Transition pathways in HILDA — an analysis using optimal matching and cluster analysis

Jane Fry & Clare Boulton*

Productivity Commission

Abstract

In this paper, data from HILDA's education and labour market activity calendar are used to examine transitions within and between the labour market and the education system for about 6500 working-age individuals.

The HILDA calendar is rarely used but provides an invaluable data source for analysing transitions. Activity data captured at frequent intervals are critical to identifying any 'turbulence' in individual activity patterns. The analysis of transitions can accommodate some of the more complex patterns, including 'churning' and 'stepping stones' from education to work.

For four age segments, optimal matching and cluster analysis techniques are used to identify a small number of groups — called pathways — based on similarities in the patterns of activities that individuals undertake. Some pathways are dominated by education or by work, some mostly by NILF; others show transitions (such as education to work or work to NILF). The pathways associated with work tend to predominate, except among the senior segment (where individuals tend to retire). Individual characteristics, such as gender, education and reasons for labour market withdrawal, add richness to the story for each pathway.

This paper is best viewed in colour for the charts to be meaningful.

---

* This paper is a shortened version of Fry and Boulton (2013). The authors would particularly like to thank Noel Gaston, Jenny Gordon, Lisa Gropp, Peter Harris and Patrick Jomini from the Productivity Commission and Glenda Quintini from the OECD for their extensive feedback. The analysis uses unit record data from the *Household, Income and Labour Dynamics in Australia* (HILDA) Survey. The HILDA Project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and views in this paper are those of the authors and should not be attributed to FaHCSIA, the Melbourne Institute, the OECD or the Productivity Commission.

1 Introduction

The Australian labour market is adjusting constantly to changes in labour demand and supply. The incentives to participate in the labour force, and to study, change with varying economic circumstances and life cycle factors. As some workers expand (or contract) their hours of work, others change jobs, while others cease work for various lengths of time. Some search for a job, others leave the labour force. To improve their prospects, some remain in, or return to, education. Others exhibit stable employment with no change in their labour market status over long periods of time.

The usual statistics associated with the labour market relate to a point in time, measuring participation, employment and unemployment rates for the working-age population or for particular subgroups. There are also data on engagement in study — particularly for individuals not in employment. These data reveal little about transitions between different activities: they do not account for changes in engagement with the labour market and with the education system over time.

Longitudinal data are required to analyse transitions between — and persistence in — education and labour market activities. For some people, a lot can happen in a year, while for others many years can pass before a change occurs. To capture frequent changes as well as longer term trends and one-off transitions, education and labour market activities need to be tracked often and over an extended period of time.

In this paper, ten years of data from the education and labour market calendar in HILDA are used to examine transitions within and between the labour market and the education system in Australia. Tracking activities over an extended period, produces a 'holistic' approach to transitions that accommodates some complex patterns, including 'churning' and 'stepping stones' from education to work. By examining what activities individuals undertake, when and for how long, distinct patterns — called pathways — are identified to facilitate further analysis.

The remainder of this paper is structured as follows. Section 2 outlines the conceptual approach used to study transitions and key findings from the relevant literature. Section 3 describes the data. Section 4 details the analytical techniques used to identify the pathways (optimal matching and cluster analysis; OMCA). Section 5 provides an overview of the pathways identified and describes a selection of pathways (more detailed results are given in appendix B). Section 6 concludes.

2 Transitions and pathways: the conceptual approach and literature review

There are three standard approaches to the study of transitions.

1. Some studies have focused on a transition's end point, analysing data on individual outcomes — such as labour market status or wages. Many studies of transitions from school to work adopt this approach.
2. Another approach is to consider duration or time spent in labour market states. For example, some studies analyse the duration of unemployment spells.
3. Yet another approach considers what happens over a given period. The number of times that individuals experience a particular type of transition (such as job loss) over time can provide insights into the likelihood of repeat transitions (recurrence). This is particularly useful for characterising labour market histories or actual experiences. Moreover, the numbers of individuals that experience particular transitions can also indicate the likelihood that others may experience similar transitions.

A relatively new approach studies transitions as a sequence of activities over (part of) a career or lifetime. This approach relies on data collected in the form of a diary or a calendar. Sequence analysis techniques — usually OMCA — can then be used to identify individuals with similar activity patterns.

Sequence analysis techniques can account for all transitions in the data concurrently (Pollock, Antcliff and Ralphs 2002) and consider the overall pattern, incidence, timing, duration and order of activities (Han and Moen 1999). With a better understanding of mobility (or persistence) patterns, researchers can answer a variety of questions about, among other things, the ease of finding employment, the stability of that employment, and whether non-employment is temporary or a trap (Quintini and Manfredi 2009). By considering particular experiences — such as periods spent not in employment, education or training (NEET) — in the context of an individual's labour market history, researchers can also distinguish, 'for example, between transitory "gap years" and deep disconnect from the labour market' (Dorsett and Lucchino 2012, p. 102).

Sequence analysis is not subject to some of the limitations of other methods of analysing transitions. Long sequences with many activity categories can make stochastic models (such as Markov chains) difficult to identify (King 2011). Moreover, 'it might be better to treat two sequences that differ only by a few steps as very close relatives rather than treating them as forever divided after that initial fork in the Markov path' (Anyadike Danes and McVicar 2005, p. 516).

Overall, the sequence approach has several advantages over the other approaches, not least of which is the ability to analyse all of the different activities that individuals undertake over time. In essence, the sequence approach combines the best aspects of the other three approaches but remains statistically agnostic.

Moreover, descriptive analysis of the pathways may facilitate the formation of hypotheses about the factors important for determining pathway membership. These can be tested using statistical models (see, for example, Anyadike-Danes and McVicar (2003), Corrales-Herrero and Rodríguez-Prado (2012), Dorsett and Lucchino (2012) and Quintini and Manfredi (2009)). What this suggests is a complementary, rather than alternative, role for OMCA to more conventional statistical analysis. According to Halpin (2010, p. 367), sequence analysis 'gives a holistic perspective', providing context to econometric analysis.

Transition pathways literature

A growing literature has adopted a sequence approach to study labour market transitions. Yu et al. (2012) is the only study using Australian data.

Design features

Some studies — such as Fuller (2011) and Yu et al. (2012) — relate to the entire working-age population. Most studies focus on particular groups, such as youths or older individuals, and the types of transitions relevant to their life stage, such as the initial entry into the workforce or retirement.

Studies of labour market transitions (based on the sequence approach) usually consider education, employment and NEET, often disaggregated in some way to form the activity categories for analysis. Employment can be differentiated by characteristics such as employment status (employee or self-employed), full-time or part-time, or broad occupational group. In some studies, periods of concurrent study and employment are identified (to capture apprenticeships and other types of study undertaken while working). NEET is often split between unemployment and not in the labour force (NILF). In most studies, the NILF category includes individuals who are no longer actively searching for work, those who are unable to work due to illness/injury or disability, and those who retire or temporarily withdraw from the labour force to care for young children or disabled relatives.

There is a tradeoff between the length of time covered and the frequency of the observations on those activities. Intra-year data over a short period of time show some of the turbulent (churning) patterns associated with short periods of

employment and unemployment. On the other hand, activity sequences over a long period are needed to identify longer-term patterns. Thus, tracking post-school activities over three years may not be sufficiently long to observe some youths gain stable employment. Some individuals who are long-term unemployed may be observed to secure employment, but a longer time would be required to determine whether this employment continues or is transitory.

Key findings of the transition pathways literature

A review of the literature shows that most individuals follow pathways that are strongly associated with investing in or using their human capital. The few studies focussing on the later life activities of seniors are dominated by transitions out of the labour force.

Comparing specific findings from different studies can be difficult due to the variations in their designs. However, some broad themes emerge.

Most studies using sequence analysis methods identify about five to ten pathways based on distinct patterns in individuals' activities. The pathways do not usually represent equal shares of the sample — a single pathway can dominate, in some cases representing more than 70 per cent of the sample. For example, the *Employed* pathway in Anyadike-Danes and McVicar (2005) represented 75 per cent of individuals. In that pathway, prolonged periods of employment dominated the observed patterns. In Corrales-Herrero and Rodríguez-Prado (2012), one pathway — labelled *Continuously in Full-Time Job* — represented 71 per cent of the sample.

Although the sequences vary within a pathway, each pathway has a distinct overall pattern. Descriptive labels for each pathway are usually formulated based on predominant characteristics of the sequences within that pathway, such as *Quick and Sustainable Employment* (Céreq 2005), *Work to Family Care* (Scherer 2001), *Gap Year Between Studies* (Albert Verdú and Davia 2010) and *Higher Education Dominated* (McVicar and Anyadike-Danes 2002).

While the results from these studies are not strictly comparable, table 1 demonstrates that judgments can often be made about the pathways that are associated most closely with unemployment and/or labour market withdrawal. In table 1 these pathways are described as 'red' pathways; as might be expected, they are usually few in number and represent a small percentage of the sample.

The Australian study by Yu et al. (2012) uses the annual time of survey data from HILDA. The authors focus on occupational mobility defined at a broad level. As a result, most pathways are characterised by many years of employment in a

particular broad occupation. The pathway associated with prolonged periods of NILF (averaging more than seven years) is dominated by individuals who had likely retired.

Characteristics of the studies include (table 1): the number and types of activities to be tracked over time (section 3), the timeframe and data frequency and the groups identified for analysis.

In this study, relatively finely defined activities are captured at monthly intervals over a ten-year period to permit both short-term 'turbulence' and longer-term patterns to emerge.

To capture different 'life stages' the working-age population is divided into four segments: youths (aged 15–24), young adults (aged 25–39), mature adults (aged 40–54) and seniors (aged 55–64). These life stages are likely to affect the types of activity patterns observed for different age groups, as well as their interpretation.

Table 1 **Summary characteristics of selected studies from the transition pathways literature**

Segment	*Location*	*Study*	*Sample period*	*Sequence length*[a]	*Frequency*	*Activities*	*Pathways*	*'Red' pathways*[b]
Youths	Britain	Dorsett & Lucchino (2012)	1991–2008	5 yrs	Monthly	Ed, Emp, U, NILF	8	1
Youths[c]	Britain	Anyadike-Danes & McVicar (2010)	1985–2000	13 yrs	Annual	Ed, Emp, U, NILF	10	2
Youths[d]	Britain	Anyadike-Danes & McVicar (2005)	1985–2000	13 yrs	Annual	Ed, Emp, U, NILF	6	2
Youths	Britain	Scherer (2001)	1985–1996	5 yrs	Monthly	Ed, Emp, U, NILF	12	3
Youths	Northern Ireland	McVicar & Anyadike-Danes (2002)	1993–1999	6 yrs	Monthly	Ed, Emp, Jobless	5	1
Youths	EU	Quintini & Manfredi (2009)	1994–2001	5 yrs	Monthly	Ed, Emp, U, NILF	9	3
Youths	EU	Brzinsky-Fay (2007)	1993–2000	5 yrs	Monthly	Ed, Emp, U, NILF	8	2
Youths	France	Céreq (2005)	2001–2004	3 yrs	Monthly	Ed, Emp, U, NILF	6	3
Youths	Germany	Scherer (2001)	1985–1996	5 yrs	Monthly	Ed, Emp, U, NILF	12	3
Youths	Spain	Albert Verdú & Davia (2010)	early 2000s	3 yrs	Monthly	Ed, Emp, U, NILF	6	1
Youths	Spain	Corrales-Herrero & Rodríguez-Prado (2012)	2001–2004	4 yrs	Monthly	Ed, Emp, U, NILF	7	2
Youths	US	Quintini & Manfredi (2009)	1980s	5 yrs	Monthly	Ed, Emp, U, NILF	9	3
Youths	US	Quintini & Manfredi (2009)	1997–2005	5 yrs	Monthly	Ed, Emp, U, NILF	9	2
Adults	China	Lin (2013)	2003[e]	25 yrs	Annual	Ed, Emp (by occ.), U, NILF	4	1
Adults[d]	Germany	Kogan (2007)	1995–2000	6 yrs	Annual	Emp (by occ.), U, NILF	10	2
Seniors	Britain	Fasang (2010)	1990–2005	8 yrs	Monthly	Emp, U, NILF (by income source)	8	1
Seniors	Germany	Fasang (2010)	1990–2005	8 yrs	Monthly	Emp, Jobless (by income source)	7	1
Seniors	US (part)	Han & Moen (1999)	1994-95[e]	(various)	Annual	Emp (by occ.), Jobless, Retired	5	1
All	Canada	Fuller (2011)	2001–2005	4 yrs	Monthly	Ed, Emp, Jobless	6–7	1
All	Australia	Yu et al. (2012)	2001–2009	9 yrs	Annual	Ed, Emp (by occ.), U, NILF	10	1

[a] In some cases the sequence length is shorter than the sample period depending on how individuals are selected for analysis. [b] Although those studies are not all strictly comparable, for the purposes of this summary a judgment has been made as to the number of pathways that were associated mostly with unemployment and/or inactivity. These have been labelled 'red' pathways. Ideally, these pathways would not include inactivity associated with childrearing for women, disability or prolonged illness, and retirement for seniors. [c] Women only. [d] Men only. [e] Annual activity recall contained in single year of survey.

Sources: As listed.

3 Data

The HILDA calendar is a rich source of information about activities that occur between the annual interviews. Calendar data are collected for each third of a month back to the start of the previous financial year. Since the interviews occur sometime after July, this creates overlaps in the calendar data. These overlaps were removed by only extracting the calendar data in each wave for the previous financial year.[2]

Five activity categories are constructed:

1. study only, defined as enrolled in a school or educational course and not in employment (includes holidays while enrolled)
2. study and employment (includes holidays)
3. employment only (employed but not enrolled in study)
4. unemployment only (not employed, not enrolled, but looking for work)
5. not in the labour force (NILF) only (not employed, not looking for work and not enrolled in education).[3]

The three employment and study categories are included to capture the different ways that individuals develop and use their human capital at different stages of life.

The calendar data are then aggregated to monthly data, using a simple hierarchy when more than one activity category is recorded in a month.[4] This process of aggregating the data retains most of their richness, but reduces their number to a level that is practical for analysis.

---

2 See appendix A in Fry and Boulton (2013) for details.

3 This set of activities is limited by the data. For example, while calendar information is collected on whether each job undertaken is full-time or part-time, reliably tracking this information when there are multiple jobs across waves is impossible. Retirement, illness or disability, and child-rearing activities could be separately identified from other activities that result in an individual leaving the labour force. However, in HILDA, such 'reasons' for any periods of NILF between interviews can only be imputed using non-calendar information pertaining to the activities at the time of the interview.

4 The hierarchy is as follows. Employment dominates, followed by unemployment then NILF. Thus any month third in employment translates to employment for that month. If there is no employment during the month, unemployment will be recorded if there is at least one month third in that state. A month of NILF can only result if all corresponding month thirds are spent in NILF. In a similar way, study dominates non-study. Thus the activity category 'study only' translates to at least one third of the month in education, and no employment during the month.

After excluding individuals with incomplete calendars, the final dataset tracks monthly activities from July 2000 to June 2010 for 6566 individuals aged 15–64 in 2001: 877 youths; 2289 young adults; 2354 mature adults and 1046 seniors.[5] Sample weights were not applied because the analytical techniques used do not support weighting.

4 Optimal matching and cluster analysis

Sequences of activities can be described by the proportion of time spent in each activity or by the temporal ordering of activities. A richer picture appears when both the duration and order of activities are considered simultaneously. When used in tandem, optimal matching and cluster analysis (OMCA) provide a systematic way to compare sequences and group those with similar patterns. These techniques are data driven and exploratory (describing what is) rather than confirmatory (formulating statistical models and testing hypotheses). Each technique is explained below.

Optimal matching

Optimal matching (OM) is used to compare sequences of events, activities or states across time. The use of OM is a relatively recent development in the labour market transitions literature, compared to other fields such as biology. OM measures similarity by matching two sequences in an optimal way by quantifying (dis)similarity on a numerical scale. It does not measure similarity using just two or three categories (for example, exactly the same, completely different or somewhat different). Instead, OM allows the analyst to examine a large number of sequences with multiple activities in many (pattern) variations and quantify 'how different' the sequences are. By measuring how close each sequence is to another, all sequences can be ranked according to their similarity. More specifically, the technique identifies the number and type of changes that must be made to one sequence to make it identical to another sequence.

5 In general, it appears that individuals who follow pathways that are characterised by a significant amount of time spent NILF or studying are more likely to leave the sample relative to individuals in pathways that involve a significant proportion of time working. These non-work pathways are therefore likely to be more prevalent than they appear in the data.

Transformations

In transforming sequences, three types of changes are permitted: substitutions (swap one activity for another), deletions and insertions. Deletions and insertions are used to align sequences that are similar over different intervals. A realignment can be effected by an insertion in one sequence or a deletion in the other (the two operations are called an indel).

Each operation is given a 'cost'; the distance between two sequences is the sum of these costs (box 1).[6] If there are different ways to transform the sequences, resulting in different total costs, the distance between the two sequences is given by the minimum total cost.

Box 1 **Calculating the distance between two activity sequences**

To illustrate how OM works, consider two short hypothetical activity sequences, for Jane and Clare. Jane was not in the labour force (NILF) for the first month but started a new job in the second month, but that job was not a good fit for her and she spent the third month looking for a new job. She was employed in the second job for the fourth and fifth months. Clare, on the other hand, was employed at the beginning of the first month, but lost her job at the end of the second month. It took her a month to find another job, and after working in that job for two months she decided to study while continuing to work. So the two women have the same activities in only one of the five months (month 4).

	Month 1	Month 2	Month 3	Month 4	Month 5
Jane	NILF	Work only	Unemployment	Work only	Work only
Clare	Work only	Unemployment	Work only	Work only	Work & study

Clare's sequence can be transformed to be identical to Jane's by substituting the four months of activities that differ. Superficially, Jane and Clare's sequences appear dissimilar. It takes 4 substitutions, or equivalently, 4 insertions and 4 deletions (8 indels) to align them, as shown in transformation A below.

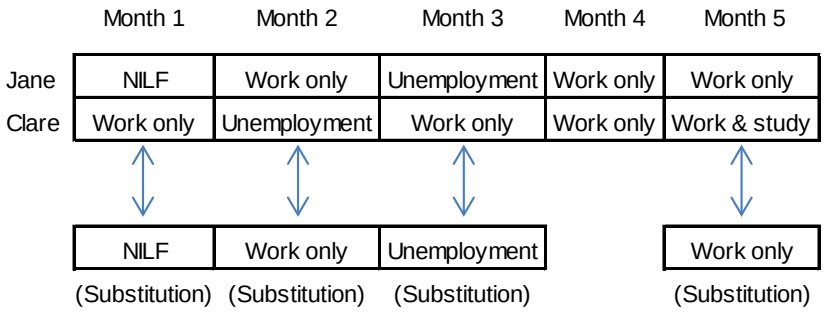
To keep things simple, no preference is given to a particular type of change. Each indel operation 'costs' 1 and each substitution 'costs' 2 (as a substitution is equivalent to an insertion at one point in a sequence and a deletion at the next point along the same sequence). So regardless of whether 4 substitutions or 8 indels are used to transform Clare's sequence to match Jane's, it costs 8 to do so.

(Continued next page)

6 The costs of substitutions can differ according to the pair of activities concerned (discussed below).

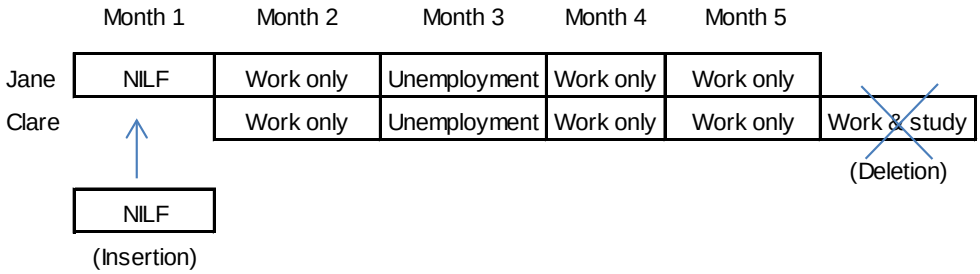
Box 1 (continued)

Transformation A (4 substitutions; or 8 indels)



An alternative transformation can be used to align the sequences. If 'NILF' is inserted at the first month of Clare's sequence and 'work & study' is deleted in month 5, then this aligns the two sequences for the remaining four months without requiring any substitutions. The distance, or cost of transforming Clare's sequence into Jane's, is now 2 indels:

Transformation B (2 indels)



Transformation B uses 2 indels compared to transformation A, which uses the equivalent of 8 indels. Therefore, B is a less costly way to achieve alignment of the sequences. With transformation B, Clare's sequence is a quarter of the distance from Jane's that was indicated with transformation A. OM would determine transformation B is the optimal way to match the sequences, and that there is a distance of 2 between Jane and Clare's sequences.

To give some intuition, the maximum distance between two sequences with five observation periods, and substitution costs of 2, would be 10. For example, this is the distance between a sequence that comprises five months of work and a sequence with five months of NILF.

Source: Adapted from Quintini and Manfredi (2009).

Setting substitution costs

The cost of substituting one activity for another must be specified before OM can be performed (Brzinsky-Fay and Kohler 2010).[7] According to Martin and Wiggins (2011, p. 389), '[s]ubstitution cost specification is a sore point in the world of OM. It is neither obvious nor agreed among researchers which principles should guide it'.[8]

The literature recognises three approaches to setting substitution costs: default; theoretically-based; and data-driven. In the default approach, a substitution between any pair of activities is assigned the same cost — for most studies that cost is 2.[9] The other two approaches specify substitution costs in a square matrix, in which the elements give the cost of substituting each type of activity for every other type. The costs can differ according to the pair of activities concerned. However, as a cost represents the distance between two activities, the matrix is symmetric — that is, the distance from activity A to activity B is the same as from B to A.

In the theoretically-based approach, substitution costs are set in a subjective manner to show (for each activity) which activities ought to be 'closer' or 'more distant' according to theory. A data-driven approach uses the transitions in the data to determine how close the activities are, and is considered to be the most neutral method (Anyadike-Danes and McVicar 2010; Hollister 2009). As a result, the data-driven approach is becoming the more common choice in the literature and is used to set substitution costs for this analysis.[10]

Costs in this study

For each age group, the substitution cost matrix was determined using the probability of each type of transition from the dataset. That is, the substitution cost

---

7 The costs of indels must also be specified before the OM stage of sequence analysis can be performed. This is discussed in Fry and Boulton (2013).

8 However, there does appear to be consensus on the importance of conducting sensitivity analysis using different costs. Many papers report that results are robust to changing substitution costs (Corrales-Herrero and Rodríguez-Prado 2012).

9 In Stata, and in other statistical programs, the default is to set the costs of indels to 1 and substitutions to 2, so that an insertion and a deletion are assigned the same cost as a substitution. Since all substitutions have the same cost under the default settings, all types of activities are equally distant from each other.

10 Fry and Boulton (2013) discuss the alternative approaches to setting substitution costs in more detail.

is the inverse of the average probability of a transition between activities A and B.[11] The substitution costs differ according to the pair of activities concerned. That is, they are the lowest for pairs of activities that have the greatest number of transitions between them.[12] However, in this analysis, substitution costs are held constant throughout the length of the sequence; an activity's position in the sequence has no bearing on the cost of a substitution for another activity.[13]

Distances between each pair of sequences are formed into a matrix, so that pairs of different sequences have relatively larger distances than pairs of similar sequences (Corrales-Herrero and Rodríguez-Prado 2012). These distances are then used to group similar sequences together.

Cluster analysis

In the context of sequence analysis, cluster analysis identifies groups of individuals with similar activity sequences, based on the distances from OM. In this paper, cluster analysis is hierarchical and agglomerative: it starts with all individuals in their own cluster (or group) and iteratively and progressively combines the clusters until all individuals in the sample are in a single cluster. Diagnostic information is used to determine the point at which the appropriate number of clusters is formed. In this study, the clusters are referred to as a pathways.

The clustering process

The distance calculated from OM is the starting point for the cluster analysis. In this analysis, the clustering method used is Ward's method. Ward's method calculates the change in the Error Sum-of-Squares (ESS) that would result from each pair of clusters being combined. The ESS is the sum of the squared distances of each individual sequence to the centre of the cluster. The ESS is therefore a measure of homogeneity of a cluster. Pairs of clusters are selected for joining according to the minimum increase in ESS. This process is repeated until there is one cluster. Diagnostic output can then be used to determine where in the clustering process the

[11] Actually 2 – Pr(A to B) – Pr(B to A), which preserves symmetry in the substitution costs. Each probability is based on the transition matrices for each time point.

[12] Fry and Boulton (2013) show the substitution cost matrix for youths and compare it to the matrix for seniors.

[13] A less restrictive approach would be to allow these costs to vary over time. For example, the distance between employment and unemployment could be smaller when the unemployment rate is high and larger when it is low (Lesnard 2010). However, Lesnard's (2006) extension to OM that uses dynamic substitution costs is not used in this analysis as the importance of the dynamic element has yet to be established empirically (Hollister 2009).

appropriate number of clusters is formed (discussed below). Ward's measure is favoured because it is commonly used in the literature on labour market transitions and, according to Corrales-Herrero and Rodríguez-Prado (2012, p. 3783), it results in the 'most homogeneous clusters'.

The progressive clustering process is shown in a dendrogram (box 2). The joining together of two clusters is shown by a horizontal line, with a measure related to distance between clusters shown on the vertical scale. The dendrogram is used to decide how many clusters there should be.

Determining the number of clusters

There are several approaches to identifying the number of clusters. In deciding the number of clusters in between the upper and lower limits, a balance between the variation within, and the variation between, clusters needs to be struck. Having too few clusters risks having too much variation within the clusters making it difficult to identify what the sequences in a cluster have in common. On the other hand, having too many clusters makes it difficult to identify what distinguishes sequences in one cluster from those in another.

It is the degree of subjectivity in determining the number of clusters in the data that is the most criticised aspect of cluster analysis. The example in box 2 illustrates the subjectivity in choosing the number of clusters. Either two or three clusters are possible, but three clusters are chosen with the aid of the dendrogram.

To reduce the degree of subjectivity in the analysis, the number of clusters was determined using a dendrogram in combination with descriptive analysis of the clusters to examine their 'plausibility'. For youths, the dendrogram indicated five clusters. For each the other age segments, the dendrograms suggested four clusters. These groups were found to be plausible for each segment and descriptive analysis readily provided a story for each cluster.

Box 2 **Interpreting dendrograms**

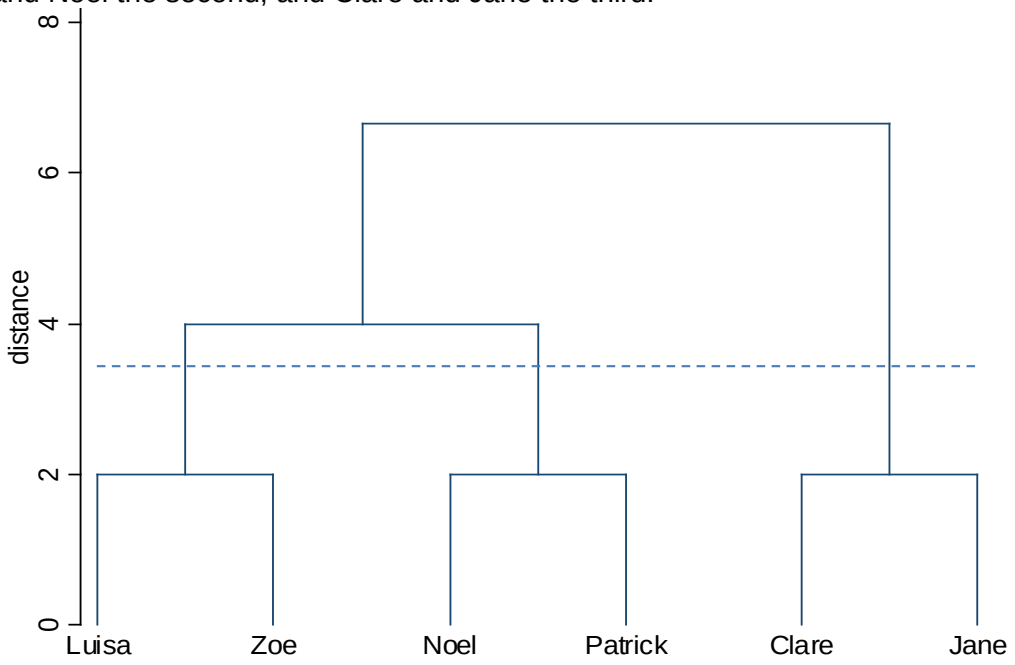
Consider four sequences in addition to Jane and Clare from box 1. OM calculated that the minimum distance between Jane and Clare is 2 (2 indels). Patrick and Noel are also a distance of 2 apart (one substitution), but both are further than 2 away from Jane or Clare. Likewise, it would take 2 indels to align Luisa and Zoe's sequences, and neither is closer than 2 to Jane, Clare, Patrick or Noel. So Jane and Clare are closest to each other, as are Patrick and Noel, and Luisa and Zoe.

	Month 1	Month 2	Month 3	Month 4	Month 5
Jane	NILF	Work only	Unemployment	Work only	Work only
Clare	Work only	Unemployment	Work only	Work only	Work & study
Patrick	Work only	Work only	Work only	Work only	NILF
Noel	Work only	Work only	Work only	Work only	Work only
Luisa	Study only	Work & study	Work only	Work only	Work only
Zoe	Work & study	Work only	Work only	Work only	Work only

(Continued next page)

Box 2 (continued)

The first stage of agglomerative clustering is where each individual is in their own cluster. This is shown along the horizontal axis in the dendrogram below. Along the vertical axis the distance between the clusters is shown, which at this stage is zero. In the second stage, those individuals whose trajectories are the shortest distance apart are clustered together. At a distance of 2, Luisa and Zoe form the first cluster, Patrick and Noel the second, and Clare and Jane the third.



The heterogeneity of each of the clusters is calculated using Ward's method. The increase in within-cluster variation that would result from joining the third cluster with either of the other clusters would be greater than the increase in variation resulting from combining clusters one and two. Therefore at the third round, there are just two clusters with Patrick, Noel, Luisa and Zoe in the first, and Jane and Clare in the second. In the final round the two clusters are joined to form a single cluster.

Drawing a horizontal line at different distances can help determine the number of clusters in the data. The dendrogram shows that because the three pairs of sequences are equally alike, and therefore join together at the same distance, there can be no more than three clusters. One cluster would have too much variation, and no pattern would be discernable. The six cluster solution would not reduce the variability in the data. So the choice is between two and three clusters.

A dashed line is drawn at a distance where there are three clusters in the data. This was done because the distance between the three and two cluster solutions, 2, is not much less than the distance between the two and one cluster solution. The three cluster solution is the first solution that is stable over a reasonable distance.

Labelling and describing the clusters

Clusters are often labelled in a descriptive way. In the education and labour market transitions literature, clusters are referred to as pathways, which picks up on the sequencing of activities. Though arbitrary, the pathway labels are derived from a commonly observed pattern of activities (for example, the *NILF to Work* pathway for young adults).

5 The results

This section presents an overview of the 17 pathways found, followed by the interpretation of a relatively prominent pathway for youths (*Churning with Work*) and a pathway displaying a common transition among young adults (*NILF to Work*).

An overview of the pathways identified

For each age segment, four or five pathways were identified (table 2).[14] Some pathways with similar characteristics, such as those labelled *Work* or *Prolonged NILF*, appear in several age segments. Within each segment, the most prevalent pathways (those followed by the largest numbers in that segment) were those associated with work, except among seniors for whom the *NILF* pathway was most prevalent. Pathways associated with education — *Education to Work*, *Work and Study to Work*, and *Work, with or without Study* — were identified for youths and, in the latter case, young adults (although this does not mean that individuals in other pathways or age segments did not undertake any education).

[14] For presentation purposes, each pathway has been given a descriptive label according to the general pattern that characterises the experiences of most individuals in that particular pathway. These labels are unrelated to the analytical techniques used to generate the pathways.

Table 2 **Pathways and prevalence by age segment**

Per cent of age segment[a]

Pathway	*Youths*	*Young adults*	*Mature adults*	*Seniors*	*All segments*[b]
Education to Work	8.2				1.1
Work and Study to Work	13.9				1.9
Churning with Work	51.7				6.6
Work, with or without Study[c]	16.8	12.4			6.6
Work		63.4	69.2	21.5	50.4
Prolonged NILF	9.5	12.0	13.3		10.2
NILF				51.2	8.2
NILF to Work		12.2	8.2		7.2
Work to NILF			9.2		3.3
Early Work to NILF				13.5	2.1
Later Work to NILF				13.8	2.2
Total	100.0	100.0	100.0	100.0	100.0

[a] For some age segments, pathway shares may not add to 100 due to rounding. [b] Pathway share (combining all four age segments) in the total sample. [c] This pathway is likely to represent transitions from education to work for youths, and (subsequent) re-entry to education for young adults.

Source: Authors' estimates based on HILDA waves 1–10.

In general, the pathways show distinct patterns. Some individuals spent ten years continuously in the same activity. For example, although still of working age, 8 per cent of the sample were already retired in 2000 (last column, table 2). Another 51 per cent spent most of their time in work; comprising 21 per cent (not shown) who were always in employment and 30 per cent (not shown) who undertook other activities some of the time. Other key patterns were:

- 7 per cent of individuals had a tenuous attachment to the labour force, churning frequently in and out of the labour force
- 5 per cent moved from education to work
- 4 per cent returned to education and spent long periods studying while working
- 10 per cent spent several years outside the labour force (mostly raising children)
- 7 per cent transitioned to work, having been outside the labour force
- 8 per cent transitioned to retirement (table 2).

Figure A.1 (Appendix A) shows in the form of an 'atlas', the activity patterns for all individuals in each of the 17 pathways. These plots are known as sequence index plots. Figure A.2 shows another representation which abstracts from the sequence dimension to indicate the proportion of individuals in each activity in any one period. These charts are called chronographs.

Interpreting the pathways

Associating individuals' activity patterns with their characteristics (available elsewhere in HILDA) reveals an interesting story of labour market transitions and interactions with the education system. Although events such as the financial crisis cannot be identified in these data, the sequence index plots show distinctive patterns. For example, one can clearly identify:

- school leavers and the various transitions they undertake
- working students early in the process of transitioning to work
- three churning pathways where work is interspersed with a variety of other activities
- women who leave the workforce to have babies, with some returning to the workforce
- several forms of transition to retirement and several pathways dominated by work.

Every pathway has a story. In this section, two pathways are used to illustrate the richness of the analysis of the data organised in terms of sequences. Three more pathways are found in appendix B — the *Work and Study to Work* pathway for youths, the *Work* pathway for young adults and the *Work to NILF* pathway for mature adults. The reader is referred to Fry and Boulton (2013) for the full set of results.

Youths in the Churning with Work pathway

Churning with Work is the largest pathway for the youth segment: it represents 453 individuals (51.7 per cent of all youths) (table B.4). These individuals are further advanced in their transition from education to work compared to some other youths, for example youths in the *Work and Study to Work* pathway (appendix B). The average age in 2001 is 20.2 years (table B.3).

This pathway is dominated by individuals spending considerable time in work only: 9 per cent of individuals spend all 120 months in work only and the other 91 per cent spend at least one quarter of their time (29 months) in work only. For this latter group, work periods are interspersed with other activities — this is sometimes referred to as 'churning' or turbulence in activity patterns (Bretherton 2011; Hunter, Gray and Jones 2000; Lamb 2001). For example, 77 per cent of individuals have between two and ten spells of work only over the ten-year period (not shown).

Initially, 45 per cent of individuals are in work only, 24 per cent in work and study, and 21 per cent in study only (figure B.2). Unemployment affects 4 per cent of individuals and 6 per cent are NILF. By June 2005, there has been a large shift in activities from study to work only — 83 per cent are in work only, 9 per cent combine work and study and only 2 per cent are in study only. Unemployment and NILF have become less prevalent, affecting about 6 per cent of individuals. This distribution of activities persists for the remainder of the sample period.

Individuals who are initially neither working nor looking for work are primarily engaged in study (table B.3). However, as time progresses and studies are completed, childcare becomes the main activity of those NILF.

In this pathway, education levels in 2001 are higher than in any other youth pathway: 52 per cent of individuals have already attained a medium level of education (Year 12/Certificate III/IV up to Advanced Diploma) and 13 per cent have a high level of education (Bachelor Degree or above). Individuals with these higher education levels tend to be working (with or without study). About 35 per cent of youths in this pathway are studying (with or without work) in 2001, and almost two thirds of them (64 per cent) are combining study with work (not shown). Individuals combining work and study in 2001 tend to have already attained higher levels of education than those in study only — 30 per cent in study only have medium or high education levels but 58 per cent in work and study already have medium or high education levels (not shown).

With 77 per cent of individuals undertaking study (with and/or without work) during the ten-year period (not shown), education levels rise: by 2010, 64 per cent have attained medium levels of education and 20 per cent have high levels (table B.3). This increase results mostly from individuals with Year 11 or lower completing Year 12 and/or qualifications at Certificate III/IV level. The increase in broad education levels is smaller than that observed for youths in the *Education to Work* pathway, partly because the incidence, repetition and to some extent duration of study activity are lower in this pathway. In addition, more detailed investigation of educational attainment showed that only about half of those individuals in this pathway who studied during the ten-year period increased the level of their highest qualification. As a result, in 2010 individuals in this pathway have relatively low levels of education compared to the other youth pathways (table B.3).

Over the ten years, individuals undertake an average of about three out of the five activities. Two of the activities are likely to involve work, since everyone has at least one spell (and an average of 2.9 spells) of work lasting an average of 32 months (2 years, 8 months) and 71 per cent have an average of two spells (with average duration of 9.4 months) of work and study combined (table B.4). Multiple

short spells of work (with or without study) could indicate churning or turbulence in employment, particularly if interspersed with spells of unemployment or NILF. However, periods in which there are spells of work only interspersed with work and study do not necessarily indicate interruptions to employment, although they could include job changes. Similarly, a long spell of work only (or a long spell of work and study) does not rule out job changes during that spell.

Young adults in the NILF to Work pathway

The *NILF to Work* pathway represents a relatively small proportion of young adults (about 12 per cent) (table B.8). The average age in 2001 is 32.3 years (table B.7). Women predominate (81 per cent) and the pathway is characterised by a progression from NILF to work (figure B.4). This transition is primarily characterised by women re-entering the workforce after taking time out to rear children. There is also some study and somewhat higher unemployment in 2001 than in other parts of the young adult segment. By 2010, the percentage of those in unemployment and NILF is representative of the young adult segment — about 15 per cent (table B.7).

In 2001, 61 per cent of individuals in this pathway are NILF (this general characteristic gave rise to the label of *NILF to Work*). This is mostly attributed to having children (80 per cent) (table B.7). However, a separate analysis revealed that 20 per cent of women in this pathway who have children aged less than 15 years in 2001 are employed (not shown). For women with children, the average age of the youngest child in 2001 is almost 3 years for women who are NILF but slightly older (3.6 years) for women who are employed. In 2001, the children of unemployed women have an average age of 5.1 years. Among the women with children in this pathway, about 82 per cent were either legally married or in a de facto relationship in 2001 irrespective of whether they were employed or NILF (not shown).

In 2001, 82 per cent of the women in this pathway have at least one child aged less than 15 years, and 63 per cent have at least one child aged less than 5 years. At the same time, among the mothers in this pathway, 75 per cent are NILF and 20 per cent are employed. By 2010, many of those children are no longer aged less than 15 (or, in some cases, no longer reside in the household), although births add to the number of children aged less than 10 years. As a result of these developments, in 2010, 77 per cent of women in this pathway have at least one child aged less than 15 years but only 7.5 per cent have one or more children aged less than 5 years. As their children age, many women re-enter the labour force. By 2010, 86 per cent of the women with children aged less than 15 years are employed and only 11 per cent are NILF (not shown).

By 2010, 91 per cent of individuals in this pathway have averaged 2.08 spells of NILF lasting 17 months per spell (table B.8). The NILF rate has reduced to 11 per cent (figure B.4) and study, disability, caring for others with illness or disability, and retirement have become the main reasons for NILF. As mentioned, by 2010 the proportion of young adults in work (85.3 per cent) is virtually identical to that of the entire young adult segment (85.1 per cent) (table B.7).

Although 37 per cent of individuals begin in work only (figure B.4), by 2010 nearly all have had spells of work only, averaging 2.95 spells of 22 months per spell (table B.8). In 2010, 85 per cent are employed (mostly without study) (figure B.4). In this pathway, 47 per cent of individuals were women who were NILF in 2001 and employed in 2010. These women tended to work in fairly low-skilled occupations, with 39 per cent working as carers/aides, sales assistants or cleaners/laundry workers (not shown).

Unemployment is relatively common among the individuals in this pathway. In 2001, this pathway has the highest unemployment of young adults (9.3 per cent) (table B.7). Over the next ten years, 61 per cent of individuals in this pathway average 2.4 spells of unemployment lasting seven months per spell (table B.8).

In 2001, education levels in this pathway are lower than those in the other young adult pathways. Of the young adults in this pathway 40 per cent have low levels of education, 43 per cent have medium levels of education, and 17 per cent have high levels of education (table B.7). About 60 per cent of individuals undertake some study (with or without work) during the ten-year period (not shown);[15] 44 per cent engage in an average of 1.8 spells of study only, lasting about 12 months; 43 per cent combine work with study, averaging 1.6 spells of 7.3 months duration (table B.8). The study activity leads to a relatively small increase in education levels (table B.7). This modest improvement results from 21 per cent of the individuals who had studied acquiring higher broad education levels, 18 per cent acquiring higher qualifications within the same broad education level and 61 per cent completing qualifications at the same or lower levels than they had in 2001 or undertaking but not completing higher qualifications by 2010.

6 Some concluding comments

Whether sourced from HILDA or elsewhere, calendar-style data are particularly valuable for understanding individuals' transitions in the context of their

[15] About 27 per cent of young adults in this pathway study with and without work.

longer-term experiences. OMCA techniques exploit the richness of such data to reveal overall similarities and differences in detailed patterns of activities.

Using HILDA's calendar, this paper has identified and described a set of transition pathways that characterise the activity patterns of individuals over ten years. However, there remains considerable scope for additional analysis of those pathways.

Policy makers are concerned about the long-term welfare of individuals and the problems associated with people who experience difficulties in making the transition to more stable employment, for example. In the area of transitions, there may be a priority placed on helping people avoid 'unsuccessful' outcomes — such as long-term unemployment or marginal labour force attachment among young people.

Thus, a key priority would be to develop a framework for distinguishing 'successful' and 'failing' outcomes. Multivariate analysis could then be used to identify which pathways and types of activity patterns are associated with 'success' or 'failure'. For example, using wave 11 of HILDA to represent 'outcomes', and then to examine econometrically the association between pathways and outcomes.

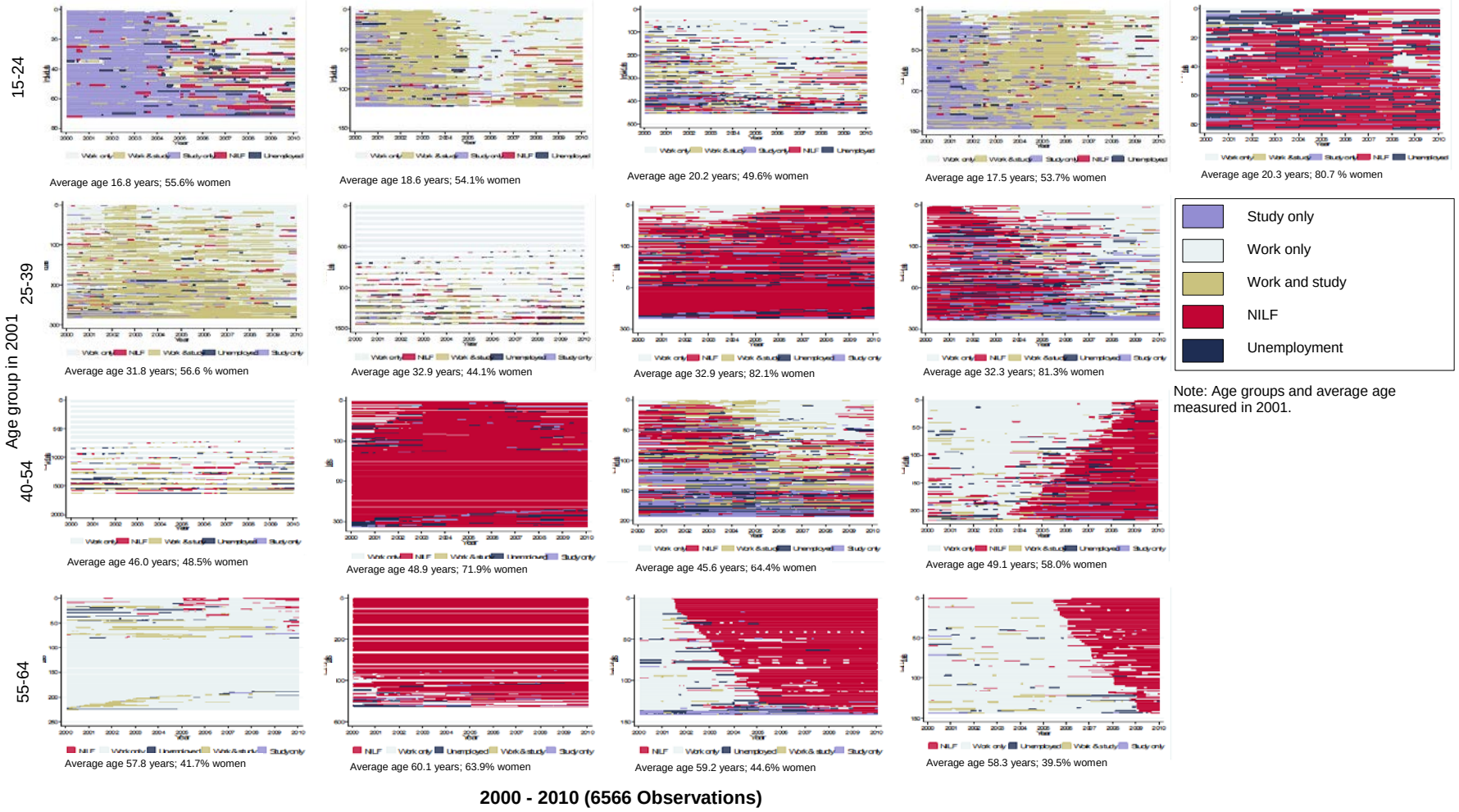
Another extension could be to identify the characteristics of individuals that are associated with (or perhaps determine) the pathway they follow. In particular, the interaction between health, disability and labour market outcomes could be considered. Such an analysis could inform the targeting of strategies to reduce the risk of individuals following pathways to 'failure'.

To summarise, HILDA's calendar — which is surprisingly under-exploited — gives a rich picture of individuals' activities over time. The calendar's length and detail makes it an invaluable information source that should be preserved and utilised.[16]

[16] To encourage others to use HILDA's calendar — which can appear overwhelming at first glance — the Stata code used to extract and manipulate the variables into continuous, individual-level sequences over ten years is available on the Productivity Commission's website, www.pc.gov.au. Instructions for implementing OMCA are included.

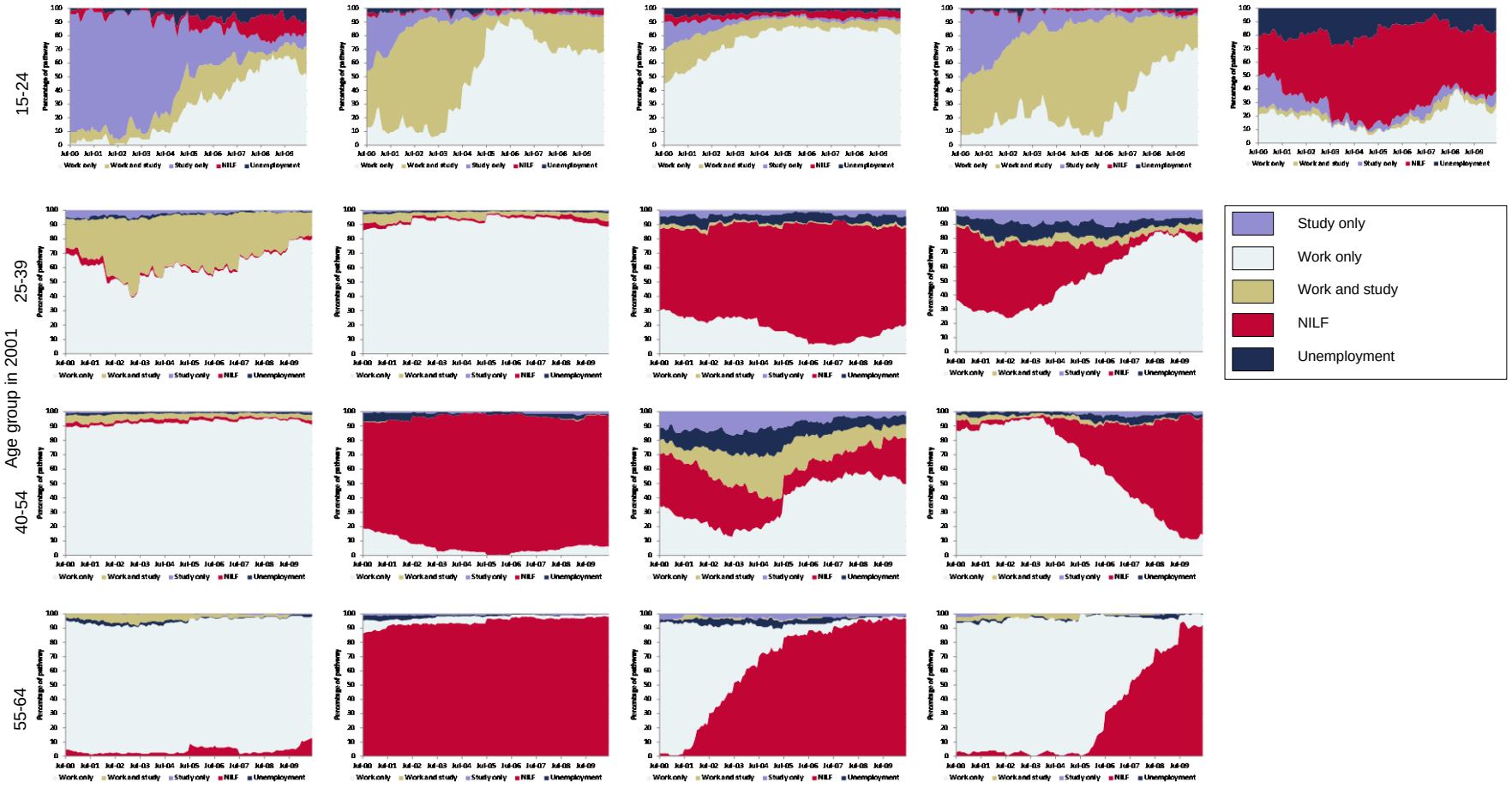
Appendix A: Australian Atlas of Education and Labour Market Transitions

Figure A.1 **Australian atlas of education and labour market transition pathways — individual sequences**



Source: Authors' estimates based on HILDA waves 1–10.

Figure A.2 **Australian atlas of education and labour market transition pathways — activity prevalence**



2000 - 2010 (6566 Observations)

Source: Authors' estimates based on HILDA waves 1–10.

Appendix B: Selected pathways

Youths in the *Work and Study to Work* pathway

In the *Work and Study to Work* pathway, most individuals are making the transition from education to work (figure B.1 below). In 2001, the average age is 18.6 years (table B.1 below), with most individuals aged 15–21 (not shown). Initial education levels are low to medium, with the majority of individuals having attained Year 11 or Year 12. Those with high education levels represent 8 per cent of the group.

Most youths in this pathway start off in study only (particularly the under 18s, who are still at school) or combine work and study (predominantly those aged 18 or over who have left school having completed Year 12). For these older individuals, there are some whose area of study is related to their current occupation (such as those with apprenticeships). This is not the case for most, who work in hospitality, retail or labouring while studying for a degree.

The remainder — typically aged 18 or over — are not studying and tend to be in work. This might be the first job in their chosen vocation after leaving the education system with mid-level qualifications (Year 12), or could represent employment while taking a 'gap year' from study. Very few individuals who are not studying are unemployed or NILF.

By 2003, the average age is 20.6 years and most individuals are aged 18 or over. The overwhelming majority are combining work and study (figure B.1). Their work may or may not be in their chosen field.

In 2005, the group is now in their early- to mid-20s and the balance of activities shifts from education to work as post-school qualifications are likely to be completed. For those who have finished their initial study and are in work only, their jobs are more likely to relate to their acquired qualifications.

From 2007, about one quarter of individuals who have been in work only for one or two years re-enter the education system and combine work and study. This could represent higher level qualifications (beyond entry level for the chosen vocation) undertaken for career advancement.

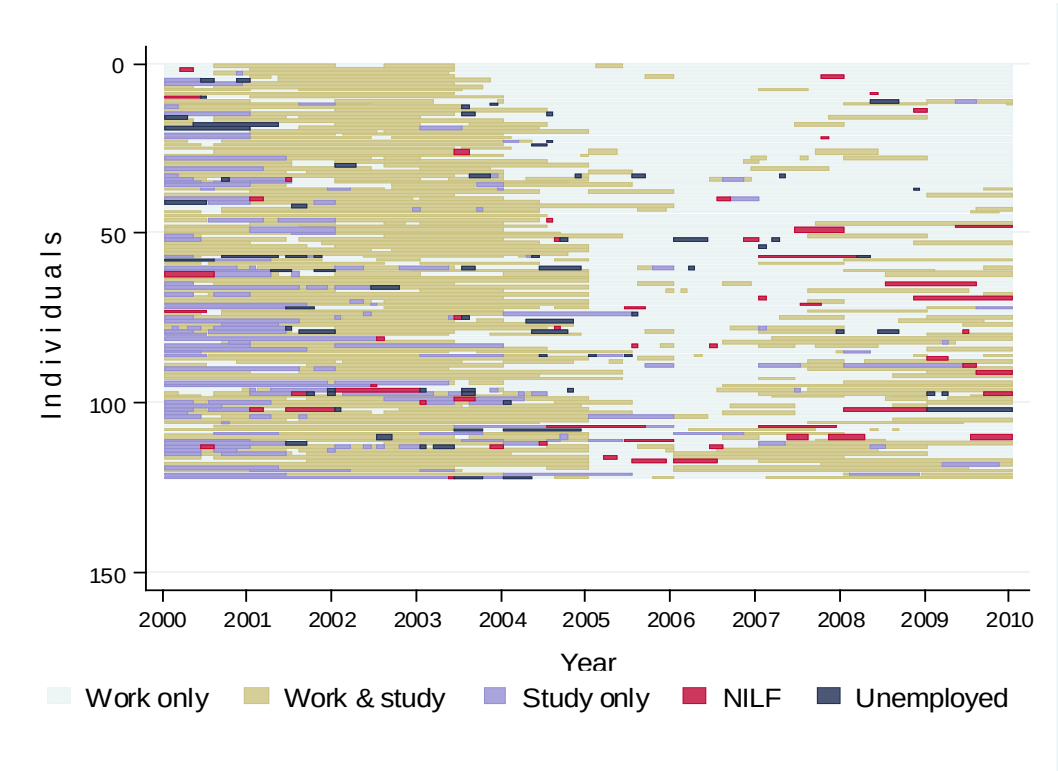
By 2010, when the group is aged between 24 and 33 (28.6 years on average), most individuals are in work (with or without study) and education levels are much higher than in 2001. Over half of the group having attained at least a Bachelor Degree (table B.1). Those who are not working tend to be women taking time out from the labour force to raise children.

Over the ten-year period, education levels have risen because everyone in this pathway has spent some time studying. They also spent some time combining work and study, averaging about three spells of 15.89 months per spell (table B.2 below). Over 60 per cent also spent time in study only averaging two spells of seven months each. Everyone spent some time in work only, averaging about three spells of 20 months each. Any experience of unemployment or NILF tended to be of short duration —three to four months on average — although some individuals had repeated spells. By 2010, the share of individuals in this pathway who are unemployed is very low (1.64 per cent compared to the 4.45 per cent for all youth pathways) (table B.1).

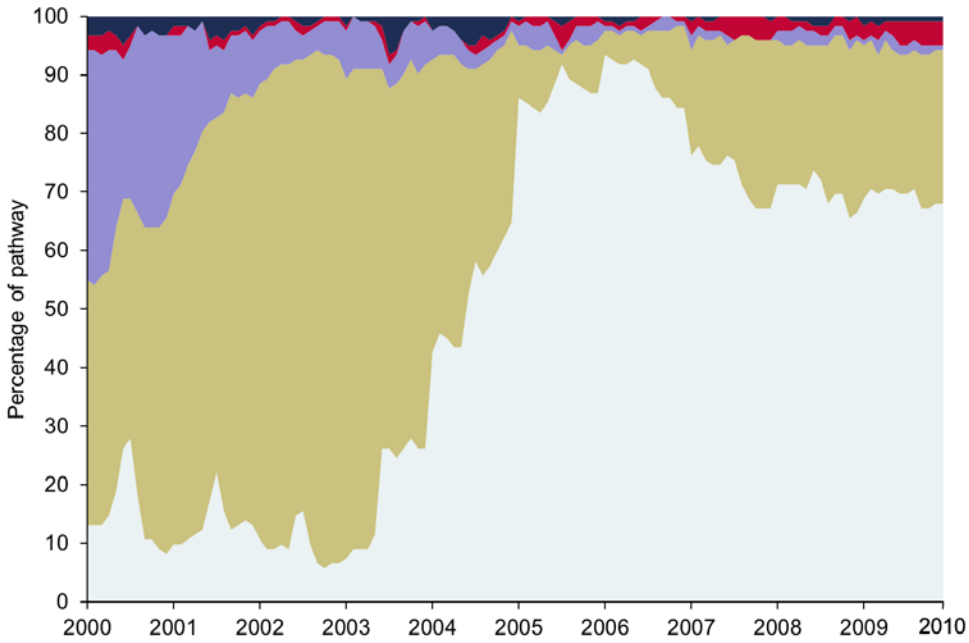
Individuals in this pathway gravitate to the major cities (66.4 per cent in 2001 and 72.1 per cent in 2010). This is likely to be associated with them studying at universities and then remaining in the major cities.

Figure B.1 **Activities in the *Work and Study to Work* pathway for youths**

(A) Activity sequences by individual



(B) Activities by monthly share



Data source: Authors' estimates based on HILDA waves 1–10.

Table B.1 **Selected characteristics of youths in the *Work and Study to Work* pathway**

Characteristic	*Measure*	*WS–W pathway*	*All youth pathways*
In 2001:			
age	years, average	18.6	19.2
gender	% female	54.10	54.39
locality (remoteness)	% major city	66.39	61.35
highest level of education[a]	% high	8.20	9.23
	% medium	53.28	43.90
	% low	38.52	46.86
unemployed	% U	6.56	9.81
NILF (incl. marginally attached)	% NILF	22.95	27.37
of which:			
home duties/childcare	% of NILF	3.57	16.67
study	% of NILF	75.00	64.58
marginally attached to labour force	% of NILF	17.86	11.67
other reasons[b]	% of NILF	3.57	7.08
In 2010:			
locality (remoteness)	% major city	72.13	62.03
highest level of education[a]	% high	50.83	30.55
	% medium	46.72	55.41
	% low	2.46	14.02
unemployed	% U	1.64	4.45
NILF (incl. marginally attached)	% NILF	6.56	12.20
of which:			
home duties/childcare	% of NILF	75.00	66.36
study	% of NILF	25.00	13.08
marginally attached to labour force	% of NILF	0.00	5.61
other reasons[b]	% of NILF	0.00	14.95

[a] Educational attainment has been classified as high (Bachelor's degree or above), medium (Year 12, Certificate III/IV, Diploma or Advanced Diploma) and low (Certificate I/II, Certificate not defined, Year 11 or lower). [b] May include caring for others with illness/disability, retirement/voluntarily inactive, own illness/disability (rendering the individual temporarily or permanently unable to work), travel/holidays/leisure activities, working in an unpaid voluntary job, or other (unspecified) activities.

Source: Authors' estimates based on HILDA waves 1–10.

Table B.2 **Selected characteristics of activity patterns in the *Work and Study to Work* pathway (youths)[a]**

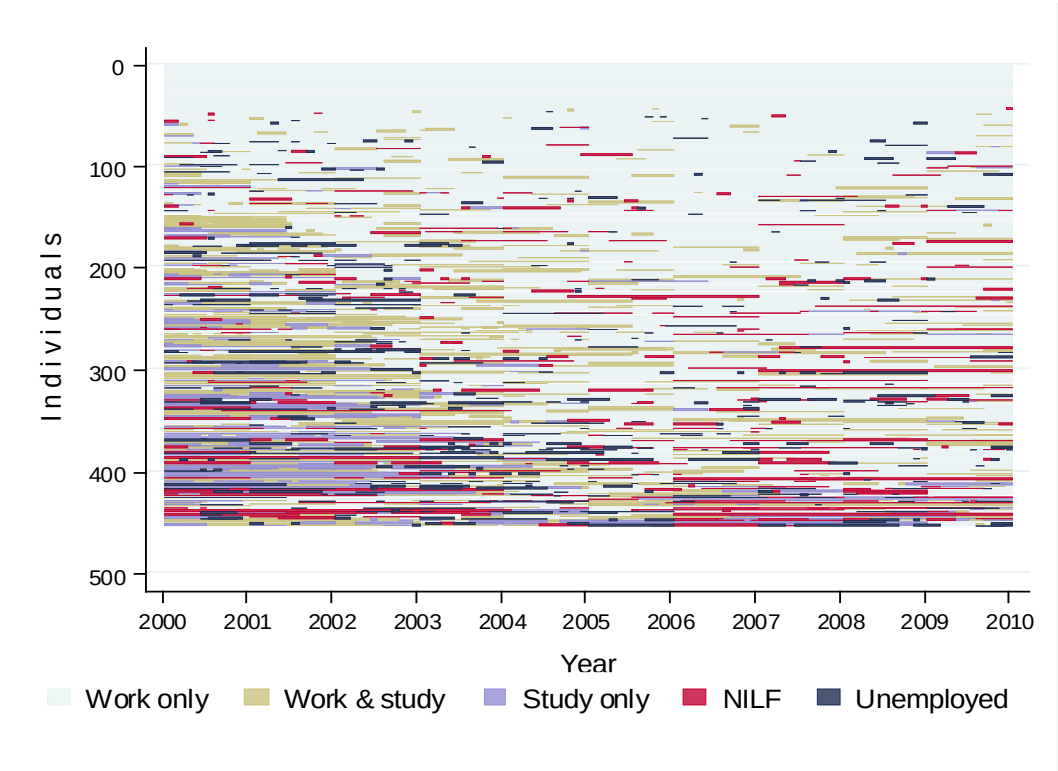
	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	49.84	39.83	7.40	1.56	1.37
Share of path with at least one spell of the activity	100.00	100.00	62.30	34.43	32.79
Conditional on at least one spell of the activity:					
average number of spells	2.98	3.01	2.00	1.31	1.75
average length of spell (months)	20.10	15.89	7.13	4.15	2.87

[a] Pathway size 122 (13.9 per cent of youths); average number of activities 3.30.

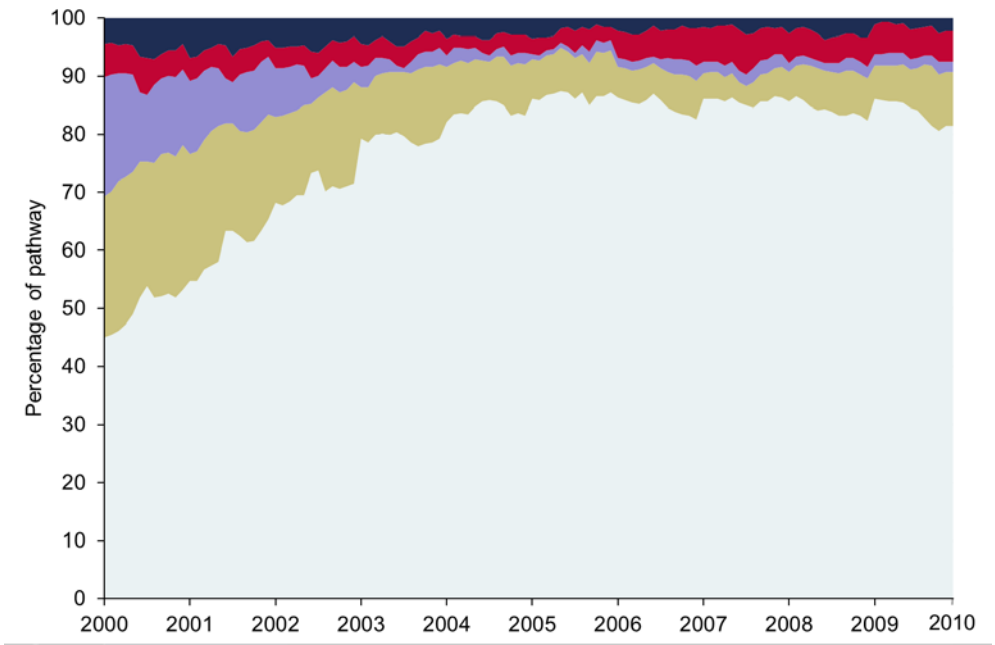
Source: Authors' estimates based on HILDA waves 1–10.

Figure B.2 **Activities in the *Churning with Work* pathway for youths**

(A) Activity sequences by individual



(B) Activities by monthly share



Data source: Authors' estimates based on HILDA waves 1–10.

Table B.3 **Selected characteristics of youths in the *Churning with Work* pathway**

Characteristic	*Measure*	*CW pathway*	*All youth pathways*
In 2001:			
age	years, average	20.2	19.2
gender	% female	49.67	54.39
locality (remoteness)	% major city	58.94	61.35
highest level of education[a]	% high	12.58	9.23
	% medium	51.88	43.90
	% low	35.54	46.86
unemployed	% U	8.39	9.81
NILF (incl. marginally attached)	% NILF	13.69	27.37
of which:			
home duties/childcare	% of NILF	16.13	16.67
study	% of NILF	56.45	64.58
marginally attached to labour force	% of NILF	14.52	11.67
other reasons[b]	% of NILF	12.90	7.08
In 2010:			
locality (remoteness)	% major city	57.17	62.03
highest level of education[a]	% high	19.65	30.55
	% medium	63.80	55.41
	% low	16.55	14.02
unemployed	% U	3.09	4.45
NILF (incl. marginally attached)	% NILF	8.39	12.20
of which:			
home duties/childcare	% of NILF	65.79	66.36
study	% of NILF	13.16	13.08
marginally attached to labour force	% of NILF	10.53	5.61
other reasons[b]	% of NILF	10.53	14.95

[a] Educational attainment has been classified as high (Bachelor's degree or above), medium (Year 12, Certificate III/IV, Diploma or Advanced Diploma) and low (Certificate I/II, Certificate not defined, Year 11 or lower). [b] May include caring for others with illness/disability, retirement/voluntarily inactive, own illness/disability (rendering the individual temporarily or permanently unable to work), travel/holidays/leisure activities, working in an unpaid voluntary job, or other (unspecified) activities.

Source: Authors' estimates based on HILDA waves 1–10.

Table B.4 **Selected characteristics of activity patterns in the *Churning with Work* pathway (youths)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	76.93	11.15	4.46	4.24	3.23
Share of path with at least one spell of the activity	100.00	71.08	38.85	40.84	43.71
Conditional on at least one spell of the activity:					
average number of spells	2.93	2.01	1.65	1.78	2.31
average length of spell (months)	31.54	9.37	8.36	6.98	3.84

[a] Pathway size 453 (51.7 per cent of youths); average number of activities 2.94.

Source: Authors' estimates based on HILDA waves 1–10.

Young adults in the *Work* pathway

This is the largest pathway for young adults, containing 63 per cent of individuals in the age segment (table B.6 below). In 2001 individuals in this pathway are aged 25–39 years with an average age of 32.9 years (table B.5 below).[17] Although women represent 55 per cent of the young adult segment, in this pathway individuals are more likely to be men (44 per cent are women).

The *Work* pathway mostly comprises individuals who have made the transition from education to work.[18] It is what might be termed a 'traditional' study and work history for individuals in their prime working age. Employment is clearly the main focus, although there are also some experiences of unemployment or withdrawal from the workforce (figure B.3 below). Individuals returning to education typically combine study with work, as they are more likely to have financial and family responsibilities. For example, in 2001, 58 per cent of individuals in this pathway already had parenting responsibilities for children aged less than 18 years. Among those with parenting responsibilities, over 90 per cent were employed and almost half (45 per cent) of those employed were women (not shown).

In 2001, education levels in this pathway reflect the average for all young adults (table B.5). Education levels are higher in this pathway than in the *Prolonged NILF*

[17] In this pathway, 42 per cent of individuals are aged 35 or over, compared to 34 per cent for the other young adult pathways in 2001 (not tabulated).

[18] Individuals in this pathway can be considered as having made the transition to work. From age 25, the only individuals who are yet to complete the transition from education to work are those who have not yet left the education system. In 2001, they are likely to be at the younger end of the age segment (25–27 years) and studying at graduate or postgraduate level, having already completed a Bachelor Degree. Only 1 per cent of young adults in the *Work* pathway meet these criteria.

pathway but lower than in the *Work with or without Study* pathway. In 2001, 24 per cent of these young adults have a low educational level, 47 per cent have a medium level of education, and 29 per cent have high levels of education. Compared to the education levels in the youth segment, the starting levels for young adults in this pathway (aged 25–39 years, with an average of 32.9 years) are similar to the final levels of education for the youth segment (aged 24–33 years with an average age of 29.2 years in 2010) (see table B.3).

Work is the dominant activity in this pathway. Initially, 86 per cent of individuals are in work only and a further 6 per cent combine work and study and there is not much change in the prevalence of work in ten years (figure B.3).

For many individuals there is also continuity of 'being in employment', even though there might be changes in jobs and employers. About one third of individuals in this pathway are in work only throughout the ten years. For the remaining two thirds, work only is the dominant activity, followed by work and study (not shown).

Individuals in this pathway average about two spells of work only, lasting 54 months (4 years, 6 months) per spell. Forty per cent of individuals also have an average of 1.63 spells of work and study lasting 7.4 months per spell (table B.6).

Broad education levels in the *Work* pathway do not change much over the ten years despite 41 per cent of individuals studying at some point, mostly while working. This is because 75 per cent of individuals in this pathway who studied completed a qualification at the same or lower level than their highest qualification, or began a higher level qualification that was either still being undertaken in 2010 or was abandoned earlier. Of the 25 per cent of 'students' who completed a higher level qualification, less than half increased the broad education profile for the pathway (not shown).

NILF and unemployment are far less common experiences among individuals in this pathway compared to the other young adult pathways. In any given month, less than 2 per cent of individuals are unemployed and up to 5.3 per cent are NILF (figure B.3).[19] Over the ten-year period, 29 per cent of individuals in this pathway are NILF and 23 per cent are unemployed. These individuals have about 1.6 spells of NILF and/or unemployment lasting an average of six months per NILF spell and three months per unemployment spell (table B.6). This is much less than the two pathways characterised by significant NILF activity, where most young adults

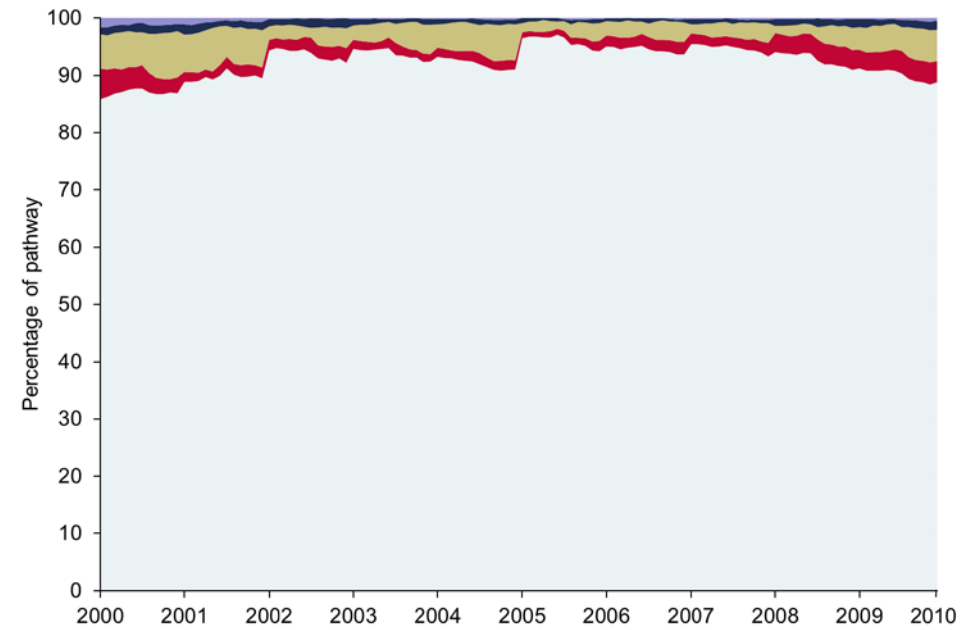
[19] In contrast, in the early years in the *NILF to Work* pathway, up to 14 per cent of young adults are unemployed and more than 50 per cent are NILF (figure B.4). In the *Prolonged NILF* pathway, 4–8 per cent of young adults are unemployed and 60–90 per cent are NILF (not shown).

spend significant time in NILF and unemployment, and averaging two spells of each.

Figure B.3 **Activities in the *Work* pathway for young adults**

(A) Activity sequences by individual

(B) Activities by monthly share



Data source: Authors' estimates based on HILDA waves 1–10.

Table B.5 **Selected characteristics of young adults in the *Work* pathway**

Characteristic	*Measure*	*W pathway*	*All young adult pathways*
In 2001:			
age	years, average	32.9	32.7
gender	% female	44.15	54.78
locality (remoteness)	% major city	65.36	65.31
highest level of education[a]	% high	28.78	27.27
	% medium	47.38	46.30
	% low	23.83	26.42
unemployed	% U	1.79	3.23
NILF (incl. marginally attached)	% NILF	2.69	18.26
of which:			
home duties/childcare	% of NILF	69.23	75.60
study	% of NILF	15.38	8.13
marginally attached to labour force	% of NILF	7.69	2.87
other reasons[b]	% of NILF	7.69	13.40
In 2010:			
locality (remoteness)	% major city	62.40	62.12
highest level of education[a]	% high	30.17	30.62
	% medium	49.37	47.40
	% low	20.45	21.98
unemployed	% U	1.65	2.40
NILF (incl. marginally attached)	% NILF	4.27	12.54
of which:			
home duties/childcare	% of NILF	61.29	61.67
study	% of NILF	4.84	5.92
marginally attached to labour force	% of NILF	8.06	2.79
other reasons[b]	% of NILF	25.81	29.62

[a] Educational attainment has been classified as high (Bachelor's degree or above), medium (Year 12, Certificate III/IV, Diploma or Advanced Diploma) and low (Certificate I/II, Certificate not defined, Year 11 or lower). [b] May include caring for others with illness/disability, retirement/voluntarily inactive, own illness/disability (rendering the individual temporarily or permanently unable to work), travel/holidays/leisure activities, working in an unpaid voluntary job, or other (unspecified) activities.

Source: Authors' estimates based on HILDA waves 1–10.

Table B.6 **Selected characteristics of activity patterns in the *Work* pathway (young adults)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	92.42	3.98	0.34	2.27	0.99
Share of path with at least one spell of the activity	100.00	39.94	6.68	28.93	22.73
Conditional on at least one spell of the activity:					
average number of spells	2.06	1.63	1.31	1.59	1.55
average length of spell (months)	53.79	7.35	4.65	5.93	3.36

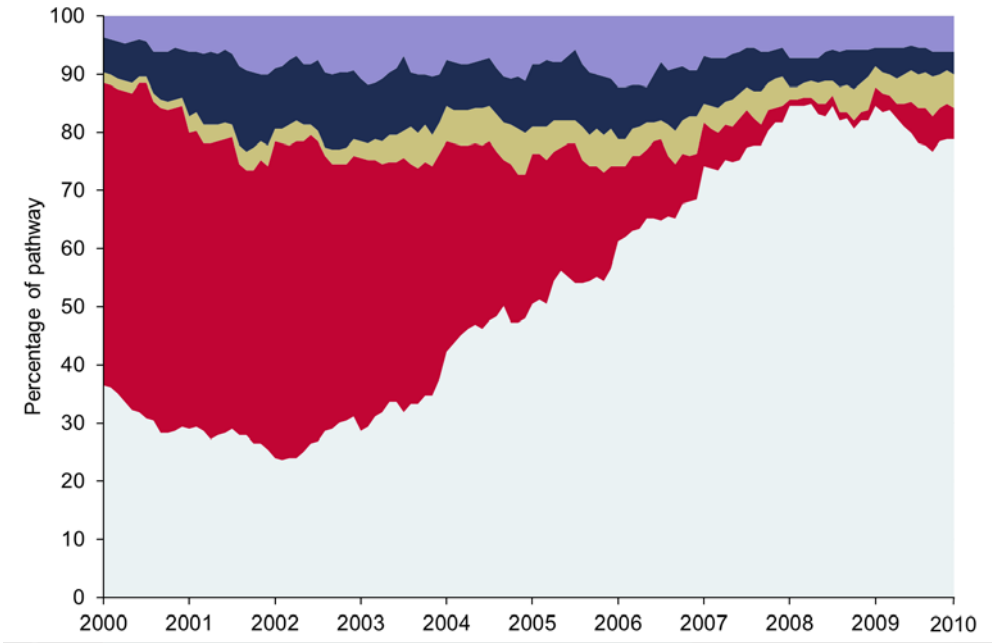
[a] Pathway size 1452 (63.4 per cent of young adults); average number of activities 1.98.

Source: Authors' estimates based on HILDA waves 1–10.

Figure B.4 **Activities in the *NILF to Work* pathway for young adults**

(A) Activity sequences by individual

(B) Activities by monthly share



Data source: Authors' estimates based on HILDA waves 1–10.

Table B.7 **Selected characteristics of young adults in the *NILF to Work* pathway**

Characteristic	*Measure*	*N–W pathway*	*All young adult pathways*
In 2001:			
age	years, average	32.3	32.7
gender	% female	81.36	54.78
locality (remoteness)	% major city	65.59	65.31
highest level of education[a]	% high	16.85	27.27
	% medium	43.36	46.30
	% low	39.79	26.42
unemployed	% U	9.32	3.23
NILF (incl. marginally attached)	% NILF	60.57	18.26
of which:			
home duties/childcare	% of NILF	80.47	75.60
study	% of NILF	7.69	8.13
marginally attached to labour force	% of NILF	1.78	2.87
other reasons[b]	% of NILF	10.05	13.40
In 2010:			
locality (remoteness)	% major city	63.44	62.12
highest level of education[a]	% high	21.87	30.62
	% medium	46.96	47.40
	% low	31.18	21.98
unemployed	% U	3.58	2.40
NILF (incl. marginally attached)	% NILF	11.11	12.54
of which:			
home duties/childcare	% of NILF	38.71	61.67
study	% of NILF	32.26	5.92
marginally attached to labour force	% of NILF	3.23	2.79
other reasons[b]	% of NILF	25.81	29.62

[a] Educational attainment has been classified as high (Bachelor's degree or above), medium (Year 12, Certificate III/IV, Diploma or Advanced Diploma) and low (Certificate I/II, Certificate not defined, Year 11 or lower). [b] May include caring for others with illness/disability, retirement/voluntarily inactive, own illness/disability (rendering the individual temporarily or permanently unable to work), travel/holidays/leisure activities, working in an unpaid voluntary job, or other (unspecified) activities.

Source: Authors' estimates based on HILDA waves 1–10.

Table B.8 **Selected characteristics of activity patterns in the *NILF to Work* pathway (young adults)[a]**

	Work	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	52.56	4.20	7.70	27.05	8.49
Share of path with at least one spell of the activity	98.92	43.37	44.44	91.40	61.29
Conditional on at least one spell of the activity:					
average number of spells	2.95	1.60	1.76	2.08	2.37
average length of spell (months)	21.59	7.28	11.83	17.08	7.00

[a] Pathway size 279 (12.2 per cent of young adults); average number of activities 3.39.

Source: Authors' estimates based on HILDA waves 1–10.

Mature adults in the *Work to NILF* pathway

This pathway represents about 9 per cent of mature adults (aged 40–54 years in 2001) (table B.10 below). The *Work to NILF* pathway is the 'oldest' pathway for mature adults, with an average age of 49.1 years in 2001 (table B.9 below). Individuals in this pathway are in the latter stages of their working lives, and many progress from work to NILF. Some individuals spend small periods of time in other activities (figure B.5 below).

In 2001, over 90 per cent of individuals are employed (figure B.5). Unemployment and NILF are low (4.6 and 4.2 per cent, respectively). In the latter case, the percentage in NILF is much lower than the average for all mature adults (17.7 per cent). Education levels are lower compared to most other mature adult pathways — 44 per cent of individuals have low levels of education (table B.9).

All individuals in this pathway spend some time in work only, averaging 2.1 spells of 35.8 months (about 3 years) in length (table B.10). These spells are more likely to occur in the earlier part of the observation period.

From 2004, when the average age is about 52 years and the maximum is 57 years, the share of individuals in NILF begins to increase sharply (figure B.5). This is partially explained by ageing, although it is unclear why NILF increases so sharply at a time when the economy was growing so strongly. (For example, there were no major policy changes around this time.) By 2010, 83 per cent are NILF and about half of these individuals are more than 60 years old. Consistent with the age profile of individuals in this pathway, most of the increasing NILF activity is due to retirement from the workforce (60 per cent). Other reasons for leaving the labour force include illness, injury, disability or care obligations (22 per cent) (not shown).

Nearly everyone in this pathway spends some time in NILF, averaging 1.74 spells of 21.4 months per spell.[20]

Accompanying this story of the transition to retirement is one of geographic movement. In 2001, 59 per cent of individuals reside in major cities, although 15 per cent of them subsequently move to regional or remote areas (most likely as they retire). This reduces the proportion in this pathway living in major cities to 50 per cent by 2010 (table B.9).

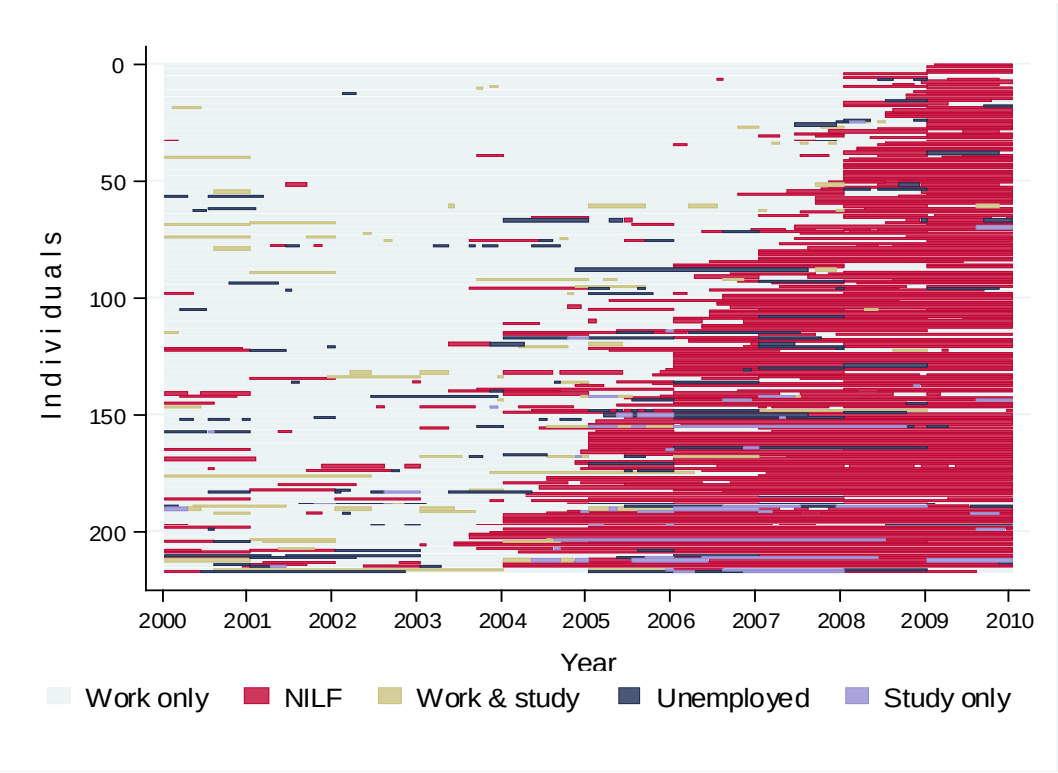
Time spent in other activities are infrequent and tend to last about six months on average. The limited incidence and short duration of study means that education levels remain quite low. Most higher level qualifications require significantly more than six months to complete and are unlikely to be undertaken by individuals in a later stage of their life cycle.[21]

[20] Individual sequences only cover a ten-year period. As a result, the first spell has an unknown start date (left censored) and the last spell has an unknown end date (right censored). Since the analysis does not adjust for censoring, the length of these spells will be underestimated (Fry and Boulton 2013).

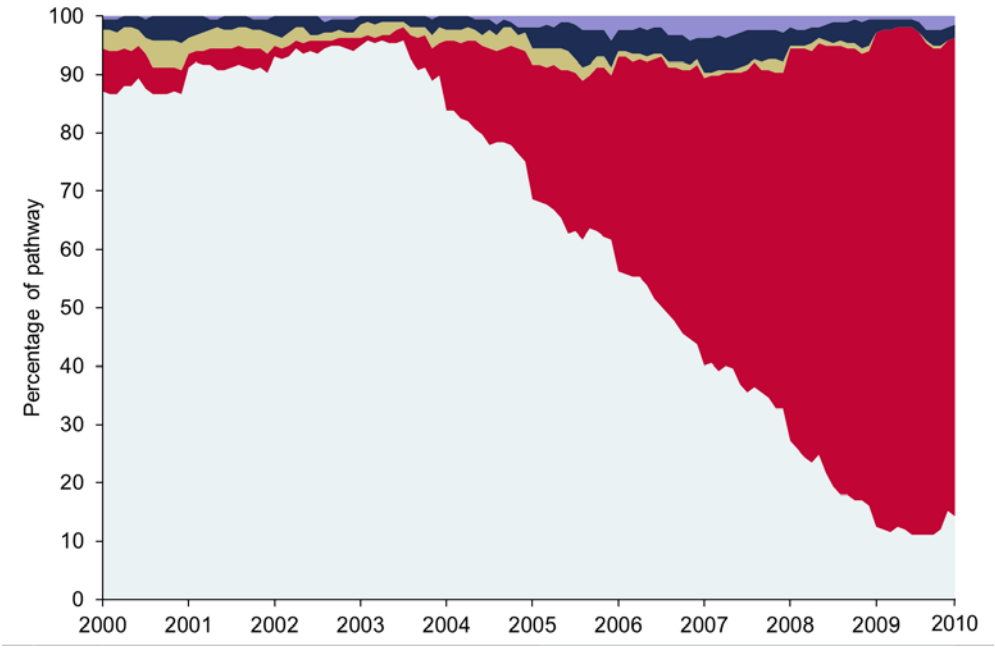
[21] However, of the 65 individuals in this pathway who studied, 6 spent two to four years continuously in study (with or without work) and acquired higher qualifications.

Figure B.5 **Activities in the *Work to NILF* pathway for mature adults**

(A) Activity sequences by individual



(B) Activities by monthly share



Data source: Authors' estimates based on HILDA waves 1–10.

Table B.9 **Selected characteristics of mature adults in the *Work to NILF* pathway**

Characteristic	*Measure*	*W–N pathway*	*All mature adult pathways*
In 2001:			
age	years, average	49.1	46.7
gender	% female	58.06	53.82
locality (remoteness)	% major city	58.53	60.62
highest level of education[a]	% high	21.66	24.39
	% medium	34.1	40.06
	% low	44.23	35.56
unemployed	% U	4.61	3.53
NILF (incl. marginally attached)	% NILF	4.15	17.67
of which:			
home duties/childcare	% of NILF	22.22	43.51
study	% of NILF	11.11	4.33
marginally attached to labour force	% of NILF	11.11	3.37
other reasons[b]	% of NILF	55.55	48.80
In 2010:			
locality (remoteness)	% major city	50.23	58.33
highest level of education[a]	% high	21.66	25.53
	% medium	37.78	42.78
	% low	40.56	31.68
unemployed	% U	2.30	1.91
NILF (incl. marginally attached)	% NILF	82.95	26.89
of which:			
home duties/childcare	% of NILF	11.67	17.69
study	% of NILF	0.56	1.11
marginally attached to labour force	% of NILF	1.11	1.42
other reasons[b]	% of NILF	86.67	79.77

[a] Educational attainment has been classified as high (Bachelor's degree or above), medium (Year 12, Certificate III/IV, Diploma or Advanced Diploma) and low (Certificate I/II, Certificate not defined, Year 11 or lower). [b] May include caring for others with illness/disability, retirement/voluntarily inactive, own illness/disability (rendering the individual temporarily or permanently unable to work), travel/holidays/leisure activities, working in an unpaid voluntary job, or other (unspecified) activities.

Source: Authors' estimates based on HILDA waves 1–10.

Table B.10 **Selected characteristics of activity patterns in the *Work to NILF* pathway (mature adults)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	63.12	1.87	1.13	30.77	3.11
Share of path with at least one spell of the activity	100.00	23.04	13.82	99.08	31.80
Conditional on at least one spell of the activity:					
average number of spells	2.12	1.58	1.60	1.74	1.88
average length of spell (months)	35.81	6.15	6.13	21.43	6.23

[a] Pathway size 217 (9.2 per cent of mature adults); average number of activities 2.68.

Source: Authors' estimates based on HILDA waves 1–10.

References

Albert Verdú, C. and Davia, M. 2010, 'Education and labour market transitions amongst compulsory education graduates and school dropouts', *Regional and Sectoral Economic Studies*, vol. 10, no. 3, pp. 5–22.

Anyadike-Danes, M. and McVicar, D. 2003, *Parallel Lives: Birth, Childhood and Adolescent Influences on Career Paths*, Northern Ireland Economic Research Centre Paper, Belfast, Northern Ireland.

—— and —— 2005, 'You'll never walk alone: childhood influences and male career path clusters', *Labour Economics*, vol. 12, pp. 511–30.

—— and —— 2010, 'My brilliant career: characterizing the early labour market trajectories of British women from Generation X', *Sociological Methods & Research*, vol. 38, no. 3, pp. 482–512.

Bretherton, T. 2011, *Understanding the Undertow: Innovative Responses to Labour Market Disadvantage and VET*, National Vocational Education and Training Research and Evaluation (NVETRE) Program Report, NCVER, Adelaide.

Brzinsky-Fay, C. 2007, 'Lost in transition? Labour market sequences of school-leavers in Europe', *European Sociological Review*, vol. 23, no. 4, pp. 409–22.

—— and Kohler, U. 2010, 'New developments in sequence analysis', *Sociological Methods & Research*, vol. 38, no. 3, pp. 359–64.

Céreq (Centre d'Études et de Recherches sur les Qualifications) 2005, *Quand l'école est finie ... Premiers pas dans la vie active de la Génération 2001* (When school is over ... First steps in the active life of Generation 2001), Céreq, France.

Corrales-Herrero, H. and Rodríguez-Prado, B. 2012, 'Characterizing Spanish labour pathways of young people with vocational lower-secondary education', *Applied Economics*, vol. 44, no. 29, pp. 3777–92.

Dorsett, R. and Lucchino, P. 2012, 'Beyond school-leaving age: the first five years', in ACEVO (the Association of Chief Executives of Voluntary Organisations (UK)) Commission on Youth Unemployment, *Youth Unemployment: The Crisis We Cannot Afford*, ACEVO, London, pp. 101–10.

Fasang, A. 2010, 'Retirement: Institutional Pathways and Individual Trajectories in Britain and Germany', *Sociological Research Online*, vol. 15, no. 2, http://www.socresonline.org.uk/15/2/1.html (accessed 28 June 2013).

Fry, J. and Boulton, C. 2013, *Prevalence of Transition Pathways in Australia*, Productivity Commission Staff Working Paper, Canberra.

Fuller, S. 2011, *Immigrant Employment Trajectories and Outcomes in the First Settlement Years: A Sequence-Oriented Approach*, Working Paper no. 11–07, Centre of Excellence for Research on Immigration and Diversity, Metropolis British Columbia, Canada.

Halpin, B. 2010, 'Optimal matching analysis and life-course data: the importance of duration', *Sociological Methods & Research*, vol. 38, no. 3, pp. 365–88.

Han, S.-K. and Moen, P. 1999, 'Clocking out: temporal patterning of retirement', *American Journal of Sociology*, vol. 105, no. 1, pp. 191–236.

Hollister, M. 2009, 'Is optimal matching suboptimal?', *Sociological Methods & Research*, vol. 38, no. 2, pp. 235–64.

Hunter, B., Gray, M. and Jones, R. 2000, 'An analysis of data from the longitudinal survey of ATSI job seekers, Topic 1: labour market participation patterns and pathways to Indigenous employment', Report to the Department of Employment, Workplace Relations and Small Business, Centre for Aboriginal Economic Policy Research, Australian National University, Canberra.

King, T. 2011, 'A framework for analysing social sequences', *Quality and Quantity*, September, pp. 1–25, http://link.springer.com/article/10.1007/s11135-011-9510-5 (accessed 11 October 2012).

Kogan, I. 2007, 'A study of immigrants' employment careers in West Germany using the sequence analysis technique', *Social Science Research*, vol. 36, no. 2, pp. 491–511.

Lamb, S. 2001, *The Pathways from School to Further Study and Work for Australian Graduates*, Longitudinal Surveys of Australian Youth (LSAY) Research Report no. 19, ACER, Melbourne.

Lesnard, L. 2006, *Optimal Matching and Social Sciences*, CREST-INSEE Working Paper, Institut National de la Statistique et des Etudes Economiques, Paris, http://halshs.archives-ouvertes.fr/docs/00/05/59/16/PDF/Lesnard_2006-Optimal_Matching_and_Social_Sciences-p.pdf (accessed 16 July 2013).

—— 2010, 'Setting cost in Optimal Matching to uncover contemporaneous socio-temporal patterns', *Sociological Methods & Research*, vol. 38, no. 3, pp. 389–419.

Lin, Q. 2013, 'Lost in transformation? The employment trajectories of China's cultural revolution cohort', *The ANNALS of the American Academy of Political and Social Science*, vol. 646, March, pp. 172–93.

McVicar, D. and Anyadike-Danes, M. 2002, 'Predicting successful and unsuccessful transitions from school to work using sequence methods', *Journal of the Royal Statistical Society Series A*, vol. 165, part 2, pp. 317–34.

Martin, P. and Wiggins, R. 2011, 'Optimal matching analysis', in Vogt, W. P. and Williams, M. (eds), *The SAGE Handbook of Innovation in Social Research Methods*, SAGE Publications, London, pp. 385–408.

Pollock, G., Antcliff, V. and Ralphs, R. 2002, 'Work orders: analysing employment histories using sequence data', *International Journal of Social Research Methodology*, vol. 5, no. 2, pp. 91–105.

Quintini, G. and Manfredi, T. 2009, *Going Separate Ways? School-to-Work Transitions in the United States and Europe*, OECD Social, Employment and Migration Working Paper no. 90, OECD, Paris.

Scherer, S. 2001, 'Early career patterns: a comparison of Great Britain and West Germany', *European Sociological Review*, vol. 17, no. 2, pp. 119–44.

Yu, S., Bretherton, T., Schutz, J. and Buchanan, J. 2012, *Understanding the Nature of Vocations Today: Exploring Labour Market Pathways*, National Vocational Education and Training Research Program Working Paper, NCVER, Adelaide.