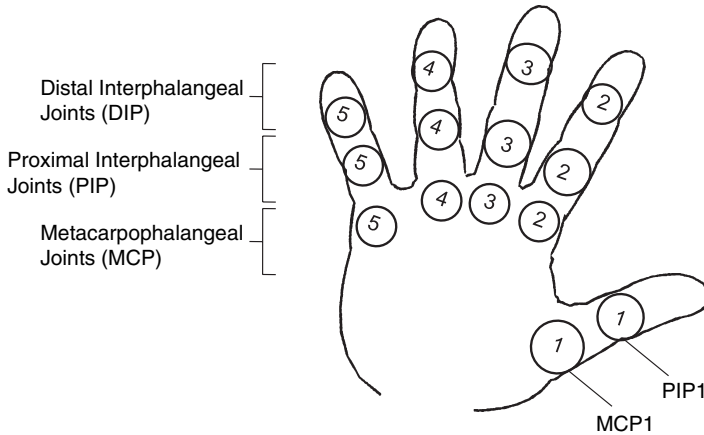


Fig. 1. Diagram showing the 14 hand joints and their type

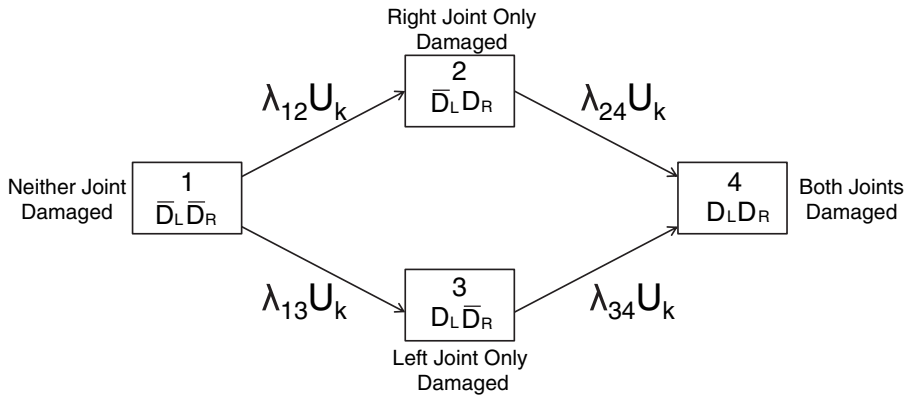


Fig. 2. Diagram of the multistate model for damage at a joint location, with random effect

(excluding the wrist joints) was extracted from the Toronto Psoriatic Arthritis Clinic’s database. Fig. 1 shows a picture of the 14 joints in a hand. In it, three types of joint are seen: the distal interphalangeal, the proximal interphalangeal and the metacarpophalangeal. All three joint types are usually found on a digit, except for the thumb where the distal interphalangeal joint is absent. There are 28 hand joints in total.

For the purposes of this paper, we use data extracted from 517 of the 790 patients in the clinic who entered before the beginning of 2007. These 517 patients correspond to those who, at entry to the clinic, had no clinical damage in any of the hand joints on either the left or right hands and therefore information on all these patients is comparable. The clinical damage processes at the hand joint level may be observed in either of two states at a clinic visit (‘not damaged’ or ‘damaged’). However, once a hand joint enters the damaged state it remains there *ad infinitum*, i.e. the damaged state is absorbing, and all the individual joints’ damage processes are irreversible. Disease activity at a specific hand joint is defined by two joint-specific processes: a tender-only joint process and an effused joint process. Both these two activity processes are reversible. The tender joint process can move back and forth between a state of ‘no stress pain’ and a state of ‘tenderness only’. The effused joint process can move between states of ‘no swelling’ and ‘effusion’.

The multistate model for our first set of analyses is depicted diagrammatically in Fig. 2 and is a four-state model for damage in each of the 14 pairs of hand joints. The four states of this multistate model are defined as

- (a) state 1, damage in neither hand, (\bar{D}_L, \bar{D}_R) ,
- (b) state 2, damage in the right hand only, (\bar{D}_L, D_R) ,
- (c) state 3, damage in the left hand only, (D_L, \bar{D}_R) , and
- (d) state 4, damage in both hands, (D_L, D_R) .

This is a multistate model at a specific joint location in both left and right hands. Moreover, it is composable, comprising separate damage subprocesses for the left and right hands at the specific joint location, and local dependences between the subprocesses will be reflected in relationships between the transition intensities of this multistate model. In addition, this model does not allow transitions directly from state 1 to state 4. This is a necessary constraint to ensure composability but is not very restrictive for models with transitions in continuous time. Transitions between state 2 and state 3 are not allowed since damage is irreversible. Furthermore, because the 14 multistate processes within a patient should be more similar than across patients, we introduce subject-specific random effects into the model. These act multiplicatively on the baseline transition intensities of the 517 patients. The random effects, U_k , $k = 1, \dots, 517$, are assumed to be distributed as independent gamma random variables with unit mean and variance θ (see equation (6) in Appendix A), which we denote by $U_k \sim \text{gamma}(1/\theta, 1/\theta)$.

To be more precise, consider the continuous time multistate process for damage, $X_k^{(l)} = \{X_k^{(l)}(t) | t \in \mathcal{T}_k\}$, at the l th joint location for the k th patient, with state space $\Omega = \{1, \dots, 4\}$ as described above, and $\mathcal{T}_k = [0, \tau_k)$ being the time interval of interest. Here the timescale is time from entry to the clinic in years. Let $\lambda_{ijk}^{(l)}(t)$ be the transition intensity function that is associated with $X_k^{(l)}$ corresponding to the transition from state i to state j , denoted as $i \rightarrow j$, where $i \neq j$ ($i = 1, 2, 3; j = 2, 3, 4$) and with transitions $1 \rightarrow 4, 2 \rightarrow 3$ and $3 \rightarrow 2$ not permissible. Let $\lambda_{0ij}^{(l)}$ be the baseline transition intensity corresponding to this $i \rightarrow j$ transition. In addition, let $\beta_{ij}^{(l)}$ be a vector of regression parameters that are associated with the predictable explanatory variable vector process $\mathbf{z}_{ijk}^{(l)}(t)$. Then the transition intensity function for the $i \rightarrow j$ transition, $i \neq j$, at the l th hand joint location for the k th patient is given by

$$\lambda_{ijk}^{(l)}\{t | \beta_{ij}^{(l)}, \mathbf{z}_{ijk}^{(l)}(t), u_k\} = u_k \lambda_{0ij}^{(l)} \exp\{\beta_{ij}^{(l)T} \mathbf{z}_{ijk}^{(l)}(t)\}, \tag{1}$$

where u_k corresponds to the subject-specific random effect. This multistate model is a generalization of Schweder’s model to incorporate random effects and is similar to the model that was proposed by Cook *et al.* (2004) for clustered progressive multistate processes. Nevertheless, this model still allows us to examine local dependences through the transition intensities. Note that we have assumed that the regression parameter vector $\beta_{ij}^{(l)}$ is joint location specific. However, it may be reasonable to set $\beta_{ij}^{(l)} = \beta_{ij}$ as the 14 joint location processes represent a similar disease progression phenomenon, and we do this subsequently. Also it may at times be sensible to constrain baseline intensities to be the same across some joint locations for parsimony, and we shall indicate when this has been done. Moreover, further constraints on regression parameters may be acceptable for simplification based on the analyses of the data.

3.1. Maximum likelihood estimation

We employ an underlying time homogeneous Markov model for panel data where patients are not under continuous follow-up and are observed only at protocol visits (Kalbfleisch and Lawless, 1985), and with baseline transition intensities modulated by the effects of explanatory

variables $\mathbf{z}(t)$ and subject-specific random effects $\{u_k\}$ as described in model (1). Suppose that the k th patient makes n_k clinic visits at times $\mathbf{t}_k = (t_{k1}, \dots, t_{kn_k})^T$ and associated with the l th joint location process for this patient are an explanatory variable vector process $\mathbf{z}_k^{(l)}(t)$ recorded at clinic visits, baseline transition intensities $\lambda_0^{(l)} = (\lambda_{012}^{(l)}, \lambda_{013}^{(l)}, \lambda_{024}^{(l)}, \lambda_{034}^{(l)})^T$ and a vector of regression parameters $\beta^{(l)}$. Then the conditional probability of observing the path that is followed by the l th joint location process of the k th patient over the n_k clinic visits, given u_k , is

$$\prod_{m=1}^{n_k-1} \mathbb{P}\{X_k^{(l)}(t_{k,m+1}) | X_k^{(l)}(t_{km}), u_k; \lambda_0^{(l)}, \beta^{(l)}, \mathbf{z}_k^{(l)}(t_{km})\}$$

and thus the overall likelihood contribution from the k th patient is given by

$$L_k(\lambda_0, \beta, \theta) = \int_0^\infty \left[\prod_{l=1}^{14} \prod_{m=1}^{n_k-1} \mathbb{P}\{X_k^{(l)}(t_{k,m+1}) | X_k^{(l)}(t_{km}), u_k; \lambda_0^{(l)}, \beta^{(l)}, \mathbf{z}_k^{(l)}(t_{km})\} \right] f_{U_k}(u_k) du_k,$$

where $\lambda_0 = (\lambda_0^{(1)T}, \dots, \lambda_0^{(14)T})^T$, $\beta = (\beta^{(1)T}, \dots, \beta^{(14)T})^T$ and $f_{U_k}(u_k)$ denotes the probability density function of the gamma(1/θ, 1/θ) distribution. (The integration over the random-effect distribution can be carried out in R (R Development Core Team, 2009) by using the `integrate` command.) Hence the likelihood function across all N individuals is given by

$$L(\lambda_0, \beta, \theta) = \prod_{k=1}^N L_k(\lambda_0, \beta, \theta). \tag{2}$$

We maximize the log-likelihood of equation (2) with respect to the baseline transition intensity parameters λ_0 , the regression parameters β and the random-effect variance θ . This is done by using the BFGS optimization routine (Broyden, 1970) through the `optim` command in R. This yields parameter estimates together with a numerically derived Hessian matrix at the estimate values from which asymptotic standard errors for the parameter estimates are obtained. Comprehensive details of the likelihood function are provided in Appendix A.

In the following subsections, we describe analyses that use the four-state model to investigate firstly the extent to which a symmetric disease pattern exists in the hands of psoriatic arthritis patients and then the relationship between activity and damage at the joint level.

3.2. Patterns—symmetric joint damage in the hands

In RA, there is a consensus among rheumatologists that damage is highly symmetric. This, however, is not so for psoriatic arthritis, where there is still uncertainty among rheumatologists concerning the extent of symmetry (Gladman, 2005, 2006). To inform this debate, at least with regard to the hands of psoriatic arthritis patients, we analyse the data from the 14 pairs of joints within each of the 517 patients by using the correlated multistate processes model that was described earlier. Table 1 presents the observed transition table for this model. We find that there were 70 169 observed transitions, of which only 687 were actual transitions between differing states, over all pairs of joints in the 517 patients, with 813 hand joints becoming damaged over the follow-up period of the study out of the possible 14476 total hand joints under investigation.

To investigate the extent of disease symmetry in the hands we set the explanatory vector $\mathbf{z}_{ijk}^{(l)}(t) = \mathbf{0}$ in equation (1) and reparameterize some of the baseline transition intensities in terms of others, i.e. we specify $\lambda_{024}^{(l)} = \lambda_{013}^{(l)} \exp(\gamma_{24})$ and $\lambda_{034}^{(l)} = \lambda_{012}^{(l)} \exp(\gamma_{34})$. Therefore a test for symmetry would require $\lambda_{012}^{(l)} < \lambda_{034}^{(l)}$ and $\lambda_{013}^{(l)} < \lambda_{024}^{(l)}$, or equivalently that $\gamma_{24} > 0$ and $\gamma_{34} > 0$, i.e. a symmetrical damage pattern implies that the tendency for a joint at a specific location to

Table 1. Observed joint transitions for the four-state multistate model

Transition from	Number of transitions to the following states:			
	(\bar{D}_L, \bar{D}_R)	(\bar{D}_L, D_R)	(D_L, \bar{D}_R)	(D_L, D_R)
(\bar{D}_L, \bar{D}_R)	66729	225	204	126
(\bar{D}_L, D_R)	0	1339	0	81
(D_L, \bar{D}_R)	0	0	1414	51
(D_L, D_R)	0	0	0	0

become damaged is increased if the contralateral joint on the other hand is earlier damaged, and this applies at all joint locations. Note that the usage of symmetry here is a rheumatological usage and does not refer to the technical use of the term symmetry in statistics to describe the invariance of cell probabilities in the permutation of the subscripts indexing these values, as when applied to describing some statistical models based on repeated categorical response data.

Table 2 presents the results from fitting the model to investigate symmetry. The parameters γ_{24} and γ_{34} are estimated to be 1.82 and 1.39 respectively with corresponding 95% confidence intervals of (1.49, 2.14) and (1.01, 1.78). Both of these confidence intervals indicate substantial departures from zero to the right and therefore present strong evidence for symmetry. This symmetry, in turn, implies that there are local dependences in both directions between the two damage subprocesses (i.e. the left and the right) of the composable multistate process.

We now consider whether the relationship between disease activity and damage, which is repeatedly found at the patient level, is seen at the individual joint level.

3.3. Regression of activity on damage at joint level

To investigate the relationship between damage and the dynamic course of disease activity at the individual joint (location) level we use the model represented by equation (1) with the explanatory variable vector process $\mathbf{z}(t)$ defined by information on activity. We define binary indicators of joint level activity, $A_L(t)$ and $A_R(t)$ for the left and right joints respectively in a pair at time t . In these variables, we make no distinction between activity in the form of tenderness and effusion. However, for transitions into a state of damage, we are interested in assessing a possible dose–response relationship between joint activity and the rate at which damage occurs. Thus we also define dynamic binary indicators for joint level tenderness only, $T_L(t)$ and $T_R(t)$, and joint level effusion (with or without tenderness but usually tender), $E_L(t)$ and $E_R(t)$, for the left (L) and right (R) hands at time t . Hence, assuming that no previous damage has occurred to either joint (state 1), models for the transition intensities out of state 1 are given by

$$\begin{aligned} \lambda_{12k}^{(l)}(t) &= u_k \lambda_{012} \exp\{\alpha_{L12} A_L^{(l)}(t) + \tau_{R12} T_R^{(l)}(t) + \varepsilon_{R12} E_R^{(l)}(t)\}, \\ \lambda_{13k}^{(l)}(t) &= u_k \lambda_{013} \exp\{\alpha_{R13} A_R^{(l)}(t) + \tau_{L13} T_L^{(l)}(t) + \varepsilon_{L13} E_L^{(l)}(t)\} \end{aligned} \tag{3}$$

where $\beta_{12} = (\alpha_{L12}, \tau_{R12}, \varepsilon_{R12})^T$, $\beta_{13} = (\alpha_{R13}, \tau_{L13}, \varepsilon_{L13})^T$ and we assume that the baseline transition intensities are constrained to be the same across the 14 hand joint locations.

When forming models for the transition intensities into state 4, we choose not to include

Table 2. Parameter estimates together with associated 95% confidence intervals (in parentheses) for the model fitted to investigate symmetry in the left and right damage processes†

Joint	$\hat{\lambda}_{012} (\times 10^{-2})$	$\hat{\lambda}_{013} (\times 10^{-2})$
MCP1	0.91 (0.65, 1.29)	0.34 (0.20, 0.57)
MCP2	0.31 (0.19, 0.52)	0.17 (0.08, 0.33)
MCP3	0.20 (0.11, 0.37)	0.36 (0.22, 0.59)
MCP4	0.62 (0.41, 0.94)	0.63 (0.42, 0.95)
MCP5	0.35 (0.21, 0.58)	0.68 (0.46, 1.02)
PIP1	0.95 (0.65, 1.38)	0.78 (0.53, 1.15)
PIP2	0.66 (0.44, 0.99)	1.16 (0.82, 1.64)
PIP3	0.67 (0.47, 0.97)	0.30 (0.19, 0.49)
PIP4	0.34 (0.21, 0.56)	0.20 (0.11, 0.37)
PIP5	0.20 (0.11, 0.37)	0.37 (0.23, 0.59)
DIP2	0.50 (0.33, 0.77)	0.57 (0.37, 0.87)
DIP3	0.41 (0.26, 0.64)	0.36 (0.23, 0.56)
DIP4	0.96 (0.66, 1.39)	0.82 (0.55, 1.22)
DIP5	0.74 (0.50, 1.10)	0.42 (0.27, 0.65)

†The estimate for the random-effect variance is $\hat{\theta} = 4.99 (3.97, 6.27)$. Estimates for the symmetry parameters are $\hat{\gamma}_{24} = 1.82 (1.49, 2.14)$ and $\hat{\gamma}_{34} = 1.39 (1.01, 1.78)$. MCP, metacarpophalangeal; PIP, proximal interphalangeal; DIP, distal interphalangeal.

information (in the form of explanatory variables) on the activity process in the opposite (damaged) joint at time t because of the dominant effect of symmetrical damage. Hence, our models for the transition intensities into state 4 are given by

$$\begin{aligned} \lambda_{24k}^{(l)}(t) &= u_k \lambda_{024} \exp\{\tau_{L24} T_L^{(l)}(t) + \varepsilon_{L24} E_L^{(l)}(t)\}, \\ \lambda_{34k}^{(l)}(t) &= u_k \lambda_{034} \exp\{\tau_{R34} T_R^{(l)}(t) + \varepsilon_{R34} E_R^{(l)}(t)\} \end{aligned} \tag{4}$$

with $\beta_{24} = (\tau_{L24}, \varepsilon_{L24})^T$, $\beta_{34} = (\tau_{R34}, \varepsilon_{R34})^T$ and, once again, we assume that the baseline transition intensities are constrained to be the same across the 14 joint locations.

Because of the panel nature of the data, we focus on the relationship between observed transitions in the damage process and the values of these activity variables (which are represented by activity or joint pain and swelling on both the left and the right hands) at the *last* clinic visit, i.e. we assume that $A_L(t)$, $A_R(t)$, $T_L(t)$, $T_R(t)$, $E_L(t)$ and $E_R(t)$ are piecewise constant between

clinic visits. Thus underlying activity is assumed to be coarsened by the timing of clinic visits. We additionally assume that these activity variables can only change ‘state’ immediately after the time of a visit to the clinic when damage information becomes available.

It is ‘biologically’ conceivable that this model, given by expressions (3) and (4), can be further constrained to allow $\alpha_{L12} = \alpha_{R13}$, $\tau_{R12} = \tau_{L13}$, $\tau_{L24} = \tau_{R34}$ and, similarly, $\varepsilon_{R12} = \varepsilon_{L13}$ and $\varepsilon_{L24} = \varepsilon_{R34}$. We applied these constraints to subsequent models in the interest of parsimony.

However, the analysis that was just described did not fully exploit the dynamic potential of the framework being used, as it only used the current visit activity information when predicting future damage. It is quite plausible that some measure(s) of the ‘activeness’ of the joint over time may better characterize the relationship between activity and damage at the joint level, since more of the history of the disease activity process can be incorporated. For illustration, in a further analysis, we added binary explanatory variables (for joints on the left and right hands) that now indicate whether a joint was ever observed active (either swollen or tender) at the present visit or any of the previous protocol visits. We included similar constraints on the explanatory variables (i.e. $\rho_{L12} = \rho_{R13}$, $\rho_{R12} = \rho_{L13}$ and $\rho_{L24} = \rho_{R34}$, where ρ is the coefficient corresponding to whether or not the joint has ever been active) and allowed only prior activity in the undamaged joint to have an effect on the corresponding transition into state 4 (i.e. $\rho_{L34} = \rho_{R24} = 0$). We considered the possibility of an effect of activity at other joints on the damage rates by including binary explanatory variables to indicate whether or not activity occurred at any other joint in the same hand and whether or not activity occurred at any other joint in the opposite hand. The addition of these explanatory variables produced no demonstrable effect and they were subsequently removed from the model. Furthermore, we also fitted this model including a four-level factor to account for joint type (levels: finger metacarpophalangeal, proximal interphalangeal and distal interphalangeal and thumb metacarpophalangeal) with each level of this variable constrained to have the same effect across all transitions. However, this was found not significantly to affect any of the transitions to damage and was subsequently removed from the model. Some previous analyses (e.g. Bond *et al.* (2007)) have suggested that the erythrocyte sedimentation rate is one of the most likely patient confounders for a relationship between activity and damage in the hand joints. Thus, we fitted a model that included an explanatory variable to represent time varying erythrocyte sedimentation rate, constrained to have the same effect across all transitions. The inclusion of this explanatory variable did not change substantially the estimated effects of the activity explanatory variables on the transitions to damage, and hence the time varying erythrocyte sedimentation rate variable, for which there was also no evidence for an effect in this model, was not included in our final model.

Table 2 shows variation in baseline transition intensities fitted to each joint location but no obvious patterns. We considered the fitting of joint-location-specific transition intensities in our final model. However, such a model would require a large number of extra parameters to be estimated together with their associated standard errors. This would be challenging computationally and the relatively low number of transitions from states 2 and 3 to state 4 (shown in Table 1) may result in unstable parameter estimates. We take the view that some assumptions must be made with regard to parameter constraints in the interests of parsimony. The inability to model differences in the baseline rates among the joint locations easily and the need to maintain sufficient parsimony are practical limitations of this model.

Tables 3 and 4 present the results of fitting the constrained model including previous activity to 510 of 517 patients included in the earlier ‘symmetry’ analysis. Seven patients were excluded owing to missing information on disease activity. In Table 3 ‘transitive’ joint refers to the joint undergoing the transition to a state of damage (i.e. to state 2, 3 or 4) and ‘opposite’ joint refers to the same joint in the opposite hand. From Table 3, where there is no previous damage in either

Table 3. Log-intensity ratio and intensity ratio estimates for activity at the individual joint level, together with associated 95% confidence intervals (in parentheses)

<i>Effect on transition to damage</i>	<i>Estimate</i>	<i>Intensity ratio</i>
<i>No previous damage in either joint</i>		
Tenderness in the transitive joint	1.01 (0.72, 1.31)	2.76 (2.06, 3.70)
Effusion in the transitive joint	1.50 (1.22, 1.77)	4.47 (3.38, 5.90)
Activity in the opposite joint	0.17 (-0.10, 0.44)	1.18 (0.90, 1.55)
Transitive joint active in the past	0.76 (0.52, 1.00)	2.14 (1.68, 2.71)
Opposite joint active in the past	0.10 (-0.15, 0.35)	1.10 (0.86, 1.41)
<i>Opposite joint damaged</i>		
Tenderness in the transitive joint	0.81 (0.41, 1.20)	2.24 (1.51, 3.32)
Effusion in the transitive joint	0.78 (0.34, 1.23)	2.19 (1.40, 3.41)
Transitive joint active in the past	0.31 (0.01, 0.62)	1.37 (1.00, 1.86)

Table 4. Baseline intensities λ and random-effect variance parameter θ , together with associated 95% confidence intervals (in parentheses)

<i>Parameter</i>	<i>Estimate</i> ($\times 10^{-2}$)	<i>95% confidence interval</i> ($\times 10^{-2}$)
λ_{012}	0.28	(0.21, 0.36)
λ_{013}	0.27	(0.21, 0.34)
λ_{024}	2.15	(1.49, 3.10)
λ_{034}	2.34	(1.58, 3.47)
θ	3.81	(2.98, 4.88)

joint (i.e. the model is currently in state 1), we observe increases in the rates of transitions into a state of damage when there is current activity (in the form of both tenderness and effusion) and when there has been some past activity in the transitive joint, compared with when no activity has been seen. We note that effusion shows a larger positive effect on transition to damage than tenderness. Conversely, activity in the opposite joint appears not to affect the transition rate to damage significantly. Where the opposite joint is already damaged we see an increase in the transition to damage both where there is current tenderness and where there is current effusion as well as where past activity has occurred in the transitive joint. We note that the effects of tenderness and effusion are similar; we no longer see an apparent dose-response relationship for the covariates representing activity in the transitive joint. This perhaps indicates that, when the opposite joint becomes damaged, the damage process in a joint is more complicated than when no such opposite damage exists and may be significantly influenced by other biological processes.

The log-intensity ratios of the six transitive association effects that were described in the previous paragraph are all positive and large, with probable differential effects observed between the comparable transitions for tender-only and the more severe effused (usually also tender) joints, ipsilaterally, where no damage has occurred to the opposite joint. The effects corresponding to

effusion tend to be larger than the comparable effects for tender-only joints, ipsilaterally, again where no damage has occurred to the opposite joint. These six association effects are all the statistically significant effects that were found and the four corresponding to current activity suggest that the link between activity and damage is local or specific to the joint on the particular hand being considered. These four associations are what we would consider as a local dependence or influence of activity on damage at a joint. We do not observe any statistically significant association, or evidence of a possibly substantive effect, of having or not having activity in a joint on one hand with having or not having damage on the contralateral joint of the other hand.

All the analyses that are reported in this subsection treated the disease activity process as explanatory variables and the damage process as the outcome. In addition, the various activity explanatory variables were treated as remaining constant between clinic visits. Such models are extremely important because the explanatory variables represent the information that is available to clinicians in managing a patient’s disease. However, if the aim is to understand the relationship between activity and damage at a more fundamental level, then it is important to recognize that the disease activity process can be a highly variable and dynamic process (especially when medication is available to treat disease activity). Therefore, with intermittently measured activity (as well as damage) and possibly highly variable gap times between clinic visits, the assumption of piecewise constancy of explanatory variables may be unsuitable for some purposes.

In the next section, we address this concern by considering a multistate model which jointly defines the activity and damage processes (i.e. both activity and damage are treated equally as outcomes).

4. Three-state model for activity–damage process

To characterize better the temporal relationship between activity and damage, we propose a new three-state model for each of the 28 phalangeal joints across the left and right hands of a psoriatic arthritis patient. The three states of this model, which combines activity and damage events, are defined as

- (a) state 1, no damage (\bar{D}) and no activity (\bar{A}) in the joint,
- (b) state 2, no damage (\bar{D}) and activity (A) in the joint, and
- (c) state 3, damage (D) in the joint.

A diagrammatic representation of this model is shown in Fig. 3. Transitions between states 1 and 2 (i.e. between the ‘activity’ states) are permitted. The damage state (i.e. state 3) is absorbing and therefore follow-up for a specific joint ends when damage occurs. Thus, observation of the multistate process for a joint is stopped either when damage occurs or at the last visit to the clinic.

Correlation between the 28 multistate joint processes within the hands of a psoriatic arthritis patient is expected and this within-patient correlation needs to be incorporated in the model. However, because of the reversibility of transitions between activity states, it is computationally difficult to introduce subject-specific random effects into this model as done for the four-state model of Section 4. Whereas Cook *et al.* (2004) achieved this with a simple discrete distribution for the random effects in the progressive multistate model setting, we instead adopt a ‘generalized estimating equation’ approach to estimation, in which we assume a working independence correlation structure for the 28 correlated multistate hand joint processes within a patient, and then calculate robust standard errors for the estimated parameters through use of

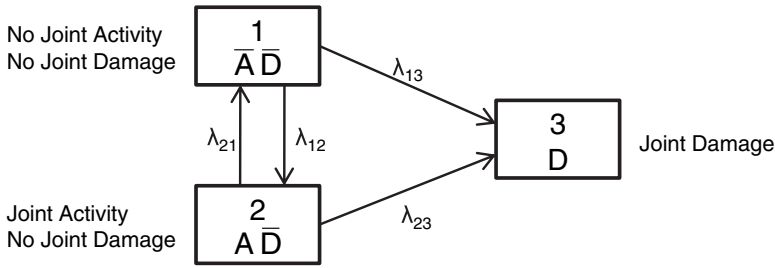


Fig. 3. Diagram of the multistate model for the activity and damage combined process

the robust information sandwich matrix (Cox, 1961; Huber, 1967; White, 1982; Royall, 1986). Our approach is similar to that employed in Lee and Kim (1998) but now for non-progressive clustered processes.

The analysis is performed by using the *msm* package (Jackson, 2008) in R, which implements a standard maximum likelihood estimation approach for multistate models with panel data (Kalbfleisch and Lawless, 1985). This package allows the user to specify explanatory variable effects to act on certain transition intensities only and yields maximum likelihood estimates for the baseline transition intensities and explanatory variable effects. We extract a score vector for each patient, calculated at the maximum likelihood estimates, together with the numerically derived Hessian matrix from the optimization in the *msm* package to obtain an asymptotic robust variance–covariance matrix for the maximum likelihood estimates, using a method presented in Cook and Lawless (2007) (page 342). This robust variance–covariance matrix accounts for the correlation between joints without the use of a subject-specific random effect.

More precisely, we consider for the *k*th patient at the *p*th phalangeal joint of the hands (*p* = 1, . . . , 28), a multistate process $X_k^{(p)} = \{X_k^{(p)}(t) | t \in \mathcal{T}_k\}$, with state space $\Omega = \{1, 2, 3\}$ as described above, and $\mathcal{T}_k = [0, T_{S_k}^{(p)} \wedge \tau_k)$ denoting the time interval of interest (measured in years) which begins at entry to the clinic and stops at the visit to the clinic where damage in the *p*th phalangeal joint is first observed, if it occurs, or else stops at the last visit to the clinic. Let $\lambda_{ijk}^{(p)}(t)$ be the transition intensity function associated with $X_k^{(p)}$ corresponding to the transition $i \rightarrow j$, where $i \neq j$ ($i = 1$ or $i = 2$; $j = 1, 2, 3$). Let $\lambda_{0ij}^{(p)}$ be the baseline transition intensity corresponding to this $i \rightarrow j$ transition. In addition, let $\beta_{ij}^{(p)}$ be a vector of regression parameters associated with the predictable explanatory variable vector process $\mathbf{z}_{ijk}^{(p)}(t)$. Then the transition intensity function for the $i \rightarrow j$ transition, $i \neq j$, at the *p*th phalangeal joint of the hands for the *k*th patient is given by

$$\lambda_{ijk}^{(p)}\{t | \beta_{ij}^{(p)}, \mathbf{z}_{ijk}^{(p)}(t)\} = \lambda_{0ij}^{(p)} \exp\{\beta_{ij}^{(p)T} \mathbf{z}_{ijk}^{(p)}(t)\}. \tag{5}$$

For parsimony and because we believe that these 28 multistate activity–damage processes describe a similar phenomenon, we assume that $\beta_{ij}^{(p)} = \beta_{ij}$ and that $\lambda_{0ij}^{(p)} = \lambda_{0ij}$ in equation (5).

The analysis of the individual joint level data which uses this three-state model is described in the next subsection. The analysis is based on the 510 patients with complete joint level information on activity and damage.

4.1. Jointly modelling activity and damage at joint level

The observed joint transition table for the three-state model that is illustrated in Fig. 3 is shown in Table 5. There are 134295 observed joint transitions among all the hand joints from the 510 patients. This corresponds to approximately nine transitions observed per patient hand joint. The actual number of observed transitions between differing states was 14856. By the end of the study period 772 hand joints were damaged, which corresponds to 5% of all 14280

Table 5. Observed joint transitions for the three-state activity–damage multistate model

Transition from	Number of transitions to the following states:		
	(\bar{A}, \bar{D})	(A, \bar{D})	D
(\bar{A}, \bar{D})	113793	6765	510
(A, \bar{D})	7319	5646	262
D	0	0	0

(= 28 × 510) hand joints. The total hand-joint-years of follow-up until damage or administrative censoring was 122008.6 years, and the observed rate of hand joint damage is six joints per 1000 hand-joint-years of follow-up.

We fit to these data the three-state model defined in equation (5), with a five-level factor variable for type of joint (i.e. metacarpophalangeal, proximal interphalangeal, distal interphalangeal, metacarpophalangeal 1 and proximal interphalangeal 1—see Fig. 1), a time-dependent explanatory variable indicating whether the joint at the previous visit to the clinic had ever been observed active (at this visit or in the past) and a time-dependent variable indicating whether or not the corresponding joint in the opposite hand is in a state of damage for $\mathbf{z}(t)$. We fitted this model with explanatory variable effects on all transitions and performed Wald tests, using the robust variance–covariance matrix, to determine whether or not effects on all transitions were required. As a result, the time-dependent explanatory variable representing past activity and the joint type factor were excluded from the equations for the transition intensities out of state 2. Additionally, we parameterized the baseline 2 → 3 transition intensity in terms of the baseline 1 → 3 transition intensity, i.e. $\lambda_{023} = \lambda_{013} \exp(\gamma)$, adjusting for explanatory variables.

Table 6 shows the results of fitting this model. We observe strong evidence to support the hypothesis that $\lambda_{023} > \lambda_{013}$. This evidence is reflected in a large γ -estimate of 3.08 (robust 95% confidence interval (1.76, 4.40)), which equates to an approximately 22 times greater transition rate to damage from an active (but not damaged) state than from a state of non-activity (and not damaged) just before the transition to damage, other factors being the same. In addition, there is evidence of the ‘ever active’ variable being positively associated with the 1 → 2 and 1 → 3 transitions. Thus a joint ever being recorded as active before the current visit increases both the baseline transition rates of making an instantaneous transition from a state of ‘no damage, no activity’ to ‘no damage, activity’ and to ‘damage’. In controlling for damage in the same joint of the opposite hand, we see that damage on the opposite hand has a significant effect only on the transitions to damage (i.e. 1 → 3 and 2 → 3), further strengthening the argument for symmetric damage seen in the previous four-state model.

The value $\exp(\gamma) = 21.72$ describes the ratio between λ_{23} and λ_{13} in the absence of damage to the opposite joint. Since the effects of damage to the opposite joint are of similar magnitude, we obtain a similar estimate of $21.72 \times 7.50/7.19 = 22.65$ for this ratio in the presence of damage to the opposite joint. This concurs to some extent with the results from the four-state model which showed that the presence of activity has a significant effect on the rate of transition to damage in a joint where the same joint in the opposite hand already exhibits damage. Furthermore, we note that there is a negligible effect of opposite joint damage on the 1 → 2 transition intensity. This suggests that the apparent symmetric effect of opposite joint damage is not mediated through an increase in disease activity but rather through another, unknown, mechanism.

Table 6. Intensity ratio and baseline intensity parameter estimates together with associated naive and robust 95% confidence intervals for the model that includes the effects of joint type, previous activity on transitions from state 1 and opposite damage on all transitions

Parameter	Estimate	95% confidence interval	
		Naive	Robust
<i>Baseline intensities ($\times 10^{-2}$)</i>			
λ_{012}	9.24	(8.19, 10.43)	(6.76, 12.63)
λ_{013}	0.14	(0.07, 0.30)	(0.04, 0.49)
λ_{021}	151.48	(146.10, 157.10)	(126.25, 181.76)
<i>Intensity ratios</i>			
$\exp(\gamma)$	21.72	(11.89, 39.69)	(5.79, 81.46)
Metacarpophalangeal on 1 \rightarrow 2	1.10	(0.99, 1.21)	(0.82, 1.47)
Metacarpophalangeal on 1 \rightarrow 3	0.37	(0.16, 0.86)	(0.05, 2.75)
Proximal on 1 \rightarrow 2	1.24	(1.12, 1.36)	(0.96, 1.59)
Proximal on 1 \rightarrow 3	1.04	(0.56, 1.94)	(0.31, 3.52)
Distal on 1 \rightarrow 2	0.86	(0.77, 0.95)	(0.63, 1.16)
Distal on 1 \rightarrow 3	2.82	(1.60, 4.96)	(0.95, 8.36)
Thumb metacarpophalangeal on 1 \rightarrow 2	1.26	(1.12, 1.42)	(0.97, 1.63)
Thumb metacarpophalangeal on 1 \rightarrow 3	2.68	(1.43, 5.04)	(0.88, 8.13)
Joint ever active on 1 \rightarrow 2	2.98	(2.84, 3.13)	(2.13, 4.17)
Joint ever active on 1 \rightarrow 3	3.07	(2.38, 3.96)	(1.27, 7.40)
Opposite joint damage on 1 \rightarrow 2	1.20	(1.01, 1.44)	(0.64, 2.27)
Opposite joint damage on 1 \rightarrow 3	7.19	(4.90, 10.54)	(2.11, 24.44)
Opposite joint damage on 2 \rightarrow 1	1.08	(0.91, 1.28)	(0.64, 1.82)
Opposite joint damage on 2 \rightarrow 3	7.50	(5.08, 11.09)	(2.40, 23.46)

Further generalization of our three-state model to incorporate piecewise constant baseline intensities resulted in some evidence for a decline in the transition rate from the state of ‘no damage, no activity’ (state 1) to the state of ‘no damage, activity’ (state 2) as the time in clinic increased. However, no evidence was found for changes over time in the transition rate from state 2 to state 1 and the two transition rates into the ‘damage’ state (state 3). Moreover the γ -estimates that were obtained when the follow-up time period was stratified into three intervals either as [0, 1), [1, 3) and [3, ∞) or [0, 5), [5, 10) and [10, ∞) were 3.155, 3.015 and 2.714 or 3.048, 2.618 and 2.807 respectively. These estimates were well within the 95% robust confidence interval of 1.76–4.40 that was reported in the previous paragraph and provided no evidence for a marked change in γ over time. Finally, there was no demonstrable modification of the estimated explanatory variable effects in Table 6 when these additional time inhomogeneous three-state models were fitted.

The analyses of this subsection again demonstrate a local dependence at the joint level between activity and damage (i.e. activity influences damage locally) and characterize more fully the temporal ordering.

5. Causality

5.1. Framework

In the section on ‘Assessing causality’ in a classic epidemiological text (Lilienfeld and Stolley, 1994) it is argued that

‘a relationship is considered causal whenever evidence indicates that the factors form part of the com-

plex of circumstances which increases the probability of the occurrence of disease and that a diminution of one or more of these factors decreases the frequency of the disease’.

Sir Austin Bradford Hill, in his 1965 Presidential address to the Section of Occupational Medicine of the Royal Society of Medicine (Hill, 1965), discussed aspects of an association that should be especially considered when attempting to infer causation in this sense from association. These aspects, which are now referred to as the Bradford Hill criteria, are

- (a) strength of association,
- (b) consistency,
- (c) specificity,
- (d) temporality,
- (e) biological gradient,
- (f) biological plausibility,
- (g) coherence,
- (h) experimental evidence (when available) and
- (i) analogy.

However, Hill never intended these ‘viewpoints’ to be ‘hard-and-fast rules of evidence that *must* be obeyed before we can accept cause and effect’ (Hill, 1965). They are not necessary and/or sufficient conditions to declare causation from an observed association, although temporality is indisputably a necessary condition as a cause must precede its effect. They are, however, useful in providing a structure when attempting to move from association to causation. In particular, they (excluding criterion (h)) have been found to be valuable in epidemiological settings.

Schweder (1970) regarded his concept of local (in)dependence as a potential aid in addressing causal questions and it is closely linked to the concept of Granger causality (Granger, 1969). We note that Granger causality was initially defined in the context of discrete time series, where suppose that there are two time series $X(t)$ and $Y(t)$, and $\mathcal{U}(t)$ represents ‘all of the information in the universe’ up to time $t \in \mathbb{Z}^+$. Then we say that $X(t)$ is Granger causal for $Y(t)$ if Y may be better predicted at time $t + 1$ given $\mathcal{U}(t)$ than given $\mathcal{U}(t) \setminus X(t)$. Otherwise we say that $X(t)$ is Granger non-causal for $Y(t)$. Local (in)dependence may be viewed somewhat as an extension of the Granger causality–non-causality concept to processes in continuous time. The asymmetry of the local independence concept makes it particularly attractive, as having one subprocess locally influencing a change in another, but the other not having any influence on the first, is precisely how we would like to characterize a causal effect of Y_j on Y_k . However, a one-sided local dependence relationship between two subprocesses of a composable Markov process is not sufficient to imply causation.

Aalen (1987), in extending the concept, stressed that local (in)dependence was a dynamic statistical approach which, by incorporating time explicitly, offers a natural way to model potential causal relationships. This is reflected in other writings which take a dynamic viewpoint including that of Arjas and Parner (2004) who not only believe that explicit accounting of the time aspect should be made because of the time ordering of cause and effect but also because often the durations between different events are an integral part of the causal problem and therefore of the analysis. This viewpoint is also seen in Commenges and Gégout-Petit (2009) who developed a general dynamical model as a framework for causal interpretation, which uses the concepts of ‘the system’ and ‘causal influence’, is centred on stochastic processes and further builds on the local dependence concept.

In our discussion of psoriatic arthritis disease progression that follows, we use both the Bradford Hill criteria and Schweder’s local (in)dependence concept to reason about whether

associations that were found in our analyses, which are based on multistate models at the individual joint level of the hands, can be considered causal. We recognize that there are other formal frameworks for thinking about causality (e.g. the decision theoretic (conditioning by intervention, do-calculus), counterfactuals (potential outcomes) and causal graphs (marginal structural models) approaches) that have been applied in the literature (Dawid, 2000; Pearl, 2000; Rubin, 1974, 1978; Robins *et al.*, 2000). However, we believe that, for our application, the Bradford Hill criteria and Schweder's local (in)dependence are more natural and helpful methods for inferring causal relationships, as our focus is not on intervention but on further mechanistic understanding of the disease. See Aalen and Frigessi (2007) for a discussion of mechanistic causality.

5.2. Information for causal inference

As described previously, recent work (Bond *et al.*, 2007; Bond and Farewell, 2009) has demonstrated that higher prior active joint counts are related to higher rates of damage development at the patient level. Also, the differential effects of tender only and the more severe effused (usually also tender) provide some indication of a dose–response relationship. The analyses of Sections 3 and 4 allow us to consider the extent to which these analyses performed at the individual joint level, and the associations that may be demonstrated, aid our understanding of disease progression and inform us additionally about causality.

In addition to the recognition that association cannot be used to prove causation, inference about causation also needs to reflect the possibility of different causal pathways. Thus, a link between activity and damage at the individual joint level need not preclude the involvement of other factors nor completely separate pathways to damage.

In Section 3.2, evidence emerged of a symmetric pattern of joint damage. For a pair of joints at the same location in the two hands, local dependences were identified in both directions between the damage subprocesses in the left and right hands. Although these results represent an important finding, these local dependences do not immediately warrant a causal explanation, as they do not produce a one-sided (local dependence) asymmetric relationship between the two damage subprocesses. It is quite plausible that the same underlying biological mechanism is driving these two damage subprocesses, although a robust biological explanation for damage symmetry has not been put forward and tested yet. However, there are some animal data to suggest neurological influence on inflammation (Chahl and Ladd, 1976; Denko and Petricevic, 1978), which, as Levine *et al.* (1987), Helliwell *et al.* (2000) and Bukhari *et al.* (2002) have discussed, may indicate that

‘a biological mechanism exists whereby afferent nerves from one joint can induce an inflammatory response in the contralateral joint by inducing the release of inflammatory mediators’

(Bukhari *et al.*, 2002). Although this explanation is somewhat speculative, it would account for the observed but non-causal relationship of damage in one hand being associated with subsequent damage in the other hand.

Thus, in exploring causal links between activity and damage at the individual joint level, it must be accepted that any link will be part of a complex of factors that influence the extent of damage seen in psoriatic arthritis patients.

However, the results in Section 3.3 do suggest

- (a) joint specificity of the relationship between activity and damage;
- (b) strong associations for the four statistically significant and biologically plausible local effects obtained and, where no previous damage has occurred in either joint of a pair,

- (c) a dose–response relationship of activity with damage at the joint level (i.e. a biological gradient).

Similar effects of tenderness and effusion, rather than a dose–response relationship, were seen where the opposite joint of a pair exhibits damage, though we note that the damage process may become more complex once the opposite joint is damaged, especially in light of the symmetric relationships that were discussed in Section 3.2. If other causal pathways influence damage where opposite damage exists then a dose–response relationship may be more difficult to detect. As indicated in Section 3.3, we would consider these relationships to represent local dependence of activity on damage although the activity process is not formally modelled. Here and in Section 4 the determining of asymmetry of local dependence relationships is not specifically addressed and is less critical because any influence of damage on subsequent activity in a joint is of less clinical interest or, at least, represents a very different clinical question. However, as Aalen argued, the dynamic perspective of Schweder's work may be most important and the analysis of Section 3.3, because of the longitudinal nature of the data, does help to characterize the temporal relationship between activity and damage.

Therefore these results do provide support for a putative causal relationship between activity and damage. It can also be argued, we feel, that, from the perspective of both biological plausibility and temporal ordering, this demonstrated relationship at the joint level offers more support for a causal link than the relationships at the patient level that were seen in previous investigations.

Results from the three-state model in Section 4 confirm that the link between activity and damage at the individual joint level is seen even in a model which more realistically reflects the highly variable nature of the activity process. This further strengthens the evidence for the association and the temporal ordering and, thus, for a causal link. Moreover, although there is some increase in the baseline transition rate from the inactive state if the joint has ever been active, from 0.0014 to $0.0014 \times 3.07 = 0.0043$, this is much less than the rate from the active state, 0.0304. Thus, the causal argument is enhanced since the disappearance of activity has a similar (but negative) effect to that of the appearance of activity. We note also that in this model, where activity is modelled in continuous time, there is no evidence to suggest that the presence of damage in the opposite joint of a pair diminishes the increase in the transition rate to damage from a state of activity when compared with that from a state of inactivity.

5.3. *Does activity cause damage?*

The determination of causality from observational data is notoriously difficult. Mathematical formalization of the causality concept, in itself, is not enough to allow a causal relationship to be inferred from an observed association. However, the presence of a mathematical framework to reason about causality is helpful, if not essential, when attempting to address clinical questions of a causal nature. The multistate modelling framework provides a good illustration of how mathematical and statistical concepts, such as local independence and local dependence, allow causal hypotheses to be framed in a way such that they can be investigated in a statistically coherent manner. Mathematical formalization together with sound application of the Bradford Hill criteria may represent a useful strategy for inferring causality from observational data.

Because causation cannot be proved with observational data, there can be debate about the extent to which causality is supported. For example, activity might be regarded as an early stage of a damage process so that their temporal ordering is simply a natural course of disease with no causal implications. However, because activity is clearly part of an inflammatory process

which is generally transient and damage is a condition which is irreversible and need not be associated with concurrent inflammation, some sort of causal link appears more reasonable. This argument is more reasonable because we have been able to demonstrate, to a considerable degree, that this damage process will occur at the joint level. Of course, if we did not find this joint level relationship, there might be other causal explanations for previous findings relating patient level associations between the extent of activity and progression of damage. These arguments must, however, be somewhat more complex and speculative and that is why we would argue that the establishment of an activity–damage relationship at the joint level ‘strengthens’ the case for a causal pathway.

Similarly, investigation of the biological gradient, or dose–response, Bradford Hill criterion can be context specific in its application. For example, most diets in the developed world contain sufficiently high levels of phenylalanine to induce neurological manifestations for individuals lacking the enzyme phenylalanine hydroxylase. However, elsewhere, a dose–response relationship will exist between the level of phenylalanine and the level of neurological problems. Thus the primary risk factor for the disease phenylketonuria is binary (genetic) when the gene is rare and a continuous dietary measure where the genetic condition is common.

One might therefore consider a ‘trigger’ hypothesis for the relationship between activity and damage. If the onset of inflammation is all that is needed to initiate a process leading to damage, then no dose–response or more complex temporal relationship with activity over time would be seen. However, conversely, if we do see such relationships then such a ‘trigger’ hypothesis is less plausible.

Previous analyses (see Bond *et al.* (2007) for a summary and further references) suggested erythrocyte sedimentation rate as the most likely patient level confounder of activity and damage relationship, and it was included in those analyses. As reported in Section 3.3, it did not have a demonstrable effect in models at the joint level. The use of random effects will mitigate the potential influence of other unmeasured confounders at the patient level. Potential confounders measured at the individual joint level are not easily envisaged on the basis of published literature on arthritic damage in primarily autoimmune arthritides.

Bond and Farewell (2009) also addressed the issue of potentially informative observation times for these data. Although some evidence for such informative observation was present, allowance for it in the study did not demonstrate any marked bias in the relationship between activity and damage.

Thus the work that is reported here represents our effort to substantiate the well-held, but yet unproven, claim that activity causes damage in psoriatic arthritis. Our systematic approach to assessing this claim, which crucially relied on the analyses of activity and damage data at the hand joint level and not only at the patient level, leads us to conclude that there is strong epidemiological evidence for a causal relationship between activity and damage, although other causal pathways may also be present. This conclusion is based on our results being consistent with the majority of the Bradford Hill criteria: specificity, strength of association, biological gradient, temporality and biological plausibility, at the joint level. Others such as consistency and analogy have been shown at the patient (systemic) level in other psoriatic arthritis populations and in RA populations respectively. Moreover our results do not, in any way that we know of, conflict with generally known facts regarding the biology and natural history of progression of disease in psoriatic arthritis patients, thus suggesting coherence. Furthermore, there are recent clinical trials that have shown that new biological therapies are effective in slowing the progression of arthritis disease. These biologics calm the inflammation of arthritis by inhibiting the components of the immune system that trigger the inflammatory response.

6. Conclusion

The study of the progression of psoriatic arthritis by using multistate models has provided an intuitive way of examining the disease process from a dynamic perspective. We considered the damage process in the individual joints as a continuous Markov process and this was modelled by using various multistate models over a series of discrete states. In our four-state model (Section 3), we considered the transition intensities relating to the damage process to be functions of the activity process at the previous visit to the clinic in each pair of hand joints. This allowed an assessment of the influence of disease activity on joint damage, while incorporating explicitly the passage of time and the inherent correlation between the joints through the use of a subject-specific random effect. Explanatory variables that are found to affect the transitions to states of damage significantly may imply a local dependence between activity and damage, in accordance with the work of Schweder. The three-state model (Section 4) allowed joint modelling of activity and damage as continuous Markov processes and provided a further insight into how the activity process may influence the damage process. The fit of this model suggested strong evidence of an effect of disease activity on joint damage, again demonstrating a local dependence between activity and damage at the individual joint level, while accounting more appropriately for the dynamic nature of both the activity and damage processes.

Overall, the use of multistate models has proved indispensable in the assessment of local (in)dependences in dynamic processes. Our application of these ideas at the joint level in psoriatic arthritis patients has allowed us to determine the extent to which evidence of local dependence between the activity and damage processes, together with the Bradford Hill criteria, permits a causal link between activity and damage. We conclude, after consideration of other possible explanations, that there is significant evidence in our analyses to suggest a causal relationship between disease activity and clinical damage in the hand joints of patients suffering from psoriatic arthritis, although the biological mechanism by which this occurs is still to be fully understood.

Our particular attempt to move from association to causation by using Schweder’s local independence and local dependence concepts and the Bradford Hill criteria is, to our knowledge, one of the first comprehensive attempts along these lines at establishing causality in a well-motivated, important and substantial medical problem. Our analyses used data from the largest cohort of psoriatic arthritis patient followed up over time in the world, with data at the most appropriate level (i.e. the joint level) for investigating the causal relationship between activity and disease progression. We recommend that this approach to inferring causality from cohort data be considered more often in the future.

Acknowledgement

This work was financially supported by the Medical Research Council, UK, funding MC_US_A030_0022.

Appendix A: Likelihood derivation for four-state model

To derive the likelihood we take advantage of the fact that our model, conditional on the random effects, is Markov. Therefore as described earlier, the contribution to the conditional likelihood, $L_k(\lambda_0, \beta, \theta|u_k)$, from the k th patient can be written as the double product of the transition probabilities, conditional on u_k , over the observed trajectories, $\{(x_{1lk}, \dots, x_{n_k lk}) : l = 1, \dots, 14\}$, followed by the joint pairs at the 14 hand joint locations of this patient through the n_k visits to the clinic, i.e.

$$L_k(\lambda_0, \beta, \theta|u_k) = \prod_{l=1}^{14} \prod_{m=1}^{n_k-1} \mathbb{P}\{X_k^{(l)}(t_{k,m+1}) = x_{m+1,lk} | X_k^{(l)}(t_{km}) = x_{mlk}, U_k = u_k; \lambda_0^{(l)}, \beta^{(l)}, \mathbf{z}_k^{(l)}(t_{km})\},$$

where the notation used is as in Section 3.

As we include only those patients who entered the psoriatic arthritis clinic in state 1 for all hand joint locations, each joint pair of every patient considered in our analysis undergoes one of a possible six observed progression paths throughout the study. These are as follows.

- (a) The joint pair remains in state 1 throughout the study.
- (b) The joint pair moves from state 1 to state 2 and remains in state 2 for the remainder of the study.
- (c) The joint pair moves from state 1 to state 3 and remains in state 3 for the remainder of the study.
- (d) The joint pair moves from state 1 to state 4 with the intermediate state, either 2 or 3, unknown.
- (e) The joint pair moves from state 1 to state 2 to state 4.
- (f) The joint pair moves from state 1 to state 3 to state 4.

Each of these possible progression paths yields a different contribution to the likelihood for each patient and, because we have a relatively straightforward progressive conditional Markov model, these transition probabilities, conditional on the subject-specific random effect, may be determined analytically.

Suppose that we have N patients under observation and we define the state $i \rightarrow$ state j transition probability for joint pair l of the k th patient, conditional on $U_k = u_k$, to be

$$p_{ijk}^{(l)}\{s, t | \mathbf{z}_k^{(l)}(s), u_k\} = \mathbb{P}\{X_k^{(l)}(t) = j | X_k^{(l)}(s) = i, U_k = u_k; \mathbf{z}_k^{(l)}(s)\},$$

for generic times t and s , with $t > s$. Additionally, we define the number of study observations for the l th joint pair of the k th patient to be $n_k^{(l)}$ of which $n_{ki}^{(l)}$ occur in state $i \in \Omega = \{1, \dots, 4\}$, and hence $\sum_{i=1}^4 n_{ki}^{(l)} = n_k^{(l)}$. Next we let the $n_{ki}^{(l)}$ ordered observation times corresponding to when joint pair l of patient k is in state i be denoted by $\mathbf{t}_{ki}^{(l)} = (t_{ki(1)}^{(l)}, t_{ki(2)}^{(l)}, \dots, t_{ki(n_{ki}^{(l)})}^{(l)})^T$.

Now, assuming that the transition intensities, conditional on the random effects, are as given in model (1); then the transition probabilities $\{p_{ijk}^{(l)}\{s, t | \mathbf{z}_k^{(l)}(s), u_k\}\}$ in our four-state model will be as shown in Appendix A.1. And the conditional likelihood contributions under the six possible observed progression paths for the l th joint pair process of the k th patient are as follows.

- (a) *Progression path 1 → 1:*

$$L_{k(1 \rightarrow 1)}^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) = \prod_{i=1}^{n_{k1}^{(l)}-1} p_{11k}^{(l)}\{t_{k1(i)}^{(l)}, t_{k1(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(i)}^{(l)}), u_k\}.$$

- (b) *Progression path 1 → 2:*

$$L_{k(1 \rightarrow 2)}^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) = \left[\prod_{i=1}^{n_{k1}^{(l)}-1} p_{11k}^{(l)}\{t_{k1(i)}^{(l)}, t_{k1(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(i)}^{(l)}), u_k\} \right] p_{12k}^{(l)}\{t_{k1(n_{k1}^{(l)})}^{(l)}, t_{k2(1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(n_{k1}^{(l)})}^{(l)}), u_k\} \\ \times \prod_{i=1}^{n_{k2}^{(l)}-1} p_{22k}^{(l)}\{t_{k2(i)}^{(l)}, t_{k2(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k2(i)}^{(l)}), u_k\}.$$

- (c) *Progression path 1 → 3:*

$$L_{k(1 \rightarrow 3)}^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) = \left[\prod_{i=1}^{n_{k1}^{(l)}-1} p_{11k}^{(l)}\{t_{k1(i)}^{(l)}, t_{k1(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(i)}^{(l)}), u_k\} \right] p_{13k}^{(l)}\{t_{k1(n_{k1}^{(l)})}^{(l)}, t_{k3(1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(n_{k1}^{(l)})}^{(l)}), u_k\} \\ \times \prod_{i=1}^{n_{k3}^{(l)}-1} p_{33k}^{(l)}\{t_{k3(i)}^{(l)}, t_{k3(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k3(i)}^{(l)}), u_k\}.$$

- (d) *Progression path 1 → 4:*

$$L_{k(1 \rightarrow 4)}^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) = \left[\prod_{i=1}^{n_{k1}^{(l)}-1} p_{11k}^{(l)}\{t_{k1(i)}^{(l)}, t_{k1(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(i)}^{(l)}), u_k\} \right] p_{14k}^{(l)}\{t_{k1(n_{k1}^{(l)})}^{(l)}, t_{k4(1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(n_{k1}^{(l)})}^{(l)}), u_k\}.$$

- (e) *Progression path 1 → 2 → 4:*

$$L_{k(1 \rightarrow 2 \rightarrow 4)}^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) = \left[\prod_{i=1}^{n_{k1}^{(l)}-1} p_{11k}^{(l)}\{t_{k1(i)}^{(l)}, t_{k1(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(i)}^{(l)}), u_k\} \right] p_{12k}^{(l)}\{t_{k1(n_{k1}^{(l)})}^{(l)}, t_{k2(1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(n_{k1}^{(l)})}^{(l)}), u_k\} \\ \times \left[\prod_{i=1}^{n_{k2}^{(l)}-1} p_{22k}^{(l)}\{t_{k2(i)}^{(l)}, t_{k2(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k2(i)}^{(l)}), u_k\} \right] p_{24k}^{(l)}\{t_{k2(n_{k2}^{(l)})}^{(l)}, t_{k4(1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k2(n_{k2}^{(l)})}^{(l)}), u_k\}.$$

(f) Progression path $1 \rightarrow 3 \rightarrow 4$:

$$L_{k(1 \rightarrow 3 \rightarrow 4)}^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) = \left[\prod_{i=1}^{n_{k1}^{(l)}-1} p_{11k}^{(l)}(t_{k1(i)}^{(l)}, t_{k1(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(i)}^{(l)}), u_k) \right] p_{13k}^{(l)}(t_{k1(n_{k1}^{(l)})}^{(l)}, t_{k3(1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k1(n_{k1}^{(l)})}^{(l)}), u_k) \times \left[\prod_{i=1}^{n_{k3}^{(l)}-1} p_{33k}^{(l)}(t_{k3(i)}^{(l)}, t_{k3(i+1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k3(i)}^{(l)}), u_k) \right] p_{34k}^{(l)}(t_{k3(n_{k3}^{(l)})}^{(l)}, t_{k4(1)}^{(l)} | \mathbf{z}_k^{(l)}(t_{k3(n_{k3}^{(l)})}^{(l)}), u_k).$$

Therefore the contribution to the conditional likelihood from joint pair l of the k th patient can be written in full generality as

$$L_k^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) = \mathbb{1}_{\{1 \rightarrow 1\}} L_{k(1 \rightarrow 1)}^{(l)}(\lambda_0, \beta, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) + \mathbb{1}_{\{1 \rightarrow 2\}} L_{k(1 \rightarrow 2)}^{(l)}(\lambda_0, \beta, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) + \mathbb{1}_{\{1 \rightarrow 3\}} L_{k(1 \rightarrow 3)}^{(l)}(\lambda_0, \beta, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) + \mathbb{1}_{\{1 \rightarrow 2 \rightarrow 4\}} L_{k(1 \rightarrow 2 \rightarrow 4)}^{(l)}(\lambda_0, \beta, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k) + \mathbb{1}_{\{1 \rightarrow 3 \rightarrow 4\}} L_{k(1 \rightarrow 3 \rightarrow 4)}^{(l)}(\lambda_0, \beta, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k)$$

where $\mathbb{1}_{\{TP\}}$ is the indicator function taking the value 1 if joint j undergoes the transition progression path TP and 0 otherwise. Thus the overall contribution to the conditional likelihood from the k th patient may be written as

$$L_k(\lambda_0, \beta, \theta | \mathbf{t}_k, \mathbf{z}_k, u_k) = \prod_{l=1}^{14} L_k^{(l)}(\lambda_0^{(l)}, \beta^{(l)}, \theta | \mathbf{t}_k^{(l)}, \mathbf{z}_k^{(l)}, u_k).$$

We account for all possible values of the random effect U_k by integrating out this term (numerically by using R). Thus the contribution to this ‘marginal’ likelihood from the k th patient is given by

$$L_k(\lambda_0, \beta, \theta | \mathbf{t}_k, \mathbf{z}_k) = \int_0^\infty L_k(\lambda_0, \beta, \theta | \mathbf{t}_k, \mathbf{z}_k, u) f_{U_k}(u) du$$

where $U_k \sim \text{gamma}(a, b)$, with $a = 1/\theta$ and $b = 1/\theta$, i.e.

$$f_{U_k}(u) = \frac{b^a u^{a-1} \exp(-bu)}{\Gamma(a)} = \frac{(1/\theta)^{1/\theta} u^{1/\theta-1} \exp(-u/\theta)}{\Gamma(1/\theta)}, \quad \forall k. \tag{6}$$

Hence the likelihood function across all N patients in the study is given by

$$L(\lambda_0, \beta, \theta) = \prod_{k=1}^N L_k(\lambda_0, \beta, \theta | \mathbf{t}_k, \mathbf{z}_k). \tag{7}$$

We maximize function (7) with respect to the baseline intensity parameters λ_0 , the regression parameters β and the variance of the random effects, θ . This is done by using the `optim` command in R, which yields parameter estimates along with a numerically derived Hessian matrix at the estimate values from which asymptotic standard errors for these estimates may be obtained.

A.1. Form of the transition probabilities

Suppose that a patient is observed at successive observation times s and t where $s < t$. We provide below the analytical form of the transition probabilities $p_{ijk}(s, t | \mathbf{z}_k, u_k)$. For convenience, we have dropped the l -superscript in what follows:

$$p_{11k}(s, t | \mathbf{z}_k, u_k) = \exp(-u_k[\lambda_{12}\{\beta_{12}, \mathbf{z}_k(s)\} + \lambda_{13}\{\beta_{13}, \mathbf{z}_k(s)\}](t - s));$$

$$p_{12k}(s, t | \mathbf{z}_k, u_k) = \frac{\lambda_{12}\{\beta_{12}, \mathbf{z}_k(s)\}}{\lambda_{12}\{\beta_{12}, \mathbf{z}_k(s)\} + \lambda_{13}\{\beta_{13}, \mathbf{z}_k(s)\} - \lambda_{24}\{\beta_{24}, \mathbf{z}_k(s)\}} \times \{\exp[-u_k \lambda_{24}\{\beta_{24}, \mathbf{z}_k(s)\}](t - s) - \exp(-u_k[\lambda_{12}\{\beta_{12}, \mathbf{z}_k(s)\} + \lambda_{13}\{\beta_{13}, \mathbf{z}_k(s)\}](t - s))\};$$

$$p_{13k}(s, t | \mathbf{z}_k, u_k) = \frac{\lambda_{13}\{\beta_{13}, \mathbf{z}_k(s)\}}{\lambda_{12}\{\beta_{12}, \mathbf{z}_k(s)\} + \lambda_{13}\{\beta_{13}, \mathbf{z}_k(s)\} - \lambda_{34}\{\beta_{34}, \mathbf{z}_k(s)\}} \times \{\exp[-u_k \lambda_{34}\{\beta_{34}, \mathbf{z}_k(s)\}](t - s) - \exp(-u_k[\lambda_{12}\{\beta_{12}, \mathbf{z}_k(s)\} + \lambda_{13}\{\beta_{13}, \mathbf{z}_k(s)\}](t - s))\};$$

$$\begin{aligned}
p_{14k}(s, t | \mathbf{z}_k, u_k) &= 1 - p_{11k}(s, t | \mathbf{z}_k, u_k) - p_{12k}(s, t | \mathbf{z}_k, u_k) - p_{13k}(s, t | \mathbf{z}_k, u_k); \\
p_{22k}(s, t | \mathbf{z}_k, u_k) &= \exp[-u_k \lambda_{24} \{\beta_{24}, \mathbf{z}_k(s)\} (t - s)]; \\
p_{24k}(s, t | \mathbf{z}_k, u_k) &= 1 - \exp[-u_k \lambda_{24} \{\beta_{24}, \mathbf{z}_k(s)\} (t - s)]; \\
p_{33k}(s, t | \mathbf{z}_k, u_k) &= \exp[-u_k \lambda_{34} \{\beta_{34}, \mathbf{z}_k(s)\} (t - s)]; \\
p_{34k}(s, t | \mathbf{z}_k, u_k) &= 1 - \exp[-u_k \lambda_{34} \{\beta_{34}, \mathbf{z}_k(s)\} (t - s)]; \\
p_{44k}(s, t | \mathbf{z}_k, u_k) &= 1.
\end{aligned}$$

Here

$$\lambda_{ij} \{\beta_{ij}, \mathbf{z}_k(s)\} = \lambda_{0ij} \exp\{\beta_{ij}^T \mathbf{z}_{ijk}(s)\}$$

and we note that β_{ij} is the subvector of β that acts on the transition intensities from state i to state j , i.e. $\beta^T = (\beta_{12}^T, \beta_{13}^T, \beta_{24}^T, \beta_{34}^T)$.

References

- Aalen, O. O. (1987) Dynamic modelling and causality. *Scand. Act. J.*, **4**, 177–190.
- Aalen, O. O., Borgan, Ø. and Gjessing, H. K. (2008) *Survival and Event History Analysis: a Process Point of View*. New York: Springer.
- Aalen, O. O., Borgan, Ø., Keiding, N. and Thormann, J. (1980) Interaction between life history events: non-parametric analysis for prospective and retrospective data in the presence of censoring. *Scand. J. Statist.*, **7**, 161–171.
- Aalen, O. O. and Frigessi, A. (2007) What can statistics contribute to a causal understanding? *Scand. J. Statist.*, **34**, 155–168.
- Aletaha, D., Funovits, J., Breedveld, F., Shar, J., Segurando, O. and Smolen, J. S. (2009) Rheumatoid arthritis joint progression in sustained remission is determined by disease activity level preceding the period of radiographic assessment. *Arth. Rheum.*, **60**, 1242–1249.
- Andersen, P. K. and Keiding, N. (2002) Multi-state models for event history analysis. *Statist. Meth. Med. Res.*, **11**, 91–115.
- Arjas, E. and Parner, J. (2004) Causal reasoning from longitudinal data. *Scand. J. Statist.*, **31**, 171–187.
- Bond, S. J. and Farewell, V. T. (2009) Likelihood estimation for a longitudinal negative binomial regression model with missing outcomes. *Appl. Statist.*, **58**, 369–382.
- Bond, S. J., Farewell, V. T., Schentag, C. T. and Gladman, D. D. (2007) Predictors for radiological damage in psoriatic arthritis: results from a single centre. *Ann. Rheum. Dis.*, **66**, 370–376.
- Broyden, C. (1970) The convergence of a class of double-rank minimisation algorithms. *J. Inst. Math. Applic.*, **6**, 76–90.
- Bukhari, M., Lunt, M., Harrison, B. J., Scott, D. G. I., Symmons, D. P. M. and Silman, A. J. (2002) Erosions in inflammatory polyarthritis are symmetrical regardless of rheumatoid factor status: results from a primary care-based inception cohort of patients. *Rheumatology*, **41**, 246–252.
- Chahl, L. A. and Ladd, J. (1976) Local oedema and general excitation of cutaneous sensory receptors produced by electrical stimulation of the saphenous nerve in the rat. *Pain*, **2**, 25–34.
- Commenges, D. (1999) Multi-state models in epidemiology. *Lifetim. Data Anal.*, **5**, 315–327.
- Commenges, D. and Gégout-Petit, A. (2009) A general dynamical statistical model with causal interpretation. *J. R. Statist. Soc. B*, **71**, 719–736.
- Cook, R. J. and Lawless, J. F. (2007) *The Statistical Analysis of Recurrent Events*. New York: Springer.
- Cook, R. J., Yi, G. Y., Lee, K.-A. and Gladman, D. D. (2004) A conditional Markov model for clustered progressive multistate processes under incomplete observation. *Biometrics*, **60**, 436–443.
- Cox, D. R. (1961) Tests of separate families of hypotheses. In *Proc. 4th Berkeley Symp. Mathematics, Probability and Statistics*, vol. 1 (ed. J. Neyman), pp. 105–123. Berkeley: University of California Press.
- Cox, D. R. (1972) The statistical analysis of dependencies in point processes. In *Stochastic Point Processes: Statistical Analysis, Theory and Applications* (ed. P. A. W. Lewis), pp. 55–66. Chichester: Wiley.
- Dawid, A. P. (2000) Causal inference without counterfactuals. *J. Am. Statist. Ass.*, **95**, 407–424.
- Denko, C. W. and Petricevic, M. (1978) Sympathetic or reflex footpad swelling due to crystal induced inflammation in the opposite foot. *Inflammation*, **3**, 81–86.
- Didelez, V. (2007) Graphical models for composable finite Markov processes. *Scand. J. Statist.*, **34**, 169–185.
- Didelez, V. (2008) Graphical models for marked point processes based on local independence. *J. R. Statist. Soc. B*, **70**, 245–264.
- Gladman, D. D. (2004) Psoriatic arthritis. In *Kelly's Textbook of Rheumatology* (eds E. D. Harris, R. C. Budd, G. S. Firestein, M. C. Genovese, J. S. Sargent, S. Ruddy and C. B. Sledge), 7th edn, pp. 1155–1164. Philadelphia: Saunders.
- Gladman, D. D. (2005) Discussion: Clinical features epidemiology, classification criteria and quality of life in psoriasis and psoriatic arthritis. *Ann. Rheum. Dis.*, **64**, suppl. II, ii24–ii25.

- Gladman, D. D. (2006) Clinical, radiological, and functional assessment in psoriatic arthritis: is it different from other inflammatory joint diseases? *Ann. Rheum. Dis.*, **65**, suppl. III, iii22–iii24.
- Gladman, D. D. and Farewell, V. T. (1999) Progression in psoriatic arthritis: role of time varying clinical indicators. *J. Rheum.*, **26**, 2409–2413.
- Gladman, D. D., Farewell, V. T. and Nadeau, C. (1995) Clinical indicators of progression in psoriatic arthritis: multivariate relative risk model. *J. Rheum.*, **22**, 675–679.
- Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438.
- Helliwell, P. S., Hetthen, J., Sokoll, K., Green, M., Marchesoni, A., Lubrano, E., Veale, D. and Emery, P. (2000) Joint symmetry in early and late rheumatoid and psoriatic arthritis. *Arth. Rheum.*, **43**, 865–871.
- Hill, A. B. (1965) The environment and disease: association or causation? *Proc. R. Soc. Med.*, **58**, 295–300.
- Hougaard, P. (1999) Multi-state models: a review. *Lifetim. Data Anal.*, **5**, 239–264.
- Huber, P. J. (1967) The behaviour of maximum likelihood estimates under nonstandard conditions. In *Proc. 5th Berkeley Symp. Mathematics Probability and Statistics*, vol. 1 (eds L. LeCam and J. Neyman), pp. 221–233. Berkeley: University of California Press.
- Husted, J. A., Gladman, D. D., Farewell, V. T. and Cook, R. J. (2001) Health-related quality of life of patients with psoriatic arthritis: a comparison with patients with rheumatoid arthritis. *Arth. Care Res.*, **45**, 151–158.
- Jackson, C. (2008) msm: multi-state Markov and hidden Markov models in continuous time. *R Package Version 0.8.1*.
- Kalbfleisch, J. D. and Lawless, J. F. (1985) The analysis of panel data under a Markov assumption. *J. Am. Statist. Ass.*, **80**, 863–871.
- Kane, D., Stafford, L., Bresnihan, B. and Fitzgerald, O. (2003) A prospective, clinical and radiological study of early psoriatic arthritis: an early synovitis clinic experience. *Rheumatology*, **42**, 1460–1468.
- Lee, E. W. and Kim, M. Y. (1998) The analysis of correlated-panel data using a continuous-time Markov model. *Biometrics*, **54**, 1638–1644.
- Levine, J. D., Goetzl, E. J. and Basbaum, A. I. (1987) Contribution of the nervous system to the pathophysiology of rheumatoid arthritis and other polyarthritides. *Rheum. Dis. Clin. Nth Am.*, **13**, 369–383.
- Lilienfeld, D. E. and Stolley, P. D. (1994) *Foundations of Epidemiology*, 3rd edn. New York: Oxford University Press.
- McHugh, N. J., Balachrishnan, C. and Jones, S. M. (2003) Progression of peripheral joint disease in psoriatic arthritis: a 5-yr prospective study. *Rheumatology*, **42**, 778–783.
- McKenna, S. P., Doward, L. C., Whalley, D., Tennant, A., Emery, P. and Veale, D. J. (2004) Development of the PsAQoL: a quality of life instrument specific to psoriatic arthritis. *Ann. Rheum. Dis.*, **63**, 162–169.
- Meria-Machado, L., de Uña-Álvarez, J., Cadarso-Suárez, C. and Andersen, P. K. (2009) Multi-state models for the analysis of time-to-event data. *Statist. Meth. Med. Res.*, **18**, 195–222.
- Molenaar, E. T. H., Voskuyl, A. E., Dinant, H. J., Bezemer, P. D., Boers, M. and Dijkmans, B. A. C. (2004) Progression of radiologic damage in patients with rheumatoid arthritis in clinical remission. *Arth. Rheum.*, **50**, 36–42.
- Mulherin, D., Fitzgerald, O. and Bresnihan, B. (1996) Clinical improvement and radiological deterioration in rheumatoid arthritis: evidence that the pathogenesis of synovial inflammation and articular erosion may differ. *Br. J. Rheum.*, **35**, 1263–1268.
- Pearl, J. (2000) *Causality: Models, Reasoning and Inference*. New York: Cambridge University Press.
- Queiro-Silva, R., Torre-Alonso, J. C., Tinture-Eguren, T. and Lopez-Lagunas, I. (2003) A polyarticular onset predicts erosive and deforming disease in psoriatic arthritis. *Ann. Rheum. Dis.*, **62**, 68–70.
- R Development Core Team (2009) *R: a Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Robins, J. M., Hernan, M. A. and Brumback, B. (2000) Marginal structural models and causal inference in epidemiology. *Epidemiology*, **11**, 550–560.
- Royall, R. M. (1986) Model robust confidence intervals using maximum likelihood estimators. *Int. Statist. Rev.*, **54**, 221–226.
- Rubin, D. B. (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educ. Psychol.*, **66**, 688–701.
- Rubin, D. B. (1978) Bayesian inference for causal effects: the role of randomization. *Ann. Statist.*, **6**, 34–58.
- Schweder, T. (1970) Composable Markov processes. *J. Appl. Probab.*, **7**, 400–410.
- Scott, D. L. (2004) Radiological progression in established rheumatoid arthritis. *J. Rheum.*, **31**, suppl. 69, 55–65.
- Siannis, F., Farewell, V. T., Cook, R. J., Schentag, C. T. and Gladman, D. D. (2006) Clinical and radiological damage in psoriatic arthritis. *Ann. Rheum. Dis.*, **65**, 478–481.
- Smolen, J. S., Han, C. and van der Heijde, D. M. F. M. (2009) Radiographic changes in rheumatoid arthritis patients attaining different disease activity states with methotrexate monotherapy and infliximab plus methotrexate: the impacts of remission and tumour necrosis factor blockade. *Ann. Rheum. Dis.*, **68**, 823–827.
- Sokoll, K. B. and Helliwell, P. S. (2001) Comparison of disability and quality of life in rheumatoid and psoriatic arthritis. *J. Rheum.*, **28**, 1842–1846.
- Weiss, N. S. (1986) *Clinical Epidemiology: the Study of the Outcome of Illness*. New York: Oxford University Press.

- Welsing, P. M. J., Landewé, R. B. M., van Riel, P. L. C. M., Boers, M., van Gestel, A. M., van der Linden Swinkels, S. and van der Heijde, D. M. F. M. (2004) The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis. *Arth. Rheum.*, **50**, 2082–2090.
- White, H. (1982) Maximum likelihood estimation under mis-specified models. *Econometrica*, **50**, 1–26.
- Wright, V. and Moll, J. M. H. (1976) Psoriatic arthritis. In *Seronegative Polyarthritits* (eds V. Wright and J. M. H. Moll), pp. 169–223. Amsterdam: North-Holland.