INVEST

INVEST Working Papers 25/2021

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

Satu Helske
Markus Keski-Säntti
Juha Kivelä
Aapo Juutinen
Antti Kääriälä
Mika Gissler
Marko Merikukka
Tea Lallukka

11.3.2021

The Inequalities, Interventions, and New Welfare State (INVEST) aims at increasing wellbeing of Finnish society during childhood, youth and early adulthood and preventing psychosocial risks compromising such development through innovative interventions. Based on cutting-edge research on the conditions and mechanisms involved at different periods of development, INVEST will evaluate and develop various universal and targeted interventions to improve the efficiency of the current welfare state institutions at critical points of the early life course. INVEST aims at providing a new model for the welfare states that is more equal, better targeted to problem groups, more anticipatory as well as economically and socially sustainable. INVEST is a Flagship project of the Academy of Finland.

ACADEMY OF FINLAND
FLAGSHIP PROGRAMME

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

Satu Helske[1], Keski-Säntti Markus[2], Juha Kivelä[2], Aapo Juutinen[2], Antti Kääriälä[2], Mika Gissler[2,3], Marko Merikukka[4], Tea Lallukka[5]

1 INVEST Research Flagship Center and Department of Social Research, University of Turku, Turku, Finland
2 Finnish Institute for Health and Welfare, Helsinki, Finland
3 Karolinska Institute, Department of Neurobiology, Care Sciences and Society, Stockholm, Sweden
4 Itla Children's Foundation, Helsinki, Finland
5 Department of Public Health, University of Helsinki, Helsinki, Finland

Corresponding author

Satu Helske

Department of Social Research, University of Turku
FI-20014 University of Turku, Finland
satu.helske@utu.fi

Abstract

Background and objectives: To extend work careers, it is important to focus on all working-aged people, including young adults. However, there is little knowledge about the stability of early careers and its life course social and health-related determinants. The aim of this study is to identify what kind of typical patterns of work participation and transitions between different statuses can be identified among young adults after their first entry into the labour market, and to examine whether the timing of entry together with parental and own socioeconomic position and health predict early work participation patterns.

Methods: 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 the end of 2015. We created sequences of monthly work participation statuses starting from the calendar month when a cohort member had their first at least six-month long spell in employment or entrepreneurship. Work participation statuses (8 categories) were collected for 60 months or until the end of the follow-up. We included altogether 51,871 individuals with an entry into paid employment, who could be followed for at least 12 subsequent months. Parental social and health-related covariates were included from national registers. We used first sequence analysis and cluster analysis to identify latent groups of individuals with similar work participation patterns. Next, we estimated a mixture Markov model that allows for 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.

Results: 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. 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 at a later age were more likely to be in continuous employment thereafter (all else equal). Regarding both own and parental social- and health-related childhood factors, latent classes show clear patterns of accumulated advantage and disadvantage: 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.

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.

1. Introduction

Due to fallen birth rates in many countries, and increased life expectancy, the proportion of people in the workforce has dramatically declined. Working life expectancy has also been increasing but remains relatively short[1]. Many studies focus on the older employees and their routes of early exit or 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, and 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. For example, exposure to physically heavy work or psychosocial stress could be linked to more adverse outcomes during sensitive periods. Accordingly, an earlier birth cohort study showed that socioeconomic background and unhealthy lifestyle are linked to early entry into paid employment and also to a lower occupational class in the first job[2]. Additionally, a recent study showed that early entry into paid employment as well as physical heaviness of work in young adulthood shape the development of unhealthy behaviours and obesity in the long run[3]. Hence, the timing of entry could be shaped by both parental and own socioeconomic 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 is linked to different labour market transitions such as unemployment during early careers and particularly the stability of employment. For example, it is unclear whether the effect of timing is linear or nonlinear in that both an early and also a very late transition would predict lower labour market attachment. Furthermore, it is unclear how the accumulation of different life-course social and health-related factors are associated with the transitions between different employment statuses in young employees.

Labour market statistics show that proportions of those 25 years or younger, as those for 55 years or older in the workforce have remained much lower as compared to midlife groups[4,5]. Thus, there is a need to extend work careers by focusing on the determinants of work participation patterns among young employees particularly, as among all the working aged. Reasons for unstable 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, e.g. be linked to both parental and own socioeconomic position and health. Particularly 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[2,6]. As mental disorders increase already among young adults, even before entering paid employment, they need to be considered when focusing on work participation among young adults[7–9]. Accordingly, recent studies using register data from Sweden have shown that among mental disorders, e.g. neurodevelopmental disorders are typically diagnosed young, and are strongly associated with subsequent work disability[10] and unemployment trajectories[11]. Additionally, a substantial part of early exit from paid employment is attributable to these disorders, thus the associations can be bidirectional with early ill health predicting unstable work participation, and similarly non-employment is a cause for ill health later on[12–14].

Early detection of modifiable risk factors of unstable work participation patterns is crucial. However, there is limited knowledge about the factors contributing to the opportunities of e.g. disabled to enter the labour market and continue working[15]. Moreover, most studies focus on one exit type, or time to exit e.g. 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. Thus, young adults are unlikely 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.

To sum, more evidence is needed, what are the paths to entry and exit in young adulthood, what kind of work participation patterns can be identified, transitions between different statuses, and what the determinants of different work participation patterns are. Importance of such life course approaches including childhood factors has been also stressed elsewhere[16,17].

We use 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 work participation as it is. Sequence analysis has been increasingly used to study the sequencing of life events and other longitudinal phenomena, including labour market participation, family formation, and health[18–21]. 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 association between predictors and different types of patterns[22].

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 e.g. health-related absences and unemployment. Furthermore, it is hypothesised that young adults who enter the labour market earlier, have more unstable work participation, which 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 to more manual work with adverse exposures. Thus, individuals who enter the labour market early are hypothesised to have less and shorter periods of sustained work participation as compared to those who begin their working life later, after longer education.

Using register data comprising an entire birth cohort and a flexible person-oriented mixture Markov modelling approach, the study therefore aims to first examine what kind of work participation patterns can be identified among young adults in their early careers, and second, how parental and own socioeconomic factors and health as well as age at entry into the labour market jointly determine such patterns. We focus on age at entry in the first long-term period of paid employment, and what kind of latent clusters can be identified in subsequent work participation after entry into paid employment and how the accumulation of a number of childhood and adolescence social and health-related determinants predict different types of work participation patterns depending on the age at entry. Additionally, the aim is in the identification of potentially sensitive periods in late adolescence and earlier that could indicate a higher risk of more unstable work participation.

2. Methods

[redacted]Data and variables

The Finnish Birth Cohort 1987 (FBC)[23,24] 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 current study 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), unemployment periods data from the Finnish Ministry of Economic Affairs and Employment, social assistance data from the Finnish Institute for Health and Welfare (THL), education level and date of death from Statistics Finland, and comprehensive school achievement data from the Finnish National Board of Education. All the register datasets were merged using personal identification numbers assigned to each Finnish citizen[25]. More precise information on the data and variables can be found on the FBC metadata web page[26].

We created sequences of monthly work participation statuses starting from the calendar month in which the individual starts their first six-month long spell in employment or entrepreneurship without being in education at the same time (at age 18 the earliest). We recorded the work participation status for the first 60 months (5 years) or until the end of the follow-up. We classify the work participation statuses into eight categories, presented in Table 1: Work participation statuses and their priority. 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, i.e., not being in employment, education, or a recipient of a benefit..

Table 1: Work participation statuses and their priority. 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, i.e., not being in employment, education, or a recipient of a benefit.

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

Altogether 59,476 children were born in Finland in 1987. Among them, we only included individuals who had had their first recorded, at least one six-month long spell in employment, by the end of the follow-up (December 2015), ending up with 54,010 individuals (91% of the cohort). We also excluded a small number of individuals due to serious disabilities or longer spells of censoring. First of all, we excluded 129 individuals who had deceased by the beginning of 2005 or during a five-year follow-up of work participation. Second, we dropped 1329 individuals who had lived abroad for more than 12 consecutive months by the end of the follow-up, as changes in their work participation status could not be followed using national registers. Furthermore, we dropped 92 individuals for having received a diagnosis for an intellectual disability and 116 individuals for receiving disability allowance at the middle or highest rate on the beginning of 2005, another 127 individuals who had deceased by the beginning of 2005 or during a five-year follow-up of work participation, and finally 1328 individuals 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.

We ended up with 52,344 individuals who had started a six-month work spell (88% of the cohort). 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 full 60 months (72% of the cohort and 79% of the target population). See section 3 for more descriptive statistics on the sample and the variables.

The timing of labour market entry was categorized into three categories: 17–18 (28.5% of the sample with a work spell of at least 12 months; note that an entry at 17 was very rare), 19–22 (50.5%), and 23–26 (21.0%). We used this categorized variable to allow for possible nonlinear relationships between the timing of entry and the work participation pattern.

Parental social and health-related covariates included five binary variables. All time-series variables were measured until the end of 2004.

- *High parental education*: whether at least one parent has a tertiary degree.
- *Nuclear family*: whether the cohort member and parents occupied the same household at the end of follow-up.

- *Social assistance*: whether the family received social assistance. Coded 1 if parents had received over 12 months of social assistance in total or at least 7 months in a single year by the end of follow-up.
- *Teenage mother*: whether the mother was a teenager when giving birth to the cohort member.
- *Parental psychiatric diagnosis*: whether a parent had a psychiatric diagnosis (ICD-9: 290–319, ICD-10: F00–99). These diagnoses were sampled from Care Register for Health Care which includes only specialised health care visits. Inpatient care visits were included in the variable formation from 1987 onwards whilst outpatient care visits from 1998 onwards.

Own social and health-related covariates included five variables:

- *Sex.*
- *Grade point average* (GPA): 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 is coded into six categories 8.50–10.00, 8.00–8.49, 7.50–7.99, 7.00–7.49, 4.00–6.99, or missing. The GPA information comes from applications to upper secondary education and is thus only available for cohort members that have applied at least once. 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).
- *Conviction*: whether they had been convicted.
- *Own teenage pregnancy*: whether they gave birth or had an abortion as a teenager.
- *Own psychiatric diagnosis*: whether the cohort member had a psychiatric diagnosis.

[redacted]Statistical methods

Sequence analysis (SA) is an exploratory data-driven approach that has become central to the life course perspective where it has been used to understand various trajectories and crucial transitions[27]. It is often used for exploratory analysis and increasingly often also used for assessing the covariate-sequence relationship. For the latter, a two-step approach is typical: first, sequence analysis 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 (either as a predictor in a linear on nonlinear regression model or as an outcome in multinomial regression). This approach, however, has been criticized for neglecting the within-cluster variation which could potentially lead to biased results if the variation is not small and random[28,29].

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 relationship of covariates and sequences. The multistate model (MSM) extends the focus of the basic event history model from one event or transition to multiple potentially recurring transitions[30,31]. 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 Markov assumption which states that the probability of a transition only depends on the current status, not on prior history. Markovian models allow for flexible analysis of life course data with multiple life statuses and large number of different transitions[22,29,32].

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

- y_{it}: observation of individual *i* at time *t*
- *s:* state (or status, e.g., employed, studying, unemployed, …)
- π_s: Probability, that a sequence starts in state *s* (initial probability)
- a_{sr}: Probability to transition from state *s* to state *r* (transition probability)

We use an extension of the basic model, the mixture Markov model (MMM), which 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 k* consists of initial probabilities k_s and transition probabilities a_{ksr}, with w_k defining the class membership probability for latent class *k*. The mixture Markov model is similar to the *latent class model* (Han, Liefbroer, & Elzinga, 2017) 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 in fact, 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 as well as initial and transition probabilities. However, allowing the same variable to influence various probabilities may hinder a proper interpretation and evaluation of its effect. Therefore, it is common practice to relate each covariate to one type of probability only[33]. We will construct a model where the covariates explain class membership probabilities w_k, 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[34]. To facilitate the analysis, we used an approach described in Helske, Helske, & Eerola (2016)[22] who use SA as a preliminary step before estimation of a mixture hidden Markov model.

We chose 10 clusters based on the SA and estimated the MMM for 10 clusters with covariates predicting latent class memberships. More detailed information on the estimation process is given in Appendix A.

[redacted] Software

All analyses were conducted with R[35]. We used the TraMineR package[36] for creating the sequences and calculating sequence dissimilarities, seqHMM[34] for estimation of Markov models and the MMM as well as for visualization of the MMM, and ggplot2[37] for other visualizations.

3. Results

[redacted] Descriptive statistics

We start by showing some descriptive statistics of the data before moving on to the analyses. **Error! Reference source not found.** 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 the most of cohort members, the first work spell is 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 tends to be fairly short, as illustrated in Figure B-1 in Appendix B.

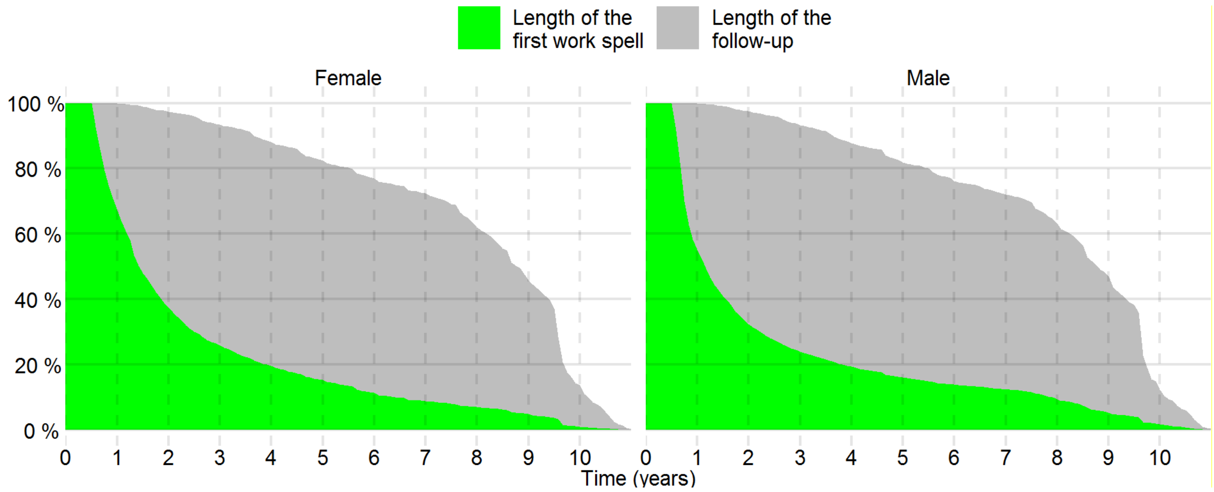


Figure 1: The length of the total follow-up after the initial entry into the labour market and the length of the first at least 6-month long work spell.

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

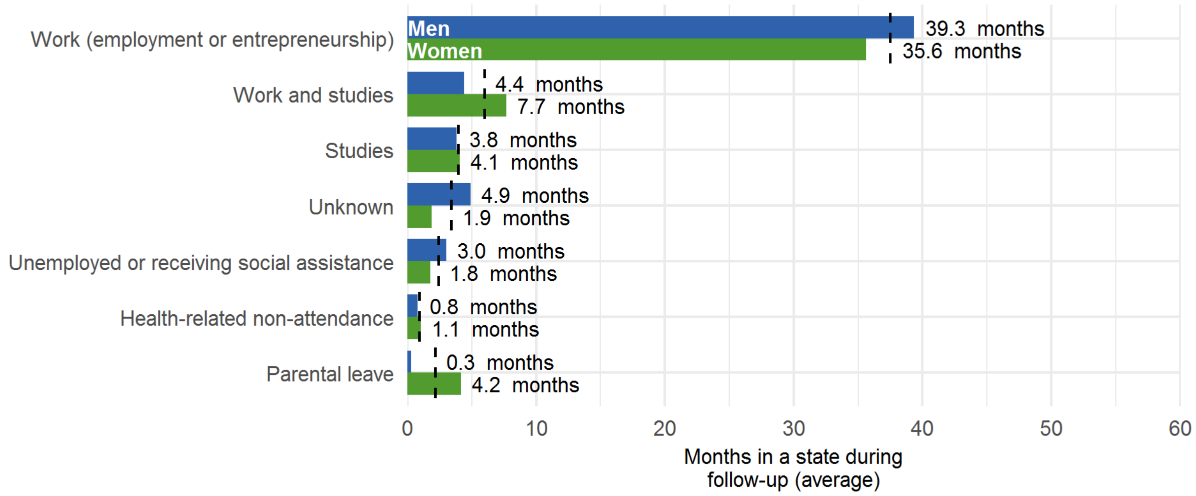


Figure 2: Mean time spent in each work participation state. The blue colour represents men and green represents women, while the dashed line shows the average for everyone.

Table 2 shows the most frequent work participation patterns in the study sample. Continuous work is the most frequent pattern, but only a small fraction, 14% of the individuals, were working continuously during the first five years after their (first) entry into the labour market. All the other patterns are much rarer and tend to be limited to temporary exits from work, most typically due to unknown status (which is frequent mainly due to military or non-military service), studying parallel to work, or being unemployed. The only permanent exit among the most frequent patterns is exiting due to parental leave. The temporal nature of exits from work can also be seen in Figure B-2 in Appendix B illustrating the numbers of transitions within individual sequences – even numbers are much more common than odd numbers.

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	7479	14.42
Work → Unknown → Work	1925	3.71
Work → Work and studies → Work	1383	2.67
Work → Health-related non-attendance → Work	1320	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	37183	71.68
Total	51 871	100

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 a half were aged 19–22; only a fifth had entered at 23–27. Men and women differ in terms of their GPA (higher for women), in having had a conviction (more often men), and their own psychiatric diagnosis (more often women).

Table 3: Descriptive statistics (frequencies and proportions) of categorical variables used in the study, for the full sample as well as separated by sex. Missing GPA refers to no or delayed application into secondary education.

	All (N = 51 871)		Male (N = 26 480)		Female (N = 25391)	
Variable	Freq	%	Freq	%	Freq	%
Age of entry: 17–18	14803	28.5	7744	29.2	7059	27.8
Age of entry: 19–22	26190	50.5	13795	52.1	12395	48.8
Age of entry: 23–27	10878	21.0	4941	18.7	5937	23.4
GPA: missing	4033	7.8	2529	9.6	1504	5.9
GPA: 4.00–6.99	13222	25.5	8782	33.2	4440	17.5
GPA: 7.00–7.49	6831	13.2	3697	14.0	3134	12.3
GPA: 7.50–7.99	7357	14.2	3702	14.0	3655	14.4
GPA: 8.00–8.49	7776	15	3520	13.3	4256	16.8
GPA: 8.50–10	12652	24.4	4250	16.0	8402	33.1
Own conviction	894	1.7	765	2.9	129	0.5
Own psychiatric diagnosis	3888	7.5	1826	6.9	2062	8.1
Own teenage pregnancy	822	1.6	0	0.0	822	3.2
Nuclear family	32845	63.3	17252	65.2	15593	61.4
Parental education	12879	24.8	6532	24.7	6347	25.0
Parental psychiatric diagnosis	7913	15.3	3999	15.1	3914	15.4
Social assistance	11158	21.5	5632	21.3	5526	21.8
Teenage mother	1599	3.1	814	3.1	785	3.1

[redacted]Latent classes of labour market participation

We will start by describing the latent classes before discussing the relationship of the latent classes and the predictors in the following section.

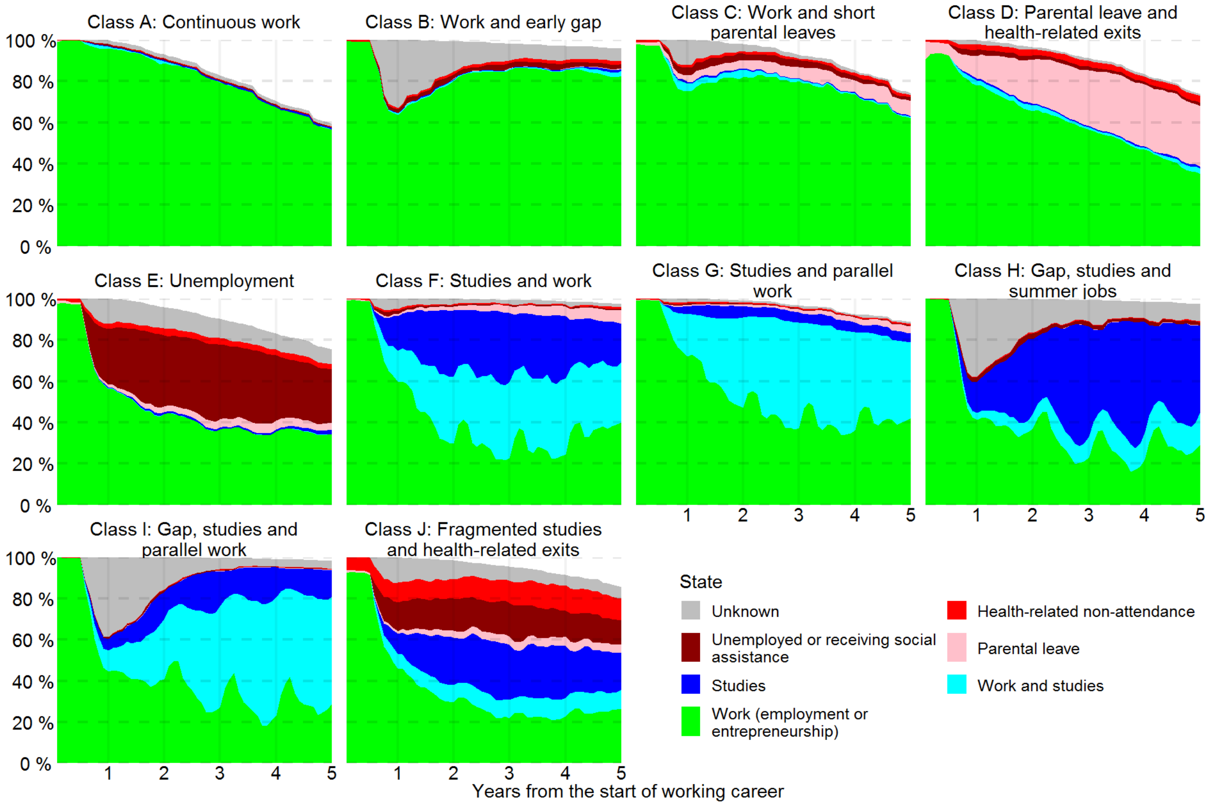


Figure 3: Latent classes from the mixture Markov model described using monthly distributions for work participation states (maximum membership assignment).

Figure 3 describes the latent classes after assigning individuals to the latent classes for which they have their highest membership probabilities. At the start, almost all individuals are in the work state (apart from a few exceptions that have a work contract but are in fact on parental leave or health-related leave of absence). Monthly transition probabilities are visualized in Figure A-4 in Appendix A. Next, we briefly describe the identified latent classes of work participation.

Class A: Continuous work (most probable class for 19% of the individuals)**.** Individuals hardly leave paid work (employment or entrepreneurship) during the follow-up – the transition probability for exiting work is 1% – but in the rare occasions that they do, they swiftly return back to work.

Class B: Work and early gap (17%)**.** In this class the individuals mainly work but often have an early gap for which we have no data. Likely for many this gap consists of military or non-military service which is mandatory by law for all men in Finland (and a voluntary option for women, quite rarely used). Another likely option is to be 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, the 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%). In this class, individuals exit work for unemployment (some are also only receiving social assistance without other main benefits and 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%)**.** Individuals in this class typically have flexible transitions between work and studies and both in parallel. Entering other states is rare.

Class G: Studies and parallel work (13%)**.** In this class, the individuals mainly study and work in parallel. They tend to have a later entry into studies and and/or their studies are short in duration. Entering other states than (full-time) work is rare.

Class H: Gap, studies and summer jobs (7%)**.** This class is characterised by recurrent transitions between studies and work, likely indicating full-time studies and summer jobs. In addition, they have spells of unknown status, often soon after first entering the labour market but also later during the follow-up. Many of these gaps likely indicate participation in military or 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%)**.** In this class the students mainly work and study in parallel. Again we observe some unknown (military-service type) gaps of unknown status, mainly soon after entering the labour market. Entering other states is rare and short in duration.

Class J: Fragmented studies and health-related exits (5%). In this class the students typically lack the usual pattern of alternating between or combining studies and work. Instead, they tend to 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.

[redacted]Predictors to early work participation patterns

Next, we turn to predictors of early work participation patterns. First, we go through how each individual predictor predicted memberships in the latent classes, after which we focus on the accumulation of advantage and disadvantage in section 3.4, showing individual predicted probabilities for some interesting exemplary cases.

Figure 4 shows estimates and 95% confidence intervals for the timing of labour market entry and cohort members’ own and parental social and health-related predictors. The estimates are interpreted similarly to the log-odds from multinomial logistic regression; the *Continuous work* class was used as the reference. Exact estimates for variable coefficients are given in supplementary material along with the rest of the estimated model parameters.

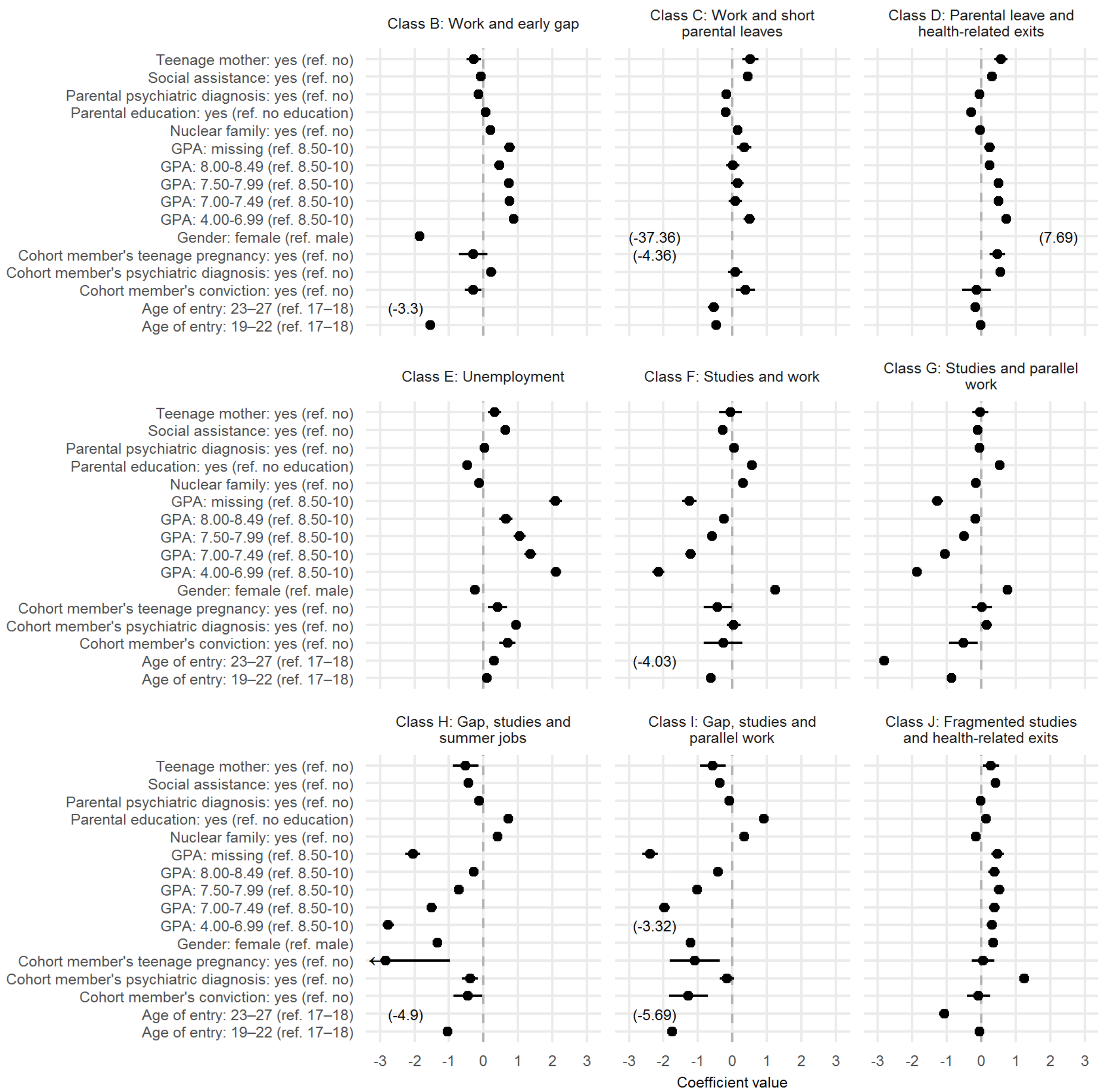


Figure 4: Estimates for log odds and their 95% confidence intervals from predictors of latent class memberships from the mixture Markov model. The reference class is Continuous work.

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 7), 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).

We did not observe nonlinear patterns between age at labour market entry and early work participation pattern in early adulthood.

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) and *Studies and parallel work* (2-fold odds). Male sex predicted higher odds for all other classes, most notably in *Work and early gap* (4-fold odds) and *Gap, studies, and summer jobs* (3-fold odds).

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 leave* class 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 (2-fold) and in the *Work and short parental leaves* class (50% increase). Not having been convicted predicted higher odds of membership in the *Work and early gap* (by 35%) and three classes of combining studies and work (G–I; by 58–360%).

Teenage pregnancy. Teenage pregnancy predicted an increase in the odds for membership by 57% in the *Parental leave and health-related exits* and by 51% in the *Unemployment* class, and reduced the odds for membership in three study classes (F, H, and I) as well as the all-male parental leave class (C).

Psychiatric diagnosis. Own 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), *Unemployment* (2.5-fold odds), and *Parental leave and health-related exits* (73% increase in odds).

Parental social and health-related predictors

Parental education. Parental education was a significant predictor of early work participation patterns, all else equal, and most importantly 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%, in the *Parental leave and health-related exits* by 35%, and in the *Work and parental leaves* by 21%.

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% and in classes *Work and early gap and Work* and *short parental leaves* by 16–22%.

Having grown up in a non-nuclear family, on the other hand, predicted higher odds for classes *Studies and parallel work, Fragmented studies and health-related exits,* and *Unemployment* by 14–17%.

Social assistance. Receiving social assistance in childhood predicts higher odds for membership in the *Unemployment* class by 88% and in classes *Work and short parental leaves*, *Fragmented studies* *and health-related exits,* and *Parental leave and health-related exits* by 36–57%. Not having received social assistance, on the other hand, predicted higher odds for membership in all classes of combining studies and work (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%, in the *Work and short parental leaves* by 68%, in the *Unemployment* class by 38%, and in the *Fragmented studies and health-related exits* by 32%. Being born to a mother who was at least 20 years of age, on the other hand, predicts higher odds for membership in the *Gap, studies and parallel work* class by 77%, in the *Gap, studies, and summer jobs* class by 68%, and in the *Work and early gap* class by 32%.

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%) and in the *Work and short parental leaves* class (by 15%).

[redacted] 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 ages at labour market entry predict different work participation patterns. We present four typical cases defined by selected risk (or protective) factors. The cases and their respective risk factors are:

- Reference case: nuclear family, intermediate GPA, low/average parental education, no risk factors;
- Accumulated advantage: nuclear family, high GPA, high parental education, no risk factors;
- “Typical” accumulated disadvantage: non-nuclear family, low GPA, social assistance, parental psychiatric diagnosis;
- Female-typical accumulated disadvantage: non-nuclear family, low GPA, social assistance, teenage pregnancy;
- Male-typical accumulated disadvantage: non-nuclear family, low GPA, social assistance, conviction.

Table C-1 in Appendix C 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 those with at least four risk factors. The sex-typical cases were chosen accordingly, but only among profiles that had differing frequencies between the two sexes.

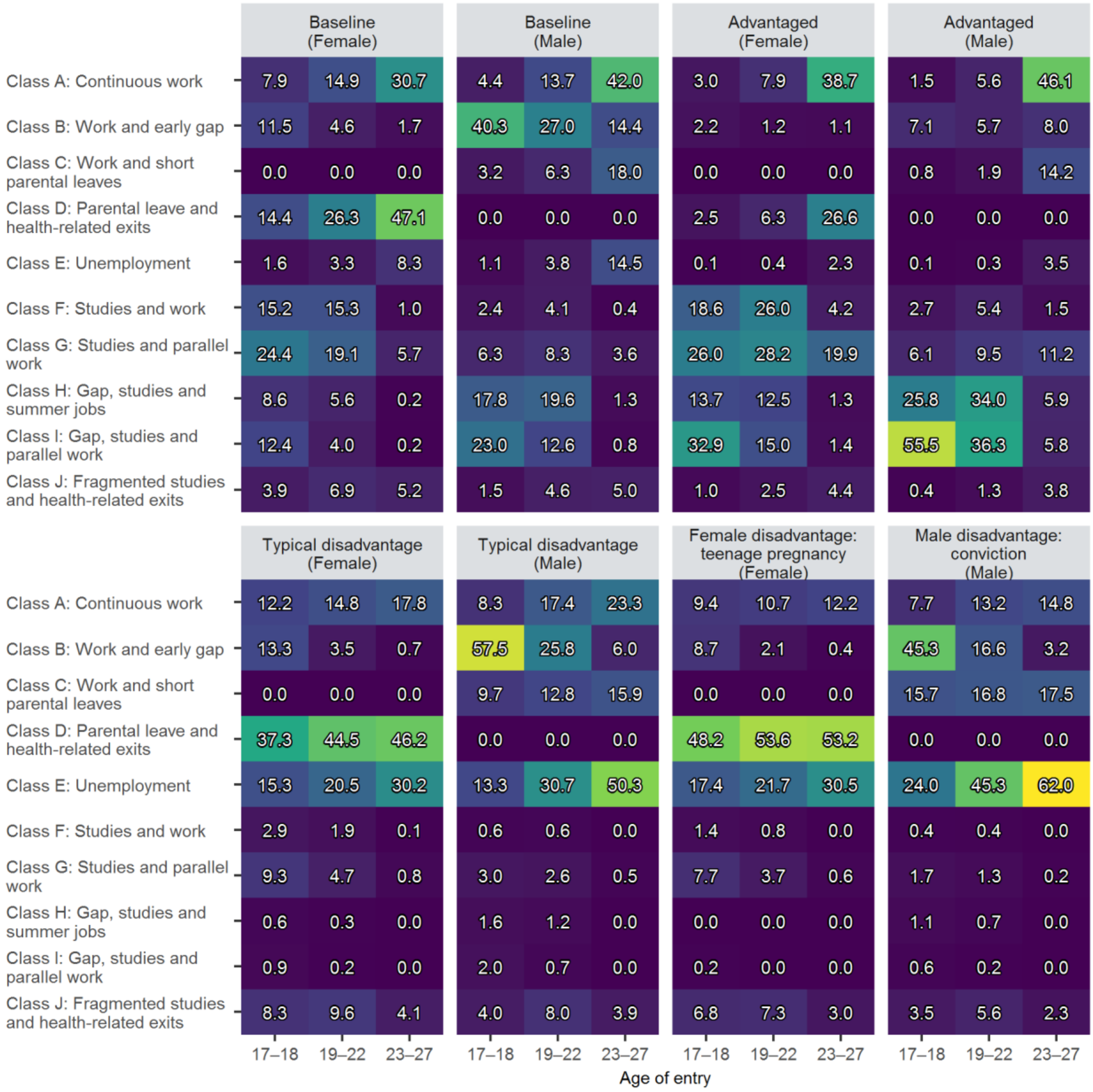


Figure 5: Probabilities for each work participation pattern (latent class) for some exemplary cases, based on the estimated mixture Markov model. The colours represent the predicted probabilities on a continuous scale from 0 (dark blue) to the highest predicted probability of 62% (yellow).

Figure 5 shows the predicted probabilities for each latent class for the example cases. For all cases, the stable employment pattern is the more likely the later the individual has entered into 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 have much higher probabilities compared to the rest of the example cases, and the probabilities are the lowest for individuals with a sex-typical disadvantaged background. Among the late entrees, the continuous work class is the most likely for women and men with an advantaged background while for women in the other groups the *Long parental leave and health-related exits* class is the most likely (43–53%) and for the disadvantaged men the most likely class is the *Unemployment* class (58–62%). For late male entrees with a reference background, there is more variation: the probabilities are about equally likely (14–18%) for classes *Work and early gap, Work and short parental leaves,* and *Unemployment.*

For the younger entrees with a reference background and especially with an advantaged background, entering studies is a more likely pattern than continuous employment. For the youngest entrees with a reference background, women's probabilities are divided between *Work and early gap, Parental leave and health-related exits*, and the four study classes of combining studies and work (of which the *Studies and parallel work* is the most likely with a 24% probability). For the youngest male entrees with a reference background, the *Work and early gap* class is the most likely (40%).

For the individuals with a disadvantaged background, the unemployment class is considerably likely, and it is the more likely the later the entry into the labour market. Women have high probabilities for the *Parental-leave and health-related exits* class. Among the men and women from disadvantaged backgrounds that do enter studies, the *Fragmented studies and health-related exits* class is the most likely for almost all groups, with the exception of the youngest female entrees, among whom the *Studies and parallel work* class is slightly more likely.

Figure C-1 in Appendix C shows predicted probabilities for cases that are similar to the reference case but where we change the value of one predictor at a time. Compared to the reference case, non-intact family or parental psychiatric diagnosis did not seem to be predicting much higher risks for unfavourable exits when appearing alone. Parental high education (when own school success was average) tended to predict very similar probabilities as high school success (when parental education was low or average). Social assistance in childhood predicted somewhat higher risks for less stable careers and fragmented studies. Teenage pregnancy (combined with low GPA, as there were no cases in the data matching the reference case) and having a teenage mother both predicted long parental leaves and short health-related exits for women; the former was a particularly strong predictor for the earlier entrees. Conviction (combined with low GPA, as it always did in these data) predicted the highest risks for fragmented employment and long-term unemployment.

4. Discussion

[redacted] 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 different routes of exit in young adulthood, i.e., in addition to unwanted, involuntary health-related exit and unemployment, we could also consider more voluntary other 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, but not complete gap years while waiting for entrance into higher education. Such gap years are fairly common in the Finnish context: university entrance is very competitive and typically requires passing entrance exams. In 2011 when the cohort members turned

24, the median age at entrance was 20 for universities and 21 for polytechnics[38], meaning that many young adults have to apply to studies multiple times (during multiple years) before being admitted - many of them work while waiting for another attempt the following year. Another typical reason for exiting unemployment 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 sequence analysis and the mixture Markov model we were able to identify different patterns in early labour market participation, to study dynamics within different types of patterns, and study the association of typical patterns and early life course determinants. A 10-class solution was selected to best and 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 nonlinear. 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 at a later age 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 cohort), and only up to age 24 for a full five-year follow-up (74% of cohort). Given that the median age for getting a lower tertiary degree was as high as 26 in Finland in 2011[38], 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 with more disadvantaged backgrounds. Men, on the other hand, often had short gaps shortly after labour market entry (likely 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 highly-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 with lower educational accomplishments and failure[39,40].

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 risk factors. This is in line with previous findings from the same data[24].

Teenage pregnancy and having a teenage mother both predicted long parental leaves and short health-related exits and frequent and long-term unemployment. The former was a particularly strong predictor for the earlier entrees. This is also in line with previous studies[6,41].

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. 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 inequality research: numerous studies have shown that parental separation is related to lower educational attainment[42–46].

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. In all, results are in line with previous research which suggest that receiving social assistance in childhood predicts higher probability for labour market exclusion[47].

[redacted]Methodological considerations

Unique datasets available for this study enable 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 datasets 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[48].

Generally, the coverage and quality of the register data used in this study are exceptionally good. The datasets 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.

Mixture Markov models 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. 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 multiple risk factors.

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 to one where the classification of the outcome is done separately from modelling the effects of the predictors, and is 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.

5. 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 socioeconomic 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 likely not a risk factor as such. It is important to target policies and actions as well as interventions, in the high risk and most disadvantaged groups in particular.

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.

References

1. Nurminen, M. *Working-Life Expectancy in Finland: Development in 2000–2009 and Forecast for 2010–2015*. (Eläketurvakeskus, 2011).

2. Halonen, J. I. *et al.* Pathways from parental mental disorders to offspring's work disability due to depressive or anxiety disorders in early adulthood—The 1987 Finnish Birth Cohort. *Depress. Anxiety* **36**, 305–312 (2019).

3. Shiri, R., Halonen, J., Serlachius, A. & Hutri-Kähönen, N. 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. *Occup Env. Med* **78**, 11 (2021).

4. Ageing and health in focus in 2012. *The World Health Organization Regional Office for Europe.* (2012). Available at: https://www.euro.who.int/en/health-topics/Life-stages/healthy-ageing/news/news/2012/01/ageing-and-health-in-focus-in-2012. (Accessed: 22nd February 2021)

5. World Health Organization. *Executive summary of the European health report 2012: Moving Europe towards health and well-being*. (2012).

6. Lallukka, T. *et al.* Determinants of long-term unemployment in early adulthood: A Finnish birth cohort study. (2019). doi:10.1016/j.ssmph.2019.100410

7. Croft, P., Blyth, F. M. & Van Der Windt, D. *Chronic Pain Epidemiology: From Aetiology to Public Health*. *Chronic Pain Epidemiology: From Aetiology to Public Health* (Oxford University Press, 2011). doi:10.1093/acprof:oso/9780199235766.001.0001

8. Farioli, A. *et al.* Musculoskeletal pain in Europe: The role of personal, occupational, and social risk factors. *Scand. J. Work. Environ. Heal.* **40**, 36–46 (2014).

9. Patel, V., Flisher, A. J., Hetrick, S. & Mcgorry, P. Series Mental health of young people: a global public-health challenge. *www.thelancet.com* **369**, (2007).

10. Virtanen, M. *et al.* Neurodevelopmental disorders among young adults and the risk of sickness absence and disability pension: a nationwide register linkage study. *Scand. J. Work. Environ. Health* **46**, 410–416 (2020).

11. Lallukka, T., Mittendorfer-Rutz, E., Ervasti, J., Alexanderson, K. & Virtanen, M. Unemployment trajectories and the early risk of disability pension among young people with and without autism spectrum disorder: A nationwide study in Sweden. *Int. J. Environ. Res. Public Health* **17**, (2020).

12. Coggon, D. *et al.* International variation in absence from work attributed to musculoskeletal illness: findings from the CUPID study. *Occup. Environ. Med.* **70**, 575 LP – 584 (2013).

13. Kela, E. &. The Statistical Yearbook of Pensioners in Finland. *Off. Stat. Finland. Soc. Prot.* (2010).

14. OECD. *Sickness, Disability and Work: Breaking the Barriers. A synthesis of findings across OECD countries.* (2010).

15. Achterberg, T. J., Wind, H., De Boer, A. G. E. M. & Frings-Dresen, M. H. W. Factors that promote or hinder young disabled people in work participation: A systematic review. *J. Occup. Rehabil.* **19**, 129–141 (2009).

16. Dunn, K. M. Extending conceptual frameworks: Life course epidemiology for the study of

back pain. *BMC Musculoskelet. Disord.* **11**, 1–11 (2010).

17. Marmot, M., Allen, J., Bell, R., Bloomer, E. & Goldblatt, P. WHO European review of social determinants of health and the health divide. *Lancet* **380**, 1011–1029 (2012).

18. Barban, N. Trajectoires familiales et santé: Une approche sous l'angle de parcours de vie. *Eur. J. Popul.* **29**, 357–385 (2013).

19. Gao, N., Ryan, M., Krucien, N., Robinson, S. & Norman, R. Paid work, household work, or leisure? Time allocation pathways among women following a cancer diagnosis. *Soc. Sci. Med.* **246**, 112776 (2020).

20. Mcketta, S., Prins, S. J., Platt, J., Bates, L. M. & Keyes, K. (No Title). (2018). doi:10.1016/j.ssmph.2018.10.003

21. Roux, J., Grimaud, O. & Leray, E. Use of state sequence analysis for care pathway analysis: The example of multiple sclerosis. doi:10.1177/0962280218772068

22. Helske, S., Helske, J. & Eerola, M. Combining Sequence Analysis and Hidden Markov Models in the Analysis of Complex Life Sequence Data. in 185–200 (Springer, Cham, 2018). doi:10.1007/978-3-319-95420-2_11

23. Paananen, R. & Gissler, M. Cohort Profile: The 1987 Finnish Birth Cohort. *Int. J. Epidemiol.* **41**, 941–945 (2012).

24. Ristikari Tiina *et al. 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]*. (2016).

25. Gissler, M. & Haukka, J. Finnish health and social welfare registers in epidemiological research. *Nor. Epidemiol.* **14**, 113–120 (2004).

26. THL. Meta Visualisation of the Finnish Birth Cohort Studies. Available at: https://fbc-studies.github.io/meta-visu/?lang=en&ds=finnish-birth-cohorts. (Accessed: 10th February 2021)

27. Gauthier, J.-A., Bühlmann, F. & Blanchard, P. Introduction: Sequence Analysis in 2014. in *Advances in Sequence Analysis: Theory, Method, Applications* 1–17 (Springer-Verlag, 2014).

28. Studer, M. WeightedCluster Library Manual: A practical guide to creating typologies of trajectories in the social sciences with R. (2013). doi:10.12682/lives.2296-1658.2013.24

29. Piccarreta, R. & Studer, M. Holistic analysis of the life course: Methodological challenges and new perspectives. *Adv. Life Course Res.* **41**, 100251 (2019).

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

31. Steele, F. Multilevel discrete-time event history analysis with applicationsto the analysis of recurrent employment transitions. *Aust. N. Z. J. Stat.* **53**, 1–20 (2011).

32. Han, S. Y., Liefbroer, A. C. & Elzinga, C. H. Mechanisms of family formation: an application of Hidden Markov Models to a life course process. *Adv. Life Course Res.* **43**, 100265 (2020).

33. Han, Y., Liefbroer, A. C. & Elzinga, C. H. Understanding social-class differences in the transition to adulthood using Markov chain models. in *Proceedings of the International Conference on Sequence Analysis and Related Methods, Lausanne, June 8-10, 2016* (eds.

Ritschard, G. & Studer, M.) 155–177 (2016).

34. Helske, S. & Helske, J. Mixture hidden Markov models for sequence data: The seqhmm package in R. *J. Stat. Softw.* **88**, (2019).

35. R Core Team. R: A language and environment for statistical computing. (2020).

36. Gabadinho, A., Ritschard, G., Müller, N. S. & Studer, M. Analyzing and Visualizing State Sequences in R with TraMineR. *J. Stat. Softw.* **40**, 1–37 (2011).

37. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, 2016).

38. Statistics Finland. *Oppilaitostilastot 2012*. (2013).

39. Breen, R. & Goldthorpe, J. H. EXPLAINING EDUCATIONAL DIFFERENTIALS. *Ration. Soc.* **9**, 275–305 (1997).

40. Heiskala, L., Erola, J. & Kilpi-Jakonen, E. Compensatory and Multiplicative Advantages: Social Origin, School Performance, and Stratified Higher Education Enrolment in Finland. *Eur. Sociol. Rev.* (2020). doi:10.1093/esr/jcaa046

41. Merikukka, M., Ristikari, T., Tuulio-Henriksson, A., Gissler, M. & Laaksonen, M. Childhood determinants for early psychiatric disability pension: A 10-year follow-up study of the 1987 Finnish Birth Cohort. *Int. J. Soc. Psychiatry* **64**, 715–725 (2018).

42. Amato, P. R. Research on divorce: Continuing trends and new developments. *Journal of Marriage and Family* (2010). doi:10.1111/j.1741-3737.2010.00723.x

43. Amato, P. R. The consequences of divorce for adults and children. *Journal of Marriage and Family* (2000). doi:10.1111/j.1741-3737.2000.01269.x

44. Bernardi, F. & Radl, J. The long-term consequences of parental divorce for children's educational attainment. *Demogr. Res.* (2014). doi:10.4054/DemRes.2014.30.61

45. McLanahan, S. & Percheski, C. Family Structure and the Reproduction of Inequalities. *Annu. Rev. Sociol.* (2008). doi:10.1146/annurev.soc.34.040507.134549

46. McLanahan, S. & Sandefur, G. Growing up with a single parent: What helps, what hurts. *Cambridge, MA Harvard Univ. Press* (1994).

47. Bäckman, O. & Nilsson, A. Childhood Poverty and Labour Market Exclusion. Findings from a Swedish Birth Cohort. *Arbetsrapport 2007* (2007).

48. Ness, R. B. Counterpoint: The Future of Innovative Epidemiology. *Am. J. Epidemiol.* **177**, 281–282 (2013).

49. Studer, M. & Ritschard, G. What matters in differences between life trajectories: a comparativereview of sequence dissimilarity measures. *J. R. Stat. Soc. Ser. A (Statistics inSociety)* **179**, 481–511 (2016).

50. Baum, L. E. & Petrie, T. Statistical inference for probabilistic functions of finite stateMarkov chains. *Ann. Math. Stat.* **67**, 1554–1563 (1966).

51. Rabiner, L. A tutorial on hidden Markov models and selected applications inspeech recognition. *Proc. IEEE* **77**, 257–286 (1989).

Author information

Affiliations

INVEST Research Flagship Center and Department of Social Research, University of Turku, Turku, Finland

Satu Helske

Finnish Institute for Health and Welfare, Helsinki, Finland

Keski-Säntti Markus, Juha Kivelä, Aapo Juutinen, Antti Kääriälä, Mika Gissler

Karolinska Institute, Stockholm, Sweden

Mika Gissler

Itla Children's Foundation, Helsinki, Finland

Marko Merikukka

Department of Public Health, University of Helsinki, Helsinki, Finland

Tea Lallukka

Contributions statements

SH and TL were responsible for the main manuscript text, MKS and JK prepared the dataset, and MKS estimated models and prepared figures. All authors (SH, MKS, JK, AJ, AK, MG, MM, and TL) reviewed the manuscript and accepted the final version.

Corresponding author

Correspondence to Satu Helske.

Additional Information

Conflict of interest

The authors declare no competing interests.

Funding

This study was supported by the Academy of Finland (Grant #330527 for TL and grant #331816 for SH) and by the Academy of Finland Flagship Programme (Grant #320162).

Figure legends

Figure 6: *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.*

Figure 7: *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.*

Figure 8: *Latent classes from the mixture Markov model described using monthly distributions for work participation states (maximum membership assignment).*

Figure 9: *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.*

Figure 10: *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).*

Tables

Table 4: *Work participation statuses and their priority. 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, i.e., not being in employment, education, or a recipient of a benefit.*

Table 5*: Most common work participation patterns and their frequencies and proportions in the study sample (omitting timing and duration of spells).*

Table 6: *Descriptive statistics (frequencies and proportions) of categorical variables used in the study, for the full sample as well as separated by sex. Missing GPA refers to no or delayed application into secondary education.*

Supplementary information

Figure S-1: *Sequence cluster results from 2 to 15 clusters using optimal matching.*

Figure S-2: *Sequence cluster results from 2 to 15 clusters using Euclidean distance.*

Figure S-3: *Length of individual follow-up.*

Table S-1: *Estimated coefficients for covariates from the mixture Markov model.*

Table S-2: *Estimated transition probabilities from the mixture Markov model.*

Table S-3: *Estimated initial probabilities from the mixture Markov model.*

Appendix

Appendix A: Model estimation

Sequence analysis

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[49]. 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)[28].

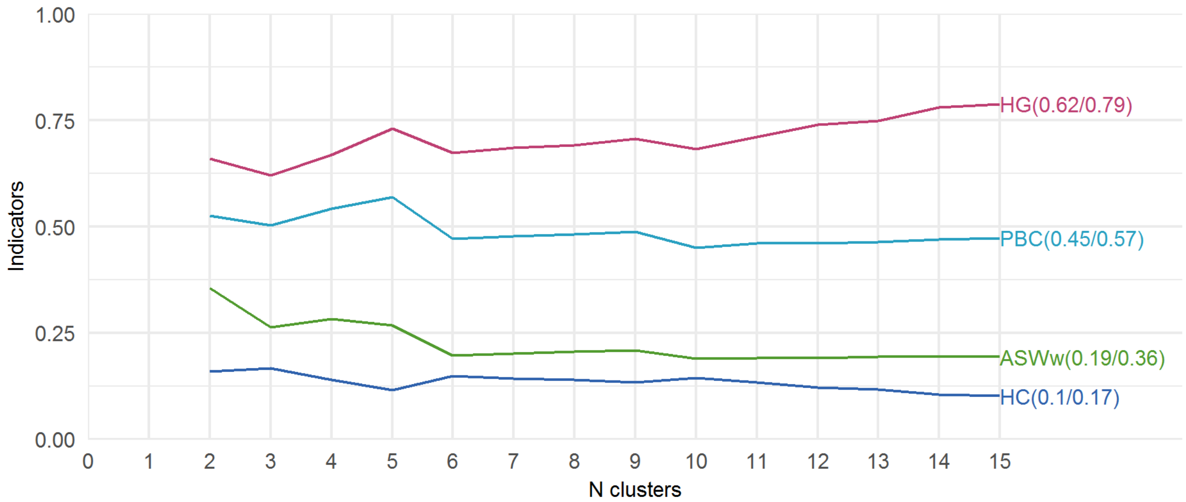


Figure A-1: Cluster quality indicators by the number of clusters.

We chose the 10-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 two to 15 for OM and Euclidean dissimilarities, see the supplementary material.

Table A-1: Description of sequence clusters.

Description	Male	Female	Total

Continuous employment	6554 (29.61%)	5155 (24.98%)	11709 (27.38%)
Studying (+ summer jobs)	3069 (13.87%)	2139 (10.36%)	5208 (12.18%)
Employed and studying	3184 (14.39%)	5332 (25.83%)	8516 (19.91%)
Unknown break from work (military service)	6189 (27.97%)	1462 (7.08%)	7651 (17.89%)
Social assistance	1555 (7.03%)	470 (2.28%)	2025 (4.73%)
Short study break from work	881 (3.98%)	1757 (8.51%)	2638 (6.17%)
Fast transition to parental leave	43 (0.19%)	2367 (11.47%)	2410 (5.63%)
Slow transition to parental leave	38 (0.17%)	1568 (7.60%)	1606 (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	22131 (100%)	20649 (100%)	42770 (100%)

Mixture Markov model

The maximum likelihood estimates of the probabilities are typically calculated with the expectation–maximization (EM) algorithm[50,51]. The EM algorithm is iterative, which means that we make a guess on 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) maximizes 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.

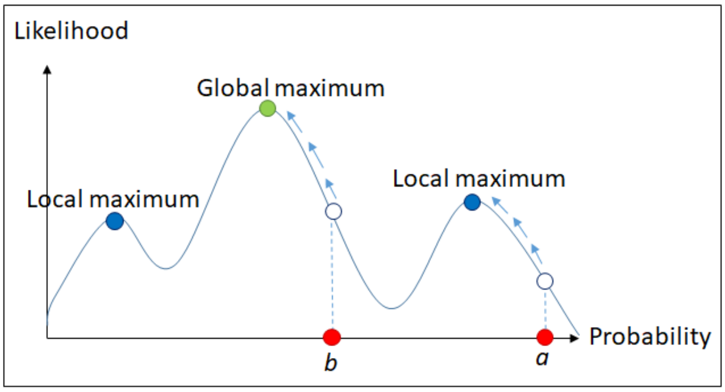


Figure A-2: Illustration of the logic of the EM algorithm and the problem of local maxima with a simple model with only one unknown probability. Using the starting value a leads to a suboptimal local maximum, while starting from value b the algorithm finds the global optimum (the model with the highest likelihood).

To facilitate the analysis, we used an approach described in Helske, Helske, & Eerola (2016)[22] who use 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 account for individuals whose work participation was recorded for full 60 months, as accounting for missing information in SA can be problematic. We then use 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 use the clusters of 60-month work participation sequences and estimate a simple Markov model independently for each cluster.
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 and transition probabilities from step 2 as meaningful starting values for the submodels of the MMM. Test between slightly modified starting values and use the value of the likelihood function for choosing the final model.

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

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

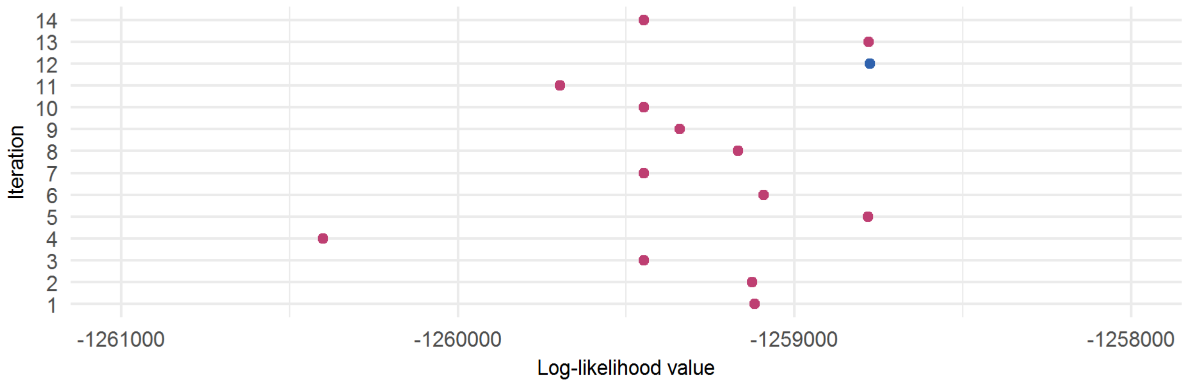


Figure A-3: Log-likelihoods of 14 estimations of the MMM using the EM algorithm, starting from different starting values. The best model with the highest likelihood (shown in blue) was found only once, but the other two with nearly similar log-likelihood values produced nearly identical models.

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**

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; e.g., 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 1 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 7-fold to the highest off-diagonal probability.

Figure A-4: Descriptions of latent classes from the mixture Markov model. 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 estimated starting probabilities for each work participation status. Transition probabilities below 0.01 and probabilities of staying in each state are omitted from the figure for clarity.

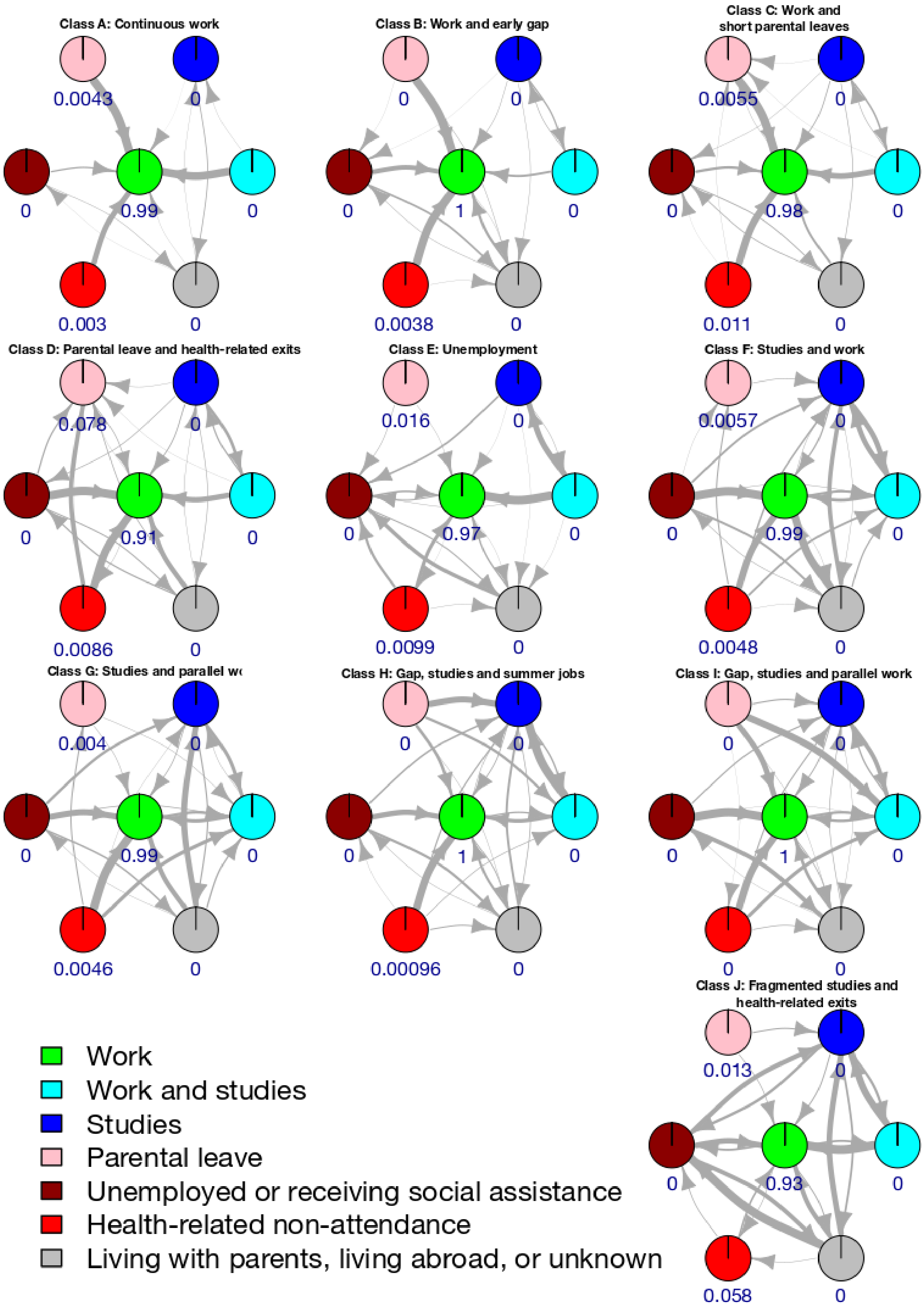


Figure A-4 shows the transition probabilities between labour market states in each latent class A–J. All estimated parameters of the mixture Markov model given as supplementary material.

Appendix B: Descriptive statistics

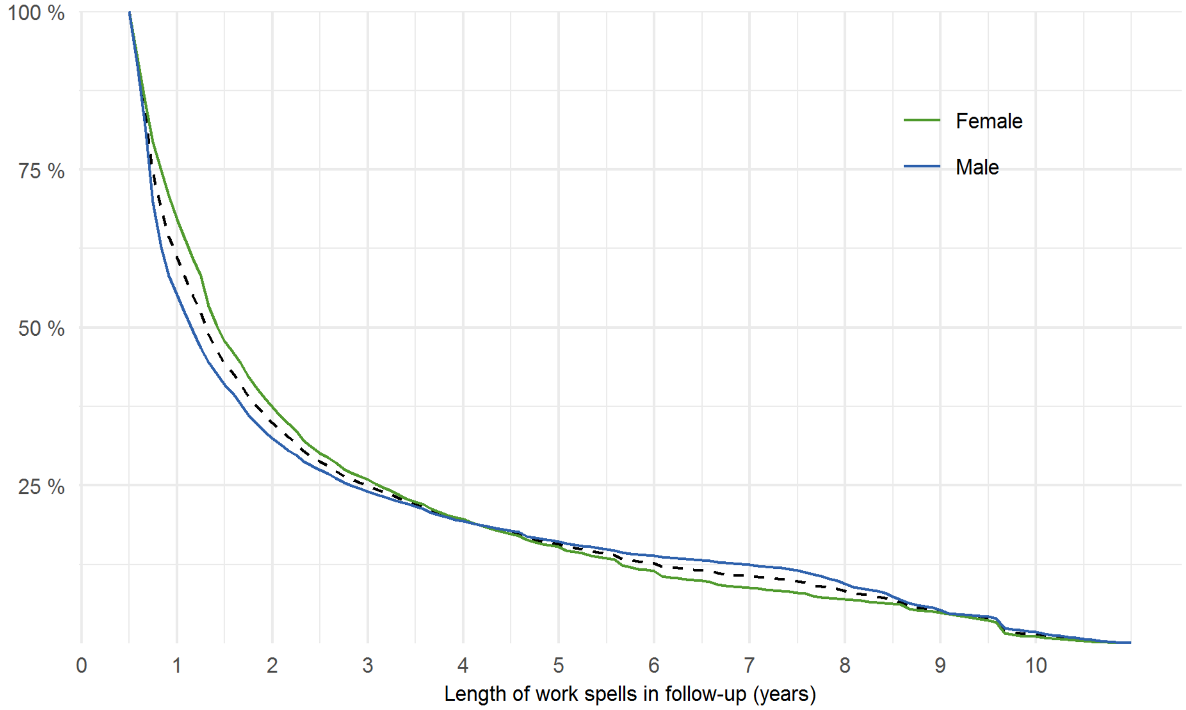


Figure B-1: Proportion of work spells of different lengths in the study sample (all work spells).

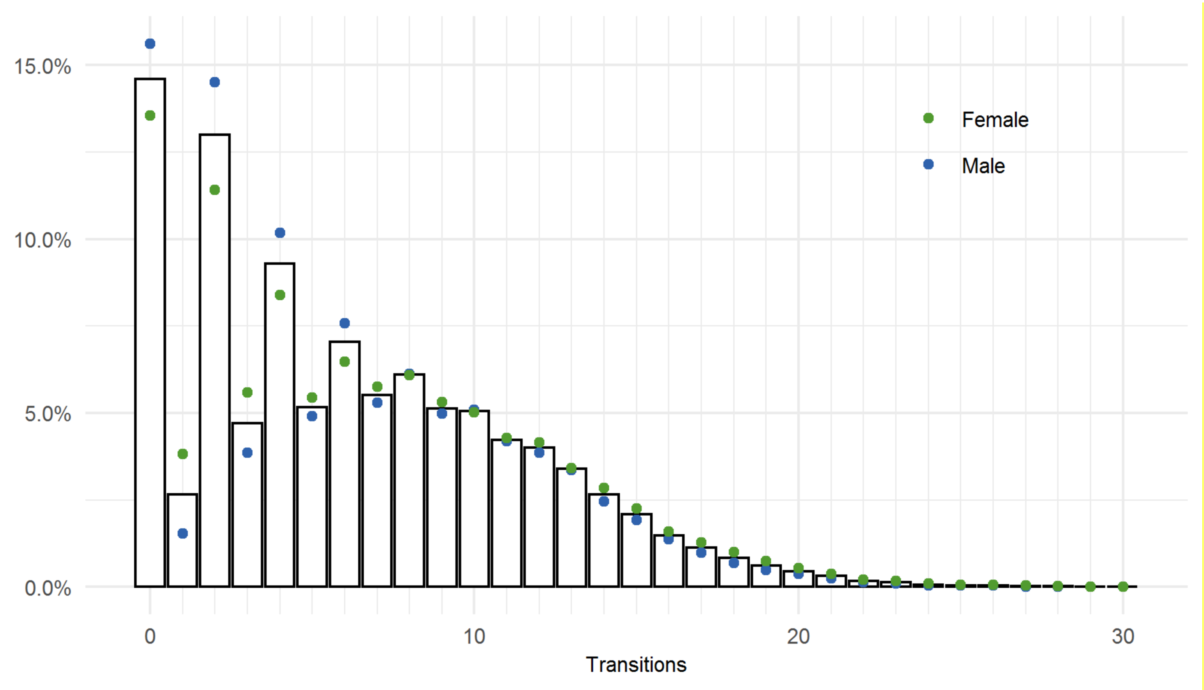


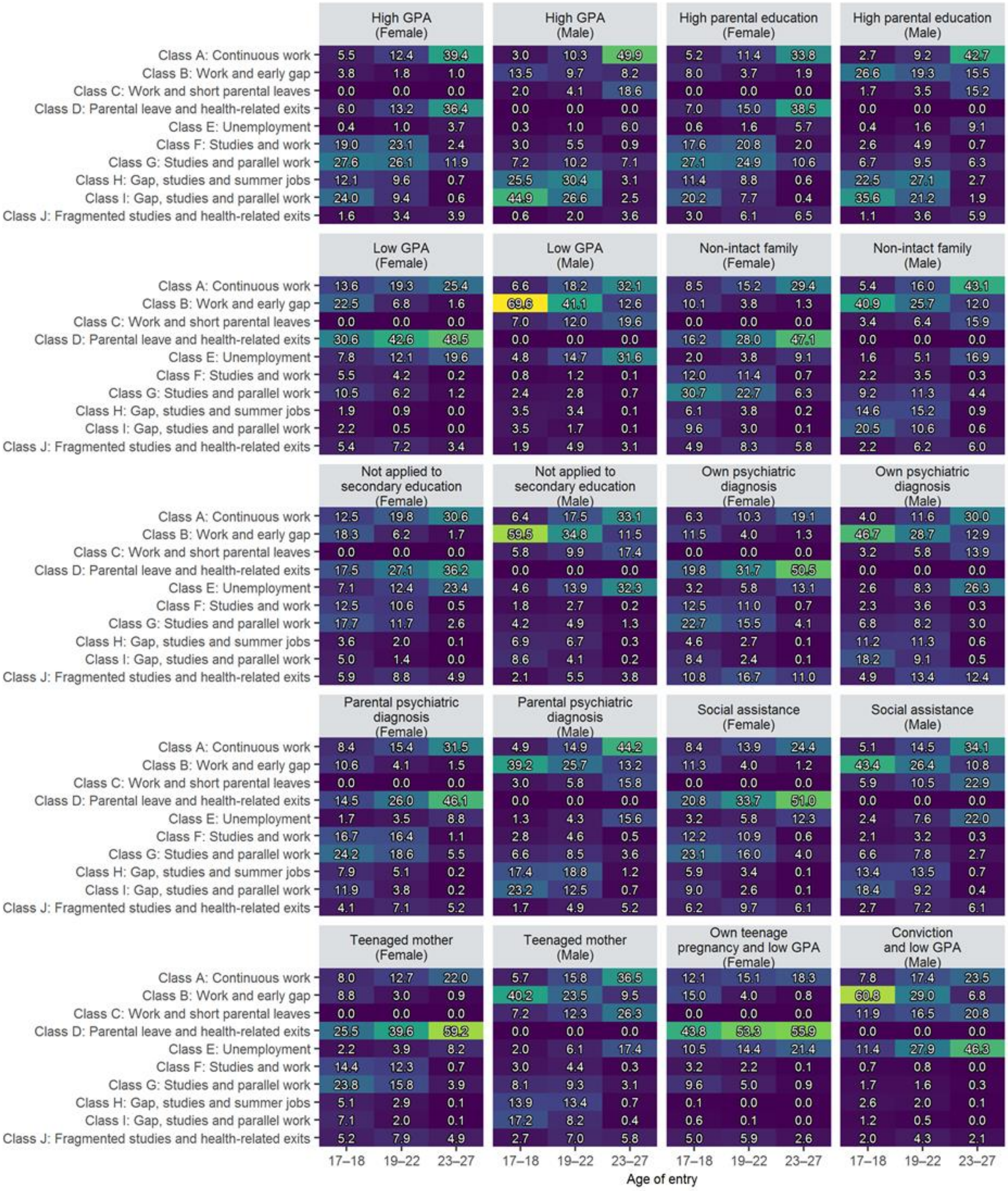
Figure B-2: Number of transitions between different work participation statuses in the study sample.

Appendix C: Variable definitions for exemplary cases for predicted probabilities

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

	Sex	GPA	Intact family	Par. educ. high	Soc. assist.	Parent psych. diag.	Own psych. diag.	Teen-aged mother	Teen-age 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

Figure C-1: Predicted probabilities for the (almost) reference case, changing one variable at a time (changed variable value shown on grey background; e.g., high GPA shows predicted probabilities for someone with GPA of 8.5 or higher and other predictors as for the reference case in table D-1).


High GPA (Female)
High GPA (Male)
High parental education (Female)
High parental education (Male)
Low GPA (Female)
Low GPA (Male)
Non-intact family (Female)
Non-intact family (Male)
Not applied to secondary education (Female)
Not applied to secondary education (Male)
Own psychiatric diagnosis (Female)
Own psychiatric diagnosis (Male)
Parental psychiatric diagnosis (Female)
Parental psychiatric diagnosis (Male)
Social assistance (Female)
Social assistance (Male)
Teenaged mother (Female)
Teenaged mother (Male)
Own teenage pregnancy and low GPA (Female)
Conviction and low GPA (Male)
Class A: Continuous work
Class B: Work and early gap
Class C: Work and short parental leaves
Class D: Parental leave and health-related exits
Class E: Unemployment
Class F: Studies and work
Class G: Studies and parallel work
Class H: Gap, studies and summer jobs
Class I: Gap, studies and parallel work
Class J: Fragmented studies and health-related exits
17–18
19–22
23–27
Age of entry