Coping with complex individual histories: A comparison of life course methods with an application to partnership transitions in Norway

Julia Mikolai[1], *University of Southampton*

Mark J. Lyons-Amos, *Institute of Education, University of London*[2]

Preliminary draft. Please do not cite without the permission of the authors!

Acknowledgement

Julia Mikolai is funded by an Economic and Social Research Council (ESRC) +3 Scholarship. Mark J. Lyons-Amos was funded by the European Research Council Starting Grant CHILDCOHAB. An early draft of this paper was presented at the Annual Meeting of the British Society for Population Studies, 9-11 September 2013, Swansea, UK and at the XXVII IUSSP International Population Conference, 26-31 August 2013, Busan, Republic of Korea. The authors would like to thank Dr Agnese Vitali for her comments and suggestions on a previous draft of this paper. The GGS data were obtained from the GGP Data Archive and refer to the publication that contains the model survey instruments: United Nations 2005. Generations & Gender Programme: Survey Instruments. New York and Geneva: UN, 2005. The authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

[1] Corresponding author, e-mail: jm1e11@soton.ac.uk

[2] Analyses undertaken whilst at ESRC Centre for Population Change, University of Southampton.

Abstract

As variation in the pattern of family life courses has increased over the past 50 years, the techniques available to analyse life course data have also expanded and research tends to be interested in explaining more complexity in the family life course. Therefore, it is necessary to extend our methodological toolkit by increasing the complexity of event history models or by applying other promising methods. The aim of this paper is to compare and contrast sequence analysis, latent class growth models, and multistate event history models, to studying the family life course. The advantages and weaknesses of each of these methods are highlighted by applying them to the same empirical problem. Using data from the first wave of the Norwegian Generations and Gender Survey from 2007/2008, changes in the partnership status of women born between 1955 and 1964 are modelled across the life course, with education as the primary covariate of interest.

Keywords: life course methodology, sequence analysis, latent class growth models, multistate event history models, Norway

Introduction

In the last half century, patterns of family life courses have changed considerably. For example, the transition to parenthood has been delayed, non-marital cohabitation and non-marital childbearing have become more common, as have union dissolution and re-partnering. These changes have generated an increased interest in the applicability of different methods for modelling life courses with their complexities. Although a number of methods are available to study the family life course, discussion is mainly limited to comparing sequence analysis and simple event history models. These papers usually conclude that using sequence analysis is more appropriate for analysing the life course from a holistic perspective (Billari, 2001b, 2005; Billari & Piccarreta, 2001, 2005; Piccarreta & Billari, 2007). However, other available holistic methods (i.e. which examine the entire family life course), such as latent class growth models and multistate event history models, can also provide holistic results and are able to address research questions which may be beyond the scope of sequence analysis.

Simple event history analysis is commonly used to examine single or competing events (Heuveline & Timberlake, 2004; Perelli-Harris & Gerber, 2011; Perelli-Harris, Sigle-Rushton, et al., 2010). These analyses vary in focus and complexity. Recent studies (Baizán, Aassve, & Billari, 2003, 2004) applied simultaneous equations models to study the determinants of several concurrent life course transitions. Others used multilevel multiprocess models to account for correlated event histories (Steele, Kallis, Goldstein, & Joshi, 2005). These "event based" approaches primarily focus on the (causal) influence of certain covariates on particular events. Although simultaneous models improve upon simple event history models by accommodating possible interdependencies between several events via modelling joint processes and unobserved heterogeneity, they are limited to studying a specific segment of the life course and evaluating one-way transitions.

Others have promoted the use of sequence analysis arguing that unlike event history models, this approach can examine the life course trajectory as a whole meaningful unit ("holistic approach") by looking for "ideal-types" of trajectories that categorise and describe different life course patterns (Billari, 2001a, 2001b, 2005; Billari & Piccarreta, 2005; Piccarreta & Billari, 2007). It is also possible to assess how different covariates influence the probability of an individual to belong to one of these "ideal-types".

Despite the availability of promising techniques from other disciplines applicable to life course research, such as latent class growth models and multistate event history models, the existing literature is mainly limited to comparing the relative merits of event history analysis (EHA) and sequence analysis (SA) (Barban & Billari, 2012; Billari, 2001a, 2005; Billari & Piccarreta, 2005; Piccarreta & Billari, 2007) with the exception of Barban and Billari (2012) who compared sequence analysis and latent transition analysis. Multistate event history models and latent class growth models, have only been recently used (Mikolai, 2013; Perelli-Harris & Lyons-Amos, 2013) for studying the family life course. These methods combine the properties of the event based and the holistic approaches by being capable of focusing on several events while accounting for their previous occurrences. In this paper, we restrict our attention to methods that examine a manifest outcome variable and we are not interested in latent transition models because such models have a substantially different focus.

The aim of this paper is twofold. First, sequence analysis, latent class growth models, and multistate event history models are compared and contrasted. Second, by applying these methods to a real life example (Norwegian women born between 1955 and 1964), the differences and similarities as well as the strengths and weaknesses of these approaches are emphasised. This example focuses on the role of education on changes in partnership status (i.e. being never partnered, transition to first cohabitation and first marriage, and the dissolution of a first cohabitation or a first marriage). This paper aims to tackle the following questions, pertinent to life course research: How can sequence analysis, latent class growth models and multistate event history models be used for studying the influence of education on partnership transitions over the early family life course? What types of research questions can be answered using these methods? And are these methods applicable to the same problems to the same extent or is one of them better than the other and if so in which situation?

The following sections briefly describe each method and explain how they operate. This is followed by a description of the specific models that this paper studies. Results for each modelling technique with the interpretation of the result are presented, and then synthesised in the concluding section of the paper.

Sequence Analysis

Sequence analysis (SA) represents each individual life course by a sequence (i.e. a character string, which indicates the order and duration of states that the individual occupied in each year). For example, the sequence SSSCCMMMM means that the respondent was single (S) for three years, cohabited (C) for two years, and was married (M) for four years. Due to the large possible number of combinations of states, usually not many individuals experience the exact same sequence. To reduce the number of sequences, Optimal Matching Analysis (OMA) is used. This approach was introduced to the social sciences by Abbott (1995).

OMA reduces the number of possible sequences by identifying how similar pairs of sequences are. Similarity is defined in terms of the number, order, and duration of states within the sequences. The algorithm calculates the similarity or dissimilarity between two sequences by taking into account three possible operations: replacement (one state is replaced by another one), insertion (an additional state is added to the sequence), and deletion (a state is deleted from the sequence). The fewer operation of any kind is needed to turn one sequence into the other, the more similar two sequences are while the more operation is needed, the more dissimilar they are. Furthermore, to each operation, a certain cost can be attached. Therefore, identifying the relative cost of all operations is critical to determining (dis)similarity between sequences. Unfortunately these require *a priori* definition by the researcher with little objective measure of the correct specification, and results can be highly sensitive to their specification (Brzinsky-Fay & Kohler, 2010). In particular, the specification of higher insertion and deletion costs tends to reduce the number of substitutions and hence the estimated distance between differing sequences. The distance between two sequences is defined by the minimum costs of the operations that is necessary to transfer one sequence into the other (Abbott & Tsay, 2000). The distances are recorded in a dissimilarity matrix.

Then, in order to find existing patterns in the data, cluster analysis is performed on this dissimilarity matrix. The aim of the cluster analysis is to minimise the chosen within cluster distance and maximise the between cluster distance. The researcher needs to specify the number of clusters to be extracted from the data either *a priori* (e.g. k-means clustering) or by using statistics based on the ratio of within/between cluster distances (Calinski–Harabasz pseudo-F index and Duda–Hart indices). Once the clusters are formed, they can be described with respect to the grouping variables. Comparison of sequences can also be based on the number of episode changes within once sequence, the length of the sequences, or the number of different events in a sequence (Brzinsky-Fay & Kohler, 2010). Furthermore, the clusters can be used both as independent and dependent variables in further analyses (although the former approach has not been widely applied).

Latent Class Growth Models

Latent Class Growth Models (LCGMs) are a form of growth curve models with the key assumption that individuals are drawn from different subpopulations (classes), and hence an overall population growth curve cannot adequately describe individual deviations, even with the additions of random effects. Similarly to SA, these models have an individual centred perspective meaning that they seek to identify relationships between individual response patterns and form groups based on these patterns (Jung & Wickrama, 2008). Growth curves are typically formed by identifying a response variable for an individual across a number of time intervals (these need not be equally spaced). Changing expected values of this response are defined by a model including parameters for an intercept and slope. The intercept and slope parameters are typically allowed to vary based not only on observed covariates (e.g. education) but also on groupings extracted from response patterns (latent classes).

The growth equation is presented in Equation 1. We define $1 \dots J$ classes, which are denoted by C_j. The response (in our application partnership state) is defined as the random variable *y*, with the growth curve for this variable defined by intercept (α) and slope parameters (β) for time *t*. Since in this example *y* is nominal, it is transformed by a link function (e.g. logit). Note that all of these parameters can vary between classes. The shape of the growth curves can be altered by the inclusion of covariate information, in this case educational attainment. In this example the parameter β_3 can alter the intercept according to the vector of dummy variables *educ* corresponding to educational level, and the slope similarly altered by β_4. Again all of these effects depend on the membership of class *j*. We note that this model can be extended to include individual level deviations from the overall population line (via a random effect) to form the more general Growth Mixture Model, although we were unable to include this in our current approach for computational reasons.

$$f(y_{tj}) = \alpha + \beta_{1,j} t + \beta_{2,j} t^2 + \boldsymbol{\beta_{3,j} educ} + \boldsymbol{\beta_{4,j} educ.t}$$

Eq. 1

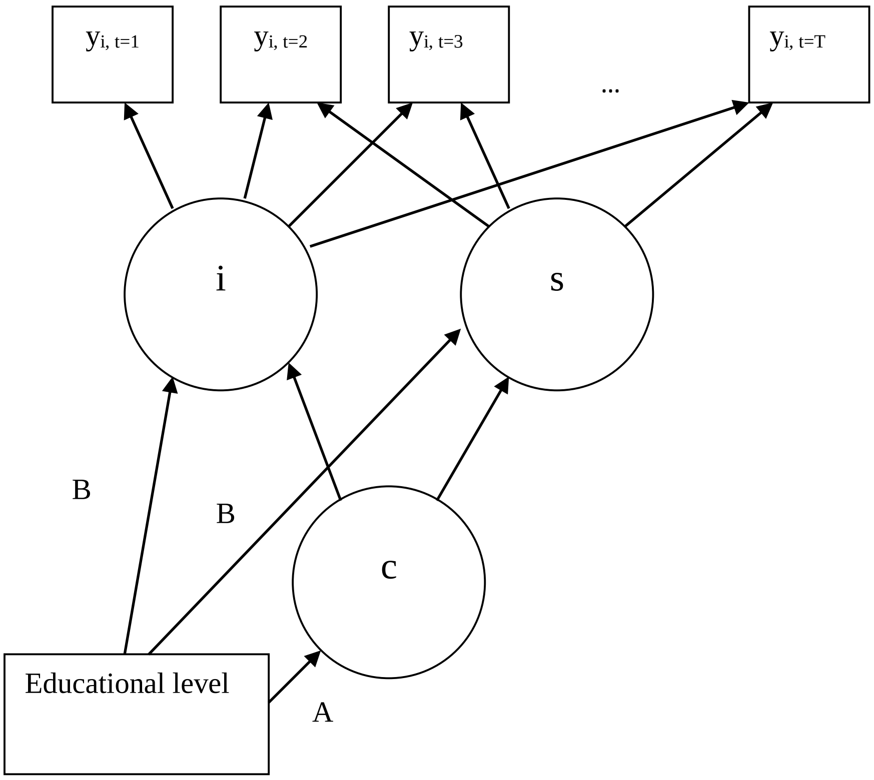
As with other latent class analysis, membership of a particular class can be determined by covariate information. This is represented as the probability π^j which is defined as $\Pr(C_j = J)$ and can depend on covariate information (in this case a vector of dummy variables representing educational attainment) in the form of Equation 2. $\boldsymbol{\gamma}$ is a vector of coefficients and again $f(\pi^j)$ is a link function, while.

$$f(\pi^j) = \boldsymbol{\gamma}.\boldsymbol{educ}$$

Eq. 2

To further facilitate interpretation, Figure 1 presents the conceptual model of LCGMs.

Figure 1. Conceptual Representation of LCGM with Covariates Altering the Growth Trajectories.

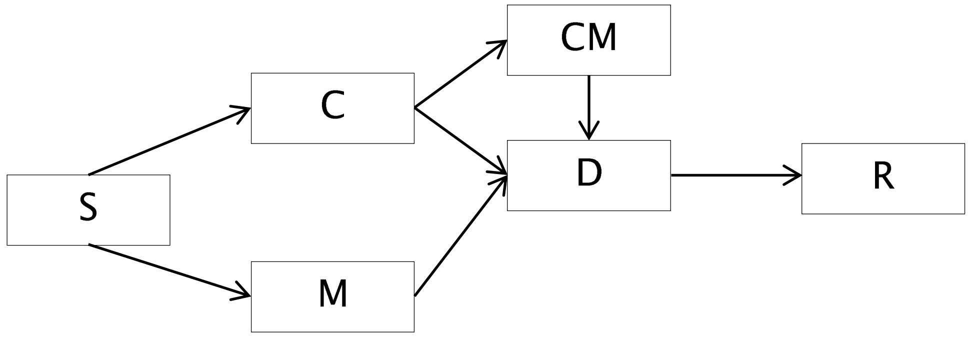


The response variable *y* forms a growth curve, described by the intercept *i* and slope *s*. The intercept and slope can vary by class *c*. These variables are all latent variables (denoted by circles). LCGMs can incorporate covariate information in two ways. First, covariates can be used to predict membership of a certain class, accounting for the probability of class membership (Wang, Hendricks Brown, & Bandeen-Roche, 2005). This is shown by line 'A' in Figure 1. This approach is comparable to sequence analysis. Where LCGMs have an advantage over SA is that it allows for covariates to alter the shape of trajectories (line 'B'). Specifically, the growth curve specified within each class is a function of covariate information and hence the trajectories will not only depend on class membership but also vary by education. An additional advantage of LCGMs is that a variety of fit statistics are available for deciding the optimal number of classes and can be validated via simulation since the estimates are model based. However, the different criteria and test statistics (such as AIC, BIC or Lo-Mendell-Rubin Likelihood Ratio Test) can lead to different and sometimes contradictory conclusions (Nylund, Asparouhav, & Muthen, 2007).

Multistate Event History Models

Multistate event history models differ from SA and LCGMs in that they do not aim to classify or group individuals. It is a variable centred approach where the main purpose is to establish statistical relationships between the independent variable(s) and several transitions. Multistate event history models are an extension of simple event history models; rather than examining one transition, this approach allows individuals to move among different states over time. These movements are assumed to be stochastic and are modelled by means of transition probabilities. Thus, multistate event history models allow for examining covariate effects on several transitions within the same model.

Figure 2. Multistate Event History Model.



Another distinct advantage of this method is the possibility to include time-varying covariates and thereby examine how the influence of a variable of interest changes over the family life course. This cannot be done using simple event history models, SA or LCGMs. The original multistate model assumes the Markov property; that is that the present behaviour of an individual is enough to predict its future behaviour (Andersen & Keiding, 2002; Hougaard, 1999). For example, it would assume that the transition probability from marriage to union dissolution is the same for all individuals irrespective of whether they have cohabited before marriage. As life course theory emphasises that earlier transitions play an important role in later transitions, this assumption is not realistic when taking a life course perspective. In order to be able to examine the partnership transitions in a dynamic way, the original Markov model can be extended. Figure 2 shows the multistate model estimated in this paper, where the following states are defined: never partnered (S), cohabitation (C), direct marriage (M), marriage that was preceded by cohabitation (CM), union dissolution (D) and re-partnering (R).

By defining the state ‘CM’, the model allows for differentiating between direct marriage and marriage that was preceded by cohabitation. Without defining such a state, the model would assume that the influence of education is the same on the transition to direct marriage and to marriage that was preceded by cohabitation. One disadvantage of multistate event history models is that as the number of states gets bigger and as individuals move along the life course, one might end up with small cell sizes and thus, or unreliable estimates of the transition hazards.

The multistate event history model is estimated as a stratified continuous-time Cox model where each transition is represented by a different stratum (de Wreede, Fiocco, & Putter, 2011; Putter, Fiocco, & Geskus, 2007; Putter, van der Hage, de Bock, Elgalta, & van de Velde, 2006). This means that we allow for each transition to have a separate baseline hazard. Covariates are incorporated as transition-specific covariates to allow for the effect of the covariates to differ across transitions. The transition hazard of individual *k* is given by:

$$\lambda_{ij}(t|\boldsymbol{Z}) = \lambda_{ij,0}(t)\exp\left(\boldsymbol{\beta}^T\boldsymbol{Z}_{ij}\right)$$

Eq. 3

where *ij* indicates a transition from state i to state *j*, $\lambda_{ij,0}(t)$ is the baseline hazard of this transition, $\boldsymbol{Z}$ is the vector of covariates at baseline and $\boldsymbol{Z}_{ij}$ is the vector of transition-specific covariates.

Data

To illustrate the similarities and differences between sequence analysis, latent class growth models, and multistate event history models, a real-life application is presented. Using data from the first wave of the Norwegian Generations and Gender Survey[3] (GGS) from 2007/2008 (N = 14,881), we examine the influence of educational attainment on changes in partnership status of women born between 1955 and 1964.

[3] This paper used the version that is available in the Harmonized Histories (Perelli-Harris, Kreyenfeld, & Kubisch, 2010).

The dataset includes extensive retrospective information on the start and end date (year and month) of up to five cohabitating and marital unions as well as union dissolutions. In the Norwegian GGS, cohabitation is defined as a co-residential relationship which lasted for at least three months. Partnership histories are reconstructed using this information. LCGMs and SA are fitted using yearly partnership information, while multistate event history models utilise monthly information.

Although the GGS provides cross-sectional weights, no longitudinal weights are available. As cross-sectional weights are only representative of the population structure in the year of the survey, the analyses presented in this paper do not incorporate weights.

Variables

Level of education. In all three models, the highest level of education at the time of the survey is measured by a categorical variable with the following categories: low (ISCED 0, ISCED 1, and ISCED 2), medium (ISCED 3 and ISCED 4), and high education (ISCED 5 and ISECD 6). High education is used as a reference category in all three models. In the multistate event history models, education is measured as a time-varying variable which is created using information on the year and month of reaching the highest level of education. We assume continuous education from age 15 and that secondary education takes 4 years while high education takes 3 years on average. Missing information (7.9%) on the year and/or month of reaching the highest level of education was imputed using information on the median age of finishing education by educational level. In LCGMs and SA, education is time-constant and indicates the highest level of education at the time of the survey.

Educational enrolment is measured by a time-varying categorical variable and indicates whether the respondent was enrolled or not (reference) in full-time education in the given month. This variable is used as a control variable only in the multistate event history models.

Modelling Strategy

This paper presents three sets of analyses. First, using sequence analysis, several groups are created based on women's yearly partnership trajectories between age 15 and 40. Women who have had similar family life experiences are expected to cluster into the same group. After performing OMA with equal costs assigned to insertion and deletion (in this instance 1), individuals are allocated to clusters based on Ward's distance. We assess the number of clusters based on two measures of average cluster linkage; the Calinski–Harabasz pseudo-F index (Calinski & Harabasz, 1974) and the Duda–Hart index (Duda & Hart, 1973). These statistics help us determine the optimal number of clusters by identifying the number of clusters which are the most distinct according to the distance matrix. Once the optimal number of clusters is established, cluster allocation is used as a response variable in a multinomial logistic regression. The models are estimated using the SQ-Ados ado for Stata 12 (Brzinsky-Fay, Kohler, & Luniak, 2006).

Then, the analysis is repeated using LCGM. Latent class growth models extract a number of classes of partnership behaviour. The number of classes is decided using a variety of fit statistics, including AIC, BIC and Sample-Size adjusted BIC. A set of 2, 3, 4, and 5 class models are explored and, for all classes, the Lo-Mendell-Rubin-Likelihood Ratio Test (LMR-LRT) is performed. This test examines the improvement in model fit for a J class model compared to a J-1 class model. In case of a 2 class model, this test is equivalent to examining whether the Latent Class Growth Model performs better than a simple Latent Growth Model, which assumes that one growth curve is enough to describe women's partnership behaviours. We do not include the analogous Bootstrap Likelihood Ratio test due to excessive computational demands. The models are estimated in Mplus 6.2 for Linux, via the iridis-3 cluster computer provided by the University of Southampton. Note that we do not explore models with more than 5 classes. Due to the specification of partnership state as a nominal variable, the implementation of this model is not part of the main Mplus language. As a result, model estimation is computationally intense due to both the difficulty of the calculations required and the volume of data to be read (the datafile needs to be expanded to person-period format). Classes are formed from yearly partnership histories and include education as a predictor of class membership as well as a covariate that can alter the partnership trajectories. This is important as a significant effect of education on the growth trajectory can be regarded as critical evidence of the importance of education in the model and ignoring this association can distort the relationship between the observed variables and class (Jung & Wickrama, 2008). To ensure convergence, the individual level variance is specified at zero around each growth curve (some classes have zero probabilities across the life course for some partnership states).

Last, we examine the influence of education on all examined partnership transitions using multistate event history analysis. The model is estimated as a continuous-time stratified Cox regression where each transition represents a stratum. To estimate this model, an augmented dataset needs to be used with one row per transition that the individual is at risk for. Women are observed from age 15, when they are never partnered until age 40, the time of the survey or the time when they experience re-partnering, whichever happens earlier (N = 7,704). As educational attainment is defined as a time-varying categorical variable, additional episode splitting is performed where an educational transition happened within an at-risk period. The models are estimated using the *mstate* package in R (de Wreede et al., 2011).

These models allow us to estimate the influence of education on each transition within the same model and to compare the influence of education across transitions.

Results

Sequence Analysis

Table 1 presents the Calinski–Harabasz and the Duda–Hart indices for 2 to 6 cluster models. On the Calinski–Harabasz and Duda–Hart indices, higher values indicate more distinct clustering, whereas for the related Duda–Hart Pseudo T-square measure, lower values are indicative of more distinct grouping. There is disagreement between these indices as to the optimal number of clusters; the Calinski–Harabasz index indicates a 3 cluster solution to be optimal, while the Duda-Hart indices indicate that both a 3 and a 4 cluster solution would be plausible. As both sets of indices show that a 3 cluster solution is plausible, we proceed with a 3 cluster model.

Table 1. Calinski–Harabasz and Duda–Hart Indices for k Cluster Specifications.

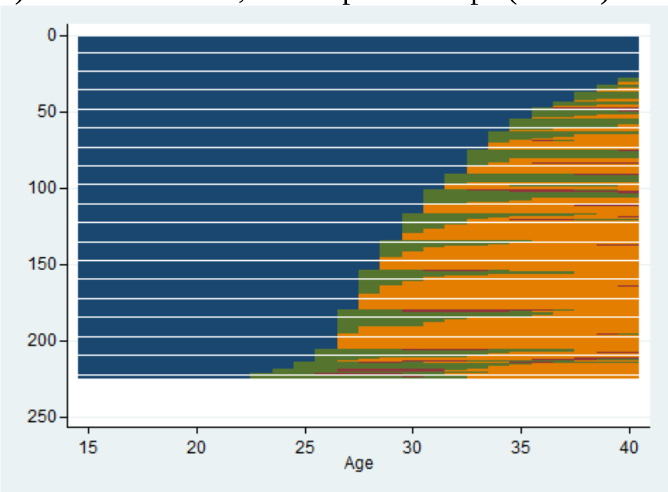
Number of Clusters (k)	Calinski–Harabasz Pseudo-F	Duda–Hart indices	
		Je(2)/Je(1)	Pseudo T-square
2	97.80	0.493	311.78
3	**210.60**	*0.9719*	*5.81*
4	144.42	**0.9921**	**1.07**
5	108.67	0.742	80.86
6	116.24	0.504	98.56

Note: Numbers in boldface indicate the best fit for the given index. Numbers in italics indicate additional plausible values.

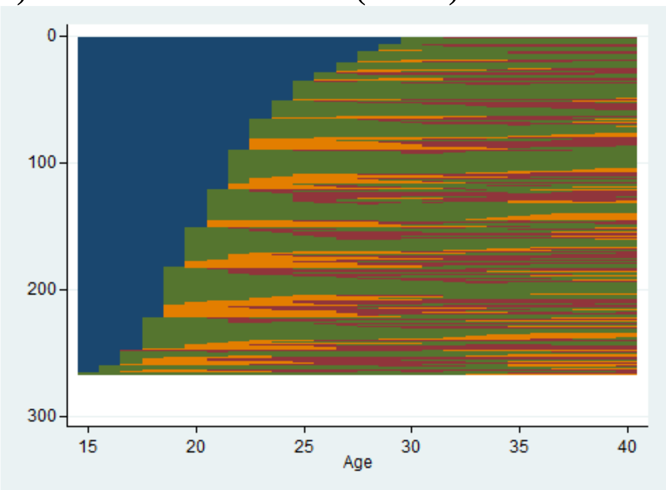
Figure 3 depicts the results of cluster analysis. The first cluster (Figure 3, panel a) is characterised by relatively late partnership formation, where the first partnership is typically cohabitation most of which translates into marriage and only some ends with union dissolution. Additionally, some women enter marriage directly. Therefore, this cluster is titled '*late, varied partnerships*'. Women who belong to the second cluster form first partnerships at a relatively young age (Figure 3, panel b). Most of these partnerships are long term cohabitation with relatively high union instability. Therefore, this group is referred to as the '*cohabitation'* cluster. The third cluster (Figure 3, panel c) is mainly characterised by early and direct marriage. Unions which start as cohabiting partnerships later translate into marriage, and most of these partnerships are stable. This cluster is, thus, named the '*(direct) marriage*' cluster.

Figure 3. Results of Sequence Analysis.

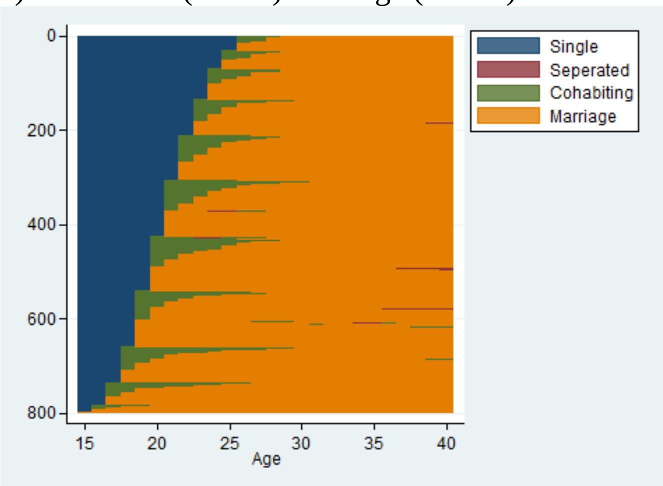
a) Cluster 1: Late, varied partnerships (17.4%)



b) Cluster 2: Cohabitation (20.7%)



c) Cluster 3: (Direct) marriage (61.9%)



After having identified these three clusters, we apply multinomial logistic regression to assess how educational attainment influences the odds of women to belong to one of these three clusters (Table 2). To facilitate the interpretation of the relative risk ratios, predicted probabilities are calculated (Figure 4). The results show that more educated women have a higher probability to belong to the first cluster (late and varied partnerships) than lower educated women. Moreover, low educated women are more likely to belong to the cohabitation cluster (cluster 2) than medium or high educated women. Finally, there are no significant differences by education in the probability of belonging to the direct marriage cluster (cluster 3).

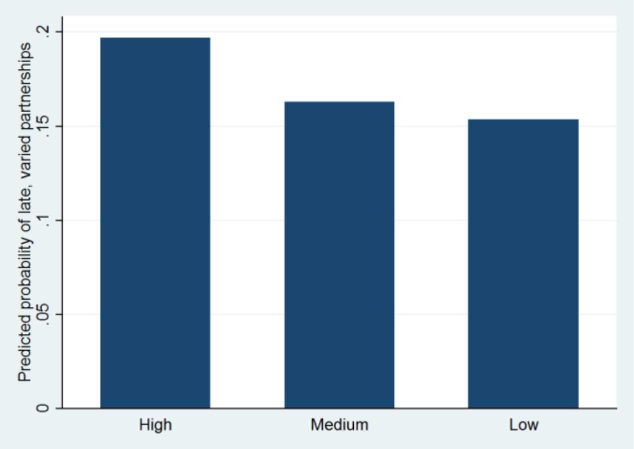
Table 2. Results of the Multinomial Logistic Regression, Regression Coefficients (with Standard Errors).

	Membership of cluster 1 vs cluster 3		Membership of cluster 2 vs cluster 3	
	Coef.	S.E.	Coef.	S.E.
Education				
High (ref)				
Medium	-0.264	0.041	0.149	0.037
Low	-0.212	0.033	0.085	0.032
Intercept	-1.133		-1.165	

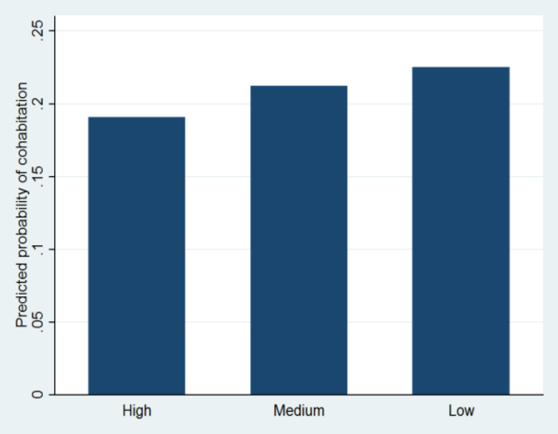
Note: **p* < .05. ***p* < .01. ****p < .001 . p<0.1*

Figure 4. Predicted Probabilities of Cluster Membership by Educational Level.

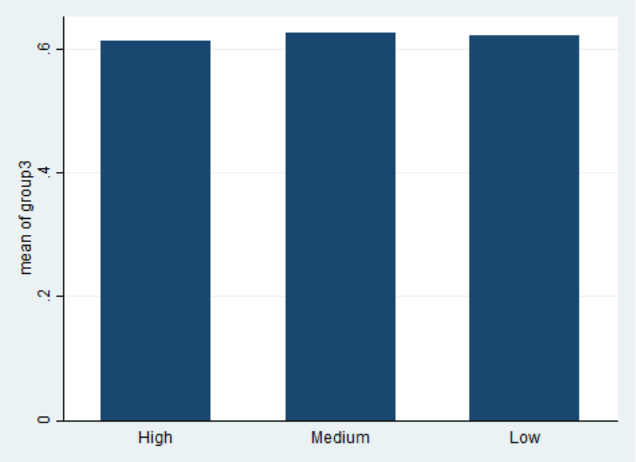
a) Cluster 1: Late, varied partnerships



b) Cluster 2: Cohabitation



c) Cluster 3: (Direct) marriage



Note: Change of scale in case of cluster 3 for visual clarity.

Latent Class Growth Models

Table 3 presents the fit statistics for 2, 3, 4 and 5 class models. The LMR-LRT p-value indicates that the 2 class model is an improvement over a 1 class model, justifying the LCGM approach. All fit statistics indicate improving model fit with the addition of higher order classes. From the examined models, the 5 class model demonstrated the best model fit based on AIC, BIC and Sample Size BIC (SSBIC) statistics. We note that the LMR-LRT indicates that a 4 class model is adequate, but select a 5 class model since this is the optimal number of classes for a greater number of fit statistics.

Table 3. Fit Statistics for 2, 3, 4 and 5 Class Models.

Number of classes (J)	AIC	BIC	SSBIC	LMR-LRT (p-value)
2	138352.929	138731.851	138588.841	0.0000
3	132500.352	133081.366	132862.085	0.0159
4	129273.584	130056.690	129761.137	**0.0210**
5	**126725.499**	**127710.697**	**127338.871**	0.1736

Note: Numbers in boldface indicate the best fit based on the given statistic.

The extracted classes are presented in Figure 5 for highly educated women. The classes are similar for low and medium educated women and the educational differences between the classes will be discussed after the description of each class. Class 1 captures early and varied partnership forms, with a rise in the probability of both cohabitation and marriage. The probability of a marriage peaks around the age of 28, and declines thereafter. The probability of cohabitation rises, plateauing at age 22, before increasing again from age 31 onwards.

Class 2 is characterised by early cohabitation, which translates to marriage only in later ages. In this class, the probability of cohabitation increases peaking at 90% at age 29. Thereafter, cohabitation is translated to marriage; the probability of which reaches 30% by age 40. There is some evidence of separation, but this consistently remains around or below 10%.

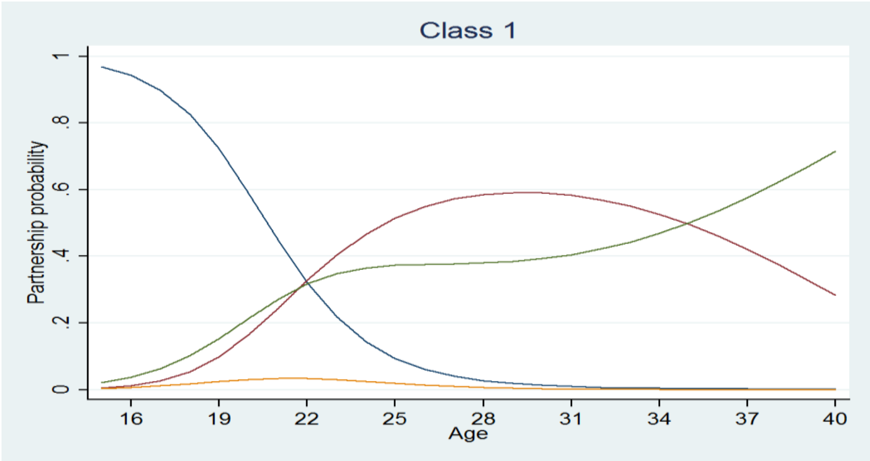
Class 3 broadly follows a traditional pattern: early and direct marriage. The probability of marriage increases rapidly and by age 25 all women are already married in this class. It is important to note that in this class, marriage is preceded by cohabitation for some women, as indicated by the small peak in cohabitation around age 20. The unions formed are, again, stable, with virtually no separation.

Class 4 represents the most ‘modern’ partnership form. There is a considerably high incidence of cohabitation before marriage, with a peak at age 25, when the probability of cohabiting is roughly 50%. Thereafter, many unions are translated into marriage, the probability of which peaks at age 31. There is some evidence of union dissolution in this class, with the probability of separation amounting to as much as 5%.

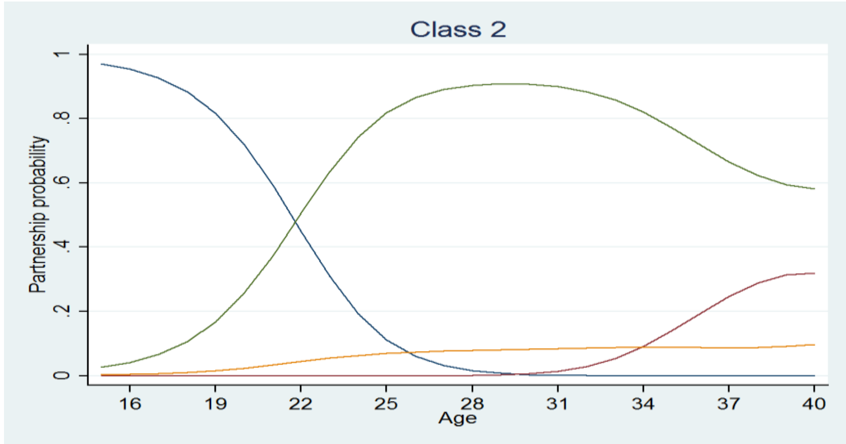
Class 5 captures a more complex pattern of late partnerships. The probability of being single does not decline until after age 25 and it never falls below 20%. After age 25, union forms are varied; both the probability of cohabitation and marriage rise to around 40% at ages 32 and 37, respectively. Finally, there is some incidence of union instability in this class at later ages.

Figure 5. Results of the 5 Class Latent Class Growth Models for Highly Educated Women.

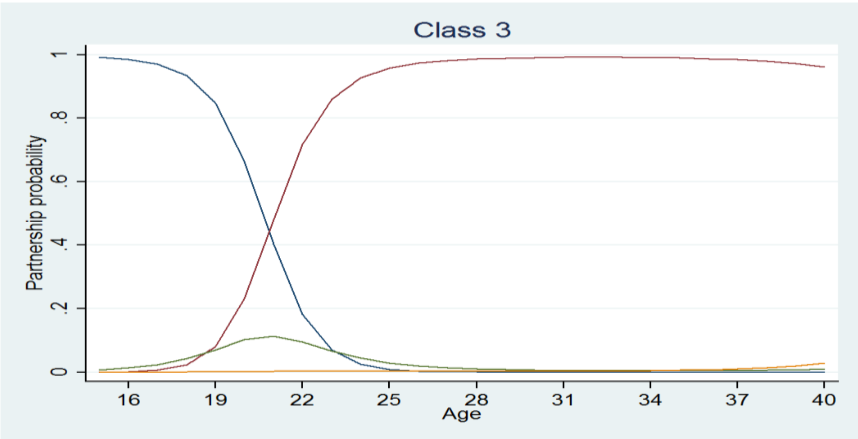
a) Class 1: Early, varied partnerships



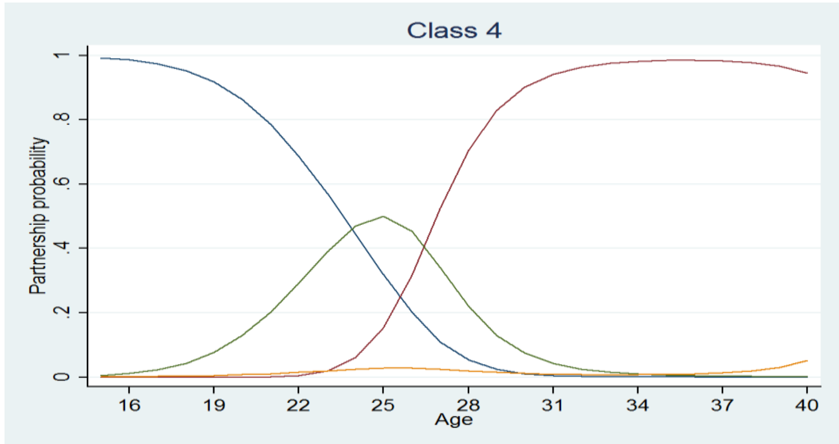
b) Early cohabitation with late translation to marriage



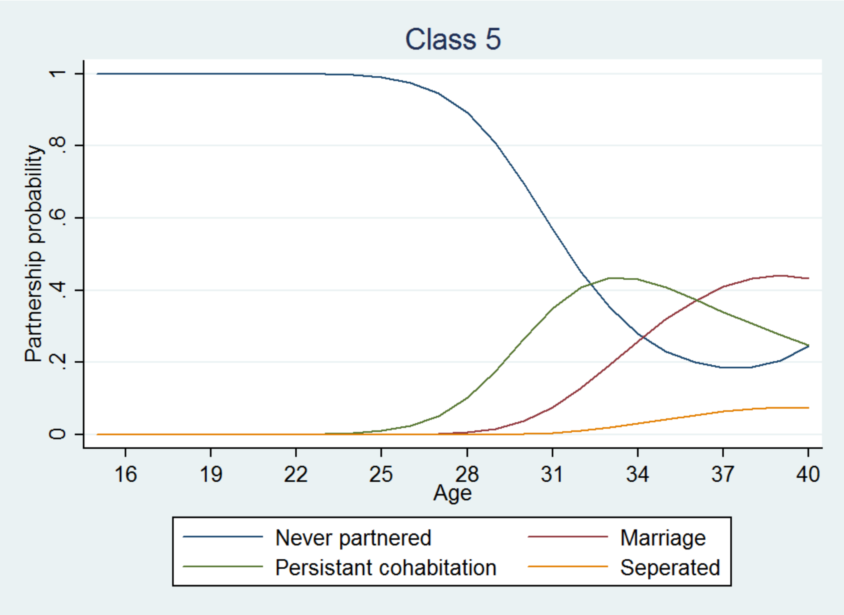
c) Slightly postponed, direct marriage



d) Marriage preceded by cohabitation



e) Class 5: Late and heterogeneous partnership forms



The predicted probabilities of class membership are presented in Table 4. This table indicates that the modal class for women with low education is class 3 (slightly postponed, direct marriage). This class is dominant, with no other class having a probability of above 0.15. Among women with medium and higher education, class 1 (early, varied partnerships) is the modal class, although the probability of membership of class 4 (marriage preceded by cohabitation) is also larger than that of the other classes. This indicates generally more complex relationship patterns among women with higher educational attainment, with some evidence of partnership postponement.

Table 4. Predicted Probability of Class Membership by Educational Level.

Class number	Educational level		
	Low	Medium	High
1	0.11	0.34	0.39
2	0.15	0.12	0.12
3	0.52	0.15	0.11
4	0.13	0.28	0.29
5	0.09	0.11	0.09

As mentioned above, the graphs on Figure 5 only depict the partnership trajectories of highly educated women. To examine how these trajectories differ among medium and low educated women, Table 5 presents the effect of education on the estimated curves by educational level and partnership state. Significant coefficients are taken as evidence of an influence of education on the timing and/or probability of a partnership behaviour within a given class. The results indicate that education does not significantly influence the probability of cohabitation and separation in Class 1. This means that if Figure 5 was reproduced for low and medium educated women, the lines representing the probability of cohabitation and separation in Class 1 would not be significantly different from those depicted for highly educated women. Additionally, education has a moderate influence on the probability of cohabitation and separation in Class 3 and Class 5.

Table 5. The Influence of Education on Partnership States by Class.

			Partnership State (ref=Marriage)					
			Single		Separated		Cohabiting	
			Intercept	Slope	Intercept	Slope	Intercept	Slope
Class	Early, varied partnership	Medium	-0.150	0.022***	-0.332	-0.034	-0.064	0.000
		Low	10.870***	-0.754***	-11.439*	1.082	3.751	-0.269
	Early cohabitation with late translation to marriage	Medium	-0.200	-0.049***	-0.368	-0.067***	0.257	-0.055***
		Low	-15.809***	1.800***	16.386***	2.163***	-14.908***	1.881***
	Slightly postponed, direct marriage	Medium	-1.456***	0.007	-0.057	-0.022*	-0.696*	-0.007
		Low	-2.173***	0.370***	2.656	0.542*	0.387	0.406***
	Marriage preceded by cohabitation	Medium	-1.500***	-0.027*	-1.439***	-0.225***	-0.946***	-0.025***
		Low	-3.845***	0.535***	-9.381***	2.970***	-2.994***	0.486***
	Late and heterogeneous partnership forms	Medium	0.193	-0.025***	0.014	0.042***	-0.356	0.008
		Low	-19.990***	0.938***	1.176	-0.955*	-9.815***	0.129

Multistate Event History Model

The results of the multistate event history model are summarised in Table 6. The findings indicate that higher educated never partnered women born between 1955 and 1964 have a higher risk of entering direct marriage than medium and low educated women when controlling for educational enrolment. Furthermore, education has a positive gradient on the transition from cohabitation to marriage; low and medium educated cohabiting women are about 60% less likely than their highly educated counterparts to marry their cohabiting partner. Finally, education has a positive gradient on the risk of re-partnering following union dissolution; low educated women have an almost 70% lower risk while medium educated women have a 54% lower risk of finding a new partner after union dissolution than highly educated women. Education does not have a significant influence on the transition to a first cohabitation, on the transition from cohabitation to union dissolution, and on the dissolution of a marital union, whether or not it was preceded by cohabitation. Additional analyses revealed that the differences in the transition risks of low and medium educated women were not significant after controlling for educational enrolment (results not shown).

Table 6. Result of the Multistate Event History Model, Hazard Ratios.

	S → C	S → M	C → CM	C → D	M → D	CM → D	D → R
Education							
Low	1.05	0.64*	0.39***	0.79	1.21	1.49	0.31***
Medium	0.92	0.68*	0.38***	1.01	0.81	1.33	0.46***
High (ref)							
Enrolment							
No (ref)							
Yes	0.64***	0.53***	0.55***	1.53**	1.71	1.14	0.97

Note: $*p < .05$. $**p < .01$. $***p < .001$

When comparing the influence of education across the different transitions, we find that its influence is the strongest in the transition from cohabitation to marriage, followed by its influence on the risks of re-partnering, and finally on direct marriage.

Conclusion and Discussion

This paper aimed to compare several methodological approaches (i.e. sequence analysis, latent class growth models, and multistate event history models) to the analysis of life course data focusing on the influence of education on partnership experiences with an application to Norwegian women born between 1955 and 1964. These methods have several similarities and differences. For example, sequence analysis and latent class growth models establish the relationship between education and the probability of belonging to certain groups (clusters or classes) based on women's partnership experiences. In our application, sequence analysis revealed three clusters based on women's partnership experiences (late, varied partnerships; cohabitation; and (direct) marriage), latent class growth models suggested the existence of five partnership classes (early, varied partnerships; early cohabitation with late translation to marriage; slightly postponed marriage; marriage preceded by cohabitation; late, heterogeneous partnerships). Multistate event history models do not classify individuals but rather examine the influence of education on every partnership transition thereby enabling us to draw conclusions about the changing influence of education over the early family life course.

As these models have different properties and approach studying the life course in a different way, it is not easy to directly compare their findings. However, by comparing the properties and results of the different techniques, we are able to make comparisons between methods with respect to their ability to address certain desirable aspects of the family life course. These are summarised in Table 7.

Table 7. Summary of the Properties of Sequence Analysis, Latent Class Growth Models, and Multistate Event History Analysis.

	SA	LCGM	Multistate Event History model
Transition intensities	(✓)	✘	✓
Classifying individuals	✓	✓	✘
Covariate information alters pattern	✘	✓	✓
Heterogeneous effect of covariates	✓	✓	✘
Computationally simple	✓	✘	✓
Changing covariate effect over the LC	✘	✘	✓
Model based	✘	✓	✓
Protection against baseline misspecification	✓	✘	✓

Note: The given method is ✓ able to, ✘ not able to or (✓) partially able to deal with this dimension of the family life course.

First, sequence analysis is best applied to research questions which attempt to describe partnership behaviours of different groups of women and the overall associations of these groups with certain covariates. This can be achieved through the method's ability to classify individuals and allow for covariates to predict women's membership in the different clusters. Overall, fitting the model does not require a lot of computing power and due the fact that the procedure is not model based, the user is protected against baseline misspecification (i.e. no baseline needs to be specified). Although not presented in this paper, the method can also calculate transition intensities between the different states. As it is not possible to condition sequences on covariate information or to allow for the incorporation of changing covariate information over the life course, this method cannot answer research questions relating to the changing influence of a variable.

Second, latent class growth models have a number of similar properties to sequence analysis. Its main advantage compared to sequence analysis is that it is able to incorporate more complicated structures by, for example, allowing for covariate information to alter the partnership trajectories. Unfortunately, the implementation of LCGMs is computationally intense and requires considerable computing power to estimate models for large datasets. Moreover, the fact that LCGMs are model based implies that a greater degree of robustness check is required particularly when estimating the shape of growth curves. On the other hand, this also means that a greater variety of fit-statistics is available than in sequence analysis, where the decision of the optimal number of clusters is more arbitrary than in LCGMs. Thus, LCGMs are most suited to studying complex research topics where the aim is to identify differences in covariate effects between groups of individuals. The present paper has demonstrated this by extracting different classes of partnership behaviour and comparing the effect of educational attainment within these classes.

Finally, although multistate event history models do not classify individuals in the same way as the previous two methods, there are a number of distinct advantages to using this method. For example, the estimation of transition intensities allows for examining several transitions over the life course within the same model as well as for estimating the changing influence of covariates over the life course by allowing for the incorporation of time-varying covariates. Neither sequence analysis, nor latent class growth models are capable of studying changing covariate effects over the life course. Additionally, the use of a stratified Cox model provides some protection against baseline misspecification. To sum up, multistate event history models can best answer research questions related specifically to changing covariate effects over the life course. For example, as this paper has shown, it can estimate the changing influence of education on the different partnership transitions over the early family life course.

References

Abbott, A. (1995). Sequence analysis: New methods for old ideas. *Annual Review of Sociology, 21*, 93-113. doi: 10.1146/annurev.soc.21.1.93

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

Andersen, P. K., & Keiding, N. (2002). Multi-state models for event history analysis. *Statistical Methods in Medical Research, 11*(2), 91-115. doi: 10.1191/0962280202SM276ra

Baizán, P., Aassve, A., & Billari, F. C. (2003). Cohabitation, marriage, and first birth: The interrelationship of family formation events in Spain. *European Journal of Population, 19*, 147-169.

Baizán, P., Aassve, A., & Billari, F. C. (2004). The interrelations between cohabitation, marriage and first birth in Germany and Sweden. *Population and Environment, 25*(531-560). doi: 10.1023/B:POEN.0000039064.65655.3b

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., 61*, 765-784.

Billari, F. C. (2001a). The analysis of early life courses: Complex descriptions of the transition to adulthood. *Journal of Population Research, 18*(2), 119-142. doi: 10.1007/bf03031885

Billari, F. C. (2001b). Sequence analysis in demographic research. *Canadian Studies in Population, 28*(2), 439-458.

Billari, F. C. (2005). Life course analysis: Two (complementary) cultures? Some reflections with examples from the analysis of the transition to adulthood. *Advances in Life Course Research, 10*(0), 261-281. doi: 10.1016/s1040-2608(05)10010-0

Billari, F. C., & Piccarreta, R. (2001). Life courses as sequences: an experiment in classification via monothetic divisive algorithms. In S. Borra, R. Rocchi, M. Vichi & M. Schader (Eds.), *Advances in Classification and Data Analysis* (pp. 351-358). Berlin and New York: Springer Verlag.

Billari, F. C., & Piccarreta, R. (2005). Analyzing demographic life courses through sequence analysis. *Mathematical Population Studies, 12*(2), 81-106. doi: 10.1080/08898480590932287

Brzinsky-Fay, C., & Kohler, U. (2010). New developments in sequence analysis. *Sociological Methods & Research, 38*(3), 359-364. doi: 10.1177/0049124110363371

Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *The Stata Journal, 6*(4), 435-460.

de Wreede, L. C., Fiocco, M., & Putter, H. (2011). mstate: An R package for the analysis of competing risks and multi-state models. *Journal of Statistical Software, 38*(7), 1-30.

Heuveline, P., & Timberlake, J. M. (2004). The role of cohabitation in family formation: The United States in comparative perspective. *Journal of Marriage and Family, 66*, 1214-1230.

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

Jung, T., & Wickrama, K. A. S. (2008). An introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Psychology Compass, 2*(1), 302-317. doi: 10.1111/j.1751-9004.2007.00054.x

Mikolai, J. (2013). *The role of education in the partnership and family formation process in Europe and the United States*. Paper presented at the Changing Families and Fertility Conference, Oslo, Norway.

Ní Bhrolcháin, M., & Beaujouan, E. (2012). Fertility postponement is largely due to rising educational enrolment. *Population Studies, 66*(3), 311-327. doi: doi:10.1080/00324728.2012.697569

Nylund, K. L., Asparouhav, T., & Muthen, B.O. . (2007). Deciding on the number of classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(4), 535-569.

Perelli-Harris, B., & Gerber, T. P. (2011). Nonmarital childbearing in Russia: Second Demographic Transition or Pattern of Disadvantage? *Demography, 48*(1), 317-342. doi: 10.1007/s13524-010-0001-4

Perelli-Harris, B., Kreyenfeld, M., & Kubisch, K. (2010). Harmonized Histories: Manual for the preparation of comparative fertility and union histories. *Rostock, MPIDR Working Paper WP-2010-011.*

Perelli-Harris, B., & Lyons-Amos, M. (2013). The heterogeneity of relationship patterns: An examination of 15 countries using latent class growth models. *CPC Working Paper, 37*.

Perelli-Harris, B., Sigle-Rushton, W., Kreyenfeld, M., Lappegård, T., Keizer, R., & Berghammer, C. (2010). The educational gradient of childbearing within cohabitation in Europe. *Population and Development Review, 36*(4), 775-801. doi: 10.1111/j.1728-4457.2010.00357.x

Piccarreta, R., & Billari, F. C. (2007). Clustering work and family trajectories using a divisive algorithm. *Journal of Royal Statistical Society, A, 170*(4), 1061-1078. doi: 10.1111/j.1467-985X.2007.00495.x

Putter, H., Fiocco, M., & Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models. *Statistics in Medicine, 26*(11), 2389-2430. doi: 10.1002/sim.2712

Putter, H., van der Hage, J., de Bock, G. H., Elgalta, Ra., & van de Velde, C. J. H. (2006). Estimation and prediction in a multi-state model for breast cancer. *Biometrical Journal, 48*(3), 366-380. doi: 10.1002/bimj.200510218

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

Wang, C. P., Hendricks Brown, C., & Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: examining the impact of a preventive intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association, 100*, 1054-1076.