Satu Helske, Fiona Steele, Katja Kokko, Eija Räikkönen, Mervi Eerola

Partnership formation and dissolution over the life course: applying sequence analysis and event history analysis in the study of recurrent events

Article (Accepted version) (Refereed)

Original citation:
Helske, Satu, Steele, Fiona, Kokko, Katja, Räikkönen, Eija and Eerola, Mervi (2015) *Partnership formation and dissolution over the life course: applying sequence analysis and event history analysis in the study of recurrent events.* Longitudinal and Life Course Studies, 6 (1). pp. 1-25. ISSN 1757-9597

DOI: 10.14301/llcs.v6i1.290

Partnership formation and dissolution over the life course: applying sequence analysis and event history analysis in the study of recurrent events

Satu Helske
University of Jyvaskyla, Finland
satu.helske@jyu.fi

Fiona Steele
London School of Economics and Political Science, UK

Katja Kokko
University of Jyvaskyla, Finland

Eija Räikkönen
University of Jyvaskyla, Finland

Mervi Eerola
University of Turku, Finland

Published in 2015 in Longitudinal and Life Course Studies 6(1), 1–25. doi:
http://dx.doi.org/10.14301/llcs.v6i1.290

Abstract

We present two types of approach to the analysis of recurrent events for discretely measured data, and show how these methods can complement each other when analysing coresidential partnership histories. Sequence analysis is a descriptive tool that gives an overall picture of the data and helps to find typical and atypical patterns in histories. Event history analysis is used to make conclusions about the effects of covariates on the timing and duration of the partnerships. As a substantive question, we studied how family background and childhood socio-emotional characteristics were related to later partnership formation and stability in a Finnish cohort born in 1959. We found that high self-control of emotions at age 8 was related to a lower risk of partnership dissolution and for women a lower probability of repartnering. Child-centred parenting practices during childhood were related to a lower risk of dissolution for women. Socially active boys were faster at forming partnerships as men.

Keywords: partnership formation, partnership dissolution, sequence analysis, event history analysis, recurrent events

1 Introduction

During the life course many events (such as marriages, child births, unemployment etc.) can occur several times to an individual. In this paper we present two approaches to the analysis of recurrent events for discretely measured data and show how these methods can complement each other when analysing coresidential partnership histories of a representative sample of Finnish men and women now in their fifties. The first method, *sequence analysis*, is a descriptive technique which we used to summarize all partner

transitions made by individuals over the whole observation period. We grouped similar histories of forming and dissolving partnerships and searched for typical and atypical patterns. In contrast, *event history analysis* is a model-based method which we used to model the probability of making a transition to or from partnership in a given time interval as a function of possibly time-varying individual characteristics. Specifically, we examined how home background and socio-emotional characteristics in childhood were related to later partnership formation and stability, whether these effects differed between women and men, and if they played a part in a tendency to repartner.

1.1 Partnerships in a life course perspective

Establishment of an intimate relationship has been recognized as one of the milestones during the transition to adulthood (e.g. Shanahan, 2000). In the past, this typically meant the start of the first and only marriage. However, the choice of union type is now no longer confined to traditional life-long marriage as cohabitation has become an integral part of family life in Western countries (Kennedy & Bumpass, 2008; Kiernan, 2001). Furthermore, it is increasingly common for people to enter a union more than once during their lives. As a result, partnership trajectories have become diverse according to the type and number of unions formed during the life course. Regarding the first union, cohabiting unions have been consistently found to be less stable than marriages (Poortman & Lyngstad, 2007) . In the case of the second and higher-order unions, the picture is more complex. In general, second unions have been shown to be as stable as the first unions, when selection based on individual characteristics is controlled for (Aassve et al., 2006; Lillard, Brien, & Waite, 1995; Poortman & Lyngstad, 2007; Steele, Kallis, Goldstein, & Joshi, 2005; Steele et al., 2006).

It is likely that second and higher-order unions differ from the first union in that they often involve individuals with more complex life histories, including multiple spells of partnerships, children from previous relationships, and the continuing influence of previous partners and their family members (Poortman & Lyngstad, 2007; Teachman, 2008). Higher-order unions also involve individuals who have learned about the process of break up. Going through this often painful process may have caused people to be more cautious the next time (Furstenberg & Spanier, 1984), which may lead to less commitment to and fewer investments in the second union compared to the first. Furthermore, marriage market conditions have also changed because people are older when they search for a partner for the second time, and therefore the pool of potential partners is more restricted (Teachman, 2008). Thus, it is likely that the factors linked to the dissolution of second and higher-order unions are not the same as those linked to the disruption of the first union.

The life course perspective (Elder, 1998) suggests that partnership transitions are interrelated with other areas of life, such as parenthood. However, empirical evidence regarding the association between partnership dissolution and having children is somewhat mixed. Earlier research has found different, even opposite, effects of having children on partnership dissolution across countries and in different family situations with regard to, for example, the number, age, and residence of children (Coppola & Di Cesare, 2008; Lillard & Waite, 1993; Lyngstad & Jalovaara, 2010; Steele, Kallis, Goldstein, & Joshi, 2005; Svarer & Verner, 2008).

1.2 Partnership transitions in context

A life course perspective suggests that decisions regarding life transitions are constrained by various contextual factors (e.g. Elder, 1998; Shanahan, 2000), as

well as by the individual's development prior to the transitions (Räikkönen, Kokko, Chen, & Pulkkinen, 2012). Our study focused on the associations between partnership transitions and individual (i.e. gender and socio-emotional behaviour) and family characteristics.

Empirical studies have demonstrated that, in general, women undergo family-related transitions for the first time at a younger age than men (e.g. Elder, 1998; Kokko, Pulkkinen, & Mesiäinen, 2009; Räikkönen et al., 2012; Ross, Schoon, Martin, & Sacker, 2009). Furthermore, the timing of family transitions may also be more closely interlinked among women than among men (Kokko et al., 2009). It has been shown that early motherhood may weaken women's subsequent attachment to the labour market (e.g. Rönkä & Pulkkinen, 1998). No such association has been found among men (Rönkä, Kinnunen, & Pulkkinen, 2000).

To the best of our knowledge, the effects of childhood socio-emotional behaviour have not been studied in previous analyses of partnership formation and dissolution. However, indirect support for the links between childhood socio-emotional behaviour and adult partnership transitions can be found in previous research. First, there is evidence that child behavioural problems predisposes individuals to earlier parenthood (e.g. Kokko, Pulkkinen, & Mesiäinen, 2009; Rönkä et al., 2000), especially among women (Kokko et al., 2009). In contrast, adaptive behaviour in childhood, such as shyness, has been shown to be related to later parenthood in men (Caspi, Elder, & Bem, 1988). Second, low self-control of emotions in childhood has been found to be a risk factor for later marital problems (Kinnunen & Pulkkinen, 2003). Third, there is evidence that high self-control of emotions in both genders, and social activity in women, contribute to favourable adult development (Pulkkinen, 2009). On

the basis of these earlier studies, we anticipated that high self-control of emotions would be connected to fewer and longer-lasting partnerships. Also, we expected that women with lower self-control of emotions and socially active men would form their first partnerships sooner.

An individual's family of origin may also influence union formation behaviours throughout adulthood. Accordingly, it has been shown that individuals who come from a less-advantaged family in terms of low socioeconomic status (SES) tend to undergo their first partnership transition at an earlier age than individuals from a high SES background, for whom the later timing of transitions is more typical (e.g. Berrington & Diamond, 2000; Rönkä et al., 2000; Ross et al., 2009; Steele, Kallis, & Joshi, 2006). Higher SES of the family of origin has also been linked to an increased risk of partnership dissolution (Bumpass, Martin, & Sweet, 1991; Lyngstad, 2006). In British cohorts, Steele et al. (2006) found that after a break-up, women from a higher SES background took longer to repartner, whereas Goldstein, Pan, and Bynner (2004) found no such effect among men. Family breakdown in childhood has been linked to earlier establishment of one's own partnership (Aassve, Burgess, Propper, & Dickson, 2006; Berrington & Diamond, 2000; Steele et al., 2006), as well as to a higher risk of partnership dissolution (Amato, 1996; Gähler, Hong, & Bernhardt, 2009; Steele et al., 2006), suggesting that union behaviours transfer at least to some extent from parents to their children.

Besides individual and family factors, the socio-historical context promotes variability in transition behaviours (e.g. Elder, 1998; Shanahan, 2000). The present study was based on longitudinal data collected for a representative sample of individuals born in Finland in 1959 (Pulkkinen, Lyyra, & Kokko, 2009; Pulkkinen & Kokko, 2010; Pulkkinen, 2009). Regarding partnership transitions

in Finland, the mean age at first marriage was 25.9 years for women and 28.1 years for men in 1986–1990 (Statistics Finland, 2010). Cohabitation before marriage or as an alternative to marriage was very popular then, just as it is now (Statistics Finland, 1994). Among women born in 1938–42, 13% had cohabited, but among women born in 1958–62, 51% had cohabited before marriage and 33% as an alternative to marriage. Since the mid-1980s, the mean age at first marriage has risen: in 2009, the mean age was 30.2 years for women and 32.5 years for men (Statistics Finland, 2010). Most men and women marry only once; in 2009 11% of married women and 12% of married men had remarried. In 2009, the total divorce rate in Finland was 50% and the mean age at the time of divorce was 41.3 years for women and 43.8 for men. Of marriages entered in 1985 39% had ended in divorce by 2009. Due to the popularity of cohabitation in Finland, in this article our definition of a partnership includes both marital and nonmarital cohabitational unions, which are treated as substitutes for each other.

2 Methods

2.1 Sample

We analysed data from the Finnish Jyväskylä Longitudinal Study of Personality and Social Development (JYLS). The study, established in 1968 by Lea Pulkkinen, includes all students from 12 randomly sampled second-grade school classes in Jyväskylä, Central Finland (Pulkkinen, 2009). All the pupils participated. The original sample consisted of 173 girls and 196 boys, of whom the majority (94%) were born in 1959. All participants were native Finns and they have been followed from age 8 to 50. During the follow-up, no systematic attrition has been found in the JYLS sample and the participants have continued to be

representative of their Finnish birth cohort (Pulkkinen, 2009; Pulkkinen & Kokko, 2010).

During two data collection phases in 2001 at age 42 and in 2009 at age 50, life history calendars (LHC; adapted from Caspi, Moffitt, Thornton, Freedman, & others, 1996; Kokko, Pulkkinen, & Mesiäinen, 2009) were used to retrospectively collect information about partnership status, children, education and work, as well as other important life events. The occurrence, timing and duration of the transitions were recorded annually first from age 15 to age 42 and later from age 42 to age 50 during interviews in which altogether 275 participants (77% of the original sample still alive at age 50) gave reports based on their memory and visual aids provided by the LHC-sheet.

The information collected with the LHCs was confirmed and complemented using other sources, such as life situation questionnaires and interviews at ages 27, 36, 42, and 50. We were able to derive almost complete partnership data between ages 15–42, but missing information due to non-response during the last phase of data collection at age 50 led to incomplete histories for 22% of the participants. The length of the follow-up varies between individuals because of the two data collection phases and small differences in their ages. Altogether 215 participants were followed for 36 years, 14 participants for 35 years, and 46 participants for only 28 years.

2.2 Variables

In addition to subjects' annual partnership histories we used information from their parenthood histories to derive a time-varying binary indicator of whether

or not the individual was a *parent* to biological or adopted children in a given year.

Socioeconomic status (SES) based on father's occupation (or mother's if she was the sole provider or had a higher status), was coded 0 if blue-collar and 1 if a white-collar worker (Pitkänen, Lyyra, & Pulkkinen, 2005).

Family structure at age 14 was coded 0 if the participant lived with both parents and 1 if the parents had divorced or a parent had died (Kokko & Pulkkinen, 2000).

Child-centred parenting was an average score of five dichotomous variables based on age 27 recollections of parenting practices and home environment (parental relationship, physical punishment, maternal supervision, relationship with the father, and *family structure*; Kokko & Pulkkinen, 2000). Missing data were imputed (Pitkänen, Kokko, Lyyra, & Pulkkinen, 2008).

Child socio-emotional behaviour at age 8 was assessed using two subscales: *social activity* and *high self-control of emotions* (including emotional stability, constructiveness, and compliance; see Kokko, Pulkkinen, Mesiäinen, & Lyyra, 2008; Pulkkinen, Kokko, & Rantanen, 2012). Each item was rated by teachers on a scale from 0 (never) to 3 (often).

2.3 Statistical methods

Sequence analysis (SA) is a model-free data-mining type of approach that provides an overview of individual sequences over the whole observation period, including the most common transitions and time spent in each

partnership state. The aim of SA is to measure pairwise (dis)similarity of the sequences, which is often followed by some kind of clustering method to find typologies of whole trajectories. *Event history analysis* (EHA; also known as survival, duration, or failure-time analysis) is used for the study of factors that influence the timing of transitions. The response variable in EHA is the duration between becoming at risk of experiencing the event of interest and the time that the event occurs.

2.3.1 Sequence analysis

SA was originally developed in bioinformatics to organize, classify, and parse protein and DNA sequence data (Durbin, Eddy, Krogh, & Mitchison, 1998). In the social sciences, Abbott introduced the use of SA in life course analysis in the mid-1980s (Abbott, 1983; Abbott, 1995; Abbott & Tsay, 2000). The basic idea in SA is to measure the distance or dissimilarity of two sequences consisting of the succession of categorical states describing the trajectories. Two major issues are essential for SA. The first concerns the composition of sequences: how many and what type of states? The second issue is related to determining the dissimilarities between the sequences: which dissimilarity measure to use and, for some measures, how to assign the 'cost' of converting one state to another? Typical steps in SA include the following: 1) creating sequences using a finite set of states; 2) choosing and implementing a method for computing pairwise dissimilarities between sequences; 3) analysing the dissimilarities (e.g. cluster analysis and/or multidimensional scaling); 4) graphical illustration and examination of sequence data.

Definition of states

Technically, the number of states does not have to be restricted (though finite), but for practical and interpretational reasons the state space is often relatively limited. Definition of the states requires careful consideration. In the present application, for example, defining divorced as single or distinguishing partnership states by the type of union instead of order would give a different viewpoint. In previous research it has been common to group all coresidential partnerships together as one state (e.g. Aassve, Billari, & Piccarreta, 2007; Gauthier, Widmer, Bucher, & Notredame, 2010; Salmela-Aro, Kiuru, Nurmi, & Eerola, 2011) or to separate marriages from cohabitations (e.g. Barban & Billari, 2012; Elzinga & Liefbroer, 2007; Piccarreta & Lior, 2010). Usually these have been combined with information on children.

We coded annual partnership states for each individual based on the *order* of the partner: 1) living single (never had a coresidential partner), 2) living with the first partner, 3) with the second partner, 4) with at least the third partner, or 5) living divorced/separated/widowed. Widowhood was very rare and thus it was merged with the other states of living without a previous partner. Transitions between the states were more restricted than in most studies of partnership sequences: only the last two could be revisited, except for the rare event of going back to a previous partner. Without separating partnerships by order it would have been difficult or even impossible to distinguish sequential partnerships.

Dissimilarities of sequences

There are several methods for measuring sequence dissimilarity, optimal matching (OM) being the most well-known (e.g. McVicar and Anyadike-Danes, 2002). In OM the goal is to find the best alignment of two sequences. Their dissimilarity is computed from the operations needed to transform one sequence into the other using insertions, deletions, and substitutions of states. Roughly, the more operations needed, the more distant the sequences are. The operations can be given different costs to reflect the amount of dissimilarity between the states. Another completely different type of approach by Elzinga is based on counting or measuring common sequence attributes such as subsequences (Elzinga, 2006; Elzinga & Liefbroer, 2007). These methods do not require defining any costs.

In the present study, we use generalized Hamming distance (Hamming, 1950; Lesnard, 2010) which compares states at the same time positions in each sequence. This performs well in our data where the observed sequence lengths vary across individuals, and where the timing of the partnership transitions is regarded as very important. To assess the closeness of two partnership histories, sequences are aligned year by year (see Example 1). Shorter sequences are complemented with missing states to achieve equal sequence lengths required to compute Hamming distances. Partnership states at each age are compared and each comparison is given a cost (see Table 1). Only the ratio of the costs is important and usually the absolute numbers have no substantive meaning; multiplying the costs by a constant does not change the results. The dissimilarity of the histories is simply the sum of the costs.

Example 1

Computing generalized Hamming distances between artificial partnership histories. The costs are given for a comparison of partnership states at each age. See Table 1 for definition of states and costs.

Age	20	21	22	23	24	25	26	27	28
Sequence 1	S	S	S	P1	P1	P1	P1	P1	P1
Sequence 2	S	S	S	S	S	S	P1	P1	*
Cost	0	0	0	2	2	2	0	0	0

Dissimilarity = 6

Age	20	21	22	23	24	25	26	27	28
Sequence 1	S	S	S	P1	P1	P1	P1	P1	P1
Sequence 3	P1	P1	P1	P1	P1	D	P2	P3	P3
Cost	2	2	2	0	0	2	2	3	3

Dissimilarity = 16

Definition of the costs depends not only on the states themselves but also on the research question of interest: which states are regarded as close and which as distant? The most common strategies have been to assign the costs based on theory or transition probabilities between the states. The latter way is automatic and has been said to reduce subjectivity (Aisenbrey & Fasang, 2010; Gauthier, Widmer, Bucher, & Notredame, 2009). However, it is not suitable for many cases such as the present study, where most of the partnership transitions are impossible and the probabilities of the transitions provide little information on the dissimilarities between the states. Setting the costs is an

ongoing debate and many modifications to the basic options have been suggested (e.g. Aisenbrey & Fasang, 2010; Gauthier et al., 2009; Halpin, 2010; Hollister, 2009; Lesnard, 2010).

Table 1: Costs for Hamming distance computations. Costs were defined to measure how distant different partnership states are regarded.

		Sequence 2					
		S	P1	P2	P3	D	*
Sequence 1	Single (S)	0	2	3	5	5	0
	1st partnership (P1)	2	0	1	3	2	0
	2nd partnership (P2)	3	1	0	2	2	0
	3rd+ partnership (P3)	5	3	2	0	2	0
	Divorced/separated (D)	5	2	2	2	0	0
	Missing (*)	0	0	0	0	0	0

We set costs that would lead to clusters that separate histories of stable and unstable partnerships from those with long periods of living single or divorced/separated. The last two were seen as distant states (cost = 5) because forming a partnership was regarded as one step in the developmental process to adulthood. Second partnerships were very common, so the cost of alignment with the first partnership state was set low (cost = 1). Aligning any state to a missing state was defined to have zero cost to ensure that sequences were grouped together according to the known parts of the histories, not with other sequences with missing information.

For the JYLS data, other dissimilarity measures including optimal matching, dynamic Hamming distance (Lesnard, 2010), the length of the longest common

subsequence, and the number of common subsequences were considered together with different cost definitions. Generalized Hamming with the costs presented in Table Error! Reference source not found. gave the most meaningful clusters and the best goodness-of-fit, as measured by the proportion of the variation explained by the clusters (pseudo coefficient of determination).

Clustering sequences

The dissimilarities between all partnership sequences are collected in a matrix that can be used to cluster similar histories together. We used Ward's agglomerative algorithm (Ward Jr., 1963). At each step, the algorithm combines the two clusters (at the first step, sequences) that minimize within-cluster variability and maximize inter-cluster variability. It is commonly used to cluster sequences since it usually produces more equal-sized clusters than other algorithms (Aisenbrey & Fasang, 2010). We also tested other clustering options but, as also found by Aassve et al. (2007), most of them (single, average, and complete linkage) resulted in one large cluster and many residual clusters with only a handful of sequences, even several clusters with only one sequence. This is not desirable for the purpose of interpretation and possible further analyses. With our dissimilarities, the "partition around medoids" method (PAM; Kaufman & Rousseeuw, 2009) was the best competitor, but not as good as Ward in terms of pseudo-R^2 (for pseudo-R^2 see Studer, Ritschard, Gabadinho, & Müller, 2011). Choosing the best number of clusters is not straightforward. Our decision was based on the dendrogram, interpretability of the clusters, and change in measures including pseudo-R^2, pseudo F (Studer et al., 2011),

Hubert's C, and Hubert's Gamma (Hubert & Arabie, 1985). See Studer (2013) for a review of measuring the quality of clustering of sequence data.

External information can be taken into account after clustering or at the clustering phase. We used regression trees (Breiman, Friedman, Olshen, & Stone, 1984) to group similar partnership histories using information on subjects' home background and socio-emotional behaviour in childhood as predictors. The idea of regression trees is to recursively partition data into clusters using values of a predictor, creating binary splits for the values of a variable for which the highest pseudo-R^2 is achieved. The tree is grown until no further significant splits (assessed through a permutation F-test) are found (Studer et al., 2011).

We studied whether sex and socio-emotional characteristics and home background during childhood predicted future partnership histories using regression tree methods with the same Hamming distances as previously.

Graphical illustrations

There are many options for graphical description of sequence data. The most common choices include cross-sectional state distribution plots and sequence index plots. State distributions plotted for each time point show the change in the prevalence of states in the course of time. Sequence index plots show the whole partnership histories for the individuals. Plotting all sequences at once in a random order is usually not very informative. Clustering eases interpretation by grouping similar histories together and multidimensional scaling or some other criterion is often used to order sequences more meaningfully.

Software

The TraMineR package in R (Gabadinho, Ritschard, Müller, & Studer, 2011) was used for the SA presented in this paper. Alternatives include TDA (Rohwer & Pötter, 2004) and the Stata packages SQ (Brzinsky-Fay, Kohler, & Luniak, 2006) and SADI (Halpin, 2014). To our knowledge, TraMineR has been the most versatile and widely used software for SA in recent years. However, the new SADI package in Stata appears to have the potential to become a strong competitor.

2.3.2 Discrete-time event history model

SA is a useful tool for obtaining an overview of histories. However, as the focus is the whole trajectory, SA cannot be used to study how the factors of interest – especially those which vary over time – are related to the timing and duration of each coresidential partnership. EHA is a highly flexible approach for the study how individual time-invariant and time-varying characteristics influence the timing of partnership transitions.

Moving in with the first partner is a milestone for an individual, but it may not be the only partnership (marriage or cohabitation) that is established during their life time. Instead of focusing only on the timing of the first partnership we can analyse the duration of all episodes of living without a partner. These are periods during which an individual is continuously "at risk" of establishing a new partnership. Individuals not living with a partner in a given time interval constitute what is referred to as the "risk set" for partnership formation. An individual's first episode starts at the beginning of the follow-up and it ends when the individual moves in with a partner for the first time or is censored because of loss to follow-up. Individuals stay out of the risk set as long as they

are living with the same partner. A new episode begins at dissolution when the individual is again "at risk" of forming a new partnership.

The durations of episodes from the same individual are likely to be correlated, which invalidates the independence assumption of standard statistical methods. This correlation is due to unmeasured time-invariant individual characteristics that affect the risk of forming any (new) partnership. The variation in the risks between individuals is generally called unobserved heterogeneity or individual frailty (e.g. Vaupel, Manton, & Stallard, 1979). Recurrent events data can be viewed as having a two-level hierarchical structure where the events are nested within individuals. These types of hierarchical data can be analysed with multilevel or random effects models (e.g. Goldstein, 2011; Raudenbush & Bryk, 2002).

Many life transitions, such as partnerships, are formed in continuous time, but it is not always possible or practical to collect data as such. Often event times are recorded in time intervals such as months or years because finer measurement (e.g. daily accuracy in a study spanning several years) would not be informative. At other times it is not possible to observe the occurrence times as frequently as would be preferred. In both cases the discrete-time model can be used as an approximation to a continuous-time model (e.g. Allison, 1982).

The two LHCs from the JYLS study contain yearly information on individuals' partnership statuses. We were interested in both the formation and dissolution of partnerships. However, annual accuracy was not always frequent enough to distinguish between consecutive partnerships. To properly define who was in the risk set of moving in with a new partner (i.e. living without a partner) at the start of a given time interval, artificial six-month intervals were created and the

partnership status of the latter part of the year changed to “single” for those who had dissolved and formed a partnership during the same year (29 cases from 24 individuals).

Random effects model for repeated partnership formation

In our annual data, a partnership beginning “at age t” occurs during the one-year interval $[t, t+1)$. Suppose that t_{ij} is the number of years for which individual j is observed in episode i, where an episode is a continuous period of time unpartnered. We form a data set with one record per year for each individual (a person-episode-period file) and define a binary indicator y_{tij} for each year $t = 1, \ldots, t_{ij}$ such that

$$y_{tij} = \begin{cases} 1 & \text{if episode } i \text{ of an individual } j \text{ ends in partnership formation at } t \\ 0 & \text{otherwise} \end{cases}$$

The discrete-time hazard function is defined as

$$p_{tij} = P\left(y_{tij} = 1 \middle| y_{t'ij} = 0 \text{ for } t' < t\right),$$

which is the conditional probability that a partnership is formed during interval t of episode i of individual j given that they have not moved in with a partner before interval t.

A logistic regression model is commonly used to model the dependence of p_{tij} on the duration unpartnered by interval t and a vector of (possibly time-varying) explanatory variables $\boldsymbol{x}_{tij}$:

$$\log\left(\frac{p_{tij}}{1-p_{tij}}\right) = \alpha'\boldsymbol{z}_{tij} + \beta'\boldsymbol{x}_{tij} + u_j,$$

where $\boldsymbol{z}_{tij}$ is a vector of functions of t and $\alpha'\boldsymbol{z}_{tij}$ defines the baseline hazard function. Polynomials and step functions are common choices for modelling the time-dependency. Unobserved variation between individuals (frailty) is

represented by u_j, which is usually assumed to follow a normal distribution $N(0, \sigma_u^2)$. The random effect shifts the log-odds of partnering up or down for the individual j while the effects of duration and covariates are assumed to be constant across individuals. Conditional on u_j, the durations of episodes for the same individual are assumed to be independent.

A similar model is specified for the risk of partnership dissolution.

A two-state model

We can extend the above model to study transitions between two (or more) states. That model considers transitions from a single state to living with a partner and the individual is dropped from observation after forming a partnership (unless they separate and re-enter the risk set). In a two-state model the durations of all episodes living with and without a partner are examined. Exit from one state implies entry to the other. Examples of the use of multistate models to study partnership transitions include Aassve et al. (2006), Goldstein et al. (2004), and Steele et al. (2006).

We denote by S_{tij} the state of individual j's ith episode at the start of interval t. Now y_{tij} is the binary indicator of a transition of either type, forming (F) or dissolving (D) a partnership. The conditional probability of a transition from state s ($s = F, D$), during interval t, given that a transition has not yet occurred in that episode, is now

$$p_{stij} = P\left(y_{tij} = 1 \middle| y_{t'ij} = 0 \text{ for } t' < t, S_{tij} = s\right),$$

and the multilevel event history model for transitions between the two states can be written as

$$\text{logit}\left(p_{stij}\right) = \alpha_s' \boldsymbol{z}_{stij} + \beta_s' \boldsymbol{x}_{stij} + u_{sj}, \quad s = F, D$$

Note that the baseline logit-hazard, covariates, coefficients, and random effects can all vary across states, as indicated by the s subscripts.

Software

Random effects models for recurrent events and multiple states can be fitted in most mainstream statistical software packages such as R, SAS and Stata, and also with more specialist software including MLwiN and Sabre. The packages may vary in the estimation procedures used, leading to differences in parameter estimates and computational times (see Steele (2011) for a detailed summary). In our study, event history models were fitted using the xtlogit procedure in Stata which implements maximum likelihood via Gauss–Hermite quadrature.

3 Results

3.1 Sequence analysis: trajectories of partnerships

Sequence analysis was used to provide an overall view of partnership histories, to obtain descriptive information on typical and atypical trajectories, and to explore how much childhood socio-emotional characteristics and family background predict future histories.

Figure 1 presents the prevalence of partnership states at each age for women and men. On average, men formed their first partnership later than women. Women spent more time living as divorced or separated than men, but from this figure we cannot see the duration of these periods.

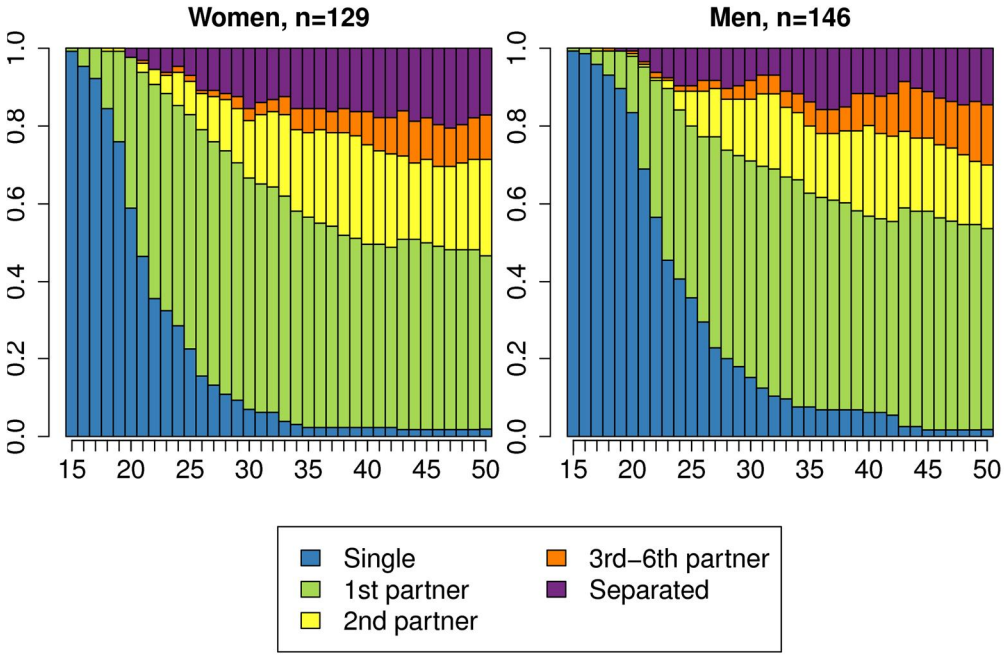

Figure 1: State distribution plots of partnership histories for women and men between ages 15–50 in JYLS data. Missing states are not included in the yearly proportions. The change in proportions at 43 is due to individuals that were lost to follow-up.

Table 2 shows the average number of years that women and men spent in each partnership state. Women had longer first and second partnerships than men, but there was a lot of variation.

Table 2: Mean and standard deviation of years spent in each partnership state since age 15 for women and men in the JYLS data.

	Women		Men	
State	Mean	S.D.	Mean	S.D.
Single	7.8	5.7	10.1	6.7
1st partnership	16.3	11.1	14.9	10.7

2nd partnership	5.2	1.6	4.3	7.4
3rd–6th partnership	1.6	5.0	1.9	5.1
Divorced/separated	4.0	5.8	3.0	5.2
Missing	1.1	2.7	1.8	3.6

Table Error! Reference source not found. shows the most frequent types of history ignoring the time spent in each state. Two out of three individuals had settled in to their first or at most second partnership. Since the transitions between states are rather limited due to several being absorbing, there are few possible histories. Except for the differences in the number of partners and dissolutions, the histories only differ by whether or not the individuals had lived alone between their partnerships. Taking account of the durations of episodes adds little additional information: the number of the JYLS participants is limited compared to the length of the follow-up so most of the sequences are unique.

Table 3: The most common partnership histories in JYLS data, when durations are omitted. S=single, P1=1st partnership, P2=2nd partnership, P3=3rd–6th partnership, D=Divorced/separated/widowed.

State	Freq.	%
S-P1	122	44.4
S-P1-D-P2	59	21.5
S-P1-D	25	9.1
S-P1-D-P2-D-P3	14	5.1
S-P1-D-P2-D	10	3.6
S	9	3.3
Total	239	86.9

3.1.1 Clustering sequences

Solutions with between 2 and 15 clusters from Ward's algorithm were studied, and the eight-cluster solution was chosen based on the criteria described in Section 2.3.1. These clusters explained 61% of the variation between the histories. Sequence index plots of the clusters are shown in Figure 2.

There were four larger clusters of relatively stable partnership histories with one or two partners that only differ in timing. Men were in the majority among those who have established a (typically long-lasting) late initial partnership, but in the "later second partnership" group the majority were women (Table 4). There emerged also two male-dominated clusters which included individuals with multiple partnerships, either earlier or later in life. Some of these individuals had experienced multiple partnerships but settled down after early adulthood, and others had not formed long-lasting partnerships at all. The last two clusters showed histories of living without a partner; some (typically women) had a partnership that ended in separation or divorce, while others (typically men) had never lived with a partner or had entered their first partnership very late.

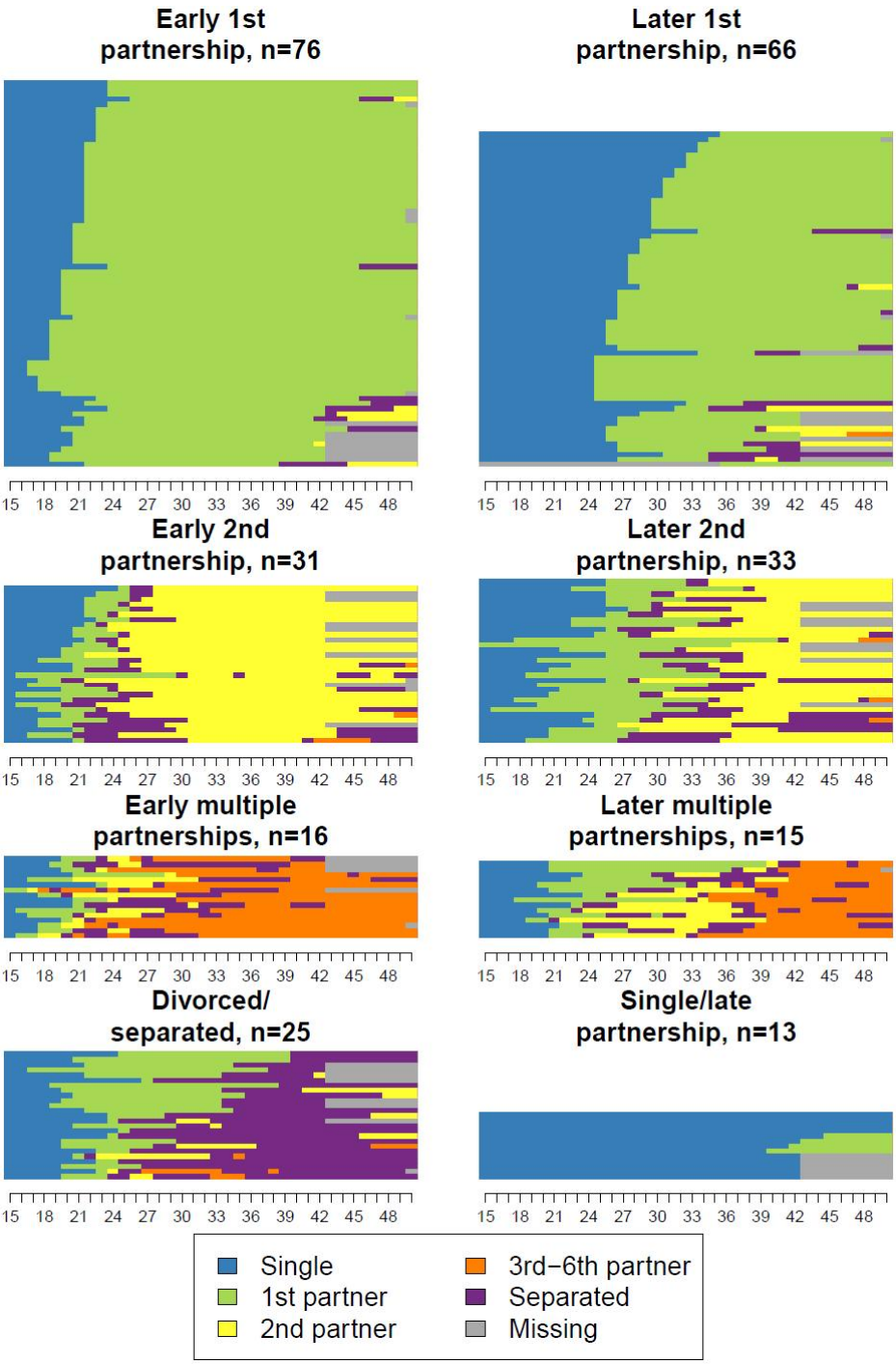

Figure 2: Eight clusters of partnership histories using generalized Hamming distances as a measure of dissimilarity and Ward's method for clustering. Multidimensional scaling was used to order sequences.

Table 4: Proportion of partnership clusters and the percentage of women.

Cluster	Size (*n*)	Size (%)	Women (%)
Earlier 1st partnership	76	27.6	55.3
Later 1st partnership	66	24.0	34.8
Early 2nd partnership	31	11.3	45.2
Later 2nd partnership	33	12.0	60.6
Earlier multiple partnerships	16	5.8	43.8
Later multiple partnerships	15	5.5	33.3
Divorced/separated	25	9.1	60.0
Single/late partnership	13	4.7	23.1
Total	275	100	46.5

3.1.2 Clustering with external information

Using the regression tree method described in Section 2.3.1, only two of the covariates were statistically significant predictors of cluster membership; these formed altogether three clusters of the data (Figure 3).

The first and the most effective split of the data was achieved with child-centred parenting (CCP). More child-centred parenting practices in the family of origin (CCP > 0.4) was related to more stable partnership histories with usually one or two partners. The second split was for the lower values of CCP and self-control of emotions (SCE). On average, individuals with lower values of CCP and SCE had more partners compared to those who also had lower values of CCP but higher SCE. Altogether grouping on CCP and SCE explained only 3.5% of the variability between the partnership histories, so most important sources of sequence variation was the timing and the number of partnerships.

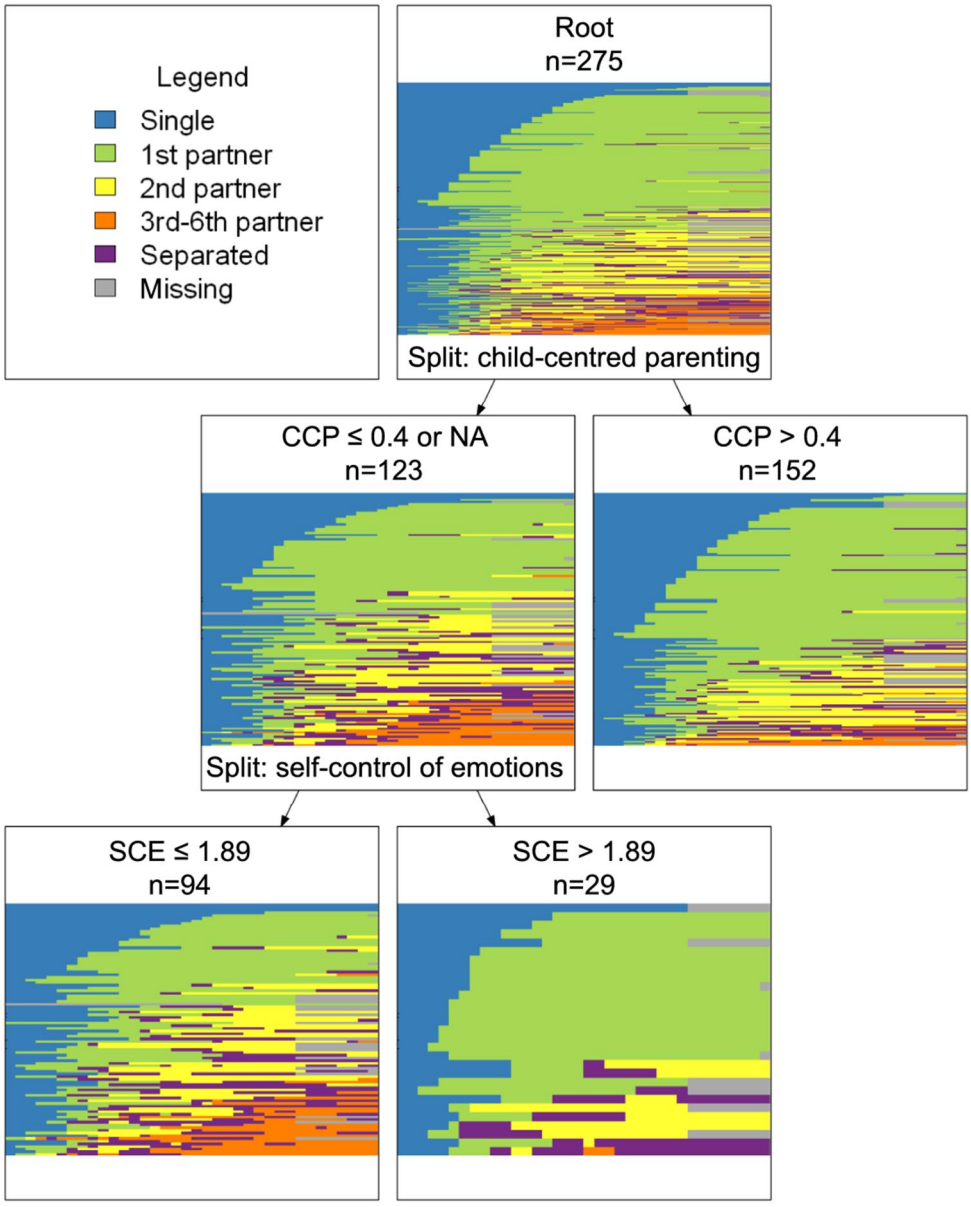

Figure 3: Regression tree of partnership histories with two significant splitting variables: child-centred parenting (CCP, scores 0–1) and high self-control of emotions (SCE, scores 0–3).

3.2 Event history analysis: transitions to and from partnerships

Event history analysis was used to examine the timing of partnership formation and dissolution and how the rate of partnership transitions depends on individual history and characteristics.

As can be seen from the partnership clusters in the previous section and again in Table 5, recurrent partnerships were common: almost a half of both women and men had established at least two partnerships (marriages or cohabitations) during the follow-up period. Third and subsequent partnerships were less common, especially among women.

Table 5: Participants in the JYLS study by sex and the number of cohabitating partnerships. Higher-order partnerships (3th–6th) are combined into one category due to their small number.

				3+ partners	
	No partners	1 partner	2 partners	Individuals	Partnerships
Women	3	66	43	17	25
Men	6	79	33	27	42

Table 6 shows the means of the age at forming partnerships, duration of partnerships and time before forming new partnerships (not accounting for right-censoring). On average, first partnerships were formed around age 22 among women and age 24 among men. The youngest formed their first partnership (cohabitation) at 15 and the oldest at 35 (women) and 45 (men). On average, a new partnership was formed 2–3 years after dissolution of the previous partnership but there was considerable variation, with a maximum duration of over 20 years.

The average duration of first partnerships that ended in dissolution during the follow-up was about 8 years. Second partnerships were of a similar length to first partnerships for women and two years shorter among men. Higher-order partnerships lasted 4–5 years on average.

Table 6: Timing of partnership events: mean ages at forming partnerships, years since dissolution before forming a new partnership, and duration of partnerships that had ended in separation in the JYLS data. Right censoring was not accounted for.

		Formation					Dissolution		
		Age		Time since diss.			Duration		
Sex	Partner	Mean	S.D	Mean	S.D.	N	Mean	S.D.	N
Female	1st	22.17	4.16			126	8.54	6.51	68
	2nd	32.07	7.91	3.45	3.36	60	8.20	6.13	25
	3rd+	36.17	7.95	2.72	2.42	25	4.38	3.13	13
Male	1st	24.30	5.20			139	8.14	7.41	74
	2nd	31.22	7.53	2.68	3.19	59	5.97	6.30	31
	3rd+	36.56	9.04	2.39	2.74	42	4.92	4.30	18

Hazards of forming first and recurrent partnerships were computed from the data. The hazard at a given age is the proportion who were newly partnered from all individuals in the risk set (those who were not living with a partner yet/anymore). The hazard function is plotted in Figure 4 using locally weighted scatterplot smoothing (lowess) to show the change in the rate of partnership formation by age. We also see that on average women formed their first partnerships earlier than men. On the other hand, those men who *had* established and dissolved their first partnerships young (before age 25) seemed

to form subsequent partnerships quicker than young women in the same situation. There was an especially high peak for teenagers, but the risk set at that age was very small. In this study, the oldest age at first partnership was 35 for women, but is some suggestion that for men the hazard of first partnership increased in their early 40s (although, again, the risk set is small).

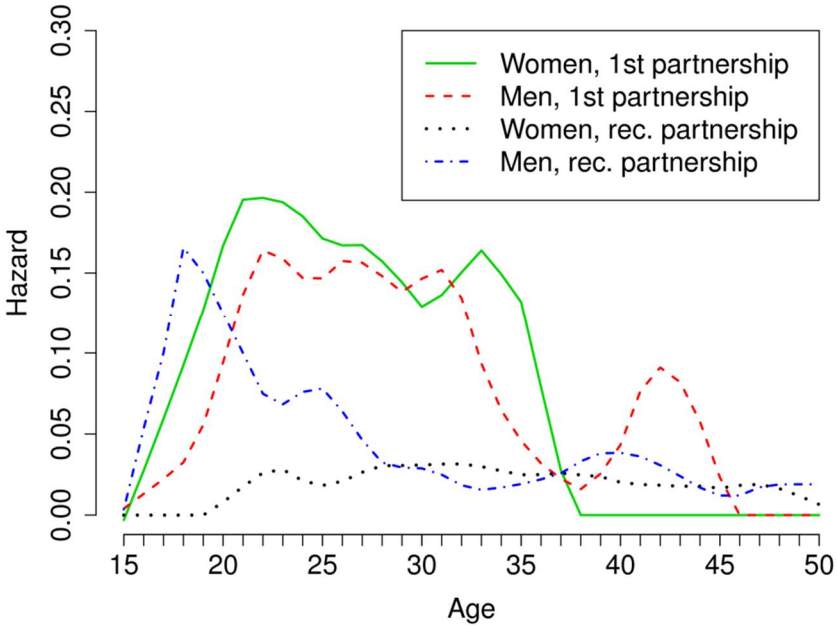

Figure 4: Hazard functions of the formation of first and recurrent partnerships for women and men. Hazards were smoothed with lowess (locally weighted scatterplot smoothing) using 20–25% of the closest points.

3.2.1 Partnership formation

Since preliminary analyses (not all shown here) revealed large differences between women and men in the timing of partnership formation and dissolution and in the factors related to these transitions, separate event history models were fitted for women and men. Based on the hazard functions shown in Figure 4, a piecewise constant function was chosen as the best representation of the baseline hazard for partnership formation. The timing of

first partnership was categorized into three periods: early (15–22 years), on-time (23–32), and late (33–50). The last category is wider than would be preferred, since it is unlikely that, for example, a 33-year-old and a 50-year-old have a same risk for establishing especially the first partnerships. However, as no women in our sample established their first partnership after age 35 it was not possible to use narrower age categories. Time since and the duration of the last partnership were also considered (using linear, quadratic, logarithmic, categorical functions of time) as well as the type of the previous partnership (marriage/cohabitation), but these variables did not show significant effects for either sex and were excluded from the models. Covariates measured in childhood were treated as time-invariant, while parenthood status and existence of previous partners were time-dependent.

We first studied the main effects of the covariates and their interactions with age and a previous partnership indicator. Interactions with age were considered to test the proportional hazards assumption, while interactions with previous partnership were tested to determine whether covariate effects differ for first and recurrent partnerships. Variables with effects that were significant at the 5% level were then tested together in one model with non-significant effects dropped one by one. None of the interactions between age and any covariate were significant.

Tables 7 and 8 show the final random effects models for partnership formation for women and men respectively. There was little evidence of unobserved heterogeneity among women (σ_u was estimated close to 0), but among men the additional of random effects led to a significant improvement in fit ($\hat{\sigma}_u$ = 0.607, significance assessed through likelihood ratio test). The "risk" of forming an initial partnership was estimated to be the highest among 23–32

year-olds for both sexes, but the differences between the age categories were small and not statistically significant at the 5% level. Among men and women who had already dissolved at least one partnership, the risk of repartnering was significantly higher among 15–22 year-olds than for the other age groups.

Table 7: Logistic model of partnership formation for women. Estimated coefficients and odds ratios (OR) are shown together with standard errors, p-values and 95% confidence intervals (CI) for the odds ratios. The last age category (33–50) was chosen as the reference category. SCE = self-control of emotions (scores 0–3), SES = socioeconomic status based on the parents' (mainly fathers') occupational status during the subject's childhood (higher/lower).

	Est.	s.e.	p	OR	OR 95% CI
Constant	−3.579	0.541	0.000		
Had previous partner(s)	1.841	0.670	0.006	6.302	(1.697,23.410)
Age 15–22	0.550	0.500	0.272	1.733	(0.650,4.620)
Age 23–32	0.975	0.507	0.054	2.651	(0.982,7.157)
Prev. partners * Age 15–22	1.883	0.716	0.009	6.571	(1.615,26.738)
Prev. partners * Age 23–32	−0.116	0.566	0.838	0.891	(0.294,2.702)
Has child(ren)	1.232	0.312	0.000	3.429	(1.861,6.318)
Prev. partners * Has child(ren)	−0.935	0.411	0.023	0.393	(0.175,0.879)
High SCE	0.025	0.138	0.856	1.025	(0.782,1.344)
Prev. partners * High SCE	−0.737	0.232	0.001	0.479	(0.304,0.754)
Higher SES	−0.058	0.208	0.782	0.944	(0.628,1.419)
Prev. partners * Higher SES	−0.889	0.394	0.024	0.411	(0.190,0.889)
Random effect SD σ_u	0.001	0.012			

Table 8: Logistic model of partnership formation for men. Estimated coefficients and odds ratios (OR) are shown together with standard errors, p-values and95% confidence intervals (CI) for odds ratios. The last age category (33–50) was chosen as the reference category.

	Est.	s.e.	p	OR	OR 95% CI
Constant	−3.410	0.480	0.000		
Had previous partner(s)	0.127	0.499	0.799	1.136	(0.427,3.023)
Age 15–22	−0.795	0.451	0.078	0.451	(0.186,1.093)
Age 23–32	0.334	0.409	0.414	1.396	(0.627,3.109)
Prev. partners * Age 15–22	2.763	0.731	0.000	15.855	(3.787,66.373)
Prev. partners * Age 23–32	0.580	0.490	0.237	1.785	(0.683,4.668)
Has child(ren)	2.849	0.370	0.000	17.275	(8.372,35.643)
Prev. partners * Has child(ren)	−2.302	0.469	0.000	0.100	(0.040,0.251)
Social activity	0.251	0.137	0.067	1.285	(0.982,1.682)
Random effect SD σ_u	0.607	0.127			

Altogether three childhood factors were associated with partnership formation: socioeconomic status (SES, Table 7), self-control of emotions (SCE, Table 7), and social activity (Table 8). Being from a higher SES family background was associated with a longer time to repartner for women. High self-control of emotions that was found to predict cluster membership in the regression tree analysis of SA was also a predictor in the event history analysis of partnership formation: women who had higher self-control of emotions at age 8 had a lower risk of forming a new partnership following a dissolution. The effect of social activity was significant at the 10% level for men: being more socially active at age 8 was associated with forming partnerships sooner. The effect was the same for first and recurrent partnerships.

Parents were faster at forming first partnerships, although only ten participants had a child before forming any coresidential partnerships. There was some evidence that fathers also formed recurrent partnerships faster compared to childless men ($\hat{\beta} = 2.849 - 2.302 = 0.548$, s.e. = 0.300, p-value = 0.068).

Child-centred parenting, which was found to be the most important covariate in the regression tree, was not a significant predictor of partnership formation for either sex after controlling for the effects of other covariates. Childhood family structure was not significant in either model after controlling for the other childhood variables.

3.2.2 Partnership dissolution

Partnership dissolutions were explored in a similar way to formations. Time was captured in the models by two different variables: the age at the start of the current partnership and the duration of the partnership. Different functional forms (linear, quadratic, logarithmic, and categorical) were studied for both variables. Covariates measured during childhood were treated as time-invariant; type of partnership (marriage/cohabitation), parenthood status, and existence of previous partners as time-dependent. Child-centred parenting and family structure (included in CCP) were correlated, which induced multicollinearity in the model for women. Both variables were considered important and included irrespective of the large standard error of CCP in the common model.

Tables 9 and 10 show the results from the event history models of partnership dissolutions for women and men respectively. The random effect standard deviations were large but non-significant. The age effect was linear and

decreasing for women. For men, the estimated effects of age and age squared formed a quadratic curve: the risk decreased until 42 years of age and then slightly increased (the age at which the hazard reached its minimum was found by taking the square root of the first derivative of the quadratic function). For men, the effect of the duration of the current partnership was linear and decreasing. For women, the risk of partnership dissolution was quadratic, increasing until 12 years into the partnership and then decreasing.

Table 9: Logistic model of partnership dissolution for women. Estimated coefficients and odds ratios (OR) are shown together with standard errors, p-values and 95% confidence intervals (CI) for the odds ratios.

	Est.	s.e.	p	OR	OR 95% CI
Constant	−1.589	0.594	0.007		
Age at partnership formation	−0.055	0.018	0.003	0.946	(0.913,0.982)
Partnership duration	0.095	0.055	0.086	1.100	(0.987,1.225)
(Partnership duration)^2	−0.004	0.002	0.046	0.996	(0.991,1.000)
Married	−1.109	0.249	0.000	0.330	(0.204,0.534)
High self-control of emotions	−0.397	0.173	0.022	0.672	(0.479,0.944)
Broken family at 14	0.532	0.248	0.032	1.702	(1.048,2.766)
Child-centred parenting	−0.636	0.476	0.182	0.529	(0.208,1.347)
Random effect SD σ_u	0.518	0.254			

Table 10: Logistic model of partnership dissolution for men. Estimated coefficients and odds ratios (OR) are shown together with standard errors, p-values and 95% confidence intervals (CI) for odds ratios.

	Est.	s.e.	p	OR	OR 95% CI
Constant	1.211	1.495	0.418		
Age at partnership formation	−0.254	0.105	0.019	0.782	(0.637,0.961)

(Age at partnership formation)^2	0.003	0.002	0.055	1.003	(1.000,1.007)
Partnership duration	−0.040	0.019	0.032	0.961	(0.926,0.997)
Broken partnership(s)	0.757	0.272	0.005	2.132	(1.252,3.630)
Has child(ren)	−0.701	0.228	0.002	0.496	(0.317,0.776)
High self-control of emotions	−0.443	0.158	0.005	0.642	(0.471,0.875)
Random effect SD σ_u	0.385	0.247			

Previous experience of dissolution increased the risk of subsequent separation or divorce among men but not among women. Married women were less likely to dissolve their partnerships compared to cohabiting women, but cohabiting and married men did not differ in their risk of dissolution. Motherhood did not change the risk of dissolution but fathers had a lower risk than men without children.

Three childhood characteristics were connected to the risk of dissolution: self-control of emotions, family disruption, and child-centred parenting. High self-control of emotions at age 8 decreased the risk of dissolution for both sexes and all partnerships, while child-centred parenting was associated with a lower risk of dissolution for women. The experience of a broken family during childhood was associated with a higher risk of partnership dissolution among women, but not men.

4 Summary and discussion

This paper had two aims: (i) to describe the use of complementary statistical methods, sequence analysis and event history analysis, in a study of recurrent events; and (ii) to apply both techniques in a study of partnership formation and dissolution over the life course.

4.1 Statistical analysis

Sequence analysis was used to build an overall picture of partnership histories from age 15 to 50. Using Ward's clustering method, eight clusters were found, which together explained over 60% of sequence variation. These differed from each other according to the number, timing, and duration of partnerships. Another clustering method, that uses external information for the division of the data, was also studied. Regression tree analysis was used to divide data into clusters based on childhood covariates. Two significant predictors of partnership histories – high self-control of emotions and child-centred parenting – were found, which altogether explained only 3.5% of the variability of partnership histories. In contrast, the three-cluster solution using Ward's method without external information resulted in $R^2 = 35\%$, which increased to 61% for the chosen eight-cluster solution. Hence, the predictive power of those covariates alone was very low, although this was to be expected as we did not account for many factors that previous studies have found to be related to partnership formation and dissolution (e.g. the presence and age of children, educational attainment, employment, income, religiosity, and health-related factors; see e.g. Aassve et al., 2006; Berrington & Diamond, 2000; Jalovaara, 2012; Lyngstad & Jalovaara, 2010; South, 2001; Steele et al., 2006). Many of these other factors are time-varying which is problematic with regression trees, and were therefore beyond the scope of the analysis. However, other life domains could be added as parallel sequences that can then be analysed with multidimensional sequence analysis methods (Gauthier et al., 2010; Müller, Sapin, Gauthier, Orita, & Widmer, 2012; Salmela-Aro et al., 2011). In a previous study, Eerola and Helske (2012) compared SA and EHA in a case of multiple parallel life domains using the same JYLS data.

Event history analysis was used to model the probability of partnership transitions between ages 15 to 50 as a function of individual (i.e., social activity and high self-control emotions) and family characteristics (i.e., child-centred home environment, SES, and structure of the family of origin). To account for dependency between the durations of repeated episodes, random effects models for partnership formations and dissolutions were fitted. For all but one model there was no statistically significant unobserved variation between individuals once the childhood variables were included in the analyses, indicating that these factors captured a substantial part of the variation in partnership formation and dissolution that is due to time-invariant characteristics. A joint model of partnership formations and dissolution (as described in Section 2.3.2) was also fitted for women and men. The idea was to study whether there was correlation between the durations of episodes of living with and without a partner, for example because individuals who separate more rapidly tend to form new partnerships sooner than individuals whose partnerships last longer (as shown by Aassve et al., 2006; Steele et al., 2006 using British data). However, our sample was too small to estimate a joint model, leading to confidence intervals of correlation estimates ranging from −1 to 1.

Sequence analysis and event history analysis provide complementary information on partnership formation. Sequence analysis is a descriptive tool that gives an overall picture of the histories and compresses them in a form that is relatively easy to interpret. Sequences are often shown as colourful lines in an index plot, from which it is – especially after clustering – easy to see the timing of important partnership transitions and the approximate duration of different episodes. Clustering helps to describe the data and to identify similar

patterns in partnership formation by providing typologies of partnership trajectories. However, choosing the number of clusters is to some extent subjective. It is therefore important to consider a range of solutions and to regard the division of life sequences into clusters as suggestive. One should also be cautious about attaching too much meaning to a cluster or a label assigned to it, as the labels given to the clusters are only approximate since borderline cases could also be assigned to other clusters. For example, in the present study most of the members of the "later 1st partnership" cluster had stayed with their first partner but there were also several members who had lived separated or with a new partner for a long time.

Analysis of individual-level event histories is better for drawing inferences about the effects of covariates on the timing of recurring partnership transitions. It can account for censoring and unobserved individual characteristics that affect the timing and duration of partnerships. However, with discretely measured recurrent events, forming the data set can be time-consuming and the size of the person-episode-period type-of-data may be large even when the number of individuals is small, leading to long estimation times when random effects models are used.

Although SA and EHA are both methods for studying longitudinal life course data, their approaches in capturing time are different in many respects and they provide versatile information on the phenomenon of interest. In SA, the focus is on the holistic pattern of the histories and analysis is retrospective in nature. In contrast, in EHA the interest lies in the transitions and the direction of inference is prospective: how much time passes before an event happens. Each episode is as important as the others, no matter how short. In SA, however, especially with the most popular alignment methods for computing

sequence dissimilarities such as OM and Hamming, small deviations from a general pattern might not be very influential. For example, in terms of our (rather restricted) state-space (Table 1), hypothetical sequences P1-P2-D-P3-D-P3-P3 and P1-P2-D-P3-P3-P3-P3 would have been regarded as very similar even though the former person had four partners and five transitions and the latter one only three partners and three transitions. The definition of the state-space also matters: had we not separated partnerships by order, distinguishing successive partnerships would have been even more difficult or indeed impossible (as with P1 and P2 in the example sequences above). In such cases, if it is important to treat each episode as distinct, other dissimilarity criteria such as those based on counting common subsequences might be better suited.

SA and EHA are, of course, not the only options suitable for studying discrete longitudinal life course data. For example, trajectory analysis (Nagin, 1999) and latent class analysis (LCA; e.g. Vermunt, Tran, & Magidson, 2008) come in the middle ground of the approaches presented in this paper by using statistical models to create homogenous clusters of similar trajectories. Semi-parametric trajectory analysis can be used for studying binary trajectories such as the histories of living single/in partnership. However, the method is not suited for categorical trajectories with more than two unordered categories. For categorical data, LCA has been used to group trajectories. The standard version of LCA does not take into account the correlation between observations measured in different time periods, but several modifications have been proposed to adjust for the temporal correlation. See Barban and Billari (2012) for a comparison of LCA to SA.

4.2 Partnership formation and dissolution

Different factors related to childhood and current life situation were found to be connected to partnership formation and dissolution for women and men. Contrary to previous research (e.g. Berrington & Diamond, 2000; Rönkä et al., 2000; Ross et al., 2009; Steele et al., 2006), we did not find a significant effect of SES of subjects' fathers on the timing of their first partnerships. In common with previous research by Goldstein et al. (2004) and Steele et al. (2006), the SES of the childhood family was also not connected to men's risk of repartnering, but women with higher SES background had a lower risk.

Many previous studies have shown an increased dissolution risk for higher-order unions, but this has been assumed to be at least partly due to selection on unobserved individual characteristics. Studies that have considered such characteristics have not found an excessive risk of dissolution for recurrent partnerships (Aassve et al., 2006; Lillard, Brien, & Waite, 1995; Poortman & Lyngstad, 2007; Steele et al., 2005; Steele et al., 2006), although few have studied men. Our finding that repartnered men had a higher risk of dissolution was in contrast to studies of British (Aassve et al. 2006) and Norwegian (Poortman and Lyngstad 2007)men, which did not find differences in the dissolution risk by partnership order.

In common with previous studies (e.g. Andersson, 2002; Liefbroer & Dourleijn, 2006; Manning, Smock, & Majumdar, 2004), married women were less likely to dissolve their partnerships (first as well as recurrent) compared to cohabiting women. In contrast, cohabiting and married men did not differ in their risks. Motherhood did not change the risk of dissolution but fathers had a lower risk compared to childless men. However, the models only accounted for having

(biological or adopted) children in general. By choosing this conceptualisation of parenthood, some information about the effects of children on the risk of dissolution of partnership is inevitably lost. Earlier research has found different, even opposite, effects of the presence, number, and age of children on partnership dissolution across countries (Coppola & Di Cesare, 2008; Lillard & Waite, 1993; Lyngstad & Jalovaara, 2010; Steele et al., 2005; Svarer & Verner, 2008).

Of the socio-emotional characteristics considered, high self-control of emotions at age 8 was the strongest explanatory variable of partnership transitions. As expected, individuals with high self-control of emotions, indicated by emotional stability and constructive and compliant behaviour (Kokko et al., 2008), had a lower risk of partnership dissolution. For women the probability of repartnering was also lower but, contrary to our expectations, there was no association with the timing of the first partnership. Furthermore, high self-control of emotions was also related to fewer and more stable partnerships for participants who had experienced less child-centred parenting practices during childhood. These results suggest that high self-control of emotions was associated with a more stable family life, even for those individuals with a less supportive family environment in childhood. It is possible that a stable partnership was a part of a cycle of good social functioning linked to child's high self-control of emotions (Pulkkinen, 2009).

In accordance with our expectations, high social activity in childhood was related to men's tendency to form first and also subsequent partnerships faster. Among women social activity was not related to the timing or pace of partnership events. This difference could be partly due to diverse forms of

social activity in boys and girls. In a previous study, Pulkkinen (1995) found that high social activity in boys was more often linked with unfavourable behaviour.

4.3 Limitations and strengths

When interpreting our results, there are some limitations that should be noted. First, our analyses considers only one age cohort of one nationality. Therefore our findings may not generalise to older and younger age cohorts and other nationalities, although many of our results were consistent with previous studies. Second, information on partnerships was gathered using the Life History Calendar (LHC), presented to the JYLS participants during the age 42 and age 50 personal interviews (in 2001 and 2009, respectively). The LHCs covered a time span from age 15 to 50. The long recall period may raise questions about the accuracy of the participants' memory and the validity of the LHC data. However, we do not consider this to be a serious flaw because prospective data on these transitions were also gathered in the JYLS study and these data have been informally used to check the validity of the LHC data (Kokko et al., 2009). Furthermore, previous studies have shown that information gathered with the LHC is reliable (Caspi et al., 1996; Freedman, Thornton, Camburn, Alwin, & Young-DeMarco, 1988). A third limitation of our study is that, in common with most other birth cohort studies where life histories are collected retrospectively, we do not have data on the childhood characteristics and partnership histories of the partners of cohort members.

The two data collection phases led to a high proportion of partnership histories that were right-censored at age 42 (the time of the first phase). We were therefore forced to use missing states for these shorter sequences, which in turn led to problems in the definition of costs in SA. Clustering results made most sense when the cost for aligning any state to a missing state was set to

zero. However, this cost setting resulted in Hamming dissimilarities that are not metric distances, as assumed by most clustering methods. Since the chosen clusters were reasonable, and in any case considered suggestive, the use of non-metric dissimilarities is most likely not very serious.

Even though the JYLS study is long and extensive, the moderate sample size imposed many restrictions in model building. For example, we were unable to model partnership formations and dissolutions jointly. Moreover, when specifying a piecewise constant baseline hazard function we were forced to use broad age intervals. We considered only a simple indicator of being a parent which did not account for the different aspects of family structure that other studies have found to be related to the risk of partnership formation and dissolution (such as the number, age, and residence of the child(ren) or blended families). We also faced challenges due to the coarse annual measurements and had to be careful when defining the risk sets: for some individuals there seemed to be no unpartnered episodes between two partnerships.

Although the use of the JYLS data imposed methodological restrictions, strengths of the data are the rich covariate information and exceptionally long period of follow-up (from age 8 to 50). This enabled the examination of childhood individual and family characteristics as precursors of partnership transitions measured up to middle-age. In particular, childhood socio-emotional characteristics have not been studied before in this context. As can be seen from the non-significant random effect variances in some of the models, we could capture a notable part of the variation due to time-invariant individual characteristics that in previous studies have simply been left to the unobserved random part. The research question concerned the effects of childhood characteristics on the timing and stability of partnerships. These childhood

measures were not used as proxies for the socio-emotional qualities of an adult. Nevertheless, a significant relationship between childhood socio-emotional characteristics and adult personality has been found in the JYLS data (Pulkkinen et al., 2012).

Another contribution of this paper was to demonstrate and compare use of SA and EHA, which to our knowledge is the first attempt to apply both methods in a study of recurrent life events.

Acknowledgements

We appreciate Professor Lea Pulkkinen's contribution to the JYLS over the years. We also thank the referees for their helpful comments and suggestions.

Satu Helske has been funded by the Jyväskylä Graduate School in Computing and Mathematical Sciences (COMAS) and the Finnish Cultural Foundation. Fiona Steele was supported by an ESRC grant for a node of the National Centre for Research Methods (RES-576-25-0032). The major funder of the Jyväskylä Longitudinal Study of Personality and Social Development (JYLS) has been the Academy of Finland, most recently through grant nos. 127125 (Pulkkinen) and 118316 and 135347 (Kokko).

References

Aassve, A., Burgess, S., Propper, C., & Dickson, M. (2006). Employment, family union and childbearing decisions in Great Britain. *Journal of the Royal*

Statistical Society: Series A (Statistics in Society), 169(4), 781-804. doi:10.1111/j.1467-985X.2006.00432.x

Aassve, A., Billari, F. C., & Piccarreta, R. (2007). Strings of adulthood: A sequence analysis of young British women's work-family trajectories. *European Journal of Population/Revue Européenne De Démographie, 23*(3-4), 369-388. doi:10.1007/s10680-007-9134-6

Abbott, A. (1983). Sequences of social events: Concepts and methods for the analysis of order in social processes. *Historical Methods, 16*(4), 129-147. doi:10.1080/01615440.1983.10594107

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology, 21*(1), 93-113. doi:10.1146/annurev.so.21.080195.000521

Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect. *Sociological Methods & Research, 29*(1), 3-33. doi:10.1177/0049124100029001001

Aisenbrey, S., & Fasang, A. E. (2010). New life for old ideas: The "second wave" of sequence analysis – Bringing the "course" Back Into the life course. *Sociological Methods & Research, 38*(3), 420-462. doi:10.1177/0049124109357532

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology, 13*(1), 61-98.

Amato, P. R. (1996). Explaining the intergenerational transmission of divorce. *Journal of Marriage and the Family, 58*(3), 628-640.

Andersson, G. (2002). Children's experience of family disruption and family formation: Evidence from 16 FFS countries. *Demographic Research, 7*(7), 343-364. doi:10.4054/DemRes.2002.7.7

Barban, N., & Billari, F. C. (2012). Classifying life course trajectories: A comparison of latent class and sequence analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 61*(5), 765-784. doi:10.1111/j.1467-9876.2012.01047.x

Berrington, A., & Diamond, I. (2000). Marriage or cohabitation: A competing risks analysis of first-partnership formation among the 1958 British birth cohort. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 163*(2), 127-151. doi:10.1111/1467-985X.00162

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees* [Inc.] Belmont, CA: Wadsworth and Brooks.

Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *Stata Journal, 6*(4), 435.

Bumpass, L. L., Martin, T. C., & Sweet, J. A. (1991). The impact of family background and early marital factors on marital disruption. *Journal of Family Issues, 12*(1), 22-42. doi:10.1177/019251391012001003

Caspi, A., Elder, G. H., & Bem, D. J. (1988). Moving away from the world: Life-course patterns of shy children. *Developmental Psychology, 24*(6), 824-831. doi:10.1037/0012-1649.24.6.824

Caspi, A., Moffitt, T. E., Thornton, A., Freedman, D., & others. (1996). The life history calendar: A research and clinical assessment method for collecting retrospective event-history data. *International Journal of Methods in Psychiatric Research, 6*(2), 101-114. doi:10.1002/(SICI)1234-988X(199607)6:2<101::AID-MPR156>3.3.CO;2-E

Coppola, L., & Di Cesare, M. (2008). How fertility and union stability interact in shaping new family patterns in Italy and Spain. *Demographic Research, 18*(4), 117-144. doi:10.4054/DemRes.2008.18.4

Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge, UK: Cambridge University Press.

Eerola, M., & Helske, S. (2012). Statistical analysis of life history calendar data. *Statistical Methods in Medical Research,* doi:10.1177/0962280212461205

Elder, G. H. (1998). The life course and human development. In W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Volume 1: Theoretical models of human development* (5th ed., pp. 939-991) Hoboken, US: Wiley.

Elzinga, C. H. (2006). Sequence analysis: Metric representations of categorical time series. *Manuscript,*

Elzinga, C. H., & Liefbroer, A. C. (2007). De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis. *European Journal of Population/Revue Européenne De Démographie, 23*(3), 225-250. doi:10.1007/s10680-007-9133-7

Freedman, D., Thornton, A., Camburn, D., Alwin, D., & Young-DeMarco, L. (1988). The life history calendar: A technique for collecting retrospective data. *Sociological Methodology, 18*, 37.

Furstenberg, F. F., Jr., & Spanier, G. B. (1984). The risk of dissolution in remarriage: An examination of Cherlin's hypothesis of incomplete institutionalization. *Family Relations,* , 433-441.

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software, 40*(4), 1-37.

Gähler, M., Hong, Y., & Bernhardt, E. (2009). Parental divorce and union disruption among young adults in Sweden. *Journal of Family Issues, 30*(5), 688-713. doi:10.1177/0192513X08331028

Gauthier, J., Widmer, E. D., Bucher, P., & Notredame, C. (2009). How much does it cost? Optimization of costs in sequence analysis of social science data. *Sociological Methods & Research, 38*(1), 197-231. doi:10.1177/0049124109342065

Gauthier, J., Widmer, E. D., Bucher, P., & Notredame, C. (2010). Multichannel sequence analysis applied to social science data. *Sociological Methodology, 40*(1), 1-38. doi:10.1111/j.1467-9531.2010.01227.x

Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Chichester, UK: John Wiley & Sons. doi:10.1002/9780470973394.indsub

Goldstein, H., Pan, H., & Bynner, J. (2004). A flexible procedure for analyzing longitudinal event histories using a multilevel model. *Understanding Statistics, 3*(2), 85-99. doi:10.1207/s15328031us0302_2

Halpin, B. (2014). *SADI: Sequence analysis tools for Stata.* Unpublished manuscript.

Halpin, B. (2010). Optimal matching analysis and life-course data: The importance of duration. *Sociological Methods & Research, 38*(3), 365-388.

Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal, 29*(2), 147-160. doi:10.1002/j.1538-7305.1950.tb00463.x

Hollister, M. (2009). Is optimal matching suboptimal? *Sociological Methods & Research, 38*(2), 235-264. doi:10.1177/0049124109346164

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2*(1), 193-218.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. Hoboken, New Jersey: John Wiley & Sons.

Kennedy, S., & Bumpass, L. (2008). Cohabitation and children's living arrangements: New estimates from the United States. *Demographic Research, 19*, 1663-1692.

Kiernan, K. (2001). The rise of cohabitation and childbearing outside marriage in Western Europe. *International Journal of Law, Policy and the Family, 15*(1), 1-21.

Kinnunen, U., & Pulkkinen, L. (2003). Childhood socio-emotional characteristics as antecedents of marital stability and quality. *European Psychologist, 8*(4), 223-237. doi:10.1027/1016-9040.8.4.223

Kokko, K., & Pulkkinen, L. (2000). Aggression in childhood and long-term unemployment in adulthood: A cycle of maladaptation and some protective factors. *Developmental Psychology, 36*(4), 463-472. doi:10.1037/0012-1649.36.4.463

Kokko, K., Pulkkinen, L., & Mesiäinen, P. (2009). Timing of parenthood in relation to other life transitions and adult social functioning. *International Journal of Behavioral Development, 33*(4), 356-365. doi:10.1177/0165025409103873

Kokko, K., Pulkkinen, L., Mesiäinen, P., & Lyyra, A. (2008). Trajectories based on post-comprehensive and higher education and their correlates and antecedents. *Journal of Social Issues, 64*(1), 59-76. doi:10.1111/j.1540-4560.2008.00548.x

Lesnard, L. (2010). Setting cost in optimal matching to uncover contemporaneous socio-temporal patterns. *Sociological Methods & Research, 38*(3), 389-419. doi:10.1177/0049124110362526

Liefbroer, A. C., & Dourleijn, E. (2006). Unmarried cohabitation and union stability: Testing the role of diffusion using data from 16 European countries. *Demography, 43*(2), 203-221. doi:10.1353/dem.2006.0018

Lillard, L. A., Brien, M. J., & Waite, L. J. (1995). Premarital cohabitation and subsequent marital dissolution: A matter of self-selection? *Demography, 32*(3), 437-457. doi:10.2307/2061690

Lillard, L. A., & Waite, L. J. (1993). A joint model of marital childbearing and marital disruption. *Demography, 30*(4), 653-681. doi:10.2307/2061812

Lyngstad, T. H. (2006). Why do couples with highly educated parents have higher divorce rates? *European Sociological Review, 22*(1), 49-60. doi:10.1093/esr/jci041

Lyngstad, T. H., & Jalovaara, M. (2010). A review of the antecedents of union dissolution. *Demographic Research, 23*(10), 257-292. doi:10.4054/DemRes.2010.23.10

Manning, W. D., Smock, P. J., & Majumdar, D. (2004). The relative stability of cohabiting and marital unions for children. *Population Research and Policy Review, 23*(2), 135-159. doi:10.1023/B:POPU.0000019916.29156.a7

McVicar, D., & Anyadike-Danes, M. (2002). Predicting successful and unsuccessful transitions from school to work by using sequence methods. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 165*(2), 317-334. doi:10.1111/1467-985X.00641

Müller, N. S., Sapin, M., Gauthier, J., Orita, A., & Widmer, E. D. (2012). Pluralized life courses? An exploration of the life trajectories of individuals with psychiatric disorders. *International Journal of Social Psychiatry, 58*(3), 266-277. doi:10.1177/0020764010393630

Nagin, D. S. (1999). Analyzing developmental trajectories: A semiparametric, group-based approach. *Psychological Methods, 4*(2), 139.

Piccarreta, R., & Lior, O. (2010). Exploring sequences: A graphical tool based on multi-dimensional scaling. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 173*(1), 165-184. doi:10.1111/j.1467-985X.2009.00606.x

Pitkänen, T., Kokko, K., Lyyra, A., & Pulkkinen, L. (2008). A developmental approach to alcohol drinking behaviour in adulthood: A follow-up study from age 8 to age 42. *Addiction, 103*(s1), 48-68. doi:10.1111/j.1360-0443.2008.02176.x

Pitkänen, T., Lyyra, A., & Pulkkinen, L. (2005). Age of onset of drinking and the use of alcohol in adulthood: A follow-up study from age 8–42 for females and males. *Addiction, 100*(5), 652-661. doi:10.1111/j.1360-0443.2005.01053.x

Poortman, A., & Lyngstad, T. H. (2007). Dissolution risks in first and higher order marital and cohabiting unions. *Social Science Research, 36*(4), 1431-1446. doi:10.1016/j.ssresearch.2007.02.005

Pulkkinen, L. (1995). Behavioral precursors to accidents and resulting physical impairment. *Child Development, 66*(6), 1660-1679. doi:10.1111/j.1467-8624.1995.tb00957.x

Pulkkinen, L. (2009). Personality – a resource or risk for successful development. *Scandinavian Journal of Psychology, 50*(6), 602-610. doi:10.1111/j.1467-9450.2009.00774.x

Pulkkinen, L., & Kokko, K. (2010). Keski-ikä elämänvaiheena [middle-age as a stage of life]. (pp. 5-13) Jyväskylä, Finland: University of Jyväskylä.

Pulkkinen, L., Kokko, K., & Rantanen, J. (2012). Paths from socioemotional behavior in middle childhood to personality in middle adulthood. *Developmental Psychology, 48*(5), 1283-1291. doi:10.1037/a0027463

Pulkkinen, L., Lyyra, A., & Kokko, K. (2009). Life success of males on nonoffender, adolescence-limited, persistent, and adult-onset antisocial pathways: Follow-up from age 8 to 42. *Aggressive Behavior, 35*(2), 117-135. doi:10.1002/ab.20297

Räikkönen, E., Kokko, K., Chen, M., & Pulkkinen, L. (2012). Patterns of adult roles, their antecedents and psychosocial wellbeing correlates among Finns born in 1959. *Longitudinal and Life Course Studies, 3*(2), 211-227.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage.

Rohwer, G., & Pötter, U. (2004). *TDA user's manual* Bochum: Ruhr-Universität Bochum.

Rönkä, A., Kinnunen, U., & Pulkkinen, L. (2000). The accumulation of problems of social functioning as a long-term process: Women and men compared. *International Journal of Behavioral Development, 24*(4), 442-450. doi:10.1080/016502500750037991

Rönkä, A., & Pulkkinen, L. (1998). Work involvement and timing of motherhood in the accumulation of problems in social functioning in young women. *Journal of Research on Adolescence, 8*(2), 221-239.

Ross, A., Schoon, I., Martin, P., & Sacker, A. (2009). Family and nonfamily role configurations in two British cohorts. *Journal of Marriage and Family, 71*(1), 1-14.

Salmela-Aro, K., Kiuru, N., Nurmi, J., & Eerola, M. (2011). Mapping pathways to adulthood among Finnish university students: Sequences, patterns, variations in family- and work-related roles. *Advances in Life Course Research, 16*(1), 25-41. doi:10.1016/j.alcr.2011.01.003

Shanahan, M. J. (2000). Pathways to adulthood in changing societies: Variability and mechanisms in life course perspective. *Annual Review of Sociology,* , 667-692.

Statistics Finland. (1994). *Suomalainen lapsiperhe [the Finnish family with children].* Helsinki, Finland: Hakapaino Oy.

Statistics Finland. (2010). *Statistical yearbook of finland 2010.* Helsinki, Finland: Tilastokeskus.

Steele, F. (2011). Multilevel discrete-time event history analysis with applications to the analysis of recurrent employment transitions.

Australian & New Zealand Journal of Statistics, 53(1), 1-20. doi:10.1111/j.1467-842X.2011.00604.x

Steele, F., Kallis, C., Goldstein, H., & Joshi, H. (2005). The relationship between childbearing and transitions from marriage and cohabitation in Britain. *Demography, 42*(4), 647-673. doi:10.1353/dem.2005.0038

Steele, F., Kallis, C., & Joshi, H. (2006). The formation and outcomes of cohabiting and marital partnerships in early adulthood: The role of previous partnership experience. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 169*(4), 757-779. doi:10.1111/j.1467-985X.2006.00420.x

Studer, M. (2013). WeightedCluster library manual: A practical guide to creating typologies of trajectories in the social sciences with R. doi:10.12682/lives. 2296-1658.2013. 24

Studer, M., Ritschard, G., Gabadinho, A., & Müller, N. S. (2011). Discrepancy analysis of state sequences. *Sociological Methods & Research, 40*(3), 471-510. doi:10.1177/0049124111415372

Svarer, M., & Verner, M. (2008). Do children stabilize relationships in Denmark? *Journal of Population Economics, 21*(2), 395-417. doi:10.1007/s00148-006-0084-9

Teachman, J. (2008). Complex life course patterns and the risk of divorce in second marriages. *Journal of Marriage and Family, 70*(2), 294-305.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16*(3), 439-454.

Vermunt, J. K., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), (pp. 373-385). Burlington, MA: Elsevier.

Ward Jr., J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236-244. doi:10.1080/01621459.1963.10500845