What young English people do once they reach school-leaving age: A cross-cohort comparison for the last 30 years

Jake Anders Department of Learning and Leadership, UCL Institute of Education, University College London, UK
jake@jakeanders.uk
Richard Dorsett LLAKES Centre for Research on Learning and Life Chances, UCL Institute of Education, University College London; National Institute of Economic and Social Research, UK

(Received January 2016 Revised June 2016) http://dx.doi.org/10.14301/llcs.v8i1.399

Abstract

This paper examines how young people's early transitions into the labour market have changed between cohorts born in 1958, 1970, 1980, and 1990. We use sequence analysis to characterise transition patterns and identify three distinct pathways in all cohorts. An 'Entering the Labour Market' group has declined significantly in size (from 91% in the earliest cohort, to 37% in the most recent), an 'Accumulating Human Capital' group has grown in its place (from 4% to 51%), but also a 'Potentially Difficult Transition' group has grown alongside this, reaching 12% in the most recent cohort. These trends appear to reflect behavioural rather than compositional changes. Females and those who are from a non-white ethnic background have gone from being more likely to be in the 'Potentially Difficult Transition' group, to being less likely. Coming from a low socioeconomic status background has remained a strong predictor of having a transition of this type across all four cohorts. These early transitions are important, not least since we show they are highly predictive of longer-term outcomes.

Keywords

School to work transitions, cross-cohort comparison, sequence analysis

Introduction

In recent years, there has been growing concern about the number of young people failing to make a successful transition from education into employment. Increasingly, this appears to be a structural, rather than cyclical, problem. We see evidence of this from that fact that although youth unemployment in the UK was falling in the late 1990s and early 2000s, it started rising again as early as 2004, long before the general downturn in the economy (OECD, 2008). This is an important issue, not least because making a successful transition from education into the labour market is important for young people's long-term economic success; periods of unemployment during these early years may have long-term scarring effects on later employment and earnings prospects (Arulampalam, 2001; Gregg, 2001; Gregg & Tominey, 2005).

In this paper, we examine how early transitions have changed over the last thirty years. We focus in particular on the group of young people whose early experiences suggest that they are making a potentially difficult transition in the sense that they neither continue in education nor do they find stable employment. We assess the changing size of this group and examine the extent to which it is possible to predict, on the basis of characteristics at the time of reaching school-leaving age, which individuals may experience a potentially difficult transition from school to work.

Our approach is to use sequence analysis (Abbott, 1995) to quantify the similarity between individuals' transitions over a period of 29 months from the September following their 16th birthday. Previous research has shown that young people's transitions into work may be highly differentiated (Fergusson, Pye, Esland, McLaughlin, & Muncie, 2000). Sequence analysis provides a means of comparing the full detail of individuals' labour market trajectories. This permits a fuller comparison than the more usual methods of studying labour market states at single point in time (Andrews & Bradley, 1997) or specific changes in states (Berrington, 2001). In taking this approach, we build on previous research that uses sequence analysis to study young people's transitions from education into the labour market (Anyadike-Danes & McVicar, 2005, 2010; Dorsett & Lucchino, 2014; Halpin & Chan, 1998; Martin, Schoon, & Ross, 2008; Quintini & Manfredi, 2009).

The major contribution of this paper is that it uses detailed survey data based on four birth cohorts, each roughly a decade apart. Previous research (Schoon, McCulloch, Joshi, Wiggins, & Bynner, 2001) has examined how transitions have changed between individuals born in 1958 and individuals born in 1970. We extend this to include also individuals born in 1980 and individuals born in 1990. This provides a major update to the existing empirical literature. By focusing on these more recent cohorts, we are able to consider individuals for whom the school to work transitions are relatively recent (at the time of writing). More specifically, the transitions of the 1980 and 1990 cohorts took place in the 1996-1999 and 2006-2009 periods, respectively, while the transitions of the 1970 and 1958 cohorts took place in the 1974-77 and 1986-89 periods, respectively. Ours is the first study to conduct cross-cohort analysis using sequence analysis over such an extended period and, by using more recent data, the results are more closely related to the present-day labour market.

This paper proceeds as follows. In the next section we describe the datasets used in this analysis. In the third section, we describe our methodological approach. Our analysis results in the identification of a typology of transition pathways, discussed in the fourth section, along with an account of how pathways have changed over time. We also examine the extent to which it is possible to use characteristics at age 16 to predict which type of transition young people will experience, focusing especially on those who make a potentially difficult transition (in the sense of not being characterised by either education participation or stable employment). We consider the extent to which these relationships have changed over the four cohorts analysed in this paper. In the fifth section, we extend the horizon over which individuals' transitions are considered, examining the extent to which the early transitions that we have considered are predictive of longer-term transitions, up to approximately age 24. The sixth section concludes.

Data

Our analysis uses information on month-by-month transitions for young people in England, starting in the September following their 16th birthdays and continuing for 29 months. This is dictated by the nature of the available data, which are drawn from four birth cohort surveys:

- The National Child Development Study (NCDS) is a longitudinal survey of all individuals born in one week in 1958. Background variables (used later to predict transitions) were taken from interviews with the participant and their parents at age 16 (NCDS Sweep 3, 1974) and activity histories assembled using recall interviews at age 23 (NCDS Sweep 4, 1981). The analysis sample has around 6,000 individuals.
- The British Cohort Study (BCS) is a longitudinal survey of individuals born in 1970. Background variables were taken from interviews with the participant and their parents at age 16 (BCS Sweep 4, 1986). Activity histories were assembled primarily using recall interviews at age 26 (BCS Sweep 5, 1996). The analysis sample contains around 8,600 individuals.
- Cohort 8 of the Youth Cohort Study (YCS) is a longitudinal survey of individuals born in 1980. Background variables were taken from interviews with the participant at age 16 (YCS Cohort 8, Sweep 1, 1996), who also provided information about their parents. Activity histories were constructed using annual interviews between ages 17 and 19. The analysis sample has around 8,700 individuals.
- The Longitudinal Study of Young People in England (LSYPE) is a longitudinal survey of individuals born in 1989-90. Background variables were taken from interviews with the participant and their parents up to and including age 16 and activity histories were constructed using annual interviews between ages 17 and 19 (LSYPE Waves 1-5, 2005-2010). The analysis sample has around 9,350 individuals.

Our methodological approach requires complete, month-by-month activity histories without any gaps. With the YCS and LSYPE, activity histories were provided with the dataset. However, with the NCDS and BCS, these needed to be constructed using the recall questions about young people's activities, along with their start and end dates. Constructing these histories required some data cleaning. This involved reconciling overlapping activity spells and, in the case of the NCDS, imputing education as the status where this was not recorded in the data but could be safely assumed.[i] Furthermore, we avoided dropping individuals missing a small number of months' activities by filling in gaps where activity status was unknown. Where there was a gap of a single month, this was imputed to have the same status as the subsequent month. Where there was a gap of two months and the same activity was recorded before and after the gap, the missing two months were imputed to also have that same status.

While these steps reduced the number of observations that were dropped, there was still some loss of sample. In the NCDS, there are partial activity histories for 9,697 individuals but the analysis (i.e. full activity history) sample is only 8,356. For the BCS, there was a partial history for 9,760 individuals but the analysis sample is 9,518. In the YCS, the respective figures are 9,265 and 8,682; and in the LSYPE they are 9,371 and 9,347.[ii] As is inevitable in longitudinal surveys, our sample is also affected by attrition. We deal with this using an inverse probability weighting strategy, applying our own weighting scheme for the NCDS and BCS, and provided weights for the YCS and LSYPE.

The monthly histories distinguish between four activity status types: employment; education; unemployment; and other inactive. The exception is the LSYPE for which unemployment and 'other inactive' are combined into a single NEET ('not in education, employment or training') status.

Other than sample loss as discussed above, the other concern with the data is that the activity histories rely on the recall of survey respondents. Paull (2002) finds evidence of recall bias in similar longitudinal data, noting that this is more likely among younger respondents and those with the most transient employment histories. As indicated above, the NCDS and BCS rely on quite long recall periods, while the problem is much reduced with the YCS and LSYPE since recall is only over the period of a single year. As such, we should bear in mind the potential increase in recorded short spells in YCS/LSYPE compared to NCDS/BCS that may be driven not by a change in behaviour, but by a change in data collection. Less worrying in a comparative sense is Paull's suggestion of bias among younger respondents; because all cohorts consider the same age group, any such bias should affect all datasets equally.

Table 1 summarises the analysis sample, showing the size of each of the cohorts, along with mean levels of those characteristics later used to predict transitions pathways. There is a good balance of the genders in the BCS, YCS and LSYPE, but males are somewhat over-represented in the case of the NCDS, suggesting that perhaps more of the female participants have been excluded from the analysis due to missing labour market histories over the period. The proportion of the sample from a minority ethnic group also changes between the cohorts, but this seems more likely to be tracking the changing ethnic composition of the population of England over this period. Similarly, we see the increased levels of parental education across the cohorts, with a rising proportion of parents having completed a degree.

Table 1 also provides an indicator of the proportion of individuals who experience a NEET spell at some point during our period of analysis. We see that 15% of the NCDS sample experienced being NEET for at least one month. This figure drops to 10% among the BCS and then rises to 25% among the YCS and 26% among the LSYPE. Overall, therefore, we see a long-term increase in the proportion of individuals who will be NEET at some point. However, given the caution above about shorter recall periods in the YCS and LSYPE potentially increasing the reporting of short spells, we might be concerned that this is, at least in part, driven by differences in data rather than capturing a real change. In the next section, we describe how we use sequence analysis to compare individuals' transition patterns. As a preliminary comment, we note that, among other advantages, this alleviates the problem of differential reporting of short spells since a change in just one month does not greatly affect the similarity between two individuals' sequences.

Table 1. Descriptive statistics of each cohort

	NCDS	BCS	YCS	LSYPE
N	8,356	9,518	8,682	9,347
Male	0.57	0.49	0.51	0.48
Non-White	0.01	0.02	0.09	0.14
Single parent family	0.07	0.04	0.15	0.25
Parent has A Levels (no degree)	0.11	0.05	0.07	0.22
Parent has a degree	0.01	0.05	0.07	0.17
Home owner occupied	0.31	0.31	0.80	0.74
Home socially rented	0.42	0.05	0.15	0.19
Living in workless household	0.06	0.03	0.08	0.13
Ever NEET?	0.15	0.10	0.25	0.26

Notes: NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis, as number excluded due to attrition was too small to model. YCS and LSYPE analysis weighted using dataset-provided attrition weights.

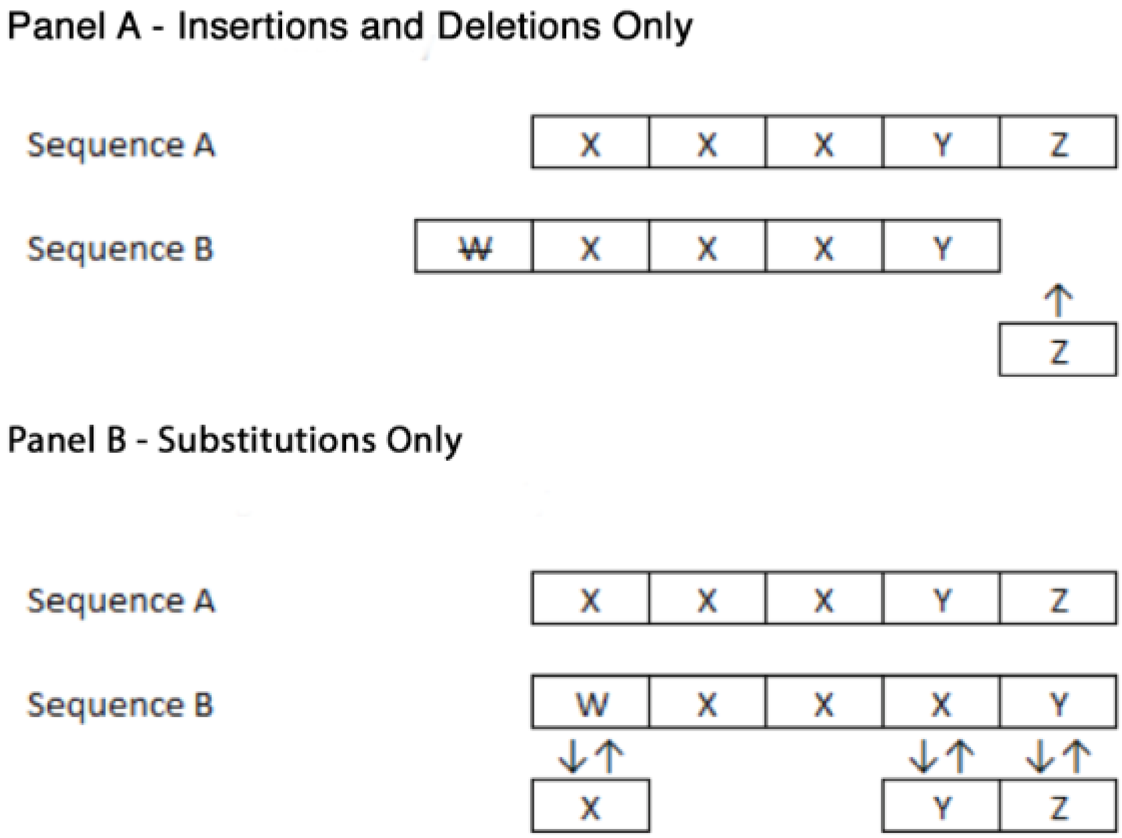
Methods

Our analytical approach involves three steps. First, we use sequence analysis to quantify dissimilarity between individuals' experiences. Second, we use these measures to identify similar-looking clusters. Third, we look at predictors of cluster membership. These steps are conducted separately for each cohort to allow more flexibility in the estimation of dissimilarity matrices and clusters; this follows the precedent set by Schoon, McCulloch, Joshi, Wiggins and Bynner (2001) and Kneale, Lupton, Obolenskaya and Wiggins (2010) in this type of work. Other issues specific to applying this method across multiple cohorts are discussed below.

Comparing individuals' transition experiences

Sequence analysis, also known as optimal matching,[iii] provides a means of quantifying the difference between activity histories (Durbin, Eddy, Krogh, & Mitchison, 1998). Most commonly, there are two broad approaches followed.[iv] The first of these is to calculate the minimum number of insertions and deletions ('indels') needed to transform one sequence into another. With monthly status data of the type used here, this means adding in, or removing, months of doing a particular activity into the sequence that the individual actually experienced. An example is shown in Figure 1 Panel A. Indels can 'warp' time (Lesnard, 2006; Martin & Wiggins, 2011); that is, the transformed histories may no longer adhere to calendar time. While this may not be an issue in some applications, it is unattractive in this case, since young people's transitions are often influenced by specific fixtures in calendar time (such as the start/end of the academic year).

Figure 1. Example of substitutions carried out to transform Sequence A into Sequence B



The second approach, followed here, is based on the number of substitutions from one state to another needed to transform one sequence into another. An example of this approach is demonstrated in Figure 1 Panel B. This approach maintains timelines within a sequence, but might in some circumstances exaggerate differences between sequences that are actually quite similar but slightly offset. Consider the example in Figure 1, which needs just one deletion and one insertion, compared with three substitutions.

To arrive at a measure of dissimilarity using this method, we must specify the 'cost' associated with each substitution; that is, how much of a change each type of substitution represents. Simplest would be to count the number of substitutions as the dissimilarity measure. However, some substitutions might involve a qualitatively more extreme change in status than others. Ideally, we would like the cost of substituting between two very similar states to be lower than that of substituting between two very different states. There are various ways of achieving this. In some applications, a matrix of substitution costs can be specified, based on prior knowledge of relevant differences between the states (Anyadike-Danes and McVicar, 2005, for example). However, arbitrary choice of substitution costs is one of the criticisms most often levelled at applications of sequence analysis (Wu, 2000). We adopt a more data-driven approach, using the inverse of the probability of transition between the two states being substituted. The less likely a transition between two states, the greater is the cost associated with a substitution of these two states. We allow these probabilities (and therefore costs) to vary over time, based on a moving average of the probability of transition in the months around the point in time at which the substitution is made. This method is referred to as calculating the Dynamic Hamming Distance (DHD) between two sequences (Lesnard, 2006). We implement this approach using the TraMineR package for R (Gabadinho, Ritschard, Müller, & Studer, 2011a).

The advantage of sequence analysis is that it provides a means of measuring the differences between individuals' histories in a way that captures their full detail. The analysis in this paper focuses on one domain - the school-to-work transition - but in principle the approach could be broadened to compare individuals' experiences across multiple domains. Pollock (2007), for instance, performs an analysis of employment, housing, partnership and child-rearing experiences. The appeal of such an approach is that it addresses the likelihood that such processes are inter-related and therefore allows a richer understanding of individuals' circumstances. The focus in this paper is necessarily more narrow. While broadening to multiple domains is possible, there are two features of our study that mean doing so may be less appropriate. First, the study considers people aged 16-18, for whom housing, partnering and child-rearing play less of a role than among the broader population. Second, and related, the time period observed for each individual is relatively short (29 months) and, while there may be several transitions for the labour market domain considered, the lower incidence of changes to housing, partnering and child-rearing status would ideally be based on a longer observation period.

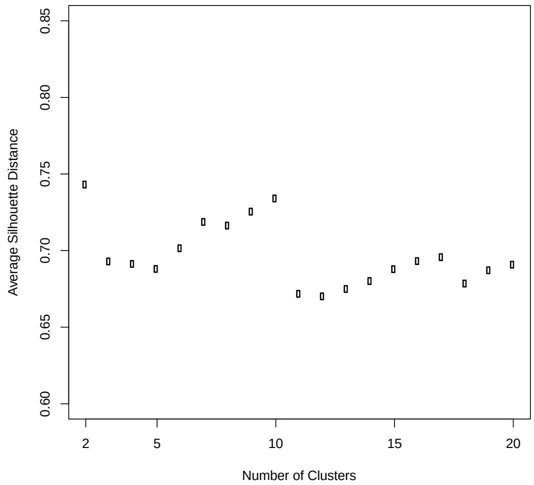
Identifying groups of individuals with similar trajectories

With the dissimilarity measures, we then carried out cluster analysis in order to group together sets of sequences that were similar. Technically, we used the non-hierarchical k-medoids/Partitioning Around Medoids (PAM) method of cluster analysis (Kaufman & Rousseeuw, 1987). The non-hierarchical approach does not impose the same constraints on cluster formation as hierarchical approaches, while k-medoids rather than k-means is more robust to outliers. PAM begins by randomly choosing the requested number of 'medoids', which are actual individuals within the dataset. All other individuals are then assigned to the cluster of the medoid to which they are most similar. There is then an iterative process of swapping current individuals selected as medoids with other potential candidates, with swaps being made where this reduces within cluster variance, until no further swaps that reduce variance are available.

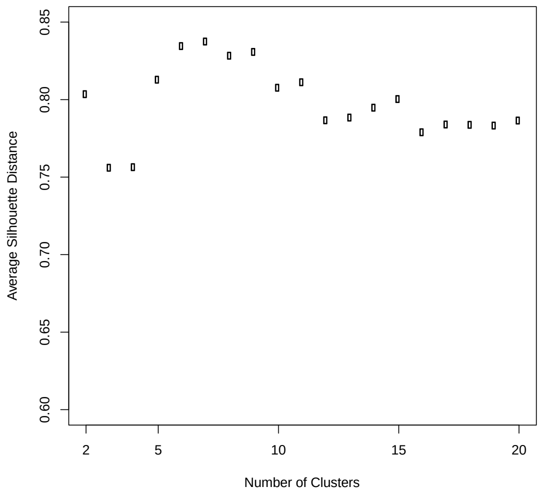
We started the analysis with two clusters and then repeated, adding one more cluster each time until there were twenty. In selecting which of these to use as our preferred cluster solution, we were guided by the average silhouette distance (Rousseeuw, 1987) as a primary diagnostic. This is reported for each of the requested solutions and for each of the four cohorts is shown in Figure 2. However, we also used some qualitative assessment of the sequences found within each cluster. Nevertheless, in all cases the average silhouette distance of the solution used is above 0.7, which is the 'rule of thumb' for the resulting cluster solution indicating that a strong structure has been found, suggested by Kaufman and Rousseeuw (1990, p. 88). Ultimately, we settle on seven cluster solutions in all of the datasets analysed.

Figure 2. Average silhouette distance of the cluster solutions

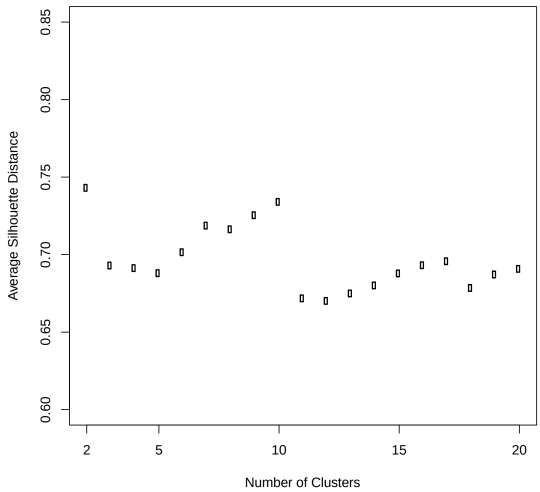
NCDS:



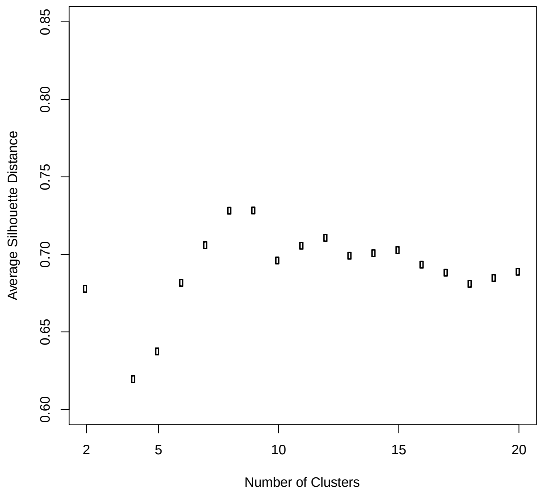
BCS:



YCS:



LSYPE:



Notes: Graphs report the average silhouette distance for each cluster solution from two to twenty clusters in each cohort. Graphs share common axes to allow comparison of the average silhouette distances in different datasets. A rule of thumb suggested by Kaufman and Rousseeuw (1990, p. 88) for "reasonable structure" is greater than 0.5 and for "strong structure" is greater than 0.7.

As noted above, we conduct this analysis separately for each cohort. This approach allows more flexibility for different types of cluster to emerge in the different cohorts, although if there are changes are in the prevalence of different types of transitions (rather than changes in the actual types of transitions) it is unlikely that the results would differ that much from a single cluster analysis across all cohorts. Furthermore, if there are such changes across cohorts, pooled analysis might result in not extracting a cluster that is important in one cohort but not across all four. While separate analysis could, in principle, make comparison a more difficult task, the next step of our approach helps to overcome this.

Grouping the clusters into substantive groups

In order to explore different kinds of transition, we choose to group together the seven clusters into three groupings on substantive grounds. We do this, rather than modelling the seven clusters separately or using the statistical approach of reducing the number of clusters sought, for three reasons:

1. Some of the clusters are very small and it would not be viable to conduct modelling using these.
2. The cluster analysis diagnostics indicated that the seven cluster solutions typically represented the strongest structure, while maintaining consistency across cohorts. Using a solution with a smaller number of clusters would not represent the strongest structure in the data.
3. Cluster solutions with a smaller number of clusters do not preserve separate clusters for transitions including significant periods of unemployment or economic inactivity. Instead, solutions with fewer clusters separate transitions by differences in the year in which individuals move from education to employment. This is unsurprising, since the number of individuals involved is much larger than the number of individuals with significant periods of unemployment or inactivity. However, there are clear substantive reasons for thinking it important to separate out individuals with substantial experience of unemployment or inactivity, especially as these groups do emerge in the cluster solutions that have the strongest structure, as indicated by average silhouette difference.

We appreciate that our process of grouping includes subjective decisions. However, we believe that the outcome represents a sound grouping based on both the structure of the data, indicated by the best-fitting cluster solutions, and the substantive importance of preserving a group in which individuals experience significant periods of unemployment and/or inactivity.

Predicting who will experience a pattern of transitions that may be a cause of concern

Being able to predict what kind of transition to the labour market individuals are likely to have, before this process has begun, is of clear potential use to policy makers. It potentially makes it easier to target support on those likely to be at risk of experiencing a pattern of transitions that might be a cause for concern. This work is in a similar spirit to that of Caspi, Entner Wright, Moffitt, and Silva (1998), who use childhood characteristics to model the probability of experiencing unemployment during the transition into the labour market. We use multinomial logistic regression models in order to assess how accurately age 16 characteristics that are common to the four datasets can predict outcomes.

Since the aim is to examine change between cohorts, we only make use of variables that can be derived to be comparable across cohorts. It is important to note that we make no claim that the associations found are causal (especially as there are relatively few available control variables to include in the regression models). The predictors we include in these models are gender, ethnicity (a dichotomous variable of white or non-white), highest parental education (specifically having achieved A-Levels or higher), housing tenure (specifically social renting or owner occupation), whether living with just one parent, and whether an individual's household is workless.

We estimate four separate models on the four datasets. Comparing these models provides evidence on how the roles of different predictive factors change, or remain the same, over time. In addition, we estimate a combined model on the pooled sample of all four datasets, which allows us to formally test whether the differences between these models are statistically significant from one another.

Comparing models from these different cohorts raises a kind of period-cohort identification problem. In other words, how do we interpret our findings? Is it that the times have changed or the population has changed? We explore this by assessing the relative importance of cohort and characteristic influences by predicting group membership for members of the LSYPE cohort using the relationships estimated for members of the NCDS cohort.

Results

Cluster solutions

We categorise the seven clusters identified in each cohort's transitions into three broader groups which we label as follows:

- 'Entering the Labour Market' includes individuals who make a relatively early entry into the labour market, leaving education and finding a seemingly stable job before or within the period of analysis;
- 'Accumulating Human Capital' includes individuals who remain in education throughout the period of analysis and are, hence, likely to have received higher education ahead of their labour market entry;
- 'Potentially Difficult Transition' includes individuals whose experience includes extended periods of unemployment or economic inactivity.

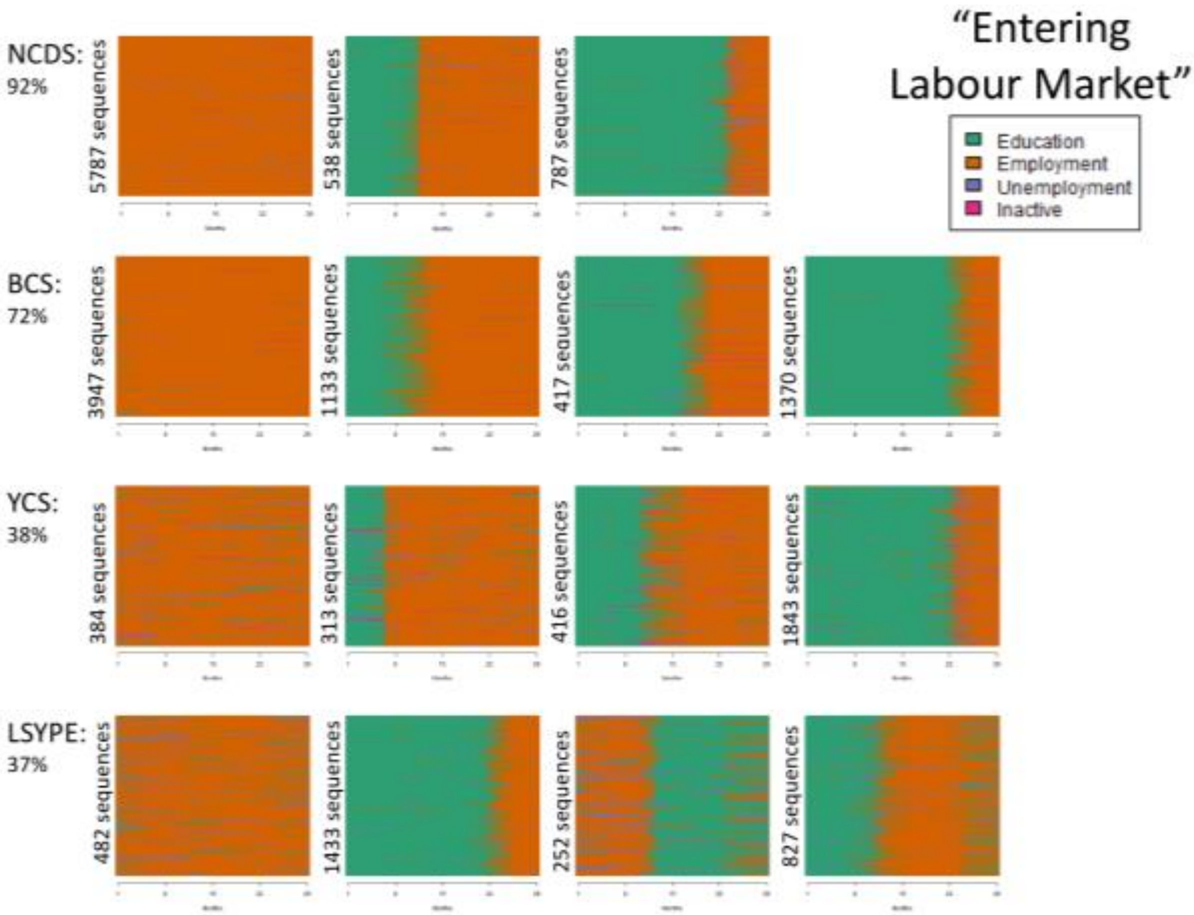
As discussed above, these are different from a directly estimated three cluster solution. Nevertheless, there is always a single cluster in the directly estimated three cluster solutions very highly correlated with the 'Accumulating Human Capital' grouping. However, the 'Entering the Labour Market' grouping tends to be split into two groups, reflecting different ages of transitions from education to employment during the period. Meanwhile, individuals we group as making a 'Potentially Difficult Transition' end up spread across two or three of the clusters in a way that is not consistent across cohorts.

While clusters that fit into these three groupings exist in each of the four cohorts, the relative sizes of these groupings have changed dramatically over time. In order to show this, we present 'index plots' of young people's labour market states over the periods considered. An index plot is a month-by-month representation of the sequence, where each horizontal line represents one young person's transition, with changes in colour showing changes in labour market state. Showing an index plot for a whole cohort is rather impenetrable, but showing plots for the clusters identified above gives a useful overview of the transitions experienced by individuals in the cluster.

Shown first, in Figure 3, is the 'Entering the Labour Market' group. These are clusters in which visual inspection reveals individuals who are either in employment throughout the period considered or who enter employment straight after education. The exception is the second LSYPE cluster which is a little more ambiguous (note that it is rather small).[v] This group has diminished significantly between the cohorts, from over 90% in the earliest to under 40% in the most recent. In addition, for those who do still follow this route, a visual inspection of the individual transitions that make up these clusters over the four cohorts suggests that earlier entry into the labour market may have become a less stable path with increasing evidence of short spells of unemployment.[vi]

Table 2 reveals the extent to which the composition of the 'Entering the Labour Market' group has changed over time. In the NCDS, the characteristics of individuals in this group essentially mirror those of the population as a whole, as one might expect for a group that makes up over 90% of the sample. However, by the BCS cohort, some differences have started to be evident. Most notably, young people whose parents' hold a degree make up only 3% of the group, compared to 5% of those in the population as a whole. The under-representation in this group of individuals with highly-educated parents persists in later cohorts too. Likewise, while there is little difference between the 'Entering the Labour Market' group and the rest of the cohort in the proportion of those who are non-white in the NCDS, by later cohorts a large gap has opened with young people with a non-white ethnic background under-represented in this group.

Figure 3. Plots of young people's individual transitions in four cohorts between the September following their 16th birthday and 29 months later: clusters placed in the 'Entering the Labour Market' group



Notes: Total number of sequences from each dataset analysed as follows: NCDS: 8,356; BCS: 9,518; YCS: 8,682; LSYPE: 9,347. Horizontal axes track months from 1 to 29.

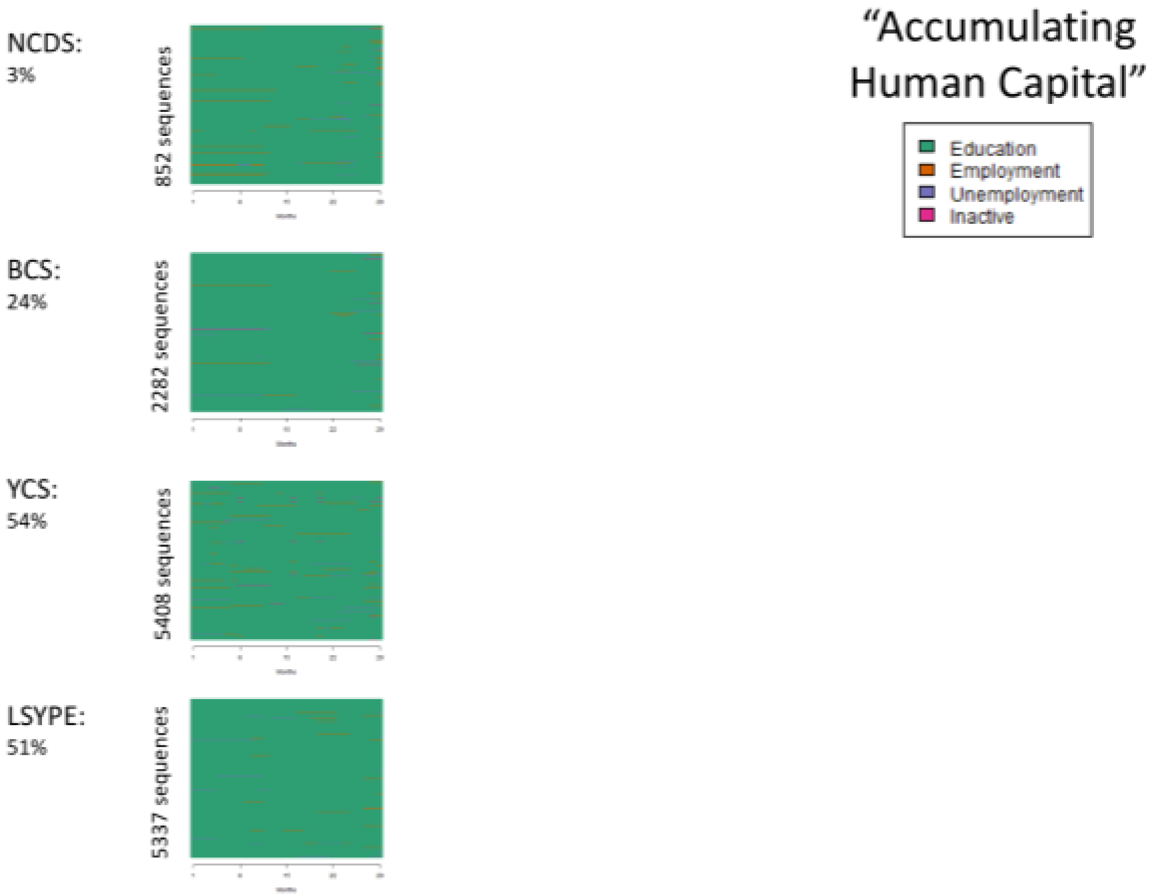
Table 2. Descriptive statistics for identified groups within each cohort

NCDS	Entering the Labour Market	Accumulating Human Capital	Potentially Difficult Transition	Overall
N	7,110	852	394	8,356
Proportion	0.91	0.04	0.05	1.00
Male	0.59	0.57	0.23	0.57
Non-White	0.01	0.01	0.02	0.01
Single parent family	0.07	0.04	0.11	0.07
Parent has A Levels (no degree)	0.10	0.28	0.06	0.11
Parent has a degree	0.01	0.08	0.00	0.01
Home owner occupied	0.31	0.51	0.13	0.31
Home socially rented	0.43	0.14	0.53	0.42
Living in workless household	0.06	0.03	0.10	0.06
BCS	**Entering the Labour Market**	**Accumulating Human Capital**	**Potentially Difficult Transition**	**Overall**
N	6867	2282	369	9518
Proportion	0.72	0.24	0.04	1.00
Male	0.50	0.48	0.34	0.49
Non-White	0.02	0.05	0.04	0.02
Single parent family	0.03	0.04	0.05	0.04
Parent has A Levels (no degree)	0.04	0.10	0.01	0.05
Parent has a degree	0.03	0.14	0.01	0.05
Home owner occupied	0.28	0.44	0.11	0.31
Home socially rented	0.05	0.03	0.10	0.05
Living in workless household	0.03	0.03	0.07	0.03
YCS	**Entering the Labour Market**	**Accumulating Human Capital**	**Potentially Difficult Transition**	**Overall**
N	2,956	5,408	318	8,682
Proportion	0.40	0.55	0.05	1.00
Male	0.49	0.51	0.55	0.51
Non-White	0.03	0.13	0.07	0.09
Single parent family	0.16	0.14	0.17	0.15
Parent has A Levels (no degree)	0.04	0.09	0.07	0.07
Parent has a degree	0.03	0.09	0.08	0.07
Home owner occupied	0.75	0.85	0.64	0.80
Home socially rented	0.20	0.10	0.28	0.15
Living in workless household	0.08	0.08	0.15	0.08
LSYPE	**Entering the Labour Market**	**Accumulating Human Capital**	**Potentially Difficult Transition**	**Overall**
N	2,994	5,399	954	9,347
Proportion	0.37	0.51	0.12	1.00
Male	0.48	0.48	0.52	0.48
Non-White	0.07	0.19	0.13	0.14
Single parent family	0.27	0.22	0.36	0.25
Parent has A Levels (no degree)	0.20	0.25	0.16	0.22
Parent has a degree	0.12	0.22	0.12	0.17
Home owner occupied	0.74	0.79	0.52	0.74
Home socially rented	0.19	0.15	0.36	0.19
Living in workless household	0.10	0.13	0.26	0.13

Notes: NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis, as number excluded due to attrition was too small to model. YCS and LSYPE analysis weighted using dataset-provided attrition weight

By contrast to the overall decline in the 'Entering the Labour Market' group, the size of the 'Accumulating Human Capital' group, shown in Figure 4, has grown significantly across the cohorts, from 4% in the earliest to around 50% in the most recent. These are clusters in which individuals remain in education throughout the period of analysis, and the growth reflects increases in both further and higher education across the cohorts analysed. Perhaps unsurprisingly, given the large increase in the size of the group, there have been differences in the average characteristics of individuals in this group. For example, the proportion of young people whose ethnicity is not white has increased from 1% in the NCDS (similar to the population as a whole) to 19% in the LSYPE (compared to 14% in the population as a whole).

Figure 4. Plots of young people's individual transitions in four cohorts between the September following their 16th birthday and 29 months later: clusters placed in the "Accumulating Human Capital" group

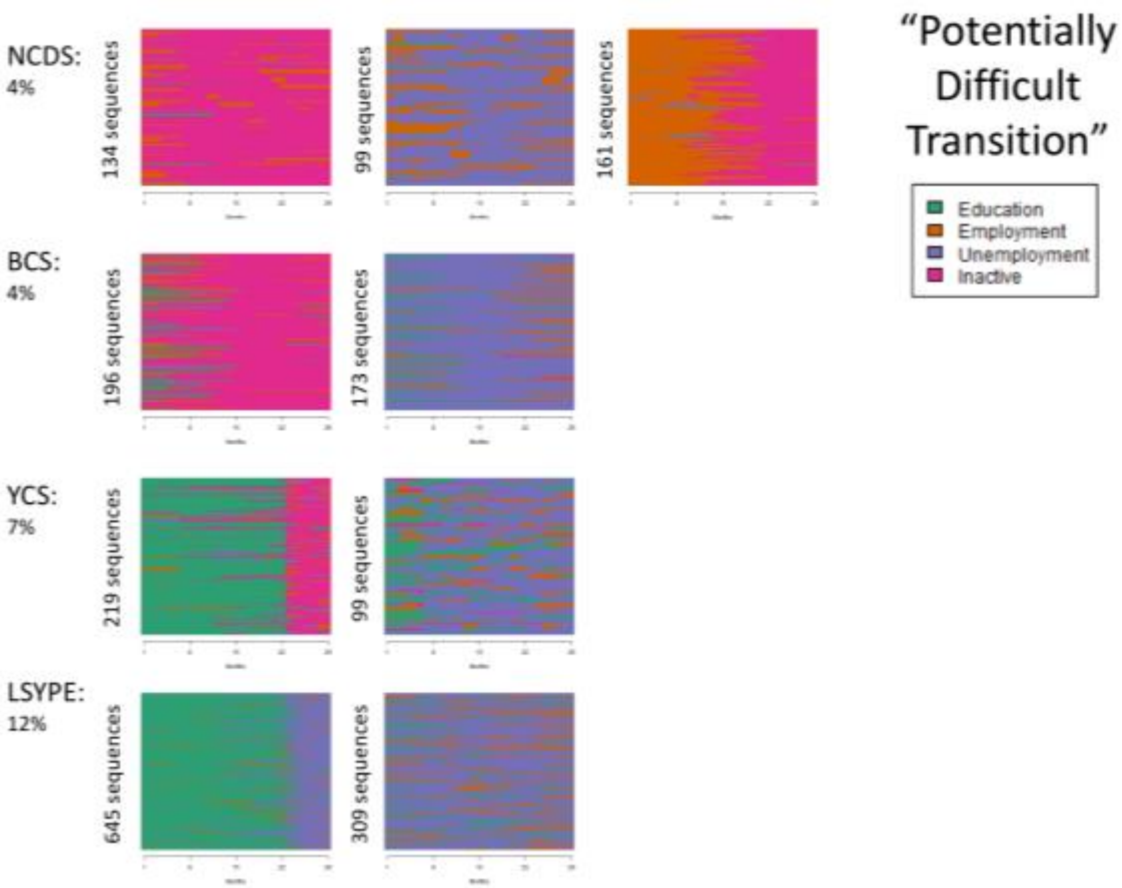


Notes: Total number of sequences from each dataset analysed as follows: NCDS: 8,356; BCS: 9,518; YCS: 8,682; LSYPE: 9,347. Horizontal axes track months from 1 to 29.

Lastly, the size of the 'Potentially Difficult Transition' group, shown in Figure 5, has also grown, although less dramatically than the 'Accumulating Human Capital' group, from 5% in the earliest cohort to 12% in the most recent. Unlike the other groups, whose respective fall and rise are relatively evenly spread through time, the growth of this group was concentrated between the 1980-born cohort and the 1990-born cohort. This group contains clusters in which individuals spend extended periods in inactivity or unemployment, seemingly not managing to settle into a job or education throughout this period of their lives. We should note that, for some, particularly where we see inactivity rather than unemployment, a transition of this type might be an active decision, for example individuals who become homemakers. As such, we should not necessarily regard all individuals in this group as a cause for concern.

Relatedly, we also see a change in the behaviour of those who go straight from education into extended inactivity (predominantly young women, especially in earlier cohorts). In earlier cohorts, individuals who experience this kind of transition move into inactivity at around age 16. However, by the later cohorts, otherwise similar looking transitions show individuals moving into inactivity at around age 18, suggesting that such individuals are more likely to receive two additional years of education in later cohorts than they were in earlier ones. There may well be benefits for these individuals from the additional human capital they gain from these two years.

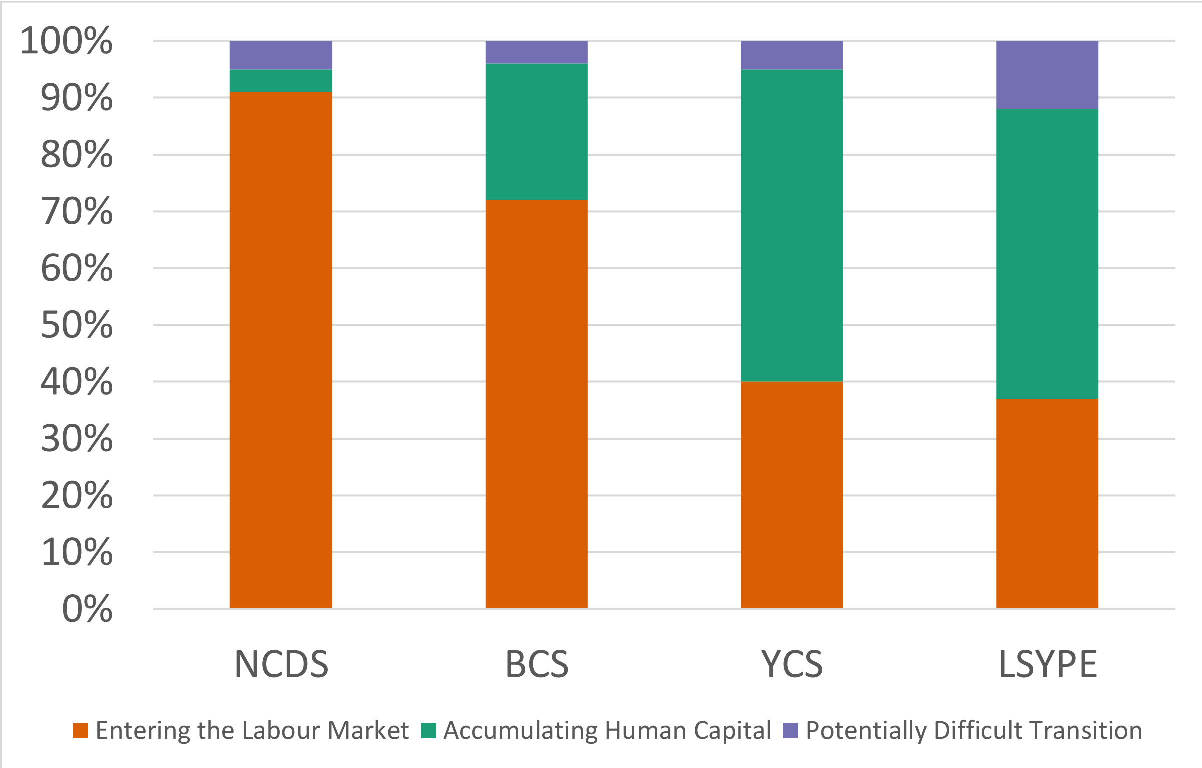
Figure 5. Plots of young people's individual transitions in four cohorts between the September following their 16th birthday and 29 months later: clusters placed in the 'Potentially Difficult Transition' group



Notes: Total number of sequences from each dataset analysed as follows: NCDS: 8,356; BCS: 9,518; YCS: 8,682; LSYPE: 9,347. Horizontal axes track months from 1 to 29. In LSYPE analysis, purple represents both Unemployment and Inactivity.

A summary graphic of the size of each grouping is reported in Figure 6. It is encouraging to note that these findings accord with those of previous analyses of this period. Most directly, while our 'Potentially Difficult Transition' group makes up 5% of the YCS cohort (born in 1980) and 12% of the LSYPE cohort (born in 1990), the size of this group in the analysis by Dorsett and Lucchino (2014) falls somewhere in between (10%), for a sample born between 1975 and 1988. Similarly, the growth in the size of the Accumulating Human Capital group tracks the well-documented trend towards increased levels of post-compulsory education.

Figure 6. Percentage of each cohort allocated to each grouping



Notes: NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis, as number excluded due to attrition was too small to model. YCS and LSYPE analysis weighted using dataset-provided attrition weights.

Predicting types of transitions

Table 3 reports the estimation results from multinomial logistic regression models of membership of a cluster in each of our three groupings.[vii] Our discussion focuses on the changing associations with the probability of experiencing a potentially difficult transition, given the particular policy interest of being able to predict transitions of this type.

The changing influence of gender and ethnicity are the most striking results, with both moving from being a significant predictor in one direction in the earliest cohort to being a significant predictor in the opposite direction by the most recent. First, in the case of ethnicity, individuals born in 1958 who are of non-white ethnicity are nearly five percentage points more likely to make a potentially difficult transition than their white peers. By contrast, in the 1990 cohort, individuals of non-white ethnicity are instead 2.5 percentage points less likely to experience a potentially difficult transition, compared to their white counterparts. Similarly, males born in 1958 are seven percentage points less likely to experience a potentially difficult transition than females from a similar background, while for the cohort born in 1990 the probability is two percentage points higher for males than females. To give some context for these results, we note that the potentially difficult transition grouping accounted for 5% of the 1958 cohort and 12% of the 1990 cohort.

Table 3. Estimated average marginal effects on the probability of an individual's membership of a cluster in each grouping (relative to the other two) from cohort-specific multinomial logistic regression models

	Entering the Labour Market				Accumulating Human Capital				Potentially Difficult Transition			
	NCDS	BCS	YCS	LSYPE	NCDS	BCS	YCS	LSYPE	NCDS	BCS	YCS	LSYPE
Non-White	-0.049**	-0.191***	-0.351***	-0.211***	0.002	0.171***	0.356***	0.237***	0.047**	0.019*	-0.005	-0.025**
	(-2.309)	(-6.927)	(-11.423)	(-12.609)	(0.197)	(6.706)	(12.017)	(15.145)	(2.500)	(1.727)	(-0.403)	(-2.277)
Male	0.071***	0.023**	-0.019	-0.001	-0.001	0.004	0.011	-0.018	-0.070***	-0.027***	0.008	0.019**
	(8.609)	(2.505)	(-1.534)	(-0.115)	(-0.298)	(0.455)	(0.860)	(-1.556)	(-9.036)	(-6.229)	(1.302)	(2.328)
Workless household	-0.023	-0.018	-0.012	-0.127***	0.006	-0.002	-0.010	0.071***	0.017	0.020**	0.022**	0.056***
	(-1.312)	(-0.730)	(-0.455)	(-6.204)	(0.603)	(-0.075)	(-0.374)	(3.488)	(1.176)	(2.404)	(2.090)	(4.855)
Lone parent	-0.003	-0.037	-0.022	0.056***	-0.008	0.017	0.029	-0.079***	0.011	0.020*	-0.007	0.023**
	(-0.153)	(-1.537)	(-1.116)	(3.950)	(-0.901)	(0.748)	(1.491)	(-5.433)	(0.745)	(1.880)	(-0.798)	(2.294)
Socially rented	0.014	0.081**	0.085**	-0.002	-0.012	-0.109***	-0.110***	-0.017	-0.002	0.029*	0.024	0.019
	(1.017)	(2.202)	(2.170)	(-0.087)	(-1.445)	(-3.159)	(-2.804)	(-0.579)	(-0.158)	(1.817)	(1.478)	(1.163)
Owner occupier	0.018	-0.022	-0.073**	-0.043*	0.027***	0.048*	0.086**	0.112***	-0.045***	-0.025*	-0.013	-0.069***
	(1.200)	(-0.750)	(-2.097)	(-1.686)	(3.689)	(1.765)	(2.517)	(4.260)	(-3.391)	(-1.649)	(-0.911)	(-4.410)
Parental A-Levels	-0.039***	-0.150***	-0.072*	-0.030**	0.048***	0.200***	0.081**	0.050***	-0.009	-0.049**	-0.009	-0.020*
	(-2.926)	(-6.294)	(-1.875)	(-2.099)	(10.748)	(12.162)	(2.149)	(3.455)	(-0.743)	(-2.304)	(-0.612)	(-1.801)
Parental degree	-0.053	-0.255***	-0.152***	-0.133***	0.096***	0.281***	0.117***	0.134***	-0.043	-0.026	0.035**	-0.000
	(-1.625)	(-12.044)	(-4.054)	(-8.250)	(15.987)	(17.954)	(3.188)	(8.430)	(-1.304)	(-1.580)	(2.490)	(-0.029)
N	8356	9518	8682	9144	8356	9518	8682	9144	8356	9518	8682	9144

Notes: Models also include regional dummy variables and missing variable dummies for the variables above. NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis, as number excluded due to attrition was too small to model. YCS and LSYPE analysis weighted using dataset-provided attrition weights. T statistics reported in parentheses. Stars indicate statistical significance: * p=0.10; ** p=0.05; *** p=0.01.

The individual coefficients on each of our proxies for socioeconomic status (SES) are not straightforward to interpret in isolation, nor do they form any particularly obvious patterns. This partially reflects the changing importance of factors such as housing tenure as indicators for SES. Instead, to illuminate the combined role of SES, Tables 4.1, 4.2 and 4.3 present the predicted probability of an individual making each type of transition (Entering the Labour Market in 4.1, Accumulating Human Capital in 4.2 and Potentially Difficult Transition in 4.3) by gender, ethnicity and two combinations of the other model characteristics chosen to be an example of a 'high SES' individual and a 'low SES' individual. A 'high SES' individual is from a two-parent household, where at least one parent works, at least one parent holds a degree, and their house is owner-occupied. Conversely, a 'low SES' individual is from a lone parent, workless household, where the parent's highest qualification is below A-Level and their home is socially rented. Taken as a whole, these combinations remain indicative of advantage and disadvantage across all four cohorts.[viii]

Table 4.3 shows that the increase in the proportion of young people in clusters categorised as 'Potentially Difficult Transition' differs across ethnic/gender combinations. White females have a 6.9% probability of making a potentially difficult transition' in the NCDS, compared with a 1.5% probability for white males and 17.5% for non-white females. By the time of the LSYPE cohort, white females have a 10.2% probability of making a potentially difficult transition, slightly lower than the 12% probability for white males and higher than the probability for non-white females, which has fallen markedly to 7.1%.

The most obvious message from the predicted probabilities is that, throughout this period, young people from more advantaged backgrounds have been less likely than those from less advantaged backgrounds to make what we classify as a potentially difficult transition. There is evidence of this gap widening over time; from 4.9 percentage points in the NCDS to 17.4 percentage points in the LSYPE.

Table 4.1. Predicted probability of membership of a cluster in the 'Entering the Labour Market' grouping, by SES, gender and ethnicity in four cohorts

White male	Low SES	High SES	Overall	N
NCDS	95.8	67.9	96.5	4,338
BCS	81.6	31.2	73.4	4,564
YCS	53.6	23.6	40.7	3,429
LSYPE	30.6	34	40	3,161
White female	Low SES	High SES	Overall	N
NCDS	86.3	65.1	91.1	3,844
BCS	76.7	30.8	72.2	4,706
YCS	56.3	25.1	42.7	4,555
LSYPE	31.6	34	40	3,201
Non-White male	Low SES	High SES	Overall	N
NCDS	90.8	65.3	93.6	103
BCS	62.7	13.5	49.2	113
YCS	21.2	6	12.5	283
LSYPE	15.9	15.8	19.8	1,291
Non-White female	Low SES	High SES	Overall	N
NCDS	69.8	61.4	80.4	71
BCS	62.7	13.4	48	122
YCS	22.7	6.5	13.3	415
LSYPE	16.1	15.6	19.5	1,491
Overall	Low SES	High SES	Overall	N
NCDS	93.2	66.8	95	8,356
BCS	79.1	30.4	72.4	9,505
YCS	51.9	22	38.5	8,682
LSYPE	28.9	31.1	36.9	9,144

Notes: Predicted probabilities from underlying regression models reported in Table 3. Models also include regional dummy variables and missing variable dummies for the variables above. NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis as number excluded due to attrition was too small to model. YCS and LSYPE analysis weighted using dataset-provided attrition weights. 'High SES' individual is from a two parent household, where at least one parent works, at least one parent holds a degree, and their house is owner occupied. 'Low SES' individual is from a lone parent, workless household, where the parent's highest qualification is below A-Level and their home is socially rented. 'Overall' are predictions based on the complete sample, not a weighted average of the 'Low SES' and 'High SES' predictions.

Table 4.2. Predicted probability of membership of a cluster in the 'Accumulating Human Capital' grouping, by SES, gender and ethnicity in four cohorts

White male	Low SES	High SES	Overall	N
NCDS	1.3	31.8	2.0	4,338
BCS	12.4	68.4	25.3	4,564
YCS	34.8	70.6	55.1	3,429
LSYPE	39.4	54.8	48	3,161
White female	Low SES	High SES	Overall	N
NCDS	1.3	33.6	2.1	3,844
BCS	11.8	68.3	25.1	4,706
YCS	34.1	70	53.8	4,555
LSYPE	42.2	56.6	49.8	3,201
Non-White male	Low SES	High SES	Overall	N
NCDS	1.4	33.9	2.1	103
BCS	27.7	86.1	49	113
YCS	68.9	90.8	84.7	283
LSYPE	61.7	76.7	71.6	1,291
Non-White female	Low SES	High SES	Overall	N
NCDS	1.1	35	2.0	71
BCS	27.7	85.8	48.3	122
YCS	68.9	90.8	84.3	415
LSYPE	64.8	78.2	73.3	1,491
Overall	Low SES	High SES	Overall	N
NCDS	1.3	32.7	2.0	8,356
BCS	12.4	68.9	25.7	9,505
YCS	37.4	72.9	57.7	8,682
LSYPE	43.9	59.1	52.4	9,144

Notes: Predicted probabilities from underlying regression models reported in Table 3. Models also include regional dummy variables and missing variable dummies for the variables above. NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis as number excluded due to attrition was too small to model. YCS and LSYPE analysis weighted using dataset-provided attrition weights. 'High SES' individual is from a two parent household, where at least one parent works, at least one parent holds a degree, and their house is owner occupied. 'Low SES' individual is from a lone parent, workless household, where the parent's highest qualification is below A-Level and their home is socially rented. 'Overall' are predictions based on the complete sample, not a weighted average of the 'Low SES' and 'High SES' predictions.

Table 4.3. Predicted probability of membership of a cluster in the 'Potentially Difficult Transition' grouping, by SES, gender and ethnicity in four cohorts

White male	Low SES	High SES	Overall	N
NCDS	2.9	0.3	1.5	4,338
BCS	5.9	0.4	1.3	4,564
YCS	11.6	5.8	4.2	3,429
LSYPE	30	11.2	12	3,161
White female	Low SES	High SES	Overall	N
NCDS	12.4	1.3	6.9	3,844
BCS	11.5	0.9	2.7	4,706
YCS	9.7	4.9	3.5	4,555
LSYPE	26.1	9.4	10.2	3,201
Non-White male	Low SES	High SES	Overall	N
NCDS	7.8	0.8	4.2	103
BCS	9.7	0.4	1.8	113
YCS	9.9	3.2	2.8	283
LSYPE	22.4	7.5	8.6	1,291
Non-White female	Low SES	High SES	Overall	N
NCDS	29	3.6	17.5	71
BCS	9.7	0.8	3.7	122
YCS	8.5	2.7	2.4	415
LSYPE	19.1	6.2	7.1	1,491
Overall	Low SES	High SES	Overall	N
NCDS	5.5	0.6	2.9	8,356
BCS	8.5	0.6	1.9	9,505
YCS	10.7	5.1	3.8	8,682
LSYPE	27.2	9.8	10.7	9,144

Notes: Predicted probabilities from underlying regression models reported in Table 3. Models also include regional dummy variables and missing variable dummies for the variables above. NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis as number excluded due to attrition was too small to model. YCS and LSYPE analysis weighted using dataset-provided attrition weights. 'High SES' individual is from a two parent household, where at least one parent works, at least one parent holds a degree, and their house is owner occupied. 'Low SES' individual is from a lone parent, workless household, where the parent's highest qualification is below A-Level and their home is socially rented. 'Overall' are predictions based on the complete sample, not a weighted average of the 'Low SES' and 'High SES' predictions.

An interesting question is whether changes in the size of the 'Potentially Difficult Transition' group are due to cross-cohort differences in composition or to cross-cohort changes in the influence of background characteristics. To explore this, we used the coefficients estimated using the NCDS to predict how group membership among individuals in the LSYPE would look had the influence of background characteristics not changed since the time of this first cohort. These predicted probabilities are reported in Table 5.1 for Entering the Labour Market, 5.2 for Accumulating Human Capital and 5.3 for Potentially Difficult Transition, in a similar way to those reported in Tables 4.1, 4.2 and 4.3. In each combination of ethnicity and gender we can compare young people's probabilities of being in each transition grouping in the NCDS, the LSYPE, and the LSYPE if the probabilities are affected by characteristics in the same way as they were in the NCDS cohort.[ix] For each ethnicity/gender combination, a comparison of the NCDS row with the 'NCDS associations/LSYPE cohort' row shows how changing composition over time affects the predicted probabilities. Similarly, a comparison of the LSYPE row with the 'NCDS associations/LSYPE cohort' row shows the changing influence of background characteristics, assuming composition is fixed.

Looking at the comparison of the NCDS probabilities with those of NCDS association on the LSYPE cohort we find that, across the full sample, the results suggest that the change in composition would be expected to, if anything, reduce the probability of making a potentially difficult transition from 2.9% to 2.2%. The biggest difference due to changing composition is among non-white females. In particular, those in the 'Low SES' group see their probability of making a potentially difficult transition fall from roughly 29% to 7.7% (among the 'High SES' group, there is no predicted change).

This implies that it is the change in the influence of background characteristics that is primarily responsible for the growth in this group. The second comparison (of the LSYPE row with the 'NCDS associations/LSYPE cohort' row) provides more detail; applying the NCDS associations to the LSYPE cohort predicts 2.2% will make a potentially difficult transition, whereas in fact 10.7% do. As such, it is the change in the relationship between characteristics and cluster memberships, rather than changes in the composition of the cohorts, that explains the growth in the proportion classified as a making a potentially difficult transition. The increased probability of making a potentially difficult transition is seen across all ethnicity/gender combinations for both high and low SES groups. However, it is among the low SES groups that the most dramatic differences are seen.

Table 5.1. Predicted probability of membership of a cluster in the 'Entering the Labour Market' grouping, by SES, gender and ethnicity for cohort born in 1989/90 and for same cohort assuming same influence of characteristics as that seen for cohort born in 1958

White male	Low SES	High SES	Overall	N
NCDS	95.8	67.9	96.5	4338
NCDS associations/LSYPE cohort	96.0	69.2	94	3161
LSYPE	30.6	34.0	40.0	3161
White female	Low SES	High SES	Overall	
NCDS	86.3	65.1	91.1	3844
NCDS associations/LSYPE cohort	86.6	66.4	90.6	3201
LSYPE	31.6	34.0	40.0.	3201
Non-White male	Low SES	High SES	Overall	
NCDS	90.8	65.3	93.6	103
NCDS associations/LSYPE cohort	91.0	66.6	92.0	1291
LSYPE	15.9	15.8	19.8	1291
Non-White female	Low SES	High SES	Overall	
NCDS	69.8	61.4	80.4	71
NCDS associations/LSYPE cohort	91.0	62.7	83.7	1491
LSYPE	16.1	15.6	19.5	1491
Overall	Low SES	High SES	Overall	
NCDS	93.2	66.8	95.0	8356
NCDS associations/LSYPE cohort	91.8	67.5	92.5	9144
LSYPE	28.9	31.1	36.9	9144

Notes: Predicted probabilities from underlying regression models reported in Table 3. Models also include regional dummy variables and missing variable dummies for the variables above. Missing value dummies are set to zero. NCDS results weighted using author's own attrition weighting scheme. LSYPE analysis weighted using dataset-provided attrition weights. 'High SES' individual is from a two parent household, where at least one parent works, at least one parent holds a degree, and their house is owner occupied. 'Low SES' individual is from a lone parent, workless household, where the parent's highest qualification is below A-Level and their home is socially rented. 'Overall' are predictions based on the complete sample, not a weighted average of the 'Low SES' and 'High SES' predictions.

Table 5.2. Predicted probability of membership of a cluster in the 'Accumulating Human Capital' grouping, by SES, gender and ethnicity for cohort born in 1989/90 and for same cohort assuming same influence of characteristics as that seen for cohort born in 1958

White male	Low SES	High SES	Overall	N
NCDS	1.3	31.8	2.0	4338
NCDS associations/LSYPE cohort	1.2	30.6	5.1	3161
LSYPE	39.4	54.8	48	3161
White female	Low SES	High SES	Overall	
NCDS	1.3	33.6	2.1	3844
NCDS associations/LSYPE cohort	1.2	32.3	5.4	3201
LSYPE	42.2	56.6	49.8	3201
Non-White male	Low SES	High SES	Overall	
NCDS	1.4	33.9	2.1	103
NCDS associations/LSYPE cohort	1.3	32.6	5.5	1291
LSYPE	61.7	76.7	71.6	1291
Non-White female	Low SES	High SES	Overall	
NCDS	1.1	35	2.0	71
NCDS associations/LSYPE cohort	1.3	33.7	5.5	1491
LSYPE	64.8	78.2	73.3	1491
Overall	Low SES	High SES	Overall	
NCDS	1.3	32.7	2.0	8356
NCDS associations/LSYPE cohort	1.2	31.8	5.3	9144
LSYPE	43.9	59.1	52.4	9144

Notes: Predicted probabilities from underlying regression models reported in Table 3. Models also include regional dummy variables and missing variable dummies for the variables above. Missing value dummies are set to zero. NCDS results weighted using author's own attrition weighting scheme. LSYPE analysis weighted using dataset-provided attrition weights. 'High SES' individual is from a two parent household, where at least one parent works, at least one parent holds a degree, and their house is owner occupied. 'Low SES' individual is from a lone parent, workless household, where the parent's highest qualification is below A-Level and their home is socially rented. 'Overall' are predictions based on the complete sample, not a weighted average of the 'Low SES' and 'High SES' predictions.

Table 5.3. Predicted probability of membership of a cluster in the "Potentially Difficult Transition", by SES, gender and ethnicity for cohort born in 1989/90 and for same cohort assuming same influence of characteristics as that seen for cohort born in 1958

White male	Low SES	High SES	Overall	N
NCDS	2.9	0.3	1.5	4338
NCDS associations/LSYPE cohort	2.8	0.3	0.9	3161
LSYPE	30	11.2	12	3161
White female	Low SES	High SES	Overall	
NCDS	12.4	1.3	6.9	3844
NCDS associations/LSYPE cohort	12.2	1.3	4.0	3201
LSYPE	26.1	9.4	10.2	3201
Non-White male	Low SES	High SES	Overall	
NCDS	7.8	0.8	4.2	103
NCDS associations/LSYPE cohort	7.7	0.8	2.5	1291
LSYPE	22.4	7.5	8.6	1291
Non-White female	Low SES	High SES	Overall	
NCDS	29	3.6	17.5	71
NCDS associations/LSYPE cohort	7.7	3.6	10.8	1491
LSYPE	19.1	6.2	7.1	1491
Overall	Low SES	High SES	Overall	
NCDS	5.5	0.6	2.9	8356
NCDS associations/LSYPE cohort	6.9	0.7	2.2	9144
LSYPE	27.2	9.8	10.7	9144

Notes: Predicted probabilities from underlying regression models reported in Table 3. Models also include regional dummy variables and missing variable dummies for the variables above. Missing value dummies are set to zero. NCDS results weighted using author's own attrition weighting scheme. LSYPE analysis weighted using dataset-provided attrition weights. 'High SES' individual is from a two parent household, where at least one parent works, at least one parent holds a degree, and their house is owner occupied. 'Low SES' individual is from a lone parent, workless household, where the parent's highest qualification is below A-Level and their home is socially rented. 'Overall' are predictions based on the complete sample, not a weighted average of the 'Low SES' and 'High SES' predictions.

Extending sequence analysis to age 24

A possible reservation about the results discussed so far is that they may not warrant particular attention since what is more important is how the school to work transitions play out in the longer run. Such a view may be justified if these early patterns do not persist. However, if they are predictive of transitions over a longer period, their importance is greatly increased.

In order to explore this, we also carried out an analysis of sequences beginning 30 months after turning 16 (i.e. following the end of the period we have been considering so far) up to approximately age 24, and compared the resulting groupings to those for the first 29 months post-16. This is only possible for the two datasets where the data are available: the NCDS and the BCS. We carry out sequence and cluster analysis on the same basis as was done for the earlier time period analyses, except that it starts 30 months after the September following their 16th birthday and continues for 69 months.[x] This time we use 14-cluster (rather than 7-cluster) solutions, reflecting the greater heterogeneity possible within longer sequences. Again, our choice of a 14-cluster solution is primarily on the basis of average silhouette distances.

We once again aggregate these clusters into our three broad groupings: Entering the Labour Market, Accumulating Human Capital and Potentially Difficult Transition. One particular challenge with conducting extended sequence analysis on the NCDS is the quality of the monthly activity data available particularly once we extend to age 24. The NCDS appears to have a rather systematic problem with gaps between different spells, which results in the loss of a substantial number of individuals from our analysis, reducing the sample size from 8,372 to 6,122. This loss seems concentrated among individuals in the 'Entering the Labour Market' group, and we suspect that this is responsible for inflating the size of the 'Accumulating Human Capital' grouping compared to that estimated in the shorter analysis. Consequently, there is a concern about the ability of the NCDS to support the longer-run analysis. The BCS analysis does not suffer from the same problem; extending to age 24 reduces the sample size only marginally (from 9,518 to 9,419). In view of this, we feel more confident about the BCS results.

In order to learn more about the relationship between the two sets of categorisations, we cross-tabulate the groupings into which individuals are placed in the shorter- (29 month) and longer-term (98 month) analyses. Considering first the NCDS, we see that a majority of individuals in the short-term groupings remain in the same grouping on the basis of the extended sequence analysis. There is also, for example, some movement from 'Entering the Labour Market' into 'Potentially Difficult Transition'. Similarly, some individuals initially classified as 'Potentially Difficult Transition' have seen a recovery by this later period. Overall, though, there is a strong correlation between the two sets of groupings. We should also note that the 'Potentially Difficult Transition' category grows primarily from individuals that were previously characterised as being 'Entering the Labour Market' and very few from the 'Accumulating Human Capital' grouping. In the BCS, the picture is much the same, except for the much-reduced size of the missing category, as discussed in the introduction to this section.

Table 6. NCDS: Cross-tabulation of groupings on basis of 16-18 sequence analysis and of groupings on basis of 18-25 sequence analysis

	18-24 Groupings				
16-18 Groupings	**ELM**	**AHC**	**PDT**	**Missing**	**Total (freq.)**
ELM	61.9	1.1	12.4	24.6	7,110
AHC	13.7	41.7	2.1	42.5	852
PDT	8.6	0.5	55.6	35.3	394
Missing	75.0	0.0	25.0	0.0	16
Total	54.6	5.2	13.4	26.9	8,372

Notes: ELM = Entering the Labour Market; AHC = Accumulating Human Capital; PDT = Potentially Difficult Transition. Reporting row proportions, except for the final (total) column, which reports frequencies.

Table 7. BCS: Cross-tabulation of groupings on basis of 16-18 sequence analysis and of groupings on basis of 18-25 sequence analysis

	18-24 Groupings				
16-18 Groupings	**ELM**	**AHC**	**PDT**	**Missing**	**Total (freq.)**
ELM	81.3	5.7	12.2	0.9	6,867
AHC	6.8	87.2	4.4	1.6	2,282
PDT	23.0	6.8	69.4	0.8	369
Total	61.1	25.3	12.6	1.0	9,518

Notes: ELM = Entering the Labour Market; AHC = Accumulating Human Capital; PDT = Potentially Difficult Transition. Reporting row proportions, except for the final (total) column, which reports frequencies.

What do we learn from this? Those who are making a potentially difficult transition in the earlier analysis are likely still to be making a potentially difficult transition on the basis of the longer run analysis: in the NCDS 85.9% of those deemed to be 'Potentially Difficult Transition' on the early basis (and for whom we can derive a longer run grouping) are placed in this group over the longer term; in the BCS the comparable figure is 70.0%. In addition, as one might expect, the longer analysis also picks up an additional number of cases that we deem to be potentially difficult transitions, on the basis of their trajectories post-29 months. However, we next explore whether this changes the risks of various observable characteristics associated with making a potentially difficult transition.

Reassuringly, in Table 8 we find a fairly similar pattern in the estimated average marginal effects of making each type of transition in this analysis as we did in the 29-month analysis, although there are unexpected or surprisingly insignificant results associated with a few characteristics for the NCDS. Nevertheless, we conclude that this suggests that while the sequence and cluster analyses themselves do not necessarily pick up all the individuals who are making a potentially difficult transition in this shorter timeframe, our shorter-run analysis nevertheless identifies the observable groups that are likely to be at greater risk.

Table 8. Estimated average marginal effects on the probability of an individual's membership of a cluster in each grouping (relative to the other two) from cohort-specific multinomial logistic regression models – Sequences to age 25

	Entering the Labour Market		Accumulating Human Capital		Potentially Difficult Transition	
	NCDS	BCS	NCDS	BCS	NCDS	BCS
Non-White	-0.013	-0.196***	-0.025	0.167***	0.037	0.028
	(-0.338)	(-5.367)	(-0.891)	(5.766)	(1.245)	(1.176)
Male	0.221***	0.128***	-0.060***	0.023***	-0.161***	-0.151***
	(18.692)	(12.545)	(-7.229)	(2.674)	(-14.557)	(-19.004)
Workless household	-0.037	-0.063**	0.007	-0.005	0.031	0.068***
	(-1.343)	(-2.263)	(0.309)	(-0.197)	(1.590)	(4.492)
Lone parent	-0.042*	-0.055**	0.035**	0.034	0.007	0.021
	(-1.764)	(-1.998)	(1.970)	(1.457)	(0.438)	(1.066)
Socially rented	-0.009	0.088**	0.015	-0.136***	-0.006	0.048**
	(-0.393)	(2.159)	(0.838)	(-3.655)	(-0.379)	(1.971)
Owner occupier	0.039*	0.013	-0.004	0.055*	-0.035**	-0.069**
	(1.707)	(0.402)	(-0.221)	(1.944)	(-2.125)	(-3.097)
Parental A-Levels	0.027	-0.166***	0.010	0.211***	-0.037***	-0.045**
	(1.558)	(-6.704)	(0.886)	(12.653)	(-2.591)	(-2.147)
Parental degree	-0.006	-0.250***	0.085***	0.348***	-0.078**	-0.098***
	(-0.157)	(-8.369)	(4.301)	(20.348)	(-2.206)	(-3.403)
N	6122	8574	6122	8574	6122	8574

Notes: Models also include regional dummy variables and missing variable dummies for the variables above. NCDS results weighted using author's own attrition weighting scheme. No weights applied to BCS analysis, as number excluded due to attrition was too small to model. T statistics reported in parentheses. Stars indicate statistical significance: * p=0.10; ** p=0.05; *** p=0.01.

Conclusions

In this paper we have used sequence analysis to analyse young people's transitions into the labour market and how these have changed over the past thirty years or so. The advantage of sequence analysis is that it allows us to consider young people's transition patterns as a whole, rather than concentrating on specific individual transitions and their timings. Our results shed new light on how the nature of young people's very early transitions have evolved, and on how the factors influencing them have changed over time.

It is well-established that young people's employment is more sensitive than older people's to the underlying strength of the economy. There are several reasons that combine to give such pro-cyclicality. For instance, firms may be more willing to lose younger workers in a recession than older workers with valuable experience. Equally, coming out of a downturn, firms may feel more prepared to recruit younger workers than older (more expensive) workers. Kahn (2010) shows that the time at which young people enter the labour market affects their subsequent outcomes. Scarring is one channel through which this operates; Gregg (2001) and Gregg & Tominey (2005) provide compelling evidence that youth unemployment can adversely affect both employment and earnings prospects as an adult.

The results in this paper are influenced by cyclicality to the extent that economic conditions prevailing at the time of reaching school-leaving age vary across cohort. However, since these cohorts span three decades, the results also pick up structural changes. Such changes reflect more than just economic influences and instead capture changes over time in, for instance, social preferences, demographics, technology and institutions. Unsurprisingly, we find a substantial shift away from early labour market entry towards gaining significant amounts of additional education or training before entering a job. However, in addition we have documented a rise in the proportion of successive cohorts that experience potentially difficult transitions, with prolonged or numerous spells not in education, employment or training. This group has grown in size from 5% of the sample born in 1958 to 12% of the sample born in 1990, with pretty much all of this growth concentrated between the 1980- and 1990-born cohorts.

Focussing on the 'Potentially Difficult Transition' group, there are two particularly striking results. First, females have gone from being more likely than males to be members of this group in early cohorts to being less likely by the more recent cohorts. One reason for this is likely to be a decline in the proportion of young women who choose to move quickly from education into an extended period of inactivity associated with homemaking or starting a family. Alongside this we find that individuals who move from education into long-term inactivity have become more likely to remain in education for two additional years (leaving education at age 18, rather than age 16) before entering inactivity.

Second, we find that young people from a non-white ethnic background go from being more likely than whites to experience a potentially difficult transition to being less likely. Across this period the non-white population of England has grown significantly in size, has diversified and has become more established. We suspect that all three of these facts have contributed to the relative improvement in the probability that individuals of non-white ethnicity experience transitions likely to be precursors of future economic prosperity.

In addition, we find that socioeconomic status, as captured through a combination of indicators, remains a powerful predictor of young people's chances of experiencing a potentially difficult transition. This is unsurprising, but underlines that it is among those from disadvantaged backgrounds where there has been the greatest increase in difficult transitions.

Lastly, we assess the extent to which the patterns seen in these early years predict longer-term outcomes. The fact that we find a high degree of correlation suggests that those likely to face ongoing difficulties in the labour market are often identifiable at a very early stage. This points to the importance of early transitions.

Acknowledgements

This research was funded by the UK Department for Business Innovation and Skills (BIS) and the Centre for Learning and Life Chances in Knowledge Economies and Societies, an ESRC-funded Research Centre (grant reference ES/J019135/1). We are grateful to Vahé Nafilyan and Laura Kirchner Sala of the Institute of Employment Studies for preparing the data. Matt Bursnall at BIS and Paolo Lucchino at NIESR provided helpful comments. The paper was presented at BIS, a meeting of the Jacob's Foundation PATHWAYS programme, and the UCL Institute of Education's Centre for Longitudinal Studies Cohort Studies Conference 2015. The usual disclaimer applies.

References

Abbott, A. (1995). Sequence Analysis: New Methods for Old Ideas. *Annual Review of Sociology, 21*, 93-113. https://doi.org/10.1146/annurev.so.21.080195.000521

Andrews, M., & Bradley, S. (1997). Modelling the Transition from School and the Demand for Training in the United Kingdom. *Economica, 64*(255), 387-413. https://doi.org/10.1111/1468-0335.00087

Anyadike-Danes, M., & McVicar, D. (2005). You'll never walk alone: Childhood influences and male career path clusters. *Labour Economics, 12*, 511-530. https://doi.org/10.1016/j.labeco.2005.05.008

Anyadike-Danes, M., & McVicar, D. (2010). My Brilliant Career: Characterizing the Early Labor Market Trajectories of British Women From Generation X. *Sociological Methods & Research, 38*(3), 482-512. https://doi.org/10.1177/0049124110362968

Arulampalam, W. (2001). Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages. *The Economic Journal, 111*(475), F585-F606. https://doi.org/10.1111/1468-0297.00664

Bell, D.N. and Blanchflower, D.G., 2011. Young people and the Great Recession. *Oxford Review of Economic Policy*, 27(2), pp.241-267. https://doi.org/10.1093/oxrep/grr011

Berrington, A. (2001). Transition to Adulthood in Britain *Transitions to Adulthood in Europe* (Vol. 10, pp. 67-102): Springer Netherlands. https://doi.org/10.1007/978-94-015-9717-3_4

Caspi, A., Entner Wright, B. R., Moffitt, T. E., & Silva, P. A. (1998). Early Failure in the Labor Market: Childhood and Adolescent Predictors of Unemployment in the Transition to Adulthood. *American Sociological Review, 63*(3), 424-451. https://doi.org/10.2307/2657557

Dorsett, R., & Lucchino, P. (2014). Explaining patterns in the school-to-work transition: An analysis using optimal matching. Advances in Life Course Research, 22(1-14). https://doi.org/10.1016/j.alcr.2014.07.002

Durbin, R., Eddy, S. R., Krogh, A., & Mitchison, G. (1998). *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/cbo9780511790492

Elzinga, C. H., & Liefbroer, A. C. (2007). De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis. *European Journal of Population, 23*, 225-250. https://doi.org/10.1007/s10680-007-9133-7

Fergusson, R., Pye, D., Esland, G., McLaughlin, E., & Muncie, J. (2000). Normalized dislocation and new subjectivities in post-16 markets for education and work. *Critical Social Policy, 20*(3), 283-305. https://doi.org/10.1177/026101830002000302

Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011a). Analyzing and Visualizing State Sequences in R with TraMineR. *Journal of Statistical Software, 40*(4), 1-37. https://doi.org/10.18637/jss.v040.i04

Gabadinho, A., Ritschard, G., Studer, M., & Müller, N. S. (2011b). Mining sequence data in R with the TraMineR package: A user's guide: University of Geneva, Switzerland. Retrieved from http://mephisto.unige.ch/pub/TraMineR/doc/TraMineR-Users-Guide.pdf

Gregg, P. (2001). The Impact of Youth Unemployment on Adult Unemployment in the NCDS. *The Economic Journal, 111*(475), F626-F653. https://doi.org/10.1111/1468-0297.00666

Gregg, P., & Tominey, E. (2005). The wage scar from male youth unemployment. *Labour Economics, 12*, 487-509. https://doi.org/10.1016/j.labeco.2005.05.004

Halpin, B., & Chan, T. W. (1998). Class Careers as Sequences: An Optimal Matching Analysis of Work-Life Histories. *European Sociological Review, 14*(2), 111-130. https://doi.org/10.1093/oxfordjournals.esr.a018230

Kahn, L. (2010) The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy, *Labour Economics*, 17(2), pp. 303-316. https://doi.org/10.1016/j.labeco.2009.09.002

Kaufman, L., & Rousseeuw, P. J. (1987). Clustering by means of Medoids *Statistical Data Analysis Based on the L1–Norm and Related Methods* (pp. 405-416): North-Holland.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. Chichester: Wiley Inter-Science. https://doi.org/10.1002/9780470316801

Kneale, D., Lupton, R., Obolenskaya, P., & Wiggins, R. D. (2010). A cross-cohort description of young people's housing experience in Britain over 30 years: An application of Sequence Analysis. DoQSS Working Paper 10-17, Institute of Education, University of London. Retrieved from http://repec.ioe.ac.uk/REPEc/pdf/qsswp1017.pdf

Lesnard, L. (2006). Optimal Matching and Social Science *Série des Documents de Travail du CREST*. Paris, France: Institut National de la Statistique et des Etudes Economiques. Retrieved from https://halshs.archives-ouvertes.fr/halshs-00008122/

Martin, P., Schoon, I., & Ross, A. (2008). Beyond Transitions: Applying Optimal Matching Analysis to Life Course Research. *International Journal of Social Research Methodology, 11*(3), 1-21. https://doi.org/10.1080/13645570701622025

Martin, P., & Wiggins, R. D. (2011). Optimal Matching Analysis. In M. Williams & W. P. Vogt (Eds.), *The SAGE Handbook of Innovation in Social Research Methods* (pp. 385-408). London: SAGE Publications Ltd. https://doi.org/10.4135/9781446268261.n22

OECD. (2008). *Jobs for Youth - United Kingdom*. Paris, France: Organisation for Economic Cooperation and Development. http://www.oecd.org/unitedkingdom/jobsforyouthdesemploispourlesjeunesunitedkingdom.htm

Paull, G. (2002). Biases in the Reporting of Labour Market Dynamics *IFS Working Paper Series*. London, UK: Institute for Fiscal Studies. Retrieved from http://www.ifs.org.uk/publications/2012

Pollock, R. (2007) Holistic trajectories: a study of combined employment, Housing and family careers by using multiple-sequence analysis. *Journal of the Royal Statistical Society*, Series A, 170(1), 167-183. https://doi.org/10.1111/j.1467-985X.2006.00450.x

Quintini, G., & Manfredi, T. (2009). Going Separate Ways? School-to-Work Transitions in the United States and Europe *OECD Social, Employment and Migration Working Paper*. Paris, France: Organisation for Economic Cooperation and Development. https://doi.org/10.1787/221717700447

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53-65. https://doi.org/10.1016/0377-0427(87)90125-7

Ryan, P., 2001. The school-to-work transition: a cross-national perspective. *Journal of economic literature*, 39(1), pp.34-92. https://doi.org/10.1257/jel.39.1.34

Schoon, I., McCulloch, A., Joshi, H. E., Wiggins, R. D., & Bynner, J. (2001). Transitions from school to work in a changing social context. *Young, 9*(1), 4-22. https://doi.org/10.1177/110330880100900102

Wu, L. L. (2000). Some Comments on Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect. *Sociological Methods and Research, 29*(1), 41-64. https://doi.org/10.1177/0049124100029001003

Endnotes

[i] For example, we made use of a separate variable reporting school leaving date and also used characteristics such as young people's highest educational qualification reported by age 23 to impute earlier education status.

[ii] We carry out a sensitivity analysis to assess the extent to which sample reduction is likely to influence our results by using the alternative approach of treating missing as a state in itself (Gabadinho, Ritschard, Studer, & Müller, 2011b, pp. 55-61). This makes little difference to our findings.
[iii] Optimal matching in this sense should not be confused with the identically named technique within the propensity score matching literature. Partly in order to avoid this ambiguity we use the term sequence analysis throughout.
[iv] Combinatorial approaches, such as those outlined by Elzinga (for example Elizinga & Liefbroer, 2007), are another alternative.
[v] This cluster contains individuals who have entered the labour market immediately at the end of compulsory schooling but return to education at a later point. This makes it slightly ambiguous if they should be classified as 'Entering the Labour Market' (since they do this but then leave again) or 'Accumulating Human Capital' (since they return to do this but are not in education throughout the period). In any case, it makes up only approximately 2.5% of the sample; little changes if it is reclassified as AHC or dropped entirely.
[vi] It is also possible, though, that some of this effect is explained by under-reporting of short spells in the NCDS/BCS, as discussed earlier.
[vii] We also fitted a single multinomial logistic regression model on the pooled sample from all cohorts, including a cohort regressor and all predictors interacted with these. These replicated the results obtained from the separate models, but allowed for inference testing of the differences between the influences of characteristics in each cohort. These significance tests are not reported in this paper but are available on request.
[viii] The distinction between high SES and low SES does not conform to any standard definition. The two groups were chosen in order to satisfy two criteria: first, that the characteristics used in the definition had a strong association with disadvantage such that it was plausible to view the high SES group as unambiguously "better off" (at least on average) than the low SES group and, second, that resulting groups were of a sufficient size to be of practical use.
[ix] The NCDS and LSYPE rows in Table 5 contain identical results to corresponding rows in Table 4 but are included for convenience of comparison.
[x] In addition, we carried out the same analysis over the whole time period (i.e. both the initial 29 months and the following 69 months) and achieved similar results to those reported later in this section.