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

Julia Mikolai[1] and Mark J. Lyons-Amos, *University of Southampton*

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 Please do not cite without the permission of the authors!

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

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. While event history analysis is commonly applied, this is not always suitable, and more holistic approaches such as sequence analysis have been proposed as alternatives. As research tends to be interested in explaining more complexity in the family life course, it is necessary to extend our methodological toolkit by increasing the complexity of event history models (multistate event history models) or applying other promising methods, such as sequence analysis or latent class growth models. The aim of this paper is to compare and contrast sequence analysis, latent class growth models, and multistate 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 for women in birth cohorts 1945-1954, 1955-1964, and 1965-1974, we model changes in partnership status across the life course, with education as the primary covariate of interest.

Introduction

In the last half century, patterns of family life courses have changed considerably. For example, the transition to parenthood is being 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 from researchers to study the life course, and especially, to find appropriate methods for modelling life courses with their complexities. Although life course theory emphasises that events that happen earlier in the life course influence later life course events, methods commonly used in life course research do not manage to completely account for this path-dependency.

Event history analysis is commonly used to examine single or multiple (competing) events (Heuveline & Timberlake, 2004; Perelli-Harris & Gerber, 2011; Perelli-Harris 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 causal relationships between certain covariates and 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. As these models do not incorporate information on the timing, sequencing, and/or duration of previous family life events, they fail to account for path-dependency and to reveal the dynamics of the family formation process.

Others have promoted the use of sequence analysis arguing that unlike event history models, this approach can examine the life course trajectories 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”. Although sequence analysis studies the family life course as a whole, it does not allow us to examine how the sequences of previous events influence the risk of a later event.

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, 2011; Billari, 2001a, 2005; Billari & Piccarreta, 2005; Piccarreta & Billari, 2007) with the exception of Barban and Billari (2011) who compared sequence analysis and latent class growth models. 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 other words, unlike simple event history analysis and sequence analysis, multistate models and latent class growth models incorporate information on the timing, sequencing, and duration of earlier family life events and as such, account for path-dependency when making predictions about events later in the family life course.

The aim of this paper is twofold. First, we compare and contrast sequence analysis, latent class growth models and multistate event history models. Second, by applying these methods to a real life example, we emphasise the differences and similarities as well as the strengths and weaknesses of these approaches. Our example focuses 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) and the role of education in this process. We aim to tackle the following questions, pertinent to researchers studying the life course: How can sequence analysis, latent class growth models and multistate event history models be used for the analysis of partnership histories? Are these methods applicable to the same problems to the same extent? Is one of these approaches 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 (SA)

Sequence analysis 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 month). For example, the sequence SSSCCMMMM means that the respondent was single (S) for three months, cohabited (C) for two months, and was married (M) for four months. 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 the duration of the 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. Similarly, the more operation is needed, the more dissimilar they are. Furthermore, to each operation, a certain cost can be attached. The distance is then 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. Cluster analysis groups individuals based on the similarity of the individual sequences. One needs to specify the number of clusters to be extracted from the data. 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 latter is not done very often).

Latent Class Growth Models (LCGMs)

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. These models differ from event history models in that they take an individual rather than a variable centred perspective. Event history models are variable centred, since they seek to establish statistical relationships between dependent and independent variables. In contrast, LCGMs – and also sequence analysis – seek to identify relationships between individual response patterns and form groups based on these (Jung & Wickrama, 2008). Figure 1 presents the conceptual Latent Class Growth Model. In this figure, 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 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 covariates can be used 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, as opposed to SA, is that a variety of fit statistics are available for deciding the optimal number of classes. However, the different criteria and test statistics (such as AIC, BIC or Lo-Mendell-Rubin Likelihood Ratio Test) can produce different results (Nylund, Asparouhav, & Muthen, 2007).

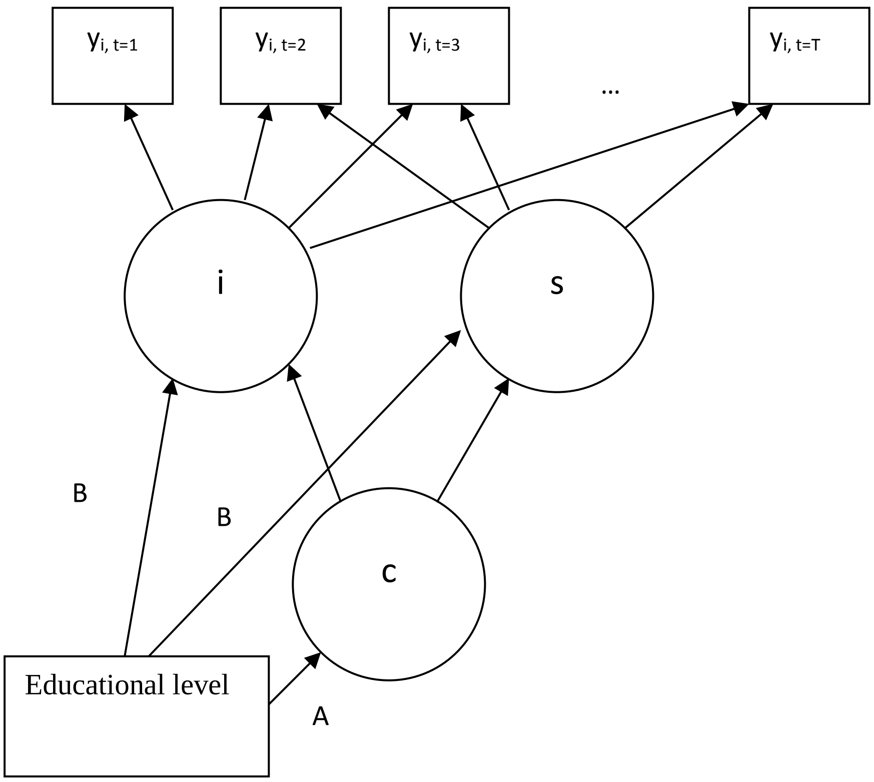


Figure 1. Conceptual representation of LCGM with covariates altering the growth trajectories.

Multistate Event History Models

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. This cannot be done by simple event history models or by sequence analysis. 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 birth 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 repartnering (R).

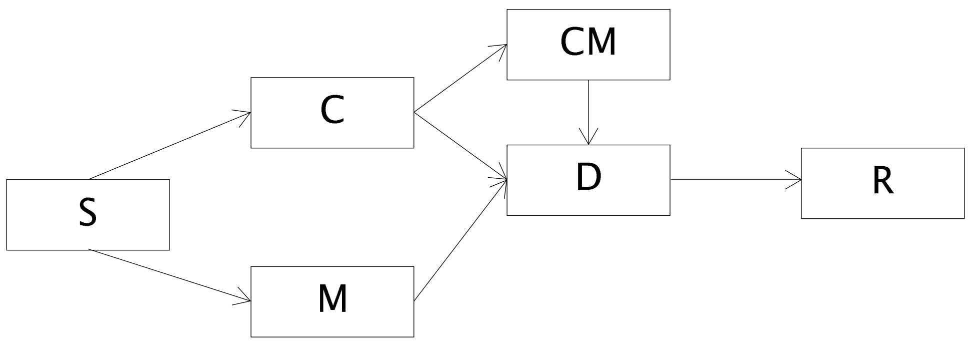


Figure 2. Multistate event history model.

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, one might end up with small cell sizes and thus, or unreliable estimates of the transition hazards.

Data

To illustrate the similarities and differences between sequence analysis, latent class growth models and simple and multistate event history models, we analyse a real-life example. Using data from the first wave of the Norwegian Generations and Gender Survey from 2007/2008 (N = 14,881), we examine the influence of educational attainment on changes in partnership status of women born between 1945 and 1974.

Models

First, using sequence analysis, we create several groups 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 the same cost assigned to insertion and deletion, we use the K-means cluster algorithm and Ward's distance to cluster the partnership sequences. The number of clusters is decided based on the difference in Log-Likelihood statistics when comparing models with different number of clusters, where the cluster is the response in an intercept only multinomial regression. Once the optimal number of clusters is established, the multinomial model is extended to incorporate the influence of education controlling for birth cohort. 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. We explore a set of 2, 3, 4 and 5 class models and also perform the Lo-Mendell-Rubin-Likelihood Ratio Test for all classes. This test examines the improvement in model fit for a J class model compared to a J-1 class model. Note that we do not explore higher order classes; due to the specification of partnership state as a nominal variable, the implementation of the model is not at present part of the main Mplus language. As a result, model estimation is computationally intensive. In case of a 2 class model, this test is equivalent to examining whether the Latent Class Growth model is performing better than a Latent Growth model. Classes are formed from yearly partnership histories, but include education as a predictor of class membership as well as a covariate that alter the partnership trajectories. To ensure convergence, we restrict the individual level variance around each growth curve to zero. The models are estimated in Mplus 6.2 for Linux, via the iridis-3 cluster computer provided by the University of Southampton.

Last, we examine the influence of education on all examined partnership histories using multistate event history analysis. The model is estimated by a stratified Cox regression where each transition represents a stratum. To estimate this model, one needs to use an augmented dataset 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 repartnering, whichever happens earlier (N = 7,988). Educational attainment is defined as a time-varying variable (low, medium (reference), high), so additional episode splitting was performed where an educational transition happened within an at risk period. The analysis is controlled for a time-varying educational enrolment (not enrolled is the reference) and birth cohort (1945-1954 (reference), 1955-1964, 1964-1974). The models are estimated using the mstate package in R. 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 the transitions.

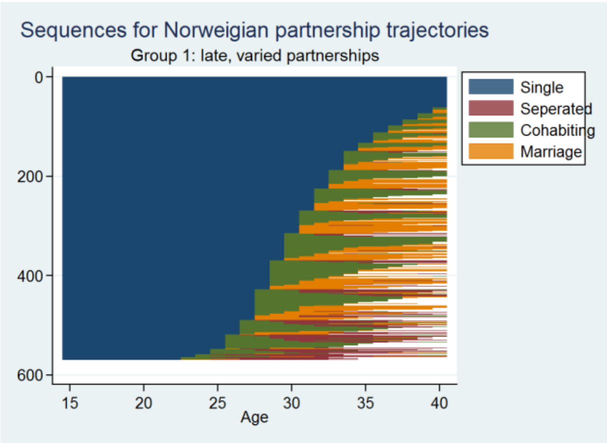
Results

Sequence Analysis

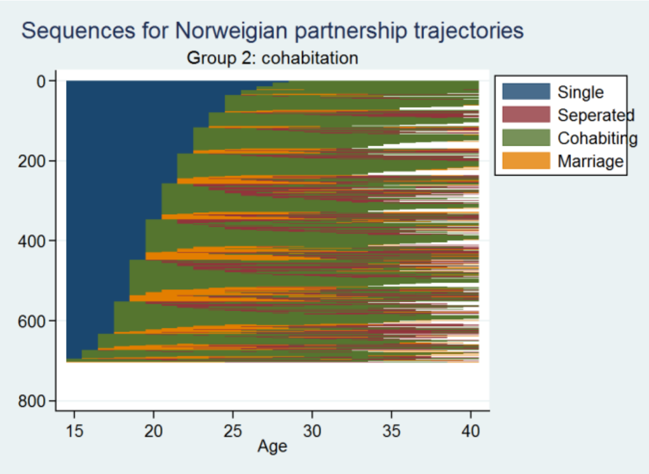
A three cluster solution fitted the data the best (Figure 3). The first cluster (Figure 3, panel a) is characterised by relatively late partnership formation, and the first partnership is typically cohabitation most of which translates into marriage and only some end with union dissolution. We named this cluster '*late, varied partnerships*'. Women who belong to the second cluster typically form a first partnership at a relatively young age (Figure 3, panel b). Most of these partnerships are long term cohabitation with relatively high union instability. Therefore, we refer to this group as the '*cohabitation*' cluster. The third cluster (Figure 3, panel c) is characterised by early and mostly direct marriage. Unions which start as cohabiting unions later translate into marriage, and most of these partnerships are very stable. This cluster is, thus, named the '*marriage*' cluster.

Figure 3. Results of Sequence Analysis.

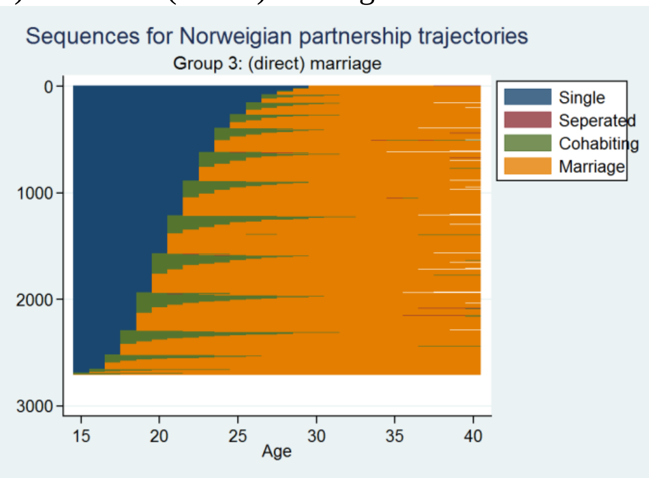
a) Cluster 1: Late, varied partnerships.



b) Cluster 2: Cohabitation



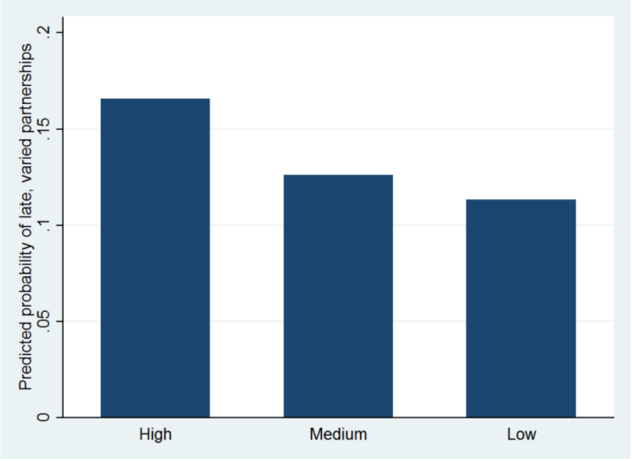
c) Cluster 3: (Direct) marriage



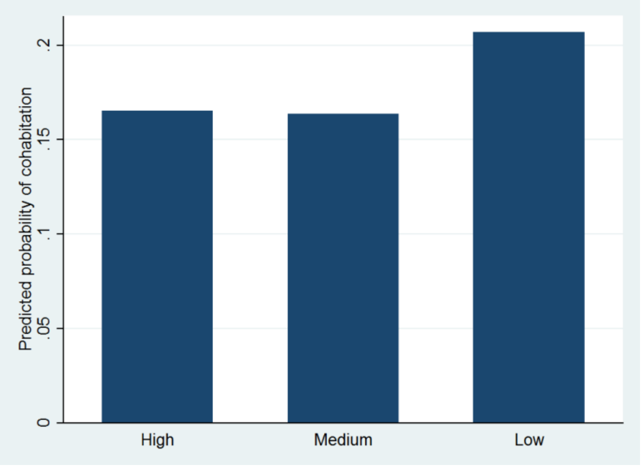
After having identified these three clusters in the data, we apply multinomial logistic regression to assess how educational attainment influences the odds of women to belong to one of these three clusters. The results of the multinomial logistic regression are shown in Table 1. To facilitate the interpretation of the relative risk ratios, we calculated predicted probabilities (Figure 4). The results show that more educated women have a higher probability to belong to the cluster with late and varied partnerships, that low educated women are more likely to belong to the cohabitation cluster than medium or high educated women and that there are no differences by education in the probability of belonging to the direct marriage cluster.

Figure 4. Predicted probabilities of cluster membership by educational level.

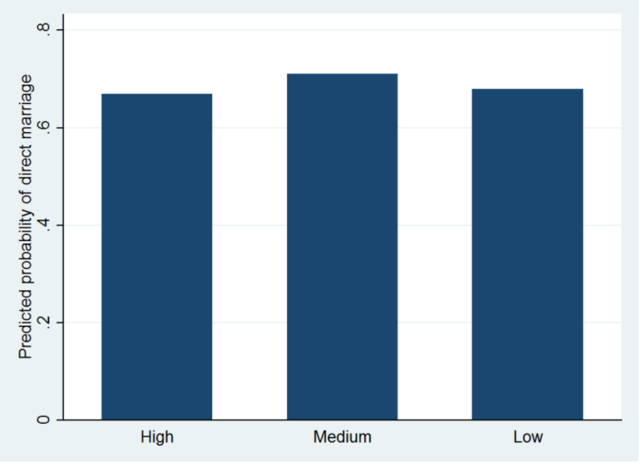
a) Cluster 1: Late, varied partnerships.



b) Cluster 2: Cohabitation



c) Cluster 3: (Direct) marriage



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

Table 1. Results of the Multinomial Logistic Regression, Regression Coefficients (with Standard Errors).

Variable	Membership of cluster 1 vs cluster 3		Membership of cluster 2 vs cluster 3	
	Coef.	S.E.	Coef.	S.E.
Education				
High (ref)				
Medium	-0.232	0.020	0.120	0.019
Low	-0.271	0.289	0.417	0.024
Cohort				
1945-54 (ref)				
1955-64	0.282	0.024	1.134	0.2754
1965-74	0.854	0.023	1.890	0.026

Latent Class Growth Model

From the examined models, the 5 class model demonstrated the best model fit based on AIC, BIC and Sample Size BIC statistics. The significant result for the 2 against 1 class model indicated the appropriateness of the LCGM over the LGM. The extracted classes for highly educated women are presented in Figure 5. The classes are similar for low and medium educated women and the educational differences between these classes will be discussed after the description of each class. Class 1 captures an early and varied partnership form, with a rise in the probability of both cohabiting and married partnership behaviours. The probability of a marriage peaks around the age of 28, and declines thereafter. Cohabitation rises, plateauing at age 22, before increasing again from age 31 onwards.

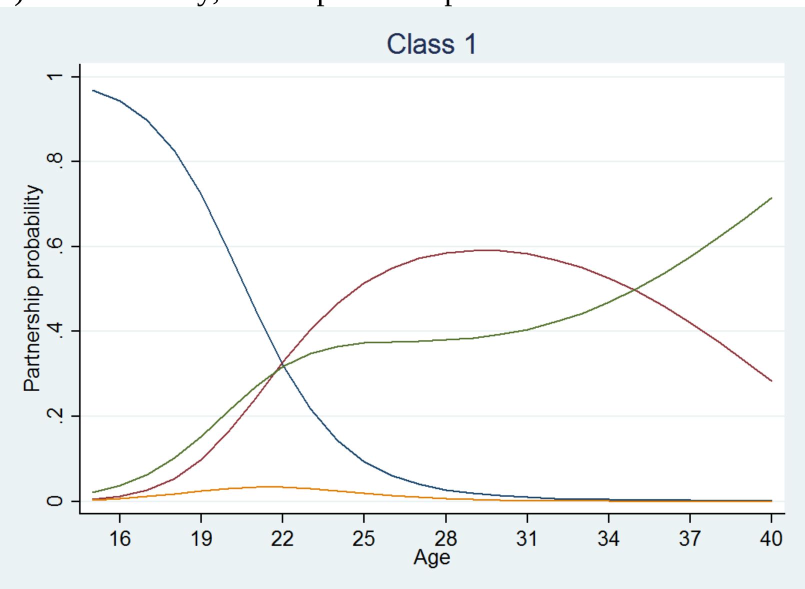
In class 2, the probability of cohabitation increases peaking at 90% at age 29. Thereafter, the probability of cohabitation is substituted for marriage which reached 30% by age 40. There is some evidence of separation, but this consistently remains around 10%. Class 3 follows a broadly traditional; marriage increases rapidly albeit at a slightly later age. In this class, all women were already married around age 25. It is important to note that in this class, marriage is preceded by cohabitation for some women, as indicated by the slight peak in cohabitation around age 20. The unions formed are, again, stable, with no separation.

Class 4 represents the most ‘modern’ marital 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, 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%.

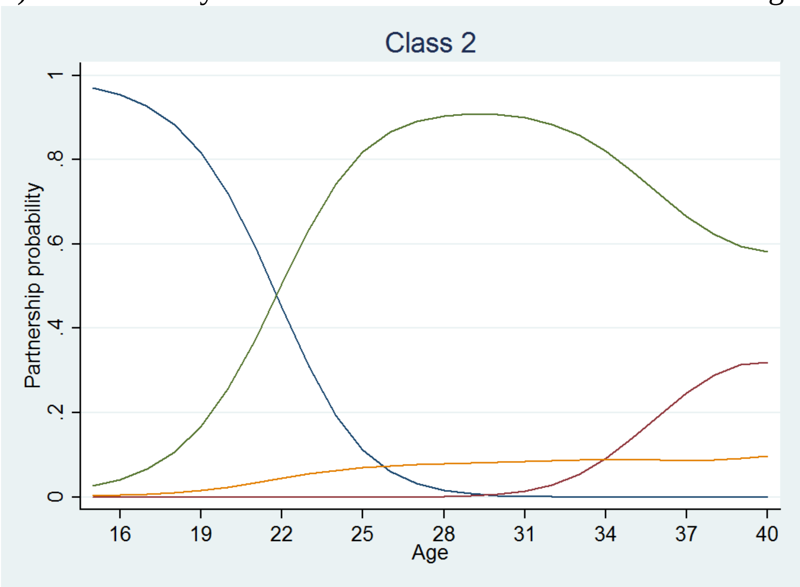
Class 5 captures a complex pattern of late partnerships. Note that the probability of being single does not decline until after age 25. After age 25, although women are likely to form a union, the probability of being single never falls below 20%. In this class, union forms are varied; there is a rise of both the probability of cohabitation and marriage to around 40% at ages32 and 37, respectively. Finally, there is again some incidence of union instability in this class.

Figure 5. Results of the 5 class LCGM for highly educated women.

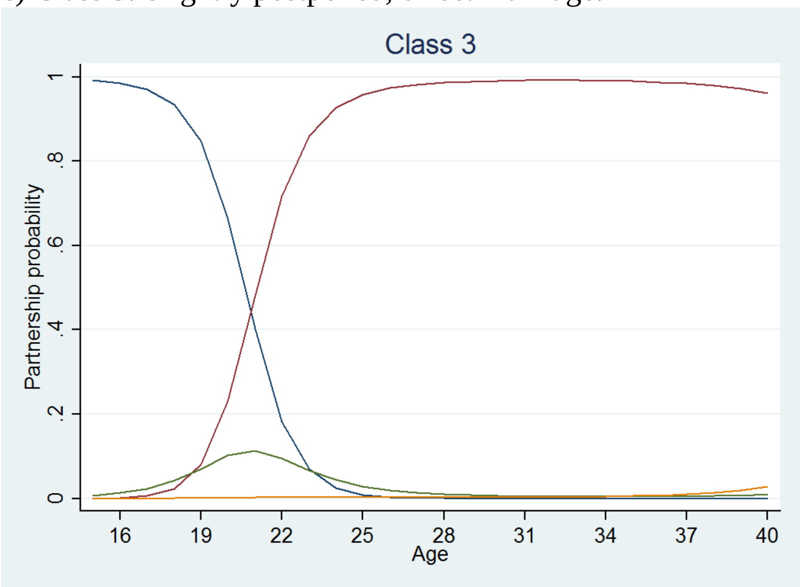
a) Class 1: Early, varied partnerships.



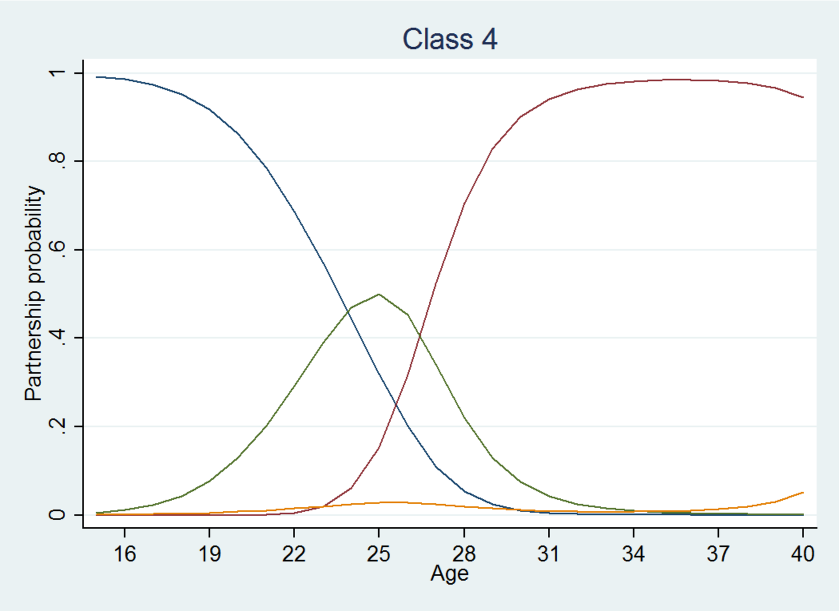
b) Class 2: Early cohabitation with late translation to marriage.



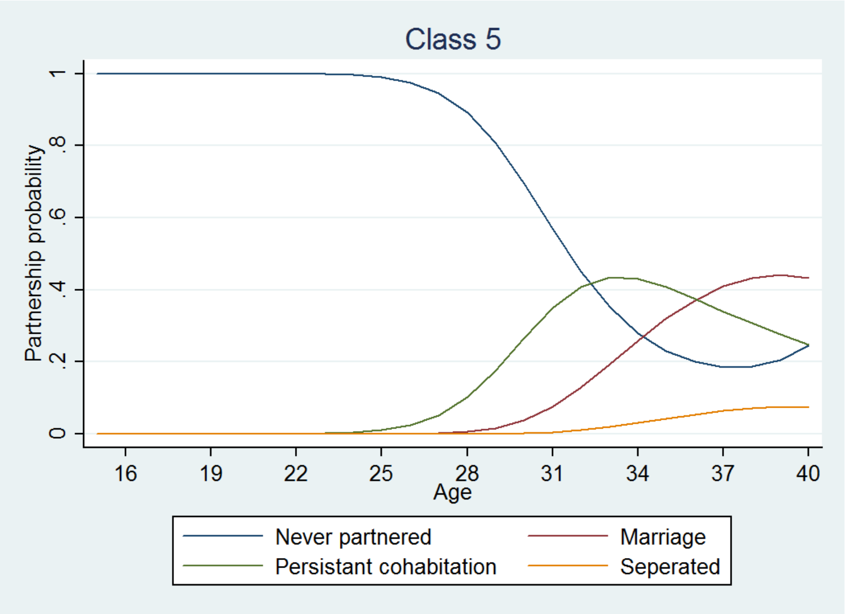
c) Class 3: Slightly postponed, direct marriage.



d) Class 4: Marriage preceded by cohabitation.



e) Class 5: Late and heterogeneous partnership forms.



As mentioned above, these graphs only depict the partnership trajectories of highly educated women. To examine how these trajectories differ among medium and low educated women, Table 2 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 within the timing of partnership within a given partnership form.

Table 2. The influence of education on partnership states by class.

			Partnership State (ref=Marriage)					
			Single		Separated		Cohabiting	
			Intercept	Slope	Intercept	Slope	Intercept	Slope
Class		Medium	-0.150	0.022***	-0.332	-0.034	-0.064	0.000
	Early, varied partnership	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

The intercept indicates the influence of the given educational category on the probability of a given partnership behaviour in a given class. The slope indicates timing differences in the given partnership behaviour in a given class by education.

Multistate Event History Model

The results of the multistate event history model are summarised in Table 3. The results indicate that higher educated women have a 30% higher risk of entering marriage than medium educated women when controlling for educational enrolment and birth cohort. Education has a positive gradient on the transition from cohabitation to marriage; highly educated cohabiting women are almost 1.7 times as likely as their medium educated counterparts to marry their cohabiting partner, while low educated women are 16% less likely to do so. Furthermore, low educated women who directly married their partner are 70% more likely to experience a divorce than medium educated women. Additionally, highly educated women who experienced cohabitation before marriage have a 40% smaller risk than medium educated women to experience a divorce. Finally, education has a positive gradient on the risk of repartnering following a union dissolution: low educated women have a 30% smaller risk of finding a new partner after union dissolution than medium educated women while high educated women have an almost 1.2 higher risk than medium educated women. Last, education does not have a significant influence on the transition to a first cohabitation and on the transition from cohabitation to union dissolution.

Table 3. 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	0.99	1.03	0.84*	0.93	1.70**	1.34	0.70**
medium (ref)							
high	0.92	1.34**	2.68***	0.89	0.86	0.60**	2.19***
Enrolment							
no (ref)							
yes	0.66***	0.53***	0.73***	1.47***	1.66**	0.81	1.19
Cohort							
1945-1954 (ref)							
1955-1964	1.74***	0.41***	0.37***	1.05	1.68**	1.03	1.48*
1965-1974	2.09***	0.18***	0.16***	1.37*	2.26***	1.25	1.96***

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

Conclusion and Discussion

This paper aimed to compare several methodological approaches to the analysis of life course data with a focus on the influence of education on partnership experiences with an application to Norwegian women. 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 4.

	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 the incorporation of changing covariate information over the life course, this method cannot answer more complex research questions.

Second, latent class growth models have a number of similar properties to sequence analysis. Its main advantages 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 are 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 previously mentioned 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 Cox model provides some protection against baseline misspecification. To sum up, multistate event history models can best answer research questions related to 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).

Barban, N., & Billari, F. C. (2011). Classifying life course trajectories: A comparison of latent class and sequence analysis. *Dodena Working Papers, No. 41*.

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. (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.

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.

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., & 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

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.