Predicting the stability of early employment with its timing and childhood social and health-related predictors : a mixture Markov model approach

Helske, Satu

2023-01-15

Helske , S , Keski-Säntti , M , Kivelä , J , Juutinen , A , Kaariala , A , Gissler , M , Merikukka , M & Lallukka , T 2023 , ' Predicting the stability of early employment with its timing and childhood social and health-related predictors : a mixture Markov model approach ' , þÿLongitudinal and Life Course Studies , vol. 14 , no. 1 , pp. 73 104 . ht

http://hdl.handle.net/10138/353726
https://doi.org/10.1332/175795921X16609201864155

Longitudinal and Life Course Studies • vol 14 • no 1 • 73–104
 • Accepted for publication 19 August 2022/First published online 29 September 2022
Online ISSN 1757-9597 • https://doi.org/10.1332/175795921X16609201864155

RESEARCH ARTICLE

Predicting the stability of early employment with its timing and childhood social and health-related predictors: a mixture Markov model approach

Satu Helske, satu.helske@utu.fi
University of Turku, Finland

Markus Keski-Säntti, markus.keski-santti@thl.fi
Juha Kivelä, Juha.Kivela@ppshp.fi
Aapo Juutinen, aapo.juutinen@thl.fi
Antti Kääriälä, antti.kaariala@thl.fi
Finnish Institute for Health and Welfare, Finland

Mika Gissler, mika.gissler@thl.fi
Finnish Institute for Health and Welfare, Finland and
Karolinska Institute, Sweden

Marko Merikukka, marko.merikukka@itla.fi
Itla Children's Foundation, Finland

Tea Lallukka, tea.lallukka@helsinki.fi
University of Helsinki, Finland

To extend work careers, it is important to focus on all working-aged people including young adults. The aim of this study was to identify typical patterns of work participation among young adults after their first entry into the labour market and to examine whether the timing of entry together with parental and own socio-economic position and health predict early work participation. More in-depth understanding of early careers and their early determinants is important to plan targeted interventions and to promote more stable work participation among young adults.

We used the Finnish Birth Cohort 1987 including data from several registers from all 59,476 children born in 1987 as well as their parents, followed until 2015. We estimated a mixture Markov model that allowed for joint identification of latent classes of labour-market attachment, estimation of labour-market transitions within classes, and prediction of class membership using childhood social and health-related determinants.

We observed that the first entry into the labour market as measured by six months in continuous employment was not a permanent entry for many, not only due to negative reasons such as unemployment and ill health but also due to more voluntary reasons such as studies. Individuals entering the labour market at a later age were more likely to be in continuous employment thereafter. More advantaged background predicted exits due to studies or – when following a late entry – stable employment, while disadvantaged background factors predicted more unstable work and long-term exits from the labour market.

Key words labour-market attachment • early careers • life course • sequence analysis • mixture Markov model

Key messages

- Using Finnish register data and mixture Markov models we studied early labour market attachment.
- We examined whether parental and own socioeconomic position and health predicted work participation.
- More advantaged background predicted stable employment or exits due to studies.
- Accumulation of disadvantage predicted unstable work and long-term exits from the labour market.

To cite this article: Helske, S., Keski-Säntti, M., Kivelä, J., Juutinen, A., Kääriälä, A., Gissler, M., Merikukka, M. and Lallukka, T. (2023) Predicting the stability of early employment with its timing and childhood social and health-related predictors: a mixture Markov model approach, *Longitudinal and Life Course Studies*, 14(1): 73–104, DOI: 10.1332/175795921X16609201864155

Introduction

Due to increased life expectancy and fallen birth rates in many countries, the proportion of people in the workforce has dramatically declined. Working-life expectancy has also been increasing but remains relatively short (Nurminen, 2011). Many studies focus on older employees and their risk factors of disability retirement. However, it is equally crucial to gain new information about entry into paid employment and the stability of early working life. For example, it is important to pinpoint risk groups of unstable working life as early as possible to target their determinants.

In young adults' lives, entry into paid employment is an important transition, which can be shaped by contextual factors, such as age at entry. Accordingly, socio-economic background and unhealthy lifestyle have been linked to an early entry into paid employment and to a lower occupational class in the first job (Halonen et al, 2019a). Additionally, early entry into paid employment and physical heaviness of work in young adulthood shape the development of unhealthy behaviours and obesity in the long run (Shiri et al, 2021). Hence, the timing of entry could be shaped by both parental and own socio-economic position and health, further contributing to subsequent health and the stability of work participation later on. However, it remains unclear how the timing of entry into paid employment together with the accumulation of different life course social and health-related determinants are linked to different labour-market transitions such as unemployment during early careers and particularly the stability of early employment.

Labour-market statistics show that the proportions of both under 25-year-olds and over 55-year-olds in the workforce have remained low compared to midlife groups (WHO, 2012a; 2012b). Thus, there is a need to extend work careers by focusing on the determinants of work participation patterns among young employees particularly. One of the reasons for late labour-market entry is entering tertiary education, which in Finland is often both delayed due to challenges in entering education and also extended due to delays in graduation (Statistics Finland, 2013). In the long run, higher education is, however, predictive of more stable labour-market participation. Reasons for other types of unstable early work participation and spells of non-employment could further be attributable to early risk factors that have a long latency. Such factors could originate already from childhood, and be linked to both parental and own socio-economic position and health. In particular, social disadvantage and early mental disorders of parents and the offspring have been linked to both work disability and long-term unemployment of young adults (Lallukka et al, 2019; Halonen et al, 2019b). As mental disorders increase among young adults, even before entering paid employment, they need to be considered when focusing on work participation among young adults (Patel et al, 2007; Croft et al, 2011; Farioli et al, 2014). Accordingly, Swedish register data shows that mental disorders, such as neurodevelopmental disorders, are typically diagnosed young and are strongly associated with subsequent work disability (Virtanen et al, 2020) and unemployment trajectories (Lallukka et al, 2020). A substantial part of early exits from paid employment is attributable to these disorders; thus the associations can be bidirectional, with early ill health predicting unstable work participation, and non-employment ill health later on (Kela, 2010; OECD, 2010; Coggon et al, 2013).

Early detection of modifiable risk factors of unstable early work-participation patterns is crucial. However, there is limited knowledge about the factors contributing to the stability of early careers, such as the opportunities for disabled people to enter the labour market and continue working (Achterberg et al, 2009). Moreover, most studies focus on one exit type or the time to exit (for instance, work disability having dichotomous outcomes). Such approaches fail to consider different statuses and transitions between statuses that more comprehensively describe the true development of work participation. Young adults are unlikely to be a homogeneous group, even when they belong to the same birth cohort, and failure to consider different actual patterns of work-participation patterns provide a rather limited picture of such patterns in young adulthood. More evidence is needed on what the paths to entry and exit from work in young adulthood are, what kind of early work-participation patterns can be identified, how likely transitions between different statuses are, and what the determinants of different work-participation patterns are. Importance of such life course approaches including childhood factors has been also stressed elsewhere (Dunn, 2010; Marmot et al, 2012).

We used sequence analysis and mixture Markov models to analyse trajectories of monthly work-participation statuses. The most important benefit of this approach is its holistic perspective: rather than focusing on specific events or transitions, this approach sheds light on the entire trajectory of early work participation. Sequence analysis has been increasingly used to study the sequencing of life events including labour-market participation, family formation and health (Barban, 2013; McKetta et al, 2018; Roux et al, 2019; Gao et al, 2020). Using mixture Markov models for such data allows us to identify typical patterns between and dynamics within individual trajectories as well as to study the associations between childhood predictors and different types of early labour-market participation patterns (Helske et al, 2018; Helske and Helske, 2019).

A key hypothesis of this study is that among a birth cohort, there are latent groups of work participation that have different social and health-related determinants that can shorten working-life duration and lead to unstable early careers due to, for example, health-related absences and unemployment. Furthermore, we hypothesise that young adults who enter the labour market earlier, have more unstable work participation; this could be explained by their more disadvantaged social background and health-related factors. Such early entry is more likely for those with no or short education, and entering on more manual work with adverse exposures.

Using register data comprising an entire birth cohort and a flexible person-oriented mixture Markov modelling approach, the study therefore aims to (1) examine what kind of work-participation patterns can be identified among young adults in their early careers and (2) how parental and own socio-economic factors and health as well as age at entry into the labour market jointly predict such patterns. We focus on age at entry in the first long-term period of paid employment and study what kind of latent classes can be identified in subsequent work participation and how the accumulation of a number of childhood and adolescence social and health-related determinants are associated with different types of work-participation patterns depending on the age at entry.

Methods

Data and variables

The Finnish Birth Cohort 1987 (henceforth FBC) (Paananen and Gissler, 2012; Ristikari et al, 2016) includes data from several registers from all children born in 1987 as well as their parents (registered mother for all children and registered father for all but 821 children). The follow-up period for the original data lasts from the perinatal period until December 2015. In this study, we use data on employment, pensions and other benefits from the Finnish Centre for Pension (ETK), study grants, unemployment benefits and parental leave data from the Social Insurance Institution of Finland (Kela), data on unemployment periods from the Finnish Ministry of Economic Affairs and Employment, data on social assistance from the Finnish Institute for Health and Welfare (THL), education level and date of death from Statistics Finland (TK), criminal convictions from Legal Register Centre (ORK) and comprehensive school achievement data from the Finnish National Board of Education (OPH). All the register data sets were merged using personal identification numbers assigned to each Finnish resident (Gissler and Haukka, 2004). Personal identification numbers are upheld by Digital and Population Data Services Agency (DVV). More precise information on the data and variables can be found on the FBC metadata web page (Finnish Institute for Health and Welfare, 2021).

From the full data set we created our study sample using information on early adulthood labour-market participation to create the outcome and information on own and parental social and health-related events from age 0 to age 17 for creating predictor variables (covariates). More precisely, for the outcome we created sequences of monthly work-participation statuses starting from the calendar month in which the individual started their first six-month spell in employment or entrepreneurship without being in education at the same time (at age 17 the earliest). We chose six months as the threshold as it was long enough to exclude summer jobs and other short employment spells, but not too long to end up dropping too many cohort members from the analysis. We recorded the work-participation status during the first five years after the first entry or for at least 12 months if they were lost to follow-up sooner. We classified the work-participation statuses into eight categories, presented in Table 1.

Altogether 59,476 children were born in Finland in 1987. We excluded 271 individuals that had received disability allowance at the middle or highest rate at the beginning of 2005 and 455 cohort members who had received a diagnosis for an intellectual disability. Furthermore, we dropped 464 individuals who had died before 2005 or during the five-year follow-up after their first entry into the labour market. Lastly, we dropped 1,980 cohort members who had lived abroad for more than 12 consecutive months during the follow-up, as changes in their work-participation status could not be followed using national registers, ending up with 56,306 cohort members.

Among them, 52,344 individuals had started a six-month work spell (88% of the full cohort and 93% of those that had not been excluded earlier due to disabilities, death or living abroad). Of them, 51,871 individuals were followed for at least 12 months after their labour market entry (87% of the cohort and 96% of the target population) and 42,770 individuals for a full 60 months (72% of the cohort and 79% of the target population).

The timing of the labour-market entry was split into three categories: age 17–18 (28.5% of the sample with a work spell of at least 12 months; entering at age 17 was very rare), age 19–22 (50.5%) and age 23–26 (21.0%). We used this categorised variable to allow for possible non-linear relationships between the timing of entry and the work-participation pattern. The cut-off points were decided based on normative educational transitions and the distribution observed in the sample.

All childhood covariates were determined from records during the time period when the cohort members were 0–17 years old (by December 2004). These covariates can be divided into own and parental social and health-related variables. The cohort members' own variables included sex, grade point average (GPA), criminal conviction, teenage

Table 1: Work-participation statuses and their priority

Status	Priority
Deceased	1
Health-related non-attendance	2
Parental leave	3
Work and studies	4
Studies	5
Work (employment or entrepreneurship)	6
Unemployed or receiving social assistance	7
Living with parents, living abroad or unknown	8

Notes: In a case of multiple coincidental statuses, the status with the higher priority was chosen as the individual's monthly status. Unknown status refers to unknown status in the registers (not being in employment or education, nor a recipient of a benefit).

pregnancy and psychiatric diagnosis. The sex of the cohort members was based on the Population Information System which is upheld by DVV. GPA is a measure of school success at the end of compulsory school around age 16. The grades range from four (lowest) to ten (highest) and the GPA variable was coded into six categories (from highest to lowest: 8.50–10.00, 8.00–8.49, 7.50–7.99, 7.00–7.49, 4.00–6.99, missing). The GPA information comes Joint Application Register managed by OPH. This register has the GPA information of the individuals who have applied to secondary level education. Most adolescents apply immediately after comprehensive school. For this reason, we recorded GPA as missing if the student had not applied for further education at all or if their application was delayed (7.8% in total). We recorded criminal record as a dichotomous variable based on the criminal records of the national central register managed by ORK and include only the cases decided in the court by the end of follow-up. Teenage pregnancy was defined as giving birth or having an abortion before age of 20 according to Medical Birth Register of THL. Psychiatric diagnoses were sampled from the Care Register for Health Care of THL which includes only specialised healthcare visits. Our follow-up includes visits with a diagnosis based on ICD-9 (in 1987–95) and ICD-10 (in 1996–2004), so we defined psychiatric diagnosis as ICD-9: 290–319 and ICD-10: F00–99. Inpatient care visits were included in the variable formation from 1987 onwards while outpatient care visits from 1998 onwards due to register limitations.

Parental social and health-related covariates included high parental education, nuclear family, social assistance, teenage mother and parental psychiatric diagnosis. High parental education was defined as whether at least one parent had a tertiary degree in the Register of Completed Education and Degrees of TK. Nuclear family is a variable based on the daily housing data from Population Information System of DVV. We defined it as a dichotomous variable that captures whether the cohort member and both parents occupied the same household at the end of the follow-up when cohort members were 17-year-olds (December 2004). Social assistance variable captured receiving social assistance, according to the Register of Social Assistance of THL. It was coded as having received social assistance either if the parents had received over 12 months of social assistance during the cohort member's childhood or if there was at least one calendar year during that time in which the family received social assistance for more than 7 months. Having a teenage mother was determined from the Medical Birth Register of THL and was defined as whether the mother was less than 20 years of age when giving birth to the cohort member. Lastly, parental psychiatric diagnoses were determined using the same ICD-code ranges as for cohort members (ICD-9: 290–319; ICD-10: F00–99) if either of the parents had received any of these diagnoses.

Statistical methods

Sequence analysis (SA) is an exploratory data-driven approach that has become central to the life course perspective for understanding various trajectories (Gauthier et al, 2014). It is typically used for exploratory analysis and for assessing the relationship between trajectories and other variables. For the latter, a two-step approach is typical: first, SA and cluster analysis are used for finding meaningful groups in the data; second, the covariate–sequence relationship is assessed by using cluster memberships as a categorical variable. This approach, however, has been criticised for neglecting the within-cluster variation, which could potentially lead to biased results if the variation is not small and random (Studer, 2013; Piccarreta and Studer, 2019).

Recently, probabilistic modelling of social sequence data has gained increasing interest. Unlike the data-mining type approach of SA, probabilistic approaches allow for the use of statistical inference to draw conclusions about the quality of the model and the covariate–sequence relationship. The multistate model (MSM) extends the focus of the basic event history model from one event or transition to multiple potentially recurring transitions (Meira-Machado et al, 2009; Steele, 2011). However, unlike SA, the MSM requires making assumptions on the dependence of transitions on time. One commonly used assumption and the approach adopted in this paper is the (first-order) Markov assumption which states that the probability of a transition only depends on the current status, not on prior history. Markovian models can be regarded as one type of MSMs that allow for flexible analysis of life course data with multiple life statuses and a large number of different transitions (Helske et al, 2018; Piccarreta and Studer, 2019; Han et al, 2020).

The Markov model estimates the probabilities for starting in each labour-market state and transitioning between the states (initial and transition probabilities). We used an extension of the basic model, the mixture Markov model (MMM), which expects that the population consists of latent classes (clusters) with varying longitudinal employment patterns and allows for different specifications of initial and transition probabilities between the classes (see Appendix A for more information on the model specification). The MMM is similar to the *latent class model* (Han et al, 2016) in that it accounts for *unobserved heterogeneity* between the sequences but is more general as it does not assume conditional independence between time points – quite the opposite, as the idea is to estimate transition probabilities from time point to time point.

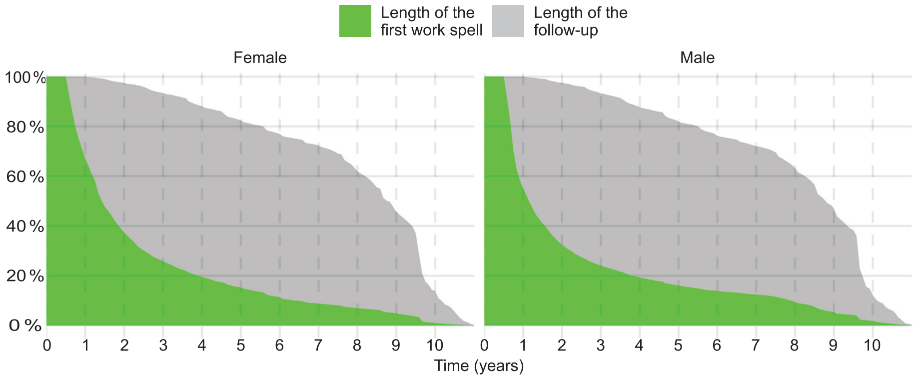
Covariates can be included in the model to explain class membership probabilities and/or initial and transition probabilities. However, allowing the same variable to influence various types of probabilities may hinder the interpretation and evaluation of its effect. Therefore, it is common practice to relate each covariate to one type of probability only (Han et al, 2016). Since our interest was to use childhood covariates to predict labour-market participation patterns, we constructed a model where the covariates explain class membership probabilities, allowing us to evaluate the relationship between the timing of the labour-market entry and further work participation and to predict labour-market participation outcomes using childhood information.

In practice, the estimation of a complex Markovian model with covariates can be time-consuming, especially if and typically when we do not know the number of latent classes beforehand (Helske and Helske, 2019). To facilitate the analysis, we used an approach described in Helske et al (2016) who used SA as a preliminary step before estimation of a mixture hidden Markov model. We chose ten clusters based on the SA and estimated the MMM with covariates predicting latent class memberships. More detailed information on the estimation process is given in Appendix A.

Software

All analyses were conducted with R (R Core Team, 2020). We used the TraMineR package (Gabadinho et al, 2011) for creating the sequences and calculating sequence dissimilarities, seqHMM (Helske and Helske, 2019) for estimation and visualisation of the MMM, and ggplot2 (Wickham, 2016) for other visualisations.

Figure 1: The length of the total follow-up after the initial entry into the labour market and the length of the first at least 6-month long work spell



Results

Descriptive statistics

Figure 1 shows the duration of the follow-up for men and women in the sample after their first entry into the labour market as well as the duration of their first at least six-month-long work episode. We see that for most of the cohort members, the first work spell was fairly short: less than two years for the majority of the cohort, while only 20% are continuously employed for four years or more. Even if accounting for all work spells, the durations were fairly short, as illustrated in Figure B-1 in Appendix B.

Figure 2 shows the mean time spent in each work-participation state. The work status was the most prevalent: about 70–78% of all months of the follow-up were spent in work (slightly more for men than women). Women spent more time in parallel work and studies as well as on parental leave while men spent more time in the unknown state (living with parents, abroad or otherwise unknown, typically military or non-military service) and in the state of unemployment or receiving social assistance.

Figure 2: Mean time spent in each work participation state. The blue colour represents men and green represents women, while the dashed line shows the average for everyone

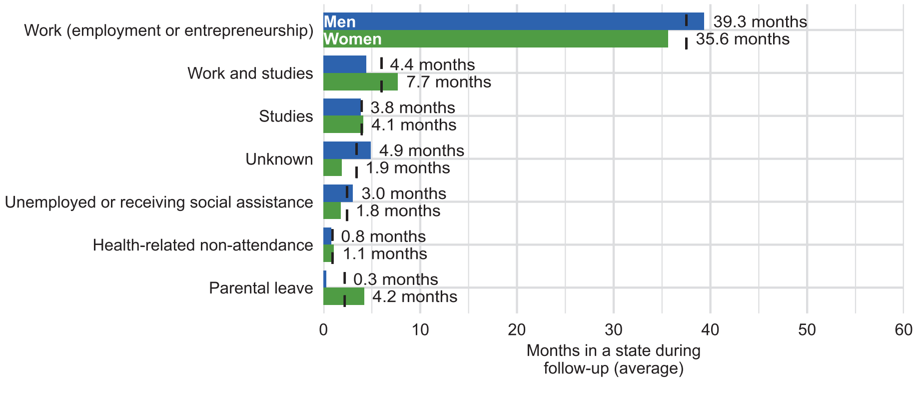


Table 2: Most common work-participation patterns and their frequencies and proportions in the study sample (omitting timing and duration of spells)

Work-participation pattern	Frequency	Proportion (%)
Work	7,479	14.42
Work → Unknown → Work	1,925	3.71
Work → Work and studies → Work	1,383	2.67
Work → Health-related non-attendance → Work	1,320	2.54
Work → Unemployed or receiving social assistance → Work	887	1.71
Work → Work and studies → Work → Work and studies →Work	605	1.17
Work → Parental leave → Work	550	1.06
Work → Parental leave	539	1.04
Other	37,183	71.68
Total	51,871	100

Notes: For example, the pattern Work → Parental leave → Work refers to all sequences where the individual exited work due to parental leave and then later returned back to work, irrespective of the durations of each specific spell.

Table 2 shows the most frequent work-participation patterns in the full study sample (before any classification), when not taking timing and duration of the spells into account. Continuous work was the most frequent pattern, but only 14% of the individuals were working continuously during the first five years after their entry into the labour market. All the other patterns were much rarer and typically described temporary exits from work, most typically due to unknown status, studying parallel to work or being unemployed. The only 'permanent' exit among the most frequent patterns was exiting due to parental leave. The transitory nature of exits from work can also to some extent be seen in Figure B-2 in Appendix B – even numbers of transitions within sequences are much more common than odd numbers (slightly less common among women due to the 'permanent' exits to parental leave).

Table 3 shows descriptive statistics of the predictors. We find that 28% of the individuals that had entered the labour market during the follow-up were aged 18 or younger and half were aged 19–22. Men and women differed in terms of their GPA (higher for women), in having had a conviction (more often men) and their own psychiatric diagnosis (more often women).

Latent classes of labour-market participation

Figure 3 describes the latent classes after assigning each individual to the latent class for which they had the highest membership probability, showing monthly prevalence of each status. At the start, almost all individuals were in the work state (apart from few exceptions that had a work contract but were on parental or health-related leave). Monthly transition probabilities are visualised in Figure A-4 in Appendix A.

Class A: Continuous work (most probable class for 19% of the individuals). Individuals hardly leave paid work – the probability for exiting work is 1% – but in the rare occasions that they do, they swiftly return to work.

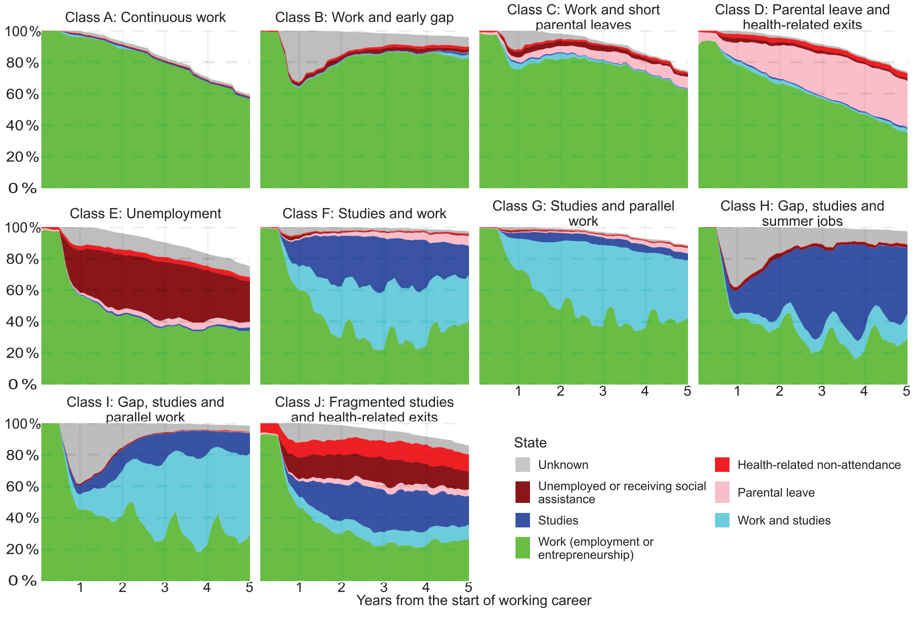
Class B: Work and early gap (17%). The individuals mainly work but often have an early gap for which we have no data. Probably, for many, this gap consists of military or non-military service which is mandatory by law for all men in Finland

Table 3: Descriptive statistics (frequencies and proportions) of categorical variables used in the study, for the full sample as well as separated by sex

	All (N = 51,871)		Male (N = 26,480)		Female (N = 25,391)	
Variable	**Freq**	**%**	**Freq**	**%**	**Freq**	**%**
Age of entry: 17–18	14,803	28.5	7,744	29.2	7,059	27.8
Age of entry: 19–22	26,190	50.5	13,795	52.1	12,395	48.8
Age of entry: 23–27	10,878	21.0	4,941	18.7	5,937	23.4
GPA: missing	4,033	7.8	2,529	9.6	1,504	5.9
GPA: 4.00–6.99	13,222	25.5	8,782	33.2	4,440	17.5
GPA: 7.00–7.49	6,831	13.2	3,697	14.0	3,134	12.3
GPA: 7.50–7.99	7,357	14.2	3,702	14.0	3,655	14.4
GPA: 8.00–8.49	7,776	15	3,520	13.3	4,256	16.8
GPA: 8.50–10	12,652	24.4	4,250	16.0	8,402	33.1
Cohort member's conviction	894	1.7	765	2.9	129	0.5
Cohort member's psychiatric diagnosis	3,888	7.5	1,826	6.9	2,062	8.1
Cohort member's teenage pregnancy	822	1.6	0	0.0	822	3.2
Nuclear family	32,845	63.3	17,252	65.2	15,593	61.4
Parental education	12,879	24.8	6,532	24.7	6,347	25.0
Parental psychiatric diagnosis	7,913	15.3	3,999	15.1	3,914	15.4
Social assistance	11,158	21.5	5,632	21.3	5,526	21.8
Teenage mother	1,599	3.1	814	3.1	785	3.1

Note: Missing GPA refers to no or delayed application into secondary education.

Figure 3: Latent classes from the mixture Markov model described using monthly distributions for work participation states (maximum membership assignment)



(and a voluntary option for women, quite rarely used). Another typical reason is being supported by parents. Other exits from work (mainly related to health and unemployment) are also a bit more probable than in the continuous work class but tend to be short in duration.

Class C: Work and short parental leaves (4%). In this male-specific class, men mainly work but they also have short exits, typically for parental leave. Early spells of unknown status are fairly common.

Class D: Parental leave and health-related exits (12%). This female-specific class is mainly characterised by exits from work to longer parental leave and short exits due to health-related non-attendance.

Class E: Unemployment (9%). Individuals exit work for unemployment (some are also only receiving social assistance without other main benefits or work-related income). Many transitions between unemployment and employment and also other short exits – mainly to parental leave or health-related or unknown status – are fairly common.

Class F: Studies and work (6%). Flexible transitions between work and studies and both in parallel. Entering other states is rare.

Class G: Studies and parallel work (13%). Individuals mainly study and work in parallel. They tend to have a later entry into studies and/or their studies are short in duration. Entering other states than work is rare.

Class H: Gap, studies and summer jobs (7%). Recurrent transitions between studies and work, probably indicating full-time studies and summer jobs. In addition, individuals have spells of unknown status, typically soon after first entering the labour market but also later during the follow-up. Many of these gaps likely indicate participation in (non-)military service or being supported by parents. Individuals rarely enter other states, but if they do, they tend to exit swiftly.

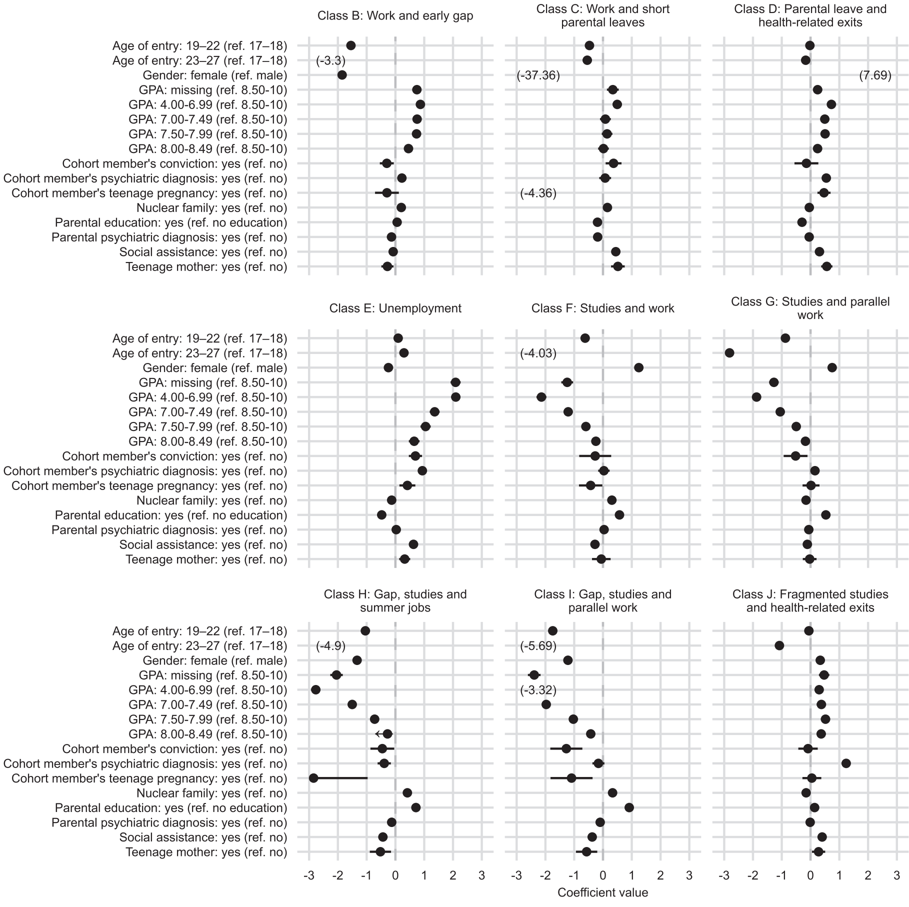
Class I: Gap, studies and parallel work (8%). Mainly work and studies in parallel. Also some unknown (military-service type) gaps of unknown status, mainly soon after first entering the labour market. Entering other states is rare and short in duration.

Class J: Fragmented studies and health-related exits (5%). Students who typically lack the usual pattern of alternating between or combining studies and work. Instead, they have spells of parental leave, unemployment or health-related exits between study spells. This is also the most likely class for individuals with long health-related non-attendance spells, sometimes even without studies.

Predictors to early work-participation patterns

Figure 4 shows estimates and 95% confidence intervals for the timing of the labour-market entry and cohort members' own and parental social and health-related predictors. The estimates are interpreted similarly to the log odds ratios from multinomial logistic regression; Class A, the *Continuous work* class, was used as the reference. Log odds ratios are difficult to interpret as such, but due to the large variation in the estimates, the figure would be illegible with odds ratios. We do, however, show odds ratios in supplementary material and in the interpretations of each variable. Exact estimates for variable coefficients are given in supplementary material along with the rest of the estimated model parameters (Table C-1 for log odds and Table C-2 for exponents).

Figure 4: Estimates for log odds and their 95% confidence intervals from predictors of latent class memberships from the mixture Markov model. The reference class is Continuous work.



Age at labour-market entry

The later the entry into the labour market, the more likely the *Continuous work* class (Class A) was and the less likely the study classes (F–I) and the *Work and early gap* class (Class B) were. A later entry (at age 23–26) also predicted somewhat lower odds for *Work and short parental leaves* (Class C), *Parental leave and health-related exits* (class D) and *Fragmented studies and health-related exits* (Class J). The only positive estimate was found in the *Unemployment* class (Class E), indicating an increase of 30% in the odds for those with a later entry rather than an early entry (all else equal).

The relationship between age at labour-market entry and early work-participation pattern in early adulthood seemed fairly linear.

Own social and health-related predictors

Sex. The *Parental leave and health-related exits* (Class D) was basically an all-female class (probabilities close to 0 for all men) and the *Work and short parental leaves* class

(Class C) an all-male class (probabilities close to 0 for all women). Women were also much more likely to be members of the *Studies and work* (3.5-fold odds, Class F) and *Studies and parallel work* (twofold odds, Class G). Male sex predicted higher odds for all other classes, most notably in *Work and early gap* (fourfold odds, Class B) and *Gap, studies, and summer jobs* (threefold odds, Class H).

GPA. Higher GPAs predicted higher odds for membership for all classes characterised by studies and work (Classes F–I), whereas lower GPAs predicted higher probabilities for all other classes. The exception was the *Work and short parental leaves* class (Class C) where the differences were statistically insignificant (and close to 0) for all GPAs between 7 and 10.

Conviction. Having been convicted predicted higher odds for membership in the *Unemployment* class (twofold, Class E) and in the *Work and short parental leaves* class (50% increase, Class C). Not having been convicted predicted higher odds of membership in the *Work and early gap* (by 35%, Class B) and three classes of combining studies and work (G–I; by 58–360%).

Teenage pregnancy. Own teenage pregnancy predicted an increase in the odds for membership by 57% in the *Parental leave and health-related exits* (Class D) and by 51% in the *Unemployment* class (Class E), and reduced the odds for membership in three study classes (F, H and I).

Psychiatric diagnosis. Own childhood psychiatric diagnosis (before age 18) predicted higher odds for membership in the same classes that also had higher odds for individuals with a low GPA, most notably for *Fragmented studies and health-related exits* (3.5-fold odds, Class J), *Unemployment* (2.5-fold odds, Class E), and *Parental leave and health-related exits* (73% increase in odds, Class D).

Parental social and health-related predictors

Parental education. Parental education was a significant predictor of early work-participation patterns, all else equal and even after controlling for own educational attainment. More specifically, parents' high education predicted higher odds of membership in all study classes (by 70–248% for Classes F–I and 15% for Class J). Parent's average or low education, on the other hand, predicted an increase in the odds of membership in the *Unemployment* class by 62% (Class E), in the *Parental leave and health-related exits* by 35% (Class D) and in the *Work and short parental leaves* by 21% (Class C).

Nuclear family. Having grown up in a nuclear family predicted higher odds of membership in classes *Studies and work*, *Gap, studies and work* and *Gap, studies and parallel work* by 36–51% (Classes F, H and I) and in classes *Work and early gap* and *Work and short parental leaves* by 16–22% (Classes B and C). Having grown up in a non-nuclear family predicted higher odds for classes *Studies and parallel work*, *Fragmented studies and health-related exits* and *Unemployment* by 14–17% (Classes G, J and E).

Social assistance. Receiving social assistance in childhood predicts higher odds for membership in the *Unemployment* class by 88% (Class E) and in classes *Work and short parental leaves*, *Fragmented studies and health-related exits* and *Parental leave and health-related exits* by 36–57% (Classes C, J and D). Not having received social assistance predicted higher odds for membership in all classes of combining studies and work (Classes F–I).

Teenage mother. Being born to a teenage mother predicted higher odds for membership in the *Parental leave and health-related exits* class by 75% (Class D), in the

Work and short parental leaves by 68% (Class C), in the *Unemployment* class by 38% (Class E) and in the *Fragmented studies and health-related exits* by 32% (Class J). Being born to a mother who was at least 20 years old predicted higher odds for membership in the *Gap, studies and parallel work* class by 77% (Class I), in the *Gap, studies and summer jobs* class by 68% (Class H) and in the *Work and early gap* class by 32% (Class B).

Psychiatric diagnosis. When accounting for all other predictors, having had a parent with a psychiatric diagnosis did not predict higher odds for membership in any of the latent classes (all positive coefficients were small and not statistically significant). However, *not* having a parent with a psychiatric diagnosis predicted slightly higher odds for membership in the *Work and early gap* class (by 20%, Class B) and in the *Work and short parental leaves* class (by 15%, Class C).

Accumulation of (dis)advantage

In this section, we look more closely into the accumulation of childhood advantage and disadvantage, by illustrating with a number of examples how different kinds of backgrounds (combinations of predictors) and different ages at the labour-market entry predicted different work-participation patterns. We present five typical cases defined by selected risk or protective factors. The cases and their respective risk/protective factors are:

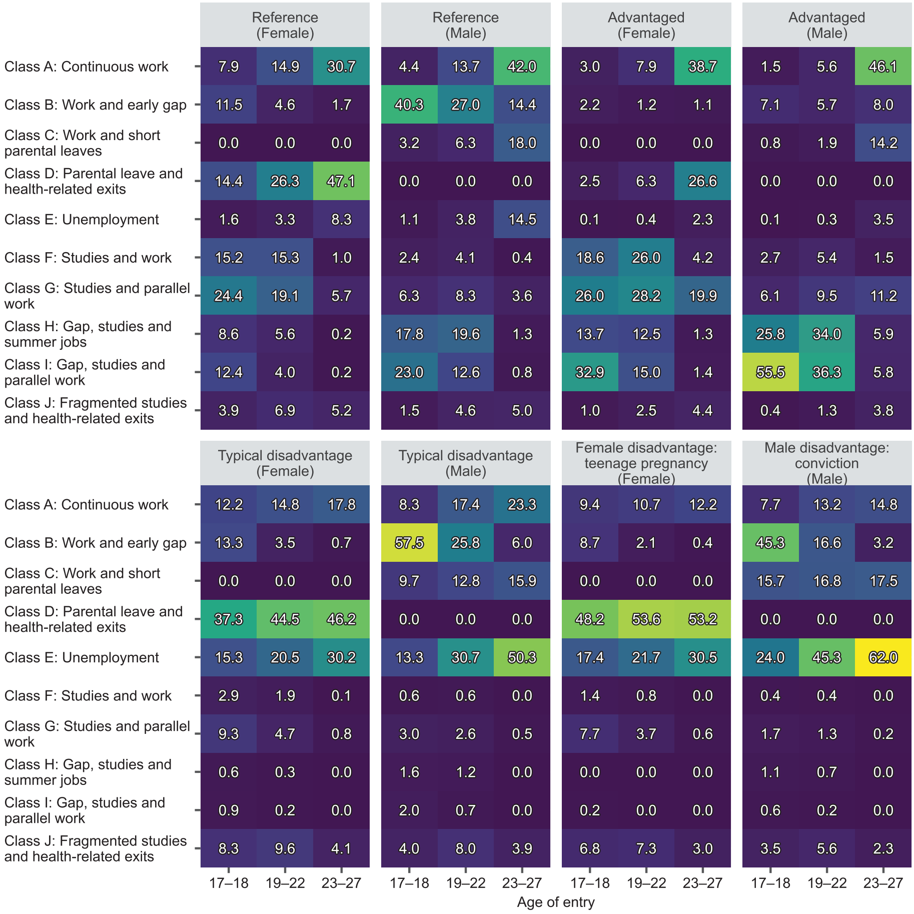
1. *Reference case* with having had a nuclear family, intermediate GPA, low/average parental education and no risk factors;
2. *Accumulated advantage* with nuclear family, high GPA, high parental education and no risk factors;
3. *'Typical' accumulated disadvantage* with non-nuclear family, low GPA, social assistance and parental psychiatric diagnosis;
4. *Female-typical accumulated disadvantage* with non-nuclear family, low GPA, social assistance and teenage pregnancy; and
5. *Male-typical accumulated disadvantage* with non-nuclear family, low GPA, social assistance, conviction.

Table D-1 in Appendix D shows the exact definitions for each example case.

All but the sex-typical cases are shown separately for women and men. The *typical accumulated disadvantage* case was chosen according to the data: it was the most frequent combination of predictors among cohort members that had at least four risk factors. The sex-typical cases were chosen accordingly, but only among profiles that had differing frequencies between the two sexes.

Figure 5 shows the predicted probabilities for each latent class for the example cases. For all cases, the continuous employment pattern (Class A) was the more likely the later the individual entered the labour market. There are considerable differences in the levels though: individuals entering the labour market after age 23 with a 'reference' background or an advantaged background had much higher probabilities compared to the disadvantaged cases, and the probabilities were the lowest for individuals with a sex-typical disadvantaged background. Among the late entrants, the continuous work class was the most likely pattern for women and men with an advantaged background while for women in the other groups the *Long parental leave and health-related exits* class was the most likely (43–53%, Class D) and for the disadvantaged men the most likely class was the

Figure 5: Probabilities for each work participation pattern (latent class) for some exemplary cases, based on the estimated mixture Markov model. The colours represent the predicted probabilities on a continuous scale from 0 (dark blue) to the highest predicted probability of 62% (yellow).



Unemployment class (58–62%, Class E). For late male entrants with a reference background, there was more variation: the probabilities were about equally likely (14–18%) for classes *Work and early gap*, *Work and short parental leaves* and *Unemployment* (Classes B, C and E). For the younger entrants with a reference background and especially with an advantaged background, entering studies was a more likely pattern than continuous employment. For the youngest entrants with a reference background, women's probabilities were divided between *Work and early gap*, *Parental leave and health-related exits* (Classes B and D), and the four study classes of combining studies and work (of which the *Studies and parallel work* (Class G) was the most likely with a 24% probability). For the youngest male entrants with a reference background, the *Work and early gap* class was the most likely (40%, Class B).

For the individuals with a disadvantaged background, the unemployment class (Class E) was likely, the more likely the later the entry into the labour market was. Women had high probabilities for the *Parental leave and health-related exits* class (Class D). Among the men and women from disadvantaged backgrounds that did enter studies, the *Fragmented studies and health-related exits* class (Class J) was the most likely for almost all groups, with the exception of the youngest female entrants, among whom the *Studies and parallel work* class (Class G) was slightly more likely.

Discussion

Main findings

This study sought to identify distinct clusters of work participation among young employees, transitions between employment statuses, and their association with age at entry into the labour market as well as their life course social and health-related determinants. Using data from an entire birth cohort, we could follow individuals from their birth until young adulthood, entry into paid employment and different work-participation patterns and transitions between employment statuses after the entry.

Labour-market entry and exit patterns were measured using register data. We could consider all relevant routes of exit in young adulthood: in addition to unwanted, involuntary health-related exit and unemployment, we could also consider more voluntary types of exit such as studying and parental leave.

First of all, we observed that entry into the labour market as measured by six months in continuous employment was not a permanent entry for many, not only due to unemployment and health-related exits but also due to studies and parental leave. With our definition we were able to omit summer jobs and other short-term employment as well as working in parallel with studies, but not complete gap years while waiting for (another) attempt to enter higher education. Such gap years are fairly common in the Finnish context: entrance into tertiary education is very competitive and for this cohort typically required passing entrance exams. In 2011 when the cohort members turned 24, the median age at entrance was 20 for universities and 21 for polytechnics (Statistics Finland, 2013), indicating that many young adults had to apply multiple times before being admitted, during which time many of them worked. Another typical reason for exiting employment in early adulthood is military and non-military service (compulsory for men, voluntary for women), of which we had indirect evidence via lack of other register data at the typical time of service.

Using SA and the MMM we were able to identify different patterns in early labour-market participation, to study dynamics within different types of patterns and to study the association of typical patterns and early life course determinants. A ten-class solution was selected to most meaningfully describe the different work-participation patterns.

Regarding the timing of labour-market entry, we were interested in whether it was predictive of future labour-market attachment and whether the relationship was linear or non-linear. During the first five years we observed a fairly linear relationship between the age at entry and early work-participation pattern: individuals entering the labour market later were more likely to be in continuous employment thereafter (all else equal). When interpreting these results, we have to keep in mind that due to the nature of our data we were only able to study individuals entering the labour market by January 2015 when most of the cohort members were aged 27 (87% of the cohort), and only up to age 24 for a full five-year follow-up (74% of the cohort).

Given that the median age for getting a lower tertiary degree was as high as 26 in Finland in 2011 (Statistics Finland, 2013), some of the cohort members had not yet entered the labour market at least partly due to ongoing studies (19% of the cohort members who were excluded due to not entering the labour market, 843 cases, were students in 2015). On the other hand, with this analytical set-up we are also missing the individuals who had yet to experience their labour-market entry at age 27 for other reasons such as long-term ill health. It is possible that the linear association would not hold for individuals entering the labour market at later ages, but this remains to be confirmed with further studies.

Regarding both own and parental social and health-related childhood factors, latent classes show clear patterns of accumulated advantage and disadvantage as hypothesised. As expected, we observed more advantaged background to be related to exits due to studies (or a combination of studies and work) or – when following a late entry – stable employment, while disadvantaged background factors were more often related to unstable work and long-term exits from the labour market.

While the continuous work pattern was about equally likely among men and women, we observed clear gendered pathways in the other groups. Women often had exits from work for long periods of parental leave – especially women from more disadvantaged backgrounds. Men, on the other hand, often had short gaps shortly after the labour-market entry (probably due to military or non-military service) and seemed to have higher risks for frequent or long-term unemployment and social assistance. We also observed a male-typical parental leave pattern consisting of short-term exits from work. When combining studies and work, women more often had flexible transitions between work and studies and a combination of both, while the pattern of full-time studying and full-time summer jobs was more common among men.

Higher GPAs were related to exiting work for studies, while lower GPAs increased the risk of exits from work for other reasons. Parent's higher education further increased the probability of entering studies while low background was related to higher risk of long-term exits from work for other reasons, even after accounting for own educational attainment in adolescence. In fact, we observed that children with average school success but highly educated parent(s) had similar predicted probabilities for labour-market participation patterns as high-achieving children from less-educated homes. This is well in line with the risk aversion theory of social mobility research that has shown that children from higher social origins are more likely than children from lower social origins to be compensated for the risks related to lower educational accomplishments and failure (Breen and Goldthorpe, 1997; Heiskala et al, 2020).

Having been convicted was a strong risk factor for men for long-term or frequent exits from employment. Most notably, exits due to unemployment and living on social assistance were very likely, especially after a late entry into the labour market and in a case of accumulation of risk factors. This is in line with previous findings from the same data (Ristikari et al, 2016).

Teenage pregnancy and having a teenage mother both predicted long parental leaves, short health-related exits, and frequent and long-term unemployment. The former was a particularly strong predictor for the earlier entrants. This is also in line with previous studies (Merikukka et al, 2018; Lallukka et al, 2019).

Own psychiatric diagnosis increased the risk of long-term exits from work, including studies without work spells in between. Parent's psychiatric diagnosis did not (independently) predict higher risks for non-participation patterns compared to continuous employment, all else equal. This finding is somewhat contrary to an earlier study from the same cohort that did find an independent effect of parent's mental disorders on offspring's psychiatric work disability and social disadvantage in adolescence that was not completely mediated by offspring's mental disorders (Halonen et al, 2019b). It is possible that this difference in findings stems from the different formulations of the outcomes – as we focus on the big picture (general and multifaceted labour-market participation pattern) we may miss some of the finer details.

Regarding other family background factors, having grown up in a nuclear family was a fairly weak predictor alone. It did, however, somewhat increase the probability of exits due to studies or exits for other reasons that are short in duration, while non-nuclear family increased the probabilities for unemployment and fragmented studies but also for later or shorter studies in parallel with work. The association of family structure and educational attainment is well in line with previous inequality research: numerous studies have shown that parental separation is related to lower educational attainment (McLanahan and Sandefur, 1994; Amato, 2000; 2010; McLanahan and Percheski, 2008; Bernardi and Radl, 2014).

Receiving social assistance in childhood predicted higher probability for exits from work for other reasons than studies, while not receiving social assistance increased the probability of exiting work for studies. The results are in line with previous research suggesting that receiving social assistance in childhood predicts higher probability for labour-market exclusion (Bäckman and Nilsson, 2007).

While some risk factors were clearly independently predictive of higher risks, this was not the case with all of them but rather the most elevated risks for less favourable labour-market participation patterns were associated with the accumulation of different risk factors.

Methodological considerations

Unique data sets available for this study enabled detailed examination of work-participation patterns during early careers among young adults and their social and health-related determinants. The use of complex, large register-based merged data sets and sophisticated person-oriented analysis methods further provided unique opportunities to produce novel evidence with implications at various levels. These innovative, novel approaches are needed and have been encouraged to be applied in the field of epidemiology (Ness, 2013).

Generally, the coverage and quality of the register data used in this study are exceptionally good. The data sets are large with statistical power to detect smaller differences and pinpoint risk groups and examine also rarer determinants of work participation. The inclusion of an entire birth cohort, to study social and health-related factors of both parents and their offspring in determining work participation is a unique strength of the study. At the same time, it is important to keep in mind that the information on services is based on usage, not need.

MMMs allowed us to examine longitudinal outcomes and multiple predictors simultaneously as well as to calculate predicted probabilities for following different types of patterns based on protective and risk factors experienced during childhood. In comparison to more traditional event history models, we were able to account for different types of exits from employment as well as the subsequent sequence of events with our approach. Furthermore, by using SA we could have been able to study more complex patterns, but this approach also comes with its own issues, most importantly the fact that the two-step process of first finding clusters with SA and cluster analysis and then taking cluster memberships as fixed at the modelling stage is often problematic and can even lead to wrong conclusions about the nature and existence of the studied relationships (for example, Piccarreta and Studer, 2019; Helske et al, 2021). With the chosen Markovian approach, we were able to study not just specific exits but a longer trajectory, often including a number of transitions, and also to take the uncertainty of the classification into account better than with other methods. For example, we were able to differentiate between, for instance, short-term and long-term exits due to parental leave (Classes C and D) and different types of patterns of exits due to studies (such as differentiating the less favourable pattern of fragmented studies from more favourable combinations of work and studies). Furthermore, we were able to study the probabilities of following different types of labour-market participation patterns by taking the uncertainty of memberships into account.

In a mixture model, the classifications of the outcome and the effects of the predictors are modelled jointly; the purpose is to find balance between an optimal partitioning of the trajectories (latent classes) and the mixing probabilities (including the effects of the predictors). This approach is different from one where the classification of the outcome is done separately from modelling the effects of the predictors, and is arguably a more correct approach for analysing heterogeneous populations when we assume that the predictors have an effect (as we do). Due to the different methodological approach, however, direct comparisons with earlier more narrowly focused studies can only be tentative. It is also important to keep in mind that the analyses here are based on prediction, not on causal estimation. Selection based on attributes other than those that were directly studied was not accounted for.

Conclusions

There are varied work-participation patterns and transitions between employment statuses among young employees. More in-depth understanding of these patterns and their determinants is important to plan targeted interventions and to promote more stable work participation among young adults.

Information about work-participation patterns and their determinants can be used in efforts to promote work participation, particularly considering young people's entry into their first job, their subsequent transitions in the labour market alongside the opportunities for re-employment, and the determinants for different patterns. If people can have a sustained work participation and more stable work careers, this also has a notable societal impact in workplaces and beyond.

On the one hand, early (and possibly also very late) entry into paid employment could reflect unstable work participating during early working-life span. This is related to low socio-economic status, low education and own health problems. On the other hand, early entry is fairly common, and in addition to more unstable labour-market participation it is also related to gap years before studies (which in turn predict better labour-market attachment in the future), so an early entry is probably not a risk factor as such. It is important to target policies and actions in the high risk and most disadvantaged groups in particular, using the information of this study. Intervention studies, for example, supporting the most disadvantaged families, teenaged mothers, children of teenaged mothers and individuals with early psychiatric problems, could confirm, if such support can be linked to later more stable working life as well.

Supplementary data

A supplementary online appendix is available for this article at: https://doi.org/10.1332/175795922X16624496452365

Funding

SH was supported by the Academy of Finland grant #331816 and by the Academy of Finland Flagship Programme (Grant #320162). TL was supported by the Academy of Finland (Grants #287488 and #319200).

Data availability

The data that support the findings of this study are not publicly available due to data protections laws and regulations. Thus, strict restrictions apply to the availability of these data, which were used solely under the permission from the national, administrative register data holders. Data are available from the register data holders upon reasonable request pending their permission.

Conflict of interest

The authors declare that there is no conflict of interest.

References

Achterberg, T.J., Wind, H., de Boer, A.G.E.M. and Frings-Dresen, M.H.W. (2009) Factors that promote or hinder young disabled people in work participation: a systematic review, *Journal of Occupational Rehabilitation*, 19(2): 129–41.

Amato, P.R. (2000) The consequences of divorce for adults and children, *Journal of Marriage and Family*, 62(4): 1269–87.

Amato, P.R. (2010) Research on divorce: continuing trends and new developments, *Journal of Marriage and Family*, 72(3): 650–66.

Bäckman, O. and Nilsson, A. (2007) *Childhood Poverty and Labour Market Exclusion: Findings from a Swedish Birth Cohort*, Working Paper No. 2007:13, Stockholm: Institute for Futures Studies.

Barban, N. (2013) Family trajectories and health: a life course perspective, *European Journal of Population*, 29(4): 357–85.

Baum, L.E. and Petrie, T. (1966) Statistical inference for probabilistic functions of finite state Markov chains, *Annals of Mathematical Statistics*, 37(6): 1554–63.

Bernardi, F. and Radl, J. (2014) The long-term consequences of parental divorce for children's educational attainment, *Demographic Research*, 30: 1653–80.

Breen, R. and Goldthorpe, J.H. (1997) Explaining educational differentials: towards a formal rational action theory, *Rationality and Society*, 9(3): 275–305.

Coggon, D. and other members of the CUPID Collaboration (2013) International variation in absence from work attributed to musculoskeletal illness: findings from the CUPID study, *Occupational and Environmental Medicine*, 70(8): 575–84.

Croft, P., Blyth, F.M. and Van Der Windt, D. (2011) *Chronic Pain Epidemiology: From Aetiology to Public Health*, Oxford: Oxford University Press.

Dunn, K.M. (2010) Extending conceptual frameworks: life course epidemiology for the study of back pain, *BMC Musculoskeletal Disorders*, 11(1): art 23, doi: 10.1186/1471-2474-11-23.

Farioli, A., Mattioli, S., Quaglieri, A., Curti, S., Violante, F.S. and Coggon, D. (2014) Musculoskeletal pain in Europe: The role of personal, occupational, and social risk factors, *Scandinavian Journal of Work, Environment and Health*, 40(1): 36–46.

Finnish Institute for Health and Welfare (2021) Meta visualisation of the Finnish birth cohort studies, https://fbc-studies.github.io/meta-visu/?lang=en&ds=finnish-birth-cohorts, (Accessed: 10 Feb 2021).

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

Gao, N., Ryan, M., Kruciena, N., Robinson, S. and Norman, R. (2020) Paid work, household work, or leisure? Time allocation pathways among women following a cancer diagnosis, *Social Science and Medicine*, 246: art 112776, doi: 10.1016/j.socscimed.2019.112776.

Gauthier, J.A., Bühlmann, F. and Blanchard, P. (2014) Introduction: sequence analysis in 2014, in P. Blanchard, F. Bühlmann and J.-A. Gauthier (eds) *Advances in Sequence Analysis: Theory, Method, Applications*, New York: Springer, pp 1–17.

Gissler, M. and Haukka, J. (2004) Finnish health and social welfare registers in epidemiological research, *Norsk Epidemiologi*, 14: 113–20.

Halonen, J.I., Virtanen, M., Ala-Mursula, L., Miettunen, J., Vaaramo, E., Karppinen, J., Kouvonen, A. and Lallukka, T. (2019a) Socioeconomic and health-related childhood and adolescence predictors of entry into paid employment, *European Journal of Public Health*, 29(3): 555–61.

Halonen, J.I., Merikukka, M., Gissler, M., Kerkelä, M., Virtanen, M., Ristikari, T., Hiilamo, H. and Lallukka, T. (2019b) Pathways from parental mental disorders to offspring's work disability due to depressive or anxiety disorders in early adulthood: the 1987 Finnish Birth Cohort, *Depression and Anxiety*, 36(4): 305–12.

Han, S.Y., Liefbroer, A.C. and Elzinga, C.H. (2016) *Understanding Social-class Differences in the Transition to Adulthood Using Markov Chain Models*, in G. Ritschard and M. Studer (eds) Proceedings of the International Conference on Sequence Analysis and Related Methods, Lausanne, June 8–10, 2016, Moscow: HSE University, pp 155–77.

Han, S.Y., Liefbroer, A.C. and Elzinga, C.H. (2020) Mechanisms of family formation: an application of Hidden Markov Models to a life course process, *Advances in Life Course Research*, 43: art 100265, doi: 10.1016/j.alcr.2019.03.001.

Heiskala, L., Erola, J. and Kilpi-Jakonen, E. (2020) Compensatory and multiplicative advantages: social origin, school performance, and stratified higher education enrolment in Finland, *European Sociological Review*, 37(2): 171–85.

Helske, S. and Helske, J. (2019) Mixture hidden Markov models for sequence data: the seqhmm package in R, *Journal of Statistical Software*, 88(1), doi: 10.18637/jss.v088.i03.

Helske, S., Helske, J. and Eerola, M. (2016) Analysing complex life sequence data with hidden Markov modelling, in *LaCOSA II : Proceedings of the International Conference on Sequence Analysis and Related Methods*.

Helske, S., Helske, J. and Eerola, M. (2018) Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data, in G. Ritschard and M. Studer (eds) *Sequence Analysis and Related Approaches. Life Course Research and Social Policies*, Vol 10, Cham: Springer, pp 185–200, doi: 10.1007/978-3-319-95420-2_11.

Helske, S., Helske, J. and Chihaya, G.K. (2021) From sequences to variables – rethinking the relationship between sequences and outcomes, SocArXiv preprint, doi: 10.31235/osf.io/srxag.

Kela (2010) The statistical yearbook of pensioners in Finland, *Official Statistics of Finland*, Social Protection. https://www.julkari.fi/bitstream/handle/10024/129272/TilastoSuomenelakkeensaajista2010.pdf?sequence=1&isAllowed=y

Lallukka, T., Kerkelä, M., Ristikari, T., Merikukka, M., Hiilamo, H., Virtanen, M., Øverland, S., Gissler, M. and Halonen, J.I. (2019) Determinants of long-term unemployment in early adulthood: a finnish birth cohort study, doi: 10.1016/j.ssmph.2019.100410.

Lallukka, T., Mittendorfer-Rutz, E., Ervasti, J., Alexanderson, K. and Virtanen, M. (2020) Unemployment trajectories and the early risk of disability pension among young people with and without autism spectrum disorder: a nationwide study in Sweden, *International Journal of Environmental Research and Public Health*, 17(7), doi: 10.3390/ijerph17072486.

Marmot, M., Allen, J., Bell, R., Bloomer, E. and Goldblatt, P. (2012) WHO European review of social determinants of health and the health divide, *The Lancet*, 380(9846): 1011–29.

McKetta, S. Prins, S.J., Platt, J., Bates, L.M. and Keyes, K. (2018) Social sequencing to determine patterns in health and work-family trajectories for U.S. women, 1968–2013, *SSM – Population Health*, 6: 301–8.

McLanahan, S. and Percheski, C. (2008) Family structure and the reproduction of inequalities, *Annual Review of Sociology*, 34: 257–76.

McLanahan, S. and Sandefur, G. (1994) *Growing Up with a Single Parent: What Helps, what Hurts*, Cambridge, MA: Harvard University Press.

Meira-Machado, L.F., de Uña-Alvarez, J., Cadarso-Suárez, C. and Andersen, P.K. (2009) Multi-state models for the analysis of time-to-event data, *Statistical Methods in Medical Research*, 18(2): 195–222.

Merikukka, M., Ristikari, T., Tuulio-Henriksson, A., Gissler, M. and Laaksonen, M. (2018) Childhood determinants for early psychiatric disability pension: a 10-year follow-up study of the 1987 Finnish Birth Cohort, *International Journal of Social Psychiatry*, 64(8): 715–25, doi: 10.1177/0020764018806936.

Ness, R.B. (2013) Counterpoint: the future of innovative epidemiology, *American Journal of Epidemiology*, 177(4): 281–2.

Nurminen, M. (2011) Working-life expectancy in Finland: development in 2000–2009 and forecast for 2010–2015, Eläketurvakeskus, https://www.julkari.fi/handle/10024/129058, (Accessed: 22 Feb 2021).

OECD (Organisation for Economic Co-operation and Development) (2010) *Sickness, Disability and Work: Breaking the Barriers – A Synthesis of Findings Across OECD Countries*, Paris: OECD Publishing, doi: 10.1787/9789264088856-en.

Paananen, R. and Gissler, M. (2012) Cohort profile: the 1987 Finnish birth cohort, *International Journal of Epidemiology*, 41(4): 941–5, doi: 10.1093/ije/dyr035.

Patel, V., Flisher, A.J., Hetrick, S. and McGorry, P. (2007) Mental health of young people: a global public-health challenge, *The Lancet*, 369(9569): 1302–13.

Piccarreta, R. and Studer, M. (2019) Holistic analysis of the life course: methodological challenges and new perspectives, *Advances in Life Course Research*, 41: art 100251, doi: 10.1016/j.alcr.2018.10.004.
R Core Team (2020) *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing, https://www.r-project.org/.
Rabiner, L. (1989) A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, 77(2): 257–86.
Ristikari, T., Törmäkangas, L., Lappi, A., Haapakorva, P., Kiilakoski, T., Merikukka, M., Hautakoski, A., Pekkarinen, E. and Gissler, M. (2016) *Suomi Nuorten Kasvuympäristönä – 25 Vuoden Seuranta Vuonna 1987 Suomessa Syntyneistä Nuorista Aikuisista [Finland as a Growth Environment for Children – 25-year Follow-up of Those Born in Finland 1987]*, Report 9/2016, Helsinki: THL.
Roux, J., Grimaud, O. and Leray, E. (2019) Use of state sequence analysis for care pathway analysis: the example of multiple sclerosis, *Statistical Methods in Medical Research*, 28(6): 1651–63.
Shiri, R., Halonen, J., Serlachius, A., Hutri-Kähönen, N., Raitakari, O.T., Vahtera, J., Viikari, J. and Lallukka, T. (2021) Work participation and physicality of work in young adulthood and the development of unhealthy lifestyle habits and obesity later in life: a prospective cohort study, *Occupational and Environmental Medicine*, 78(3): 153–9.
Statistics Finland (2013) Oppilaitostilastot 2012 [Educational institution statistics 2012], https://www.stat.fi/tup/julkaisut/tiedostot/julkaisuluettelo/ykou_opla_201200_2013_10094_net.pdf.
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.
Studer, M. (2013) *WeightedCluster Library Manual: A Practical Guide to Creating Typologies of Trajectories in the Social Sciences with R*, LIVES Working Papers No. 24, Lausanne: LIVES – Swiss Centre of Expertise in Life Course Research.
Studer, M. and Ritschard, G. (2016) What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(2): 481–511.
Virtanen, M., Lallukka, T., Kivimäki, M., Alexanderson, K., Ervasti, J. and Mittendorfer-Rutz, E. (2020) Neurodevelopmental disorders among young adults and the risk of sickness absence and disability pension: a nationwide register linkage study, *Scandinavian Journal of Work, Environment & Health*, 46(4): 410–16.
WHO (World Health Organization Regional Office for Europe) (2012a) Ageing and health in focus in 2012, https://web.archive.org/web/20210224173828/https://www.euro.who.int/en/health-topics/Life-stages/healthy-ageing/news/news/2012/01/ageing-and-health-in-focus-in-2012, (Accessed: 22 Feb 2021).
WHO (World Health Organization Regional Office for Europe) (2012b) Executive summary of the European health report 2012: Moving Europe towards health and well-being, EUR/RC62/Inf.Doc./1, https://www.euro.who.int/__data/assets/pdf_file/0017/170108/RC62-id01-Eng.pdf, (Accessed: 4 Sep 2022).
Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*, 2nd edn, New York: Springer.

Appendix

Appendix A: model estimation

Sequence analysis

Sequence analysis was used as an intermediary step to help with setting up informative starting values for the parameters of the more complex mixture Markov model. The results of SA were not used directly in the final analyses. As sequence dissimilarity measures we considered measures that are known to be sensitive to the duration of spells spent in one state, namely optimal matching (OM) and Euclidean distance (Studer and Ritschard, 2016). We then used Ward's agglomerative clustering algorithm for grouping sequences into relatively homogeneous clusters. For the choice of the dissimilarity measure, we used a sample of 15,000 sequences to reduce computation times, and then chose the measure with more meaningful clusters. Finally, we used the OM measure for clustering the full sample of 60-month sequences and chose the number of clusters by subjective assessment of meaningfulness as well as using a range of cluster quality indices suggested by Studer (2013).

We chose the ten-cluster solution as producing the most meaningful classification. Up to ten clusters the substantive meaning of each cluster was useful and each cluster was clearly distinctive, while after ten clusters the substantive differences were small. As shown by Figure A-1, the cluster quality indices did not give a clear indication of one solution being preferable to another (except at five clusters, shown as a peak – this solution, however, was not distinctive enough given the complexity of the data). As the SA part was only a preliminary step, we do not present the full results here, but only give short descriptions and proportions in Table A-1. For an illustration of the number of clusters ranging from 2 to 15 for OM and Euclidean dissimilarities, see the supplementary material.

Mixture Markov model

A simple Markov (chain) model for sequence data can be described with using the following notations and probabilities:

Figure A-1: Cluster quality indicators by the number of clusters

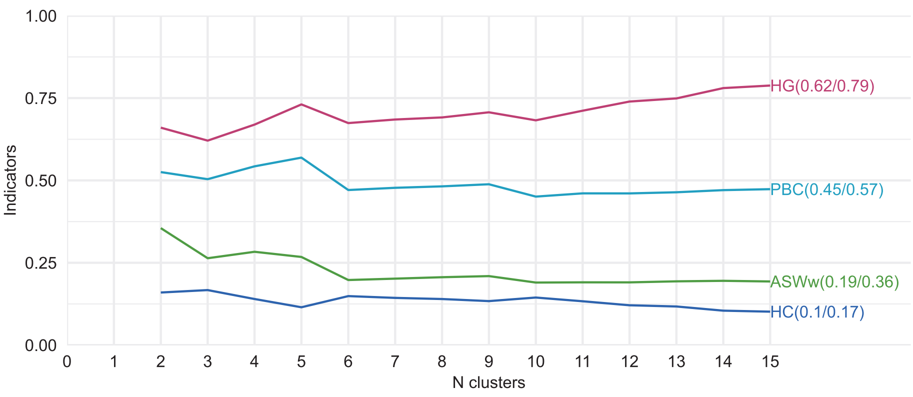


Table A-1: Description of sequence clusters

Description	**Male**	**Female**	**Total**
Continuous employment	6,554 (29.61%)	5,155 (24.98%)	11,709 (27.38%)
Studying (+ summer jobs)	3,069 (13.87%)	2,139 (10.36%)	5,208 (12.18%)
Employed and studying	3,184 (14.39%)	5,332 (25.83%)	8,516 (19.91%)
Unknown break from work (military service)	6,189 (27.97%)	1,462 (7.08%)	7,651 (17.89%)
Social assistance	1,555 (7.03%)	470 (2.28%)	2,025 (4.73%)
Short study break from work	881 (3.98%)	1,757 (8.51%)	2,638 (6.17%)
Fast transition to parental leave	43 (0.19%)	2,367 (11.47%)	2,410 (5.63%)
Slow transition to parental leave	38 (0.17%)	1,568 (7.60%)	1,606 (3.75%)
Mainly unknown or living in parental home	439 (1.98%)	174 (0.84%)	613 (1.43%)
Health-related exit from employment	179 (0.81%)	215 (1.04%)	394 (0.92%)
Total	22,131 (100%)	20,649 (100%)	42,770 (100%)

- y_{it}: observation of individual *i* at time *t*
- *s:* state (or status: employed, studying, unemployed and so on)
- $\pi(s)$: Probability, that a sequence starts in state *s* (initial probability)
- $a(s,r)$: Probability to transition from state *s* to state *r* (transition probability)

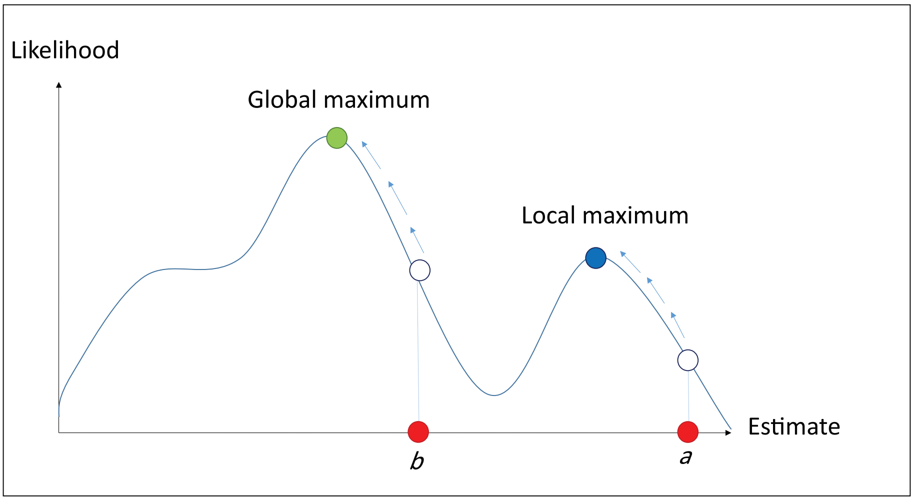
The mixture Markov model (MMM) expects that the population consists of latent classes (clusters) with varying employment patterns and allows for different specifications of initial and transition probabilities between these classes. Each *submodel* M_k consists of initial probabilities $\pi_k(s)$ and transition probabilities $a_k(s,r)$, with w_k defining the class membership probability for latent class *k*, which here depends on the covariates x_i. The (log) likelihood of the model is calculated as

$$\begin{aligned}\log L &= \sum_{i=1}^{N} \log P\left(y_i|x_i,\ M_k\right) = \sum_{i=1}^{N} \log \sum_{k=1}^{K} P\left(M_k|x_i\right) P\left(y_i|M_k\right)\\ &= \sum_{i=1}^{N} \log \sum_{k=1}^{K} P\left(\ M_k|x_i\right) P\left(y_{i1}|M_k\right) \prod_{t=2}^{T} P\left(y_{it}|y_{i(t-1)},\ M_k\right)\\ &= \sum_{i=1}^{N} \log \sum_{k=1}^{K} w_k \pi_k(y_{i1}) \prod_{t=2}^{T} a_k(y_{i(t-1)}, y_{it})\end{aligned}$$

where *N* is the number of individuals, *K* is the number of classes (submodels), *T* is the number of time points, and x_i includes the covariates for individual *i*.

The maximum likelihood estimates of the probabilities are typically calculated with the expectation–maximisation (EM) algorithm (Baum and Petrie, 1966; Rabiner, 1989). The EM algorithm is iterative, which means that we make a guess at the values of the parameters (or often use random starting values) and the algorithm then changes these values until it has found a solution that (locally) maximises the value of the likelihood function. To reduce the risk of being trapped in a poor local maximum (and ending up with a suboptimal model), estimation should be started multiple times from different starting values. See Figure A-2 for a simple illustration of the problem. The full estimation process can be extremely slow, so using meaningful starting values are often needed for finding an optimal solution in a reasonable time.

Figure A-2: Illustration of the logic of the EM algorithm and the problem of local maxima with a simple model with only one unknown probability

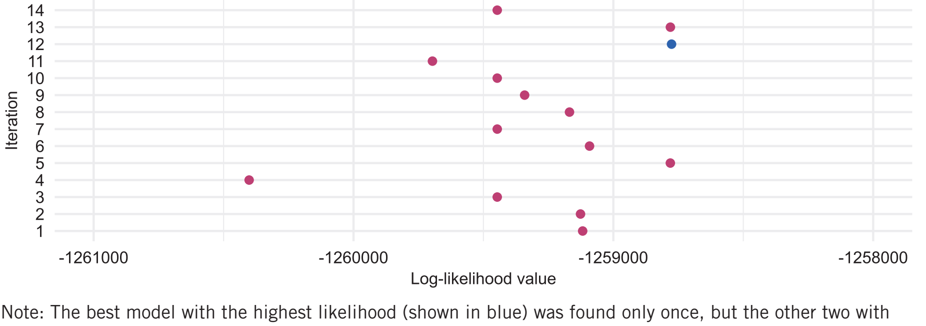


Note: Using the starting value a leads to a suboptimal local maximum, while starting from value b the algorithm finds the global optimum (the model with the highest likelihood).

To facilitate the analysis, we used an approach described in Helske et al (2016) who used SA as a preliminary step before estimation of a mixture hidden Markov model. The outline of our analytical strategy is thus as follows:

1. *Apply sequence analysis and cluster analysis* to determine the number of latent classes. At this step, we only accounted for individuals whose work participation was recorded for the full 60 months, as accounting for missing information in SA can be problematic. We then used the optimal matching algorithm for calculating sequence dissimilarities and Ward's clustering algorithm for identifying groups with similar labour-market participation trajectories.
2. *Estimate Markov models independently for each cluster* of work-participation sequences from Step 1. At this step, we used the clusters of 60-month work-participation sequences and estimated a simple Markov model independently for each cluster.

Figure A-3: Log-likelihoods of 14 estimations of the MMM using the EM algorithm, starting from different starting values



Note: The best model with the highest likelihood (shown in blue) was found only once, but the other two with nearly similar log-likelihood values produced nearly identical models.

Table A-2: Mean probability of latent classes (columns) by the most probable latent class (rows) from the mixture Markov model for work participation sequences

Latent class	A	B	C	D	E	F	G	H	I	J
A	**74.90**	10.97	2.44	8.27	1.44	0.08	1.00	0.36	0.12	0.44
B	7.93	**81.08**	3.49	1.36	3.16	0.10	0.47	0.80	1.08	0.54
C	0.85	7.83	**87.38**	0.06	1.32	0.08	1.00	0.31	0.42	0.76
D	4.97	3.30	0.00	**84.13**	2.99	0.70	2.20	0.11	0.08	1.52
E	2.17	4.53	1.68	2.76	**85.51**	0.00	0.01	0.09	0.00	3.25
F	0.01	0.12	0.02	0.95	0.00	**79.05**	3.89	6.26	7.68	2.00
G	0.26	0.76	0.51	2.23	0.00	2.85	**81.63**	0.47	9.65	1.62
H	0.67	1.06	0.20	0.17	0.09	5.38	1.05	**83.19**	5.11	3.09
I	0.04	1.44	0.14	0.11	0.00	5.49	8.69	3.51	**79.89**	0.70
J	0.64	0.99	0.77	2.07	3.73	1.92	2.48	3.57	0.73	**83.09**

3. *Estimate a MMM with covariates* using all sequences with 12–60 month follow-up and the number of clusters decided at Step 1 as the number of latent classes and the estimated initial probabilities π_s and transition probabilities a_{sr} from Step 2 as meaningful starting values for the submodels of the MMM. We tested between randomised starting values (using the default randomisation option for multiple EM restarts in the seqHMM package) and chose the final best-fitting model according to the likelihood function.

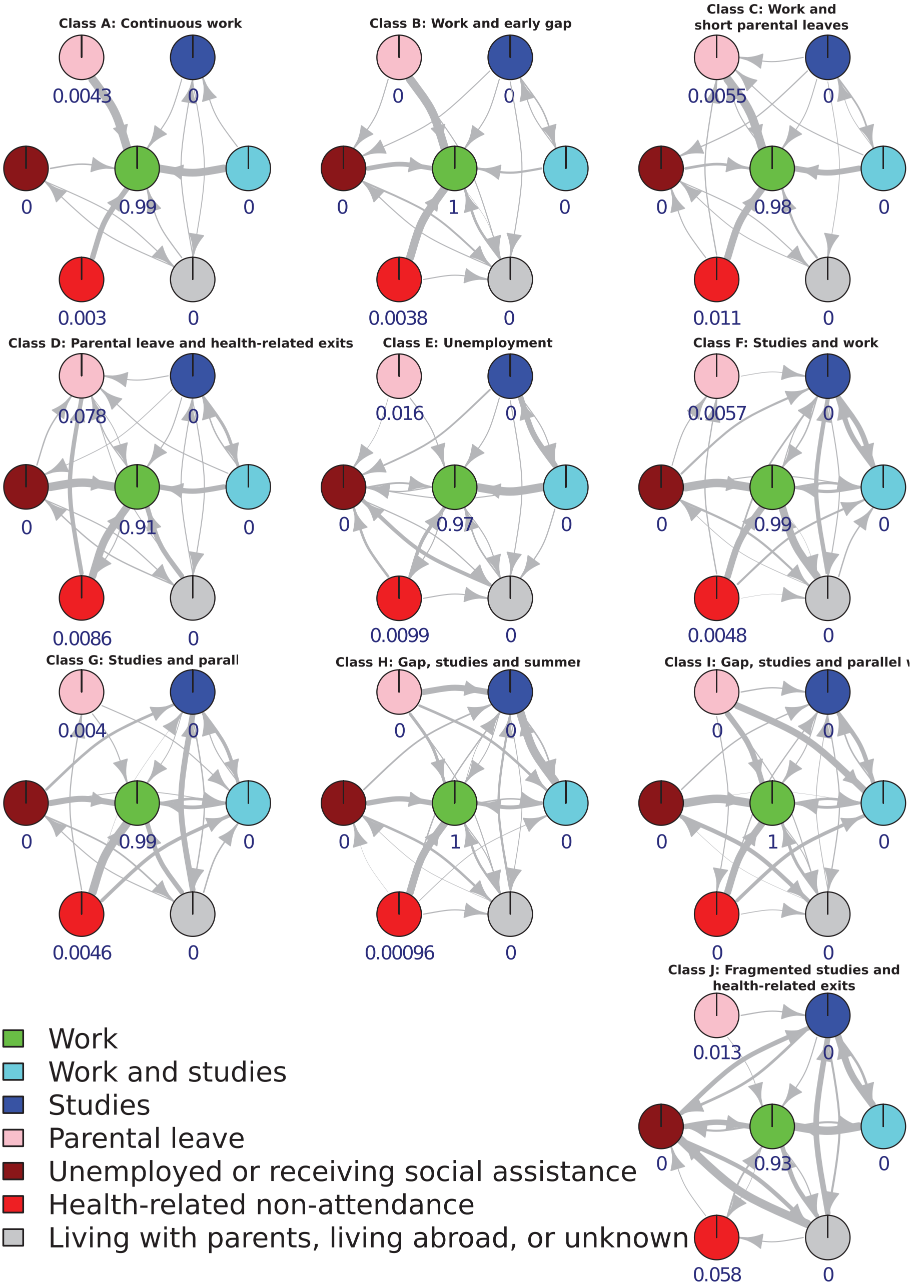
Fixing the number of submodels based on the results of Step 1 saved us from having to estimate several MMMs with different numbers of latent classes, and the probabilities estimated at Step 2 allowed us to greatly reduce the estimation time by using starting values that are plausible and probably relative close to the model with the highest likelihood.

In order to avoid being trapped in a local maximum, we estimated the model 16 times (of which 14 estimations were successful) and as per convention chose the model with the highest likelihood as the final model (Figure A-3).

Table A-2 presents the mean probabilities of each latent class by the most probable latent class. The diagonal (bold) shows the mean probability of the respective class; for example, the average membership probability for Class A (Continuous work) was 75% for individuals with the highest class probability for Class A and in general these ranged from 75% for Class A to 87% for Class C (Work and short parental leaves). High probabilities on the diagonal and low probabilities elsewhere indicate highly distinctive classes with little membership uncertainty. As can be seen, we had some non-negligible uncertainty for some clusters, but nothing too alarming. The highest off-diagonal probability was found for Class A with an 11% probability for being assigned to Class B (Work and early gap) instead, so even in the worst case the probability on the diagonal was almost sevenfold to the highest off-diagonal probability.

Figure A-4 shows the transition probabilities between labour-market states in each latent Class A–J. The numerical values for the estimated parameters of the mixture Markov model are given as supplementary material.

Figure A-4: Descriptions of latent classes from the mixture Markov model



Notes: The nodes (circles) describe the work participation states and the arrows indicate the monthly transition probabilities between them – the thicker the arrow, the higher the probability.

Numbers below the nodes are the estimated starting probabilities for each work participation status. Transition probabilities below 1% and probabilities of staying in each state are omitted from the figure for clarity.

Appendix B: Descriptive statistics

Figure B-1: Proportion of work spells of different lengths in the study sample (all work spells)

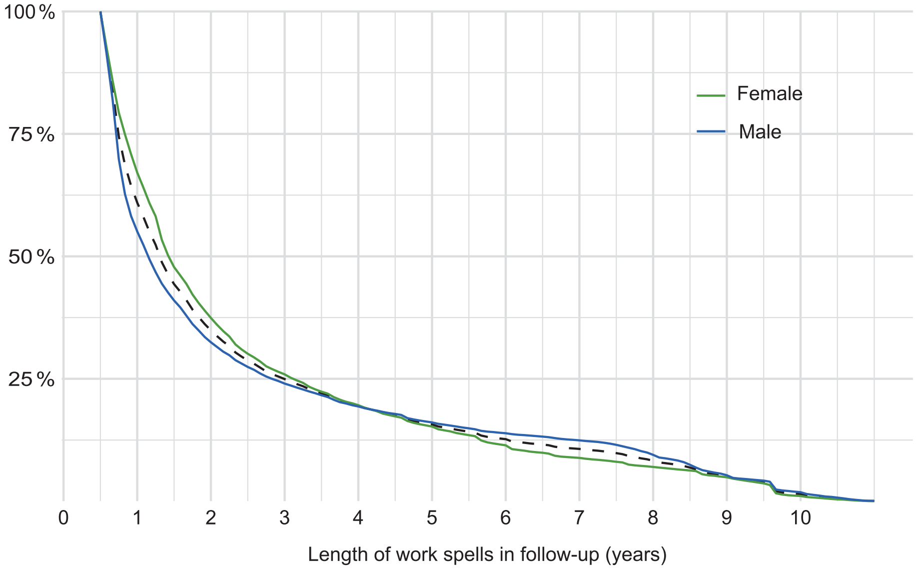
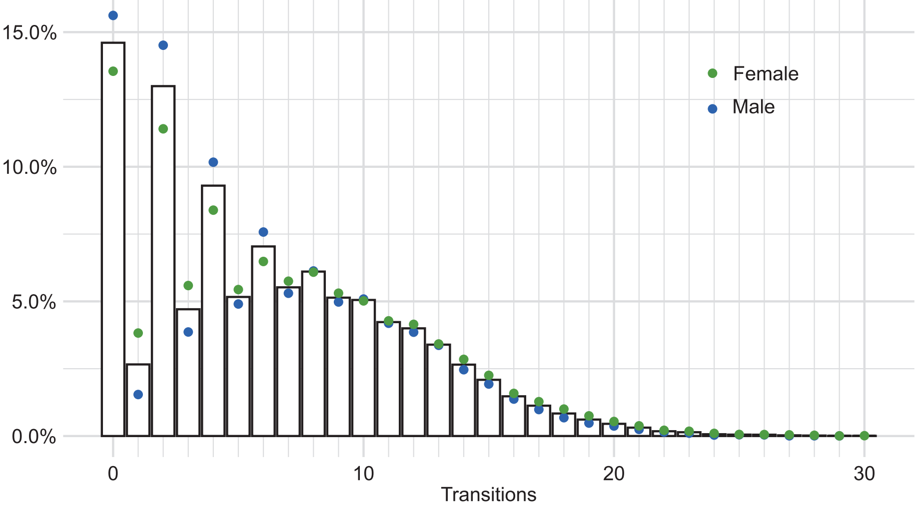


Figure B-2: Number of transitions between different work participation statuses in the study sample



Appendix C: Estimated model parameters

Table C-1: Estimates for log odds of Figure 4

Variable	B	C	D	E	F	G	H	I	J
Age of entry: 19–22 (ref. 17–18)	−1.55	−0.47	−0.02	0.09	−0.62	−0.87	−1.05	−1.75	−0.06
Age of entry: 23–27 (ref. 17–18)	−3.30	−0.55	−0.17	0.30	−4.03	−2.81	−4.90	−5.69	−1.08
Gender: female (ref. male)	−1.85	−37.36	7.69	−0.24	1.25	0.75	−1.33	−1.22	0.34
GPA: missing (ref. 8.50–10)	0.75	0.34	0.25	2.09	−1.24	−1.27	−2.05	−2.39	0.47
GPA: 4.00–6.99 (ref. 8.50–10)	0.87	0.50	0.73	2.1	−2.14	−1.87	−2.77	−3.32	0.30
GPA: 7.00–7.49 (ref. 8.50–10)	0.75	0.08	0.50	1.36	−1.21	−1.05	−1.50	−1.97	0.38
GPA: 7.50–7.99 (ref. 8.50–10)	0.73	0.14	0.51	1.05	−0.59	−0.50	−0.72	−1.03	0.52
GPA: 8.00–8.49 (ref. 8.50–10)	0.45	0.02	0.24	0.65	−0.25	−0.18	−0.28	−0.42	0.37
Cohort member's conviction: yes (ref. no)	−0.30	0.37	−0.14	0.69	−0.27	−0.52	−0.46	−1.27	−0.09
Cohort member's psychiatric diagnosis: yes (ref. no)	0.22	0.08	0.55	0.94	0.03	0.15	−0.39	−0.16	1.24
Cohort member's teenage pregnancy: yes (ref. no)	−0.30	−4.36	0.47	0.41	−0.43	0.02	−2.84	−1.09	0.05
Nuclear family: yes (ref. no)	0.20	0.15	−0.04	−0.13	0.31	−0.16	0.41	0.33	−0.15
Parental education: yes (ref. no education)	0.06	−0.19	−0.3	−0.48	0.57	0.53	0.71	0.91	0.14
Parental psychiatric diagnosis: yes (ref. no)	−0.14	−0.18	−0.05	0.03	0.04	−0.06	−0.13	−0.10	−0.01
Social assistance: yes (ref. no)	−0.08	0.45	0.31	0.63	−0.28	−0.11	−0.44	−0.38	0.40
Teenage mother: yes (ref. no)	−0.28	0.52	0.56	0.32	−0.06	−0.03	−0.52	−0.57	0.28

Table C-2: Estimates for odds of Figure 4

Variable	B	C	D	E	F	G	H	I	J
Age of entry: 19–22 (ref. 17–18)	0.21	0.62	0.98	1.10	0.54	0.42	0.35	0.17	0.94
Age of entry: 23–27 (ref. 17–18)	0.04	0.58	0.85	1.35	0.02	0.06	0.01	0	0.34
Gender: female (ref. male)	0.16	0	2193.7	0.78	3.47	2.12	0.26	0.3	1.40
GPA: missing (ref. 8.50–10)	2.11	1.41	1.28	8.09	0.29	0.28	0.13	0.09	1.60
GPA: 4.00–6.99 (ref. 8.50–10)	2.38	1.64	2.06	8.14	0.12	0.15	0.06	0.04	1.35
GPA: 7.00–7.49 (ref. 8.50–10)	2.12	1.09	1.64	3.91	0.3	0.35	0.22	0.14	1.46
GPA: 7.50–7.99 (ref. 8.50–10)	2.08	1.15	1.66	2.85	0.55	0.61	0.49	0.36	1.68
GPA: 8.00–8.49 (ref. 8.50–10)	1.57	1.02	1.28	1.92	0.78	0.84	0.76	0.65	1.45
Cohort member's conviction: yes (ref. no)	0.74	1.44	0.87	2.00	0.76	0.59	0.63	0.28	0.92
Cohort member's psychiatric diagnosis: yes (ref. no)	1.25	1.08	1.73	2.55	1.03	1.17	0.68	0.85	3.45
Cohort member's teenage pregnancy: yes (ref. no)	0.74	0.01	1.60	1.51	0.65	1.02	0.06	0.34	1.05
Nuclear family: yes (ref. no)	1.23	1.17	0.96	0.88	1.36	0.86	1.51	1.4	0.86
Parental education: yes (ref. no education)	1.06	0.83	0.74	0.62	1.77	1.7	2.04	2.49	1.15
Parental psychiatric diagnosis: yes (ref. no)	0.87	0.83	0.95	1.03	1.04	0.94	0.88	0.91	0.99
Social assistance: yes (ref. no)	0.92	1.56	1.36	1.87	0.76	0.90	0.65	0.69	1.50
Teenage mother: yes (ref. no)	0.76	1.68	1.76	1.38	0.94	0.97	0.59	0.57	1.32

Appendix D: Variable definitions for exemplary cases for predicted probabilities

Table D-1: Example cases and their covariate values.

	Sex	GPA	Intact family	Par. educ. high	Soc. assist.	Parent psych. diag.	Own psych. diag.	Teenaged mother	Teenage preg.	Con-viction
Reference	F	7.5–8	Yes	No	No	No	No	No	No	No
Reference	M	7.5–8	Yes	No	No	No	No	No	No	No
Advantaged	F	8.5–10	Yes	Yes	No	No	No	No	No	No
Advantaged	M	8.5–10	Yes	Yes	No	No	No	No	No	No
Typical disadv.	F	4–6.99	No	No	Yes	Yes	No	No	No	No
Typical disadv.	M	4–6.99	No	No	Yes	Yes	No	No	No	No
Female-type disadv.	F	4–6.99	No	No	Yes	Yes	No	No	Yes	No
Male-type disadv.	M	4–6.99	No	No	Yes	Yes	No	No	No	Yes

The full article for this supplementary appendix can be found at https://doi.org/10.1332/175795921X16609201864155