Australian Government
Productivity Commission

Prevalence of Transition Pathways in Australia

Productivity Commission
Staff Working Paper

August 2013

Jane Fry
Clare Boulton

The views expressed in this paper are those of the staff involved and do not necessarily reflect the views of the Productivity Commission.

ISBN 978-1-74037-446-0

This publication is available from the Productivity Commission website at www.pc.gov.au. If you require part or all of this publication in a different format, please contact Media and Publications.

Publications enquiries:
Media and Publications
Productivity Commission
Locked Bag 2 Collins Street East
Melbourne VIC 8003

Tel: (03) 9653 2244
Fax: (03) 9653 2303
Email: maps@pc.gov.au

General enquiries:
Tel: (03) 9653 2100 or (02) 6240 3200

An appropriate citation for this paper is:

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

JEL codes: C19, C38, I21, J21, J22, J24, J64.

The Productivity Commission

The Productivity Commission is the Australian Government's independent research and advisory body on a range of economic, social and environmental issues affecting the welfare of Australians. Its role, expressed most simply, is to help governments make better policies, in the long term interest of the Australian community.

The Commission's independence is underpinned by an Act of Parliament. Its processes and outputs are open to public scrutiny and are driven by concern for the wellbeing of the community as a whole.

Further information on the Productivity Commission can be obtained from the Commission's website (www.pc.gov.au) or by contacting Media and Publications on (03) 9653 2244 or email: maps@pc.gov.au

Contents

Tables

Acknowledgments

A number of staff from the Productivity Commission provided valuable help and advice in producing this paper. The authors would particularly like to thank Noel Gaston and Patrick Jomini for their guidance and feedback; as well as Meredith Baker, Jackson Bunting, Dominic Crowley, Matt Forbes, Geoff Gilfillan, Jenny Gordon, Lisa Gropp, Peter Harris, Patrick Laplagne, Alison McClelland, Lawrence McDonald, Leonora Risse, and Lou Will for helpful comments.

The authors also wish to thank the external referee Glenda Quintini (Directorate for Employment, Labour and Social Affairs, Organisation for Economic Co-operation and Development) for her helpful feedback.

This paper 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 external referee or the Productivity Commission.

Abbreviations

ABS	Australian Bureau of Statistics
BHPS	British Household Panel Survey
CA	Cluster analysis
ECHP	European Community Household Panel
ESS	Error Sum-of-Squares
GFC	Global financial crisis
HILDA	Household, Income and Labour Dynamics in Australia
LFS	Labour Force Survey
LSAY	Longitudinal Surveys of Australian Youth
MNL	Multinomial logit
NEET	Not in employment, education or training
NILF	Not in the labour force
OM	Optimal matching
OMCA	Optimal matching and cluster analysis
PC	Productivity Commission
TOS	Time of survey data
VET	Vocational education and training

Key points

- This paper uses longitudinal information from the calendar in the *Household, Income and Labour Dynamics in Australia* (HILDA) Survey to track monthly education and labour market activities from 2000 to 2010 for about 6500 working-age individuals. The techniques of optimal matching and cluster analysis (OMCA) are used to identify and group individuals with similar patterns of activities into 'pathways'.
- Much of the wider literature considers transitions from one activity to another (such as study to employment, or employment to retirement). OMCA applied to calendar-style data for other countries shows that there can be multiple transitions (such as reversals or repeated activities, like returning to the labour force or churning in and out of employment) and different pathways can arise with key life events (such as leaving education, family formation or retirement).
- Seventeen pathways are identified. Although each pathway contains some variation between the sequences of activities, distinct patterns can be observed.
  - For youths aged 15–24 in 2001, five pathways are identified: three associated with increasing education levels and transitions to work; one associated with churning in and out of work; and one dominated by young women withdrawing from the labour force to raise children.
  - Activity sequences for young adults aged 25–39 are grouped into four pathways: two involving work (one with increasing education); and two involving prolonged periods outside the labour force associated with raising children (with one pathway showing subsequent return to work).
  - Mature adults aged 40–54 in 2001 follow one of four pathways: one dominated by work; two dominated by women spending time outside the labour force raising children (with one return to work pathway); and one pathway associated with early retirement.
  - For seniors aged 55–64, four pathways are identified: one dominated by work; and three associated with retirement or transitions to retirement.
- Successful and unsuccessful outcomes in the labour market can be related to the pathways that individuals follow. The analysis in this paper can be a valuable input to identifying relationships between pathways and outcomes, and the individual characteristics that are associated with specific pathways. That analysis could then inform strategies to reduce the risk of unsuccessful labour market outcomes, such as prolonged unemployment.

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 circumstances change, some workers expand (or contract) their hours of work, other workers 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 individuals remain in, or return to, education. Others exhibit stable employment with no change in their labour market status over long periods of time.

Understanding the nature, prevalence and consequences of these labour market and educational transitions relates to at least four key areas of policy interest.

First, certain types of transitions can be linked to the reallocation of workers during structural adjustment in response to major events, such as the global financial crisis or sustained changes in the terms of trade.

Second, the Council of Australian Governments' *National Partnership Agreement on Youth Attainment and Transitions* is directed at supporting successful transitions from school to work in order to reduce youth unemployment and promoting engagement with the labour market. According to Youth Connections National Network (2013):

> In the current environment of an ag[e]ing population and an increasing focus on productivity and skills development; the need for all young Australians to successfully transition to the workforce and make the most of their abilities is critical. (p. 5)

The focus on youth transitions has both short- and long-term consequences, as:

> [The transition phase] … is a period in which the basis is laid for many of the personal and vocational skills that will determine individuals' labour market trajectories for decades to come, in which education and training qualifications are obtained that are valued in the labour market and that make a profound difference to life chances, and in which the basis is laid for economic returns over the life span. (Dandolopartners 2012, p. 25)

Third, an analysis of transitions can inform policies to assist adults to improve their skills in their current career or to develop the skills they require to change careers to growing occupations.

Finally, population ageing will increase the number of older Australians, some of whom might want to work beyond the traditional retirement age. Understanding this type of transition will help in the design of policies that affect individuals around retirement age.

1.1 Transitions and outcomes in Australia — what has been studied?

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.[1] However, such data reveal little about transitions between different activities. For example, 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 — various education/labour market activities.

There has been much analysis of school-to-work transitions for youths (usually aged under 25 years) in Australia. For example, researchers have defined pathways for young individuals who left school and did not obtain a tertiary qualification (Lamb and McKenzie 2001) and for those who enrolled in tertiary study (Lamb 2001), based on their main activity each week over a seven-year period in the late 1980s and early 1990s. Fitzpatrick et al. (2011) estimated the average length of time taken to find employment after leaving education in the late 1990s. Buddelmeyer and Marks (2010) estimated the effects of post-school qualifications on annual transitions of youths between labour market states.

Studies of labour market transitions beyond age 25 tend to focus on particular groups in the population. For example, Lattimore (2007) explored the likelihood and reasons why Australian men leave the workforce. Cai (2010) investigated the effects of persistence in labour market status and changes in circumstances, such as having children, on the work choices of women who were married or in de facto relationships. Transitions into and out of employment by lone and couple mothers with dependent children aged less than 15 years were studied in Baxter and Renda (2011). Different patterns of full-time and part-time employment in transitions to retirement for men and women were identified in Gilfillan and Andrews (2010) and

[1] In the youth transitions literature, it is becoming common to focus on individuals 'not in employment, education or training' (NEET).

transitions from part-time work to either full-time work or retirement (men only) in Sane (2011).

Borland and Johnston (2010) focused on the employed, exploring the effects of labour market history on the duration of employment (although that study was limited to exits from employment rather than transitions between all labour market states). Early school leavers and their re-engagement in education was the focus in Black, Polidano and Tseng (2011). Polidano and Mavromaras (2010) focused on the labour market outcomes of people with disabilities to determine the effect of completing a vocational qualification.

Much of the literature considers transitions as one-off changes from one activity to another (such as study to employment, or employment to retirement).

By examining activities over an extended period of time, the analysis in this paper accounts for multiple changes in activities (such as reversals or repeated activities, like those associated with returning to the labour force or churning in and out of employment) and identifies broader patterns — or pathways — that might be associated with different life stages.

1.2 Why pathways are important

Over an extended period, longitudinal data show sequences of activities. Rather than focussing on a single transition, analysis of a sequence 'emphasises that single events should not be isolated from each other but have to be understood in their continuity' (Aisenbrey and Fasang 2010, p. 421).

According to Pollock, Antcliff and Ralphs (2002):

> The fact that a person occupies a number of different statuses over a period of time is a fundamental aspect of life. These changes in status exist everywhere, as we accumulate qualifications, move through different family structures, and occupy a variety of labour market and occupational statuses. Focusing on employment, … it is possible to identify distinct career trajectories … This form of analysis has permitted an understanding of how people from different backgrounds progress through their employment careers. (pp. 91–2)

Not everyone follows the same sequence of activities. There are variations in what they do, when they start, how long they continue and what they did before and do next. However, patterns also have similarities. As a result, in analysing the patterns in sequences, it is helpful to group together people with the same (or a similar) pattern. That pattern then characterises a 'pathway' for the group. Optimal matching

and cluster analysis are two techniques that — when combined — are ideally suited to identifying such groups.

Grouping similar sequences into pathways can show how individuals move into and through the labour market (Yu et al. 2012), while accommodating breaks, divergences and reversals (Corrales-Herrero and Rodríguez-Prado 2012). They also show patterns of mobility that can affect later labour market outcomes (Fuller 2011).

The analysis of pathways can show, for example, which pathways are more likely to lead to jobs and which pathways are associated with individuals being disconnected from the labour market (Corrales-Herrero and Rodríguez-Prado 2012). Identifying 'less successful' pathways and the characteristics associated with individuals in those pathways, can assist in targeting intervention.

The *Household, Income and Labour Dynamics in Australia* (HILDA) Survey contains a rich source of information in its education and labour market calendar. To date, these data are underutilised even though they are extremely valuable for understanding the dynamic processes associated with, for example:

- how youths transition to employment
- how and when workers transition to retirement
- how jobseekers become discouraged.

This paper highlights the value of these data.

1.3 Roadmap

The purpose of this paper is to extend and enhance the existing analysis of transitions in Australia. This is achieved by describing the broad patterns — or pathways — and persistence in education and labour market activities during an extended period. In doing so, this paper also highlights the considerable value of a systematic analysis of the HILDA calendar information.

Chapter 2 summarises the relevant international literature on sequence analysis of labour market transitions and highlights the gap that this paper seeks to fill.

Chapter 3 begins with a brief outline of the HILDA calendar data and the methodology used to identify the various pathways followed by individuals over a ten-year period. A novel technique for comparing sequences — optimal matching — is used to calculate a measure of the similarity of each pattern to all other patterns in the data. Cluster analysis is then used to assign individuals to groups

according to the similarities in their activity patterns. Each group represents a particular transition pathway. Additional details are provided in appendices A (data) and B (methodology).

Since activity patterns (and their possible implications for policy) are likely to vary over the life cycle, the analysis is conducted separately for youths (aged 15–24), young adults (aged 25–39), mature adults (aged 40–54) and seniors (aged 55–64).

After summarising the pathways identified and their prevalence, the chapter describes six key pathways using characteristics of the activity sequences and of the individuals in those pathways. (Appendix C provides results for the remaining pathways.)

Chapter 4 indicates areas for future research.

2 Transitions as sequences

This chapter outlines the various approaches that have been used to examine transitions (section 2.1), some of the design characteristics of those studies that have adopted a sequence approach (section 2.2), and the key findings from that literature (section 2.3).

2.1 Research approaches to transitions

There are three standard approaches to the study of transitions.

First, 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.

A second approach is to consider duration or time spent in labour market states. For example, some studies analyse the duration of unemployment spells.

A third 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.[1]

A relatively new approach studies transitions over (part of) a career or lifetime as a sequence of activities. This approach relies on data collected in the form of a diary or a calendar. Sequence analysis techniques — usually optimal matching and cluster

[1] This approach can be applied to data for two time periods to compare labour market states (identify single transitions) using a simple cross-tabulation or it can be extended to trace the links between various types of (multiple) transitions over time using a Markov chain technique. See for example, Fabrizi and Mussida (2009) and Magnac (2000).

analysis (OMCA, see appendix B) — can then be used to identify individuals with similar activity patterns.[2]

The sequence approach has several advantages over the others (box 2.1), not least of which is the ability to analyse the different activities that individuals undertake.

Box 2.1 **Advantages of sequence analysis**

Most proponents of sequence analysis argue that it complements more conventional methods by providing a more holistic perspective and context for results from other types of analysis (Anyadike-Danes and McVicar 2010; Halpin 2010; King 2011; Pollock, Antcliff and Ralphs 2002; Yu et al. 2012).

More specifically, a sequence approach 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).

The approaches that focus on a single transition or point in time can miss the turbulent nature of some sequences (Fuller 2011), disregard the varying dynamics experienced by individuals (Quintini and Manfredi 2009), and ignore valuable information prior to and after the change (Brzinsky-Fay 2007; Pollock, Antcliff and Ralphs 2002).

Appendix B contains more information on the specific method of sequence analysis adopted in this paper.

The next section summarises some of the design features of studies that use sequence analysis in the context of labour market transitions (the 'transition pathways' literature), and the final section draws some implications for the sequence approach used in this paper (chapter 3).

[2] There are other techniques that could be used to analyse this type of data. The techniques typically used in the three other approaches to transitions analysis can also be used, for example.

2.2 Study characteristics in the transition pathways literature

A growing literature has adopted a sequence approach to study labour market transitions in Britain, Northern Ireland, the United States, China and parts of Europe. Yu et al. (2012) is the only study using Australian data.

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 their initial entry into the workforce or retirement.

Periods of unemployment and inactivity can adversely affect future labour market outcomes (usually termed a 'scarring effect'). With less employment experience, youths tend to be more vulnerable than others when economic conditions deteriorate. Moreover, for youths, gaining steady employment takes time as 'young workers, who have only relatively recently entered the labour market, are engaging in job shopping in order to look for the sector and the occupation that suits them best' (Bachmann and Burda 2007, pp. 13–14) and 'an extensive process of mixing and matching among workers and firms characterizes the youth labour market' (Mroz and Savage 2006, p. 261). Youths seem to be particularly vulnerable to periods of unemployment and/or inactivity when they leave education and enter the labour market for the first time. According to the Youth Connections National Network, they are 'the lowest skilled and the most vulnerable sector of the labour market' (2013, p. 10). Hence, there is an understandable emphasis on 'improving' (shortening and smoothing) school-to-work transitions.

At the other end of the age spectrum, a few studies examine older people's transitions to retirement from the labour force. Older workers' labour force participation has implications for the funding and demands on the age pension system (Productivity Commission 2005). Since the 1970s, in most developed countries there has been a trend towards earlier and less standardised retirement (Fasang 2010) — although the increasing trend of early retirement might have abated since the global financial crisis of 2008. However, retirement, in terms of exit from the labour force and reliance on the age pension (or non-labour income, such as superannuation in Australia), has become less of a one-off transition (that is, retirement as a single permanent exit from the labour force) and more of a staged process (Han and Moen 1999).

Activities

Studies of labour market transitions (based on the sequence approach) usually consider education, employment and 'not in employment, education or training' (NEET), often disaggregated in some way to form the activity categories for analysis. For example, education is sometimes split according to full-time or part-time enrolment and by level of study. 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.[3]

Typically, about 6 or 7 activity categories are defined, although some studies have up to 13 activities. The number of activities is usually driven by data availability and the research question, although the more categories there are, the more difficult it can be to distinguish high-level changes (such as job loss or finishing education) from more marginal changes (such as a change in occupation or the level of study).

Effects on pathways

Disaggregating an activity category can increase the scope for more pathways to be identified. For example, Kogan (2007) and Yu et al. (2012) distinguish employment by broad occupation and find eight out of ten pathways identified in their studies could be characterised by individuals spending most of their time employed in the same broad occupation. Were the analysis to be redone using a generic employment activity, it is likely that individuals in those eight pathways would be regrouped into fewer, larger pathways characterised by continuous employment.

Timeframes

In these types of studies there is a tradeoff between the period covered and the frequency of the observations on activities of interest.

[3] In some studies, such as Scherer (2001), looking after the home is treated as a separate activity category.

Computational complexity typically limits sequences of monthly activities to less than ten years. With much of the focus on youths and their initial entry to the labour market — which can involve many changes in a short time — most studies use shorter sequences of about three to six years. This is usually deemed sufficient to capture the transition from education to work.

However, a longer timeframe can capture late changes (such as gaining employment after prolonged periods of unemployment or inactivity — as in the *Recovery* pathway in Quintini and Manfredi 2009). Brzinsky-Fay (2007) analysed the activities of European youths over five years after the completion of general schooling. Two of the pathways identified were labelled *Failure* and *Detour*. Both pathways were characterised by prolonged unemployment over the first three years. Thereafter, the patterns diverged as individuals in the *Detour* pathway increasingly gained employment that continued for the remainder of the five-year period, while individuals in the *Failure* pathway mostly remained unemployed. If a shorter timeframe had been used for that study, it is likely that the *Detour* pathway would not have been identified.

Turbulent patterns in labour market histories cannot be captured using annual data. In the case of youths, 'the school-to-work transition process has become longer and more complex due to the growing importance of intermediate states such as temporary jobs, rapid job-changing and instability' (Corrales-Herrero and Rodríguez-Prado 2012, p. 3779). Such instability is apparent in the *In and Out* pathway in Quintini and Manfredi (2009), which shows many youths — particularly in the United States — change activity (sometimes several times) within a year.

Annual activities tend to be used in studies that adopt a longer timeframe for the sequences — between 10 and 25 years. Sometimes this choice is dictated by the data. For example, the *General Social Survey of China* (used by Lin 2013) asks respondents to recall which years they spent engaged in various employment and non-employment activities over the course of their careers. In other studies, there is a conversion from monthly to annual activities using either a representative month's activity (Schoon et al. 2001) or a most frequent activity during each year (Anyadike-Danes and McVicar 2005, 2010).

To summarise, 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 the monthly post-school activities of youths over three years may not be sufficiently long to observe some youths complete their education and gain stable employment. Some individuals who are long-term unemployed may be observed to secure employment, although a longer

time would be required to determine whether their employment continues or is transitory. In this study, monthly data for a ten-year period are used, which reduces the problem of censoring that arises when using survey data of short duration.[4]

2.3 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.[5] 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 there is variation among the sequences within any pathway, each pathway has a distinct overall pattern that differs from other pathways. 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).

For each pathway, the characteristics of individuals, such as their gender, education levels and reasons for withdrawing from the labour market, can enhance the explanation for each pathway. This is particularly important when an activity pattern can occur for different reasons. For example, the motives for being NILF, or

4 Although, using a longer timeframe does not eliminate the problem of censoring entirely.

5 All of those studies use optimal matching (or a variant of the technique) to quantify the 'distance' between sequences. Some of those studies then use cluster analysis to form groups, each representing a pathway. The remaining studies do not identify or construct pathways but use the distance variable for other types of analysis (for example, as a dependent variable in regression).

no longer actively seeking work, are likely to differ for males and females as well as the young and those of retirement age.

Although the results from these studies are not strictly comparable, table 2.1 demonstrates that for most studies judgements can be made about the pathways associated most closely with unemployment and/or labour market withdrawal. In table 2.1 these pathways are described as 'red' pathways, and for most studies they are usually few in number (and represent a small percentage of the sample).

The one available study for Australia (Yu et al. 2012) uses annual data and does not split the population by age. With its focus on occupational mobility among adults, there is a heavy emphasis on broad level occupations among the activities tracked and this affects the characteristics of the pathways identified — most pathways are characterised by many years of employment in a particular broad occupation. The one pathway associated with prolonged periods of NILF (averaging more than seven years) was dominated by individuals who had likely retired. These issues are addressed in this study (chapter 3).

Table 2.1 **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 judgement 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 Pathways and estimates of prevalence

To analyse labour market transitions as ordered sequences of activities, appropriate data are required. Pollock, Antcliff and Ralphs (2002) outline the characteristics of an ideal dataset for this purpose. Specifically, the data should be at an individual level, record successive activities, the dates and duration when they are undertaken, and cover an appropriately long observation window. The data should also contain relevant supplementary information about the individuals and the activities. This information helps explain the patterns uncovered in terms of the reasons or motivating factors behind the observed patterns, and the characteristics of individuals with those patterns.

For education and labour market activities of the working-age population in Australia, the *Household, Income and Labour Dynamics in Australia* (HILDA) Survey is the only data source with these attributes. HILDA is a large, nationally-representative household panel survey that has been running since 2001. It contains both historical information about individuals when they are first interviewed and time-varying information recorded at annual interviews. The first ten waves of data have been used in this study. Further details of HILDA are available in appendix A.

HILDA also contains a calendar where individuals record their employment, job search and education-related activities. The calendar is split into 'month thirds' (corresponding to early, middle and late in the month) and individuals are asked to recall their activities since the beginning of the previous financial year. This information permits tracking of short-duration activities that occur throughout the year rather than being restricted to activities that coincide with the annual interviews. This allows some of the more diverse experiences to be analysed.

In this study, the education and labour market activities are divided into the following five mutually exclusive and exhaustive groups:

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).[1]

Individuals with incomplete sequences of activities were removed. For those remaining, activity information was aggregated from month thirds to months (appendix A). This left a sample of 6566 individuals with activities tracked each month from July 2000 to June 2010. Sample weights have not been applied in the following analysis.

This chapter provides a summary of the pathways identified (section 3.1). Several pathways are analysed in some detail (section 3.2), with results for the remaining pathways contained in appendix C.

3.1 Pathway summary

After splitting the sample into age segments (box 3.1), optimal matching and cluster analysis (box 3.2) were used to identify groups of similar sequences within each age segment. Each group represents a pathway, and subsequent analysis of characteristics of the individuals and their sequences suggested a descriptive label for each pathway. Appendix B provides technical details.

[1] 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.

Box 3.1 **Age segments**

The working-age population includes individuals at different 'life stages'. These life stages are likely to affect the types of activity patterns observed for different age groups. For example, younger people are more likely to be engaged in prolonged periods of study than older people. Females who spend time outside the labour force may be raising children if they are young, or retired if they are older.

Since a particular activity pattern can have different interpretations (and relevance) for different age groups, the working-age population is divided into segments for analysis. These segments are defined according to age in 2001 (at the first interview):

- youths, aged 15–24 (877 individuals)
- young adults, aged 25–39 (2289 individuals)
- mature adults, aged 40–54 (2354 individuals)
- seniors, aged 55–64 (1046 individuals).

Box 3.2 **Grouping similar activity sequences using optimal matching and cluster analysis**

Sequences of activities can be described in various ways, such as by the proportion of time spent in each activity state or by the (temporal) ordering of activities. However, a richer picture becomes available when both the order and duration of activities are considered as part of a pattern.

Optimal matching and cluster analysis (OMCA) are two techniques that, when used in tandem, provide a systematic way to compare activity sequences and group together those with similar patterns. These techniques are data driven and designed to be exploratory (describing what is) rather than confirmatory (formulating statistical models and testing hypotheses). Appendix B contains more details about the techniques.

Optimal matching (OM) provides a means of quantifying the extent of pattern similarity (or difference) between large numbers of potentially quite complex sequences. Rather than using a simple ternary classification of patterns (exactly the same, completely different, anything in between), pattern similarity in OM (the distance) has a sliding scale that is determined according to the number and type of changes required to make one pattern look identical to another. Each type of change is given a particular 'cost' and the distance between two sequences represents the total of these costs for that pair of sequences. The OM procedure creates a matrix of the distances between each sequence in the dataset, with smaller distances representing greater similarity.

Cluster analysis is a technique for identifying groups in data. Using the distance matrix from OM as an input to cluster analysis provides a means of combining similar sequences together and distinguishing different sequences. The basic idea is to form groups that maximise within-cluster similarity and minimise between-cluster similarity. In this study, each cluster is referred to as a pathway.

For each age segment, four or five pathways were identified (table 3.1).[2] Some pathways with similar characteristics, such as those labelled *Work* or *Prolonged NILF*, were identified in several age segments. Within each age segment, the most prevalent pathways (that is, the pathways followed by the most individuals in that age segment) were those associated with work, except among seniors for whom the *NILF* pathway was most prevalent. All three 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 all individuals in other pathways or age segments did not undertake any education).

Table 3.1 **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	16.8	12.4			6.6[c]
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.

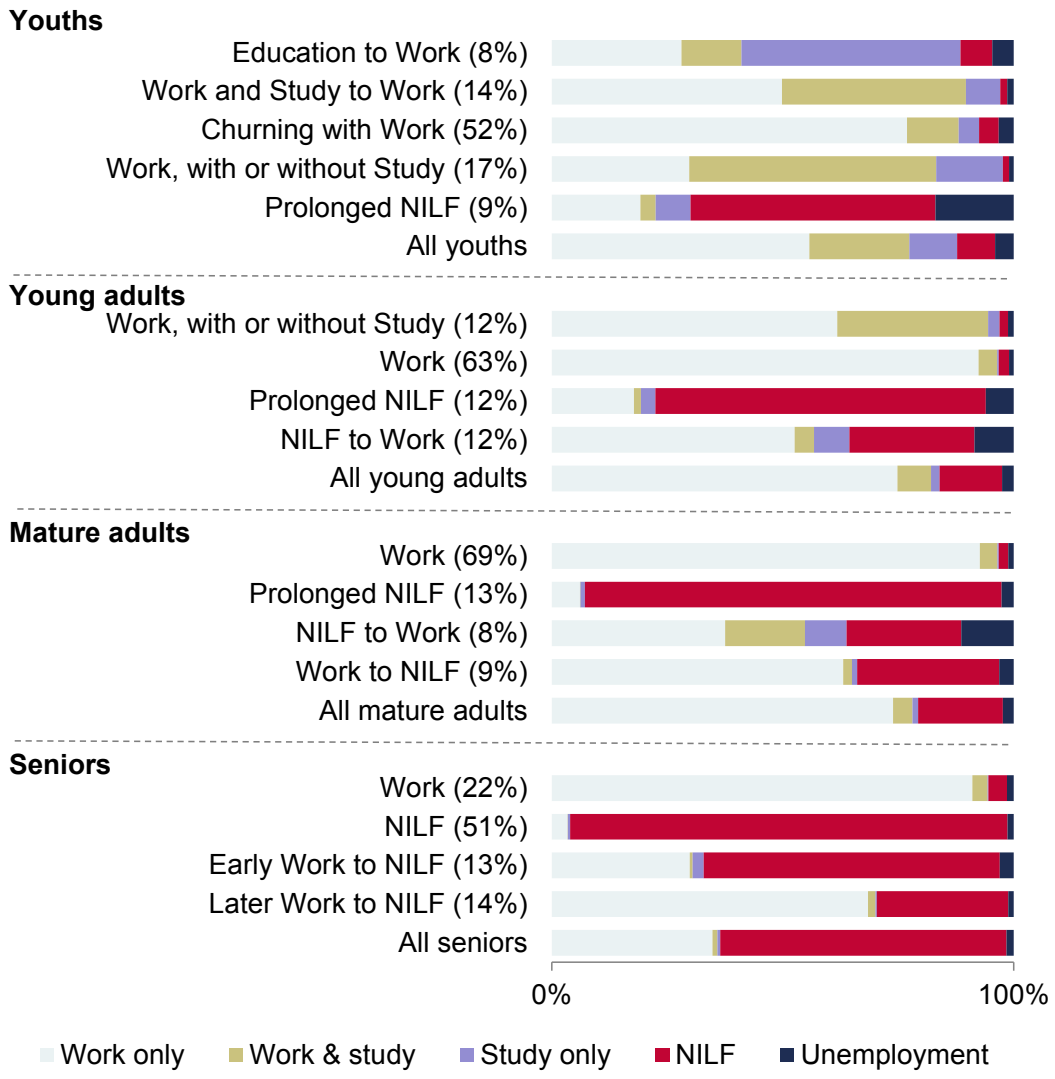
The pathways in each age segment have distinct characteristics in terms of how much of the ten-year period individuals spent in each activity (figure 3.1). More detailed results show that:

- For youths, three pathways are associated with transitions from education to work, one is characterised by work (although for many youths there is instability as spells of work are often interspersed with other activities — referred to as 'churning'), and one pathway consists of prolonged periods of NILF.

2 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.

- Of the four pathways identified for young adults, two are characterised by work, one is associated with prolonged periods of NILF and one involves transitions from NILF to work.
- Mature adults are grouped into four pathways associated with work, NILF or a transition between the two.
- The four pathways for seniors are associated with work, NILF or a transition from work to NILF.

Figure 3.1 **Average time allocation, by activity, pathway and age segment**
Per cent of time[a]



[a] The time allocation for each activity represents the total percentage of the ten-year period spent undertaking that activity, averaged over individuals in that age segment and pathway. The average time that individuals spent on an activity is not necessarily one continuous stretch and the ordering of activities shown is not related to the underlying patterns in the sequences. Pathway prevalence (table 3.1) shown in brackets.

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

More generally, the pathways showed there were several key patterns in activities among the working-age population. 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 3.1). 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 3.1).

3.2 Selected pathways

For each of the 17 pathways identified, detailed results describing characteristics of the individuals and their activity patterns provide an interesting story of labour market transitions and interactions with the education system. As an illustration, six of the pathways are considered in detail:

- youths in *Work and Study to Work* and *Churning with Work*
- young adults in *Work* and *NILF to Work*
- mature adults in *Work to NILF*
- seniors in *Work*.

Results for the other 11 pathways are contained in appendix C.

For each pathway, two charts and two tables are presented. The first type of chart — labelled *(A) Activity sequences by individual* — is called a sequence index plot and shows the data (box 3.3). Sequence index plots can show the effects of major changes in the economy and economic policies during the 2000s on the activities of individuals. Since the sequence of activities for each individual represents the same calendar period, plots of these sequences will show whether or not particular changes appear to have affected the pattern of activities at particular points in time through large common 'blocks' of activity. For example, the global financial crisis (GFC) might have affected decisions of retired seniors by inducing them to return to

work or begin looking for work from mid-2008 or 2009. Similarly, any post-GFC rise in youth unemployment should be apparent.[3] Likewise, if the introduction of the baby bonus in 2002 influenced the fertility decisions of young women, there might be an increase in withdrawal from, or prolonged absence from, the labour force after that time. Neither of these changes appears to have affected the patterns in this analysis in a significant way. This is not to say that these effects could not appear in a finer analysis of these data.

Box 3.3 **The sequence index plot**

In a sequence index plot, individuals are numbered along the vertical axis and time is shown on the horizontal axis. The activity sequence for each individual in the pathway is represented by a horizontal series of coloured markers, with each type of activity represented by a different colour: a change in colour from one period to the next represents a transition to another activity. Individual sequences are ordered according to their similarity to the most common sequence for the age segment, with those most similar along the top of the plot.

A cautionary note on sequence index plots. When there are many individuals and/or time periods, a sequence index plot can be subject to 'overplotting', as the markers for each type of activity are successively overlaid on the chart. Some information is obscured, and the visual impression gained from a sequence index plot can change with the order in which the activities are plotted. This can mean that one or two individuals spending one or two periods in a different activity are not easily seen. In the sequence index plots produced for this study, more frequently occurring activities for an age segment are plotted before less frequently occurring activities.

The second type of chart — labelled *(B) Activities by monthly share* — is called a chronograph and shows the distribution of activities across individuals for each month. The chronograph shows the share of individuals undertaking each activity each month (in essence, a time series of activity participation rates). It does not provide any information on activity duration but gives another perspective on the activities that might not be apparent from the sequence index plot.

3 Figures from the ABS *Labour Force Survey* (Cat. No. 6291.0.55.001) showed the overall unemployment rate rose from 4.1 per cent in October 2008 to 6.0 per cent in February 2009 and remained between 5 and 6 per cent until July 2010. While it is difficult to calculate unemployment rates for that time period corresponding to the youth age group defined in this study (who would have been aged between about 22 and 31 years in 2008 and 2009), unemployment rates of 20–24 year olds and 25–34 year olds rose between October 2008 and February 2009 (from 5.2 per cent to 8.8 per cent and from 4.3 per cent to 5.9 per cent, respectively).

The first table shows characteristics of individuals in a particular pathway. Characteristics of individuals in all pathways for that age segment are provided for comparison. These characteristics relate to the time at which the survey was conducted in 2001 and 2010 and do not come from the calendar. Consequently, the extent of unemployment and NILF may not appear the same as that in the chronograph (box 3.4). The second table uses the calendar data to describe characteristics of the activity sequences in the pathway, in terms of time spent in each activity, numbers of spells of each activity and spell duration.[4]

Box 3.4 **Reconciling calendar and time-of-survey data: unemployment and NILF**

In HILDA, there are two sources of information on the activities undertaken by individuals: variables for labour force status and other characteristics pertaining to the time of the survey (TOS), and variables from the employment and education calendar that refer to activities during the 12–18 months prior to the interview (appendix A).

TOS information can be used to supplement information available in the calendar. For example, calendar information can be used to identify individuals who have undertaken study between two interviews while TOS information can provide more detail on the level and area of study and whether the qualification was completed. Similarly, TOS information on the reason for being outside the labour force at the time of the interview can be used to impute reasons for other individuals who — according to the calendar — were NILF between interviews.

In this study, the prevalence of unemployment and NILF at the time of the survey can be calculated from calendar or TOS information. However, calculations using the two sources frequently give different answers. There are two reasons for this.

First, there are (subtle) differences in the split between unemployment and NILF in the two sources in HILDA. TOS labour force status uses the ABS definition of unemployment (actively seeking work, available to start work and not waiting to start a job). Individuals who were marginally attached to the labour force (that is, those who wanted to work and were either actively looking for work but were not able to start work at the time, or were not actively looking, but available to start work) are defined as NILF. In contrast, calendar information implicitly defines unemployment using a broader definition of 'looking for work'. This leads to the marginally attached being classified as unemployed rather than being NILF, inflating the prevalence of unemployment based on the calendar.

(Continued next page)

[4] Neither the number of spells nor the duration have been adjusted for censoring.

Box 3.4 (continued)

Second, there are differences that arise from the particular classification of calendar activities adopted for this study. In this analysis, time spent studying and not working has been extracted from calendar information on unemployment or NILF and tracked as a separate 'study only' category. The corresponding extraction is not done for TOS information, reducing the prevalence of unemployment and/or NILF based on the calendar.

Table 1 below shows how these differences can be reconciled for youths in the *Churning with Work* pathway. TOS information for 2001 shows 22.08 per cent of individuals in this pathway were jobless (regardless of study). The equivalent figure from the calendar data was 22.95 per cent. Excluding the marginally attached and those who study while NILF from the ABS measure of NILF in the TOS data leaves 3.97 per cent, which is close to the comparable figure of 3.75 per cent from the calendar. Based on the calendar data, 54.9 per cent of the total jobless were studying. Applying this share to the total jobless using TOS information gives 12.12 per cent who were studying while jobless, which is close to the calendar equivalent of 12.6 per cent. TOS information shows 8.39 per cent of individuals were unemployed (regardless of study). Adding the marginally attached and removing those who study while unemployed leaves 5.99 per cent, which is close to the calendar equivalent of 6.6 per cent.

With 12.12 per cent studying while jobless and 7.7 per cent studying while NILF in the TOS data, about 36 per cent of jobless students are looking for work.

Table 1 **Unemployment and NILF reconciliation example**

Measure	*TOS*	*Calendar*
ABS NILF incl. marginally attached	13.69	
Study while NILF	7.73	
Marginally attached	1.99	
ABS NILF less marginally attached and study	3.97	3.75
Share of study among all jobless		*0.549*
Study while jobless	12.12[a]	12.60
Study while unemployed	4.39	
ABS unemployment	8.39	
ABS unemployment plus marginally attached	10.38	
ABS unemployment plus marginally attached less study	5.99	6.60
Share of unemployment among jobless students	*0.362*	
Total jobless (incl. study)	22.08	22.95

[a] Estimate based on share of study among all jobless from calendar data.

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

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 3.2 below). In 2001, the average age is 18.6 years (table 3.2 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 3.2). 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 have attained at least a Bachelor Degree by this stage (table 3.2). Those who are not working tend to be women taking time out from the labour force to raise children.

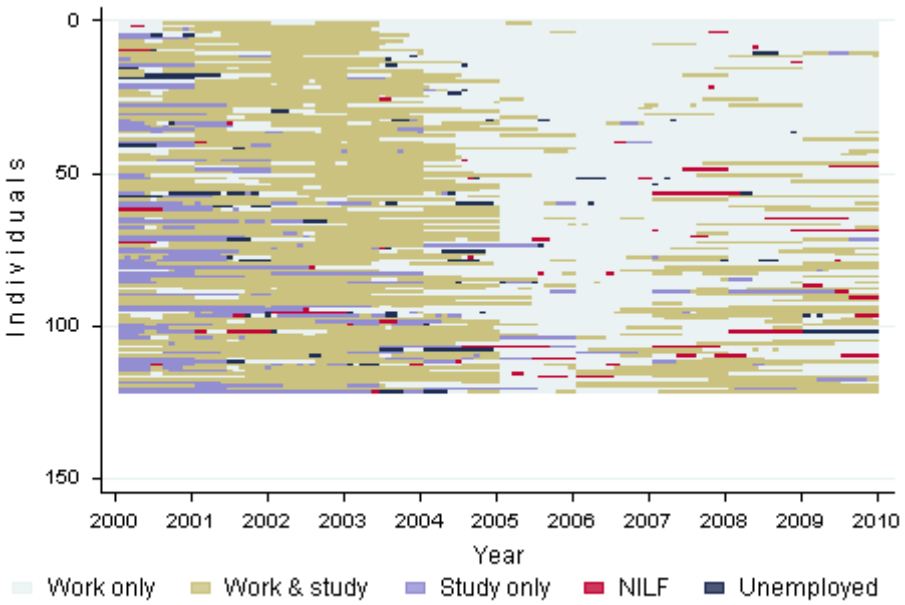
Over the ten-year period, education levels have risen because everyone in this pathway has spent some time studying. They also spent some time combining work

and study, averaging about three spells of 15.89 months per spell (table 3.3 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 3.2).

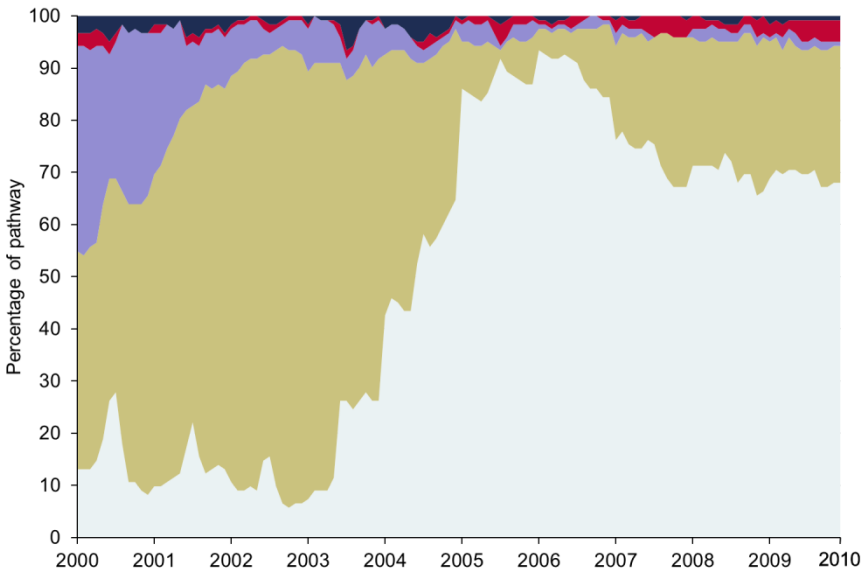
Individuals in this pathway gravitate to the major cities (66.4 per cent in 2001 and 72.1 per cent in 2010). This is likely to be associated with them studying at universities and then remaining in the major cities.

Figure 3.2 **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 3.2 **Selected characteristics of individuals in the *Work and Study to Work* pathway for youths**

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 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 3.3 **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.

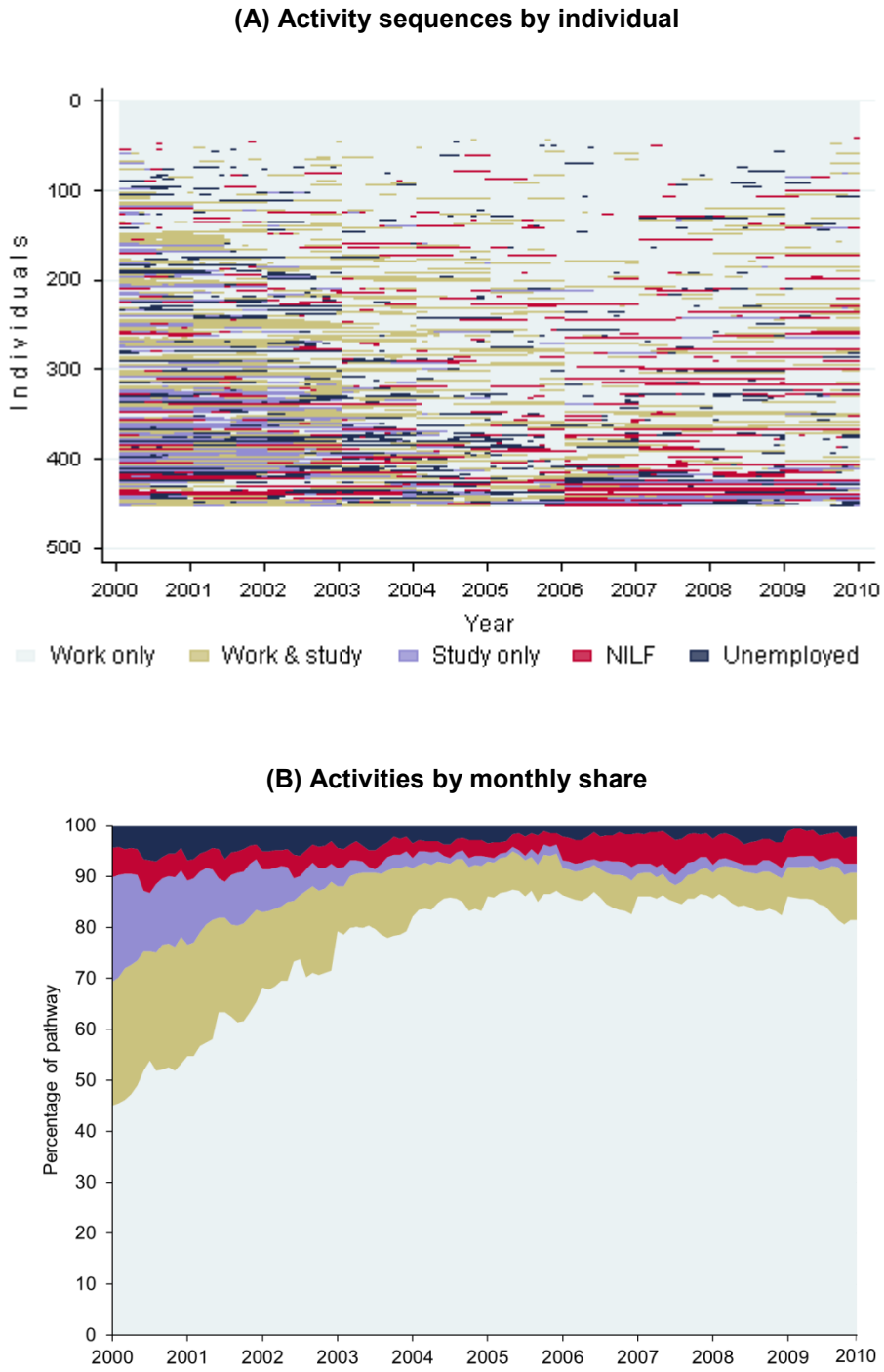
Youths in the *Churning with Work* pathway

Churning with Work is the largest pathway for the youth segment: it represents 453 individuals or 51.7 per cent of all youths (table 3.5 below). These individuals are further advanced in their transition from education to work compared to those in the *Work and Study to Work* pathway. The average age in 2001 is 20.2 years (table 3.4 below).

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 3.3 below). 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.

Figure 3.3 **Activities in the *Churning with Work* pathway for youths**



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

Individuals who are initially neither working nor looking for work are primarily engaged in study (table 3.4). 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, Diploma or 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 3.4). 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[5] 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 (box 3.5). As a result, in 2010 individuals in this pathway have relatively low levels of education compared to the other youth pathways (table 3.4).

5 That is, the likelihood of additional spells of study, at any level of education.

Box 3.5 **Study activity and changes in the profile of three broad levels of education**

Not all study activity over the ten-year period changes the broad education profile as reported in the tables. There are three main reasons why this can occur.

First, the categories in the education profile are quite broad. The categories are based on each individual's highest completed qualification and those qualifications have been grouped into 'high', 'medium' and 'low' levels. High level education comprises Bachelor Degree and above. Medium represents Year 12, Certificate III/IV, Diploma and Advanced Diploma. Low contains Certificate I/II, Certificate not defined, Year 11 and below. Therefore, some individuals can change their highest completed qualification — say, from Diploma to Advanced Diploma — without changing the broad education profile.

Second, not all qualifications undertaken are completed during this ten-year period. Some individuals discontinue study for a qualification part way through (due to 'module completion' in some vocational education and training courses, or 'dropout') and some individuals who enrol later in the period continue to study beyond 2010 and their completion status is not observed (that is, these data are censored).

Third, some individuals may complete an additional qualification at or below the level of their highest qualification.

For example, although 77 per cent of youths in the *Churning with Work* pathway reported any study activity during the ten years, only a third of those individuals contributed to the increase in the broad education level for that pathway. Another 19 per cent of those who had studied increased the level of their highest qualification but remained at the same broad education level. The study activity of the remaining individuals (48 per cent) represented study at the same level as their highest qualification, at a lower level, or study towards a higher level qualification that was either abandoned before — or ongoing at — mid-2010.

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

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 3.5). 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.

Compared to the *Work and Study to Work* pathway, a larger share of individuals spend time unemployed and/or NILF, and these experiences are also more likely to be repeated and to last longer.

Table 3.4 **Selected characteristics of individuals in the *Churning with Work* pathway for youths**

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 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 3.5 **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 3.7 below). In 2001 individuals in this pathway are aged 25–39 years with an average age of 32.9 years (table 3.6 below).[6] 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.[7] 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 3.4 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 parents were women (not shown).

In 2001, education levels in this pathway reflect the average for all young adults (table 3.6). Education levels are higher in this pathway than in the *Prolonged NILF*

6 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 shown).

7 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 3.4 above).

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 10 years (figure 3.4 below).

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 3.7).

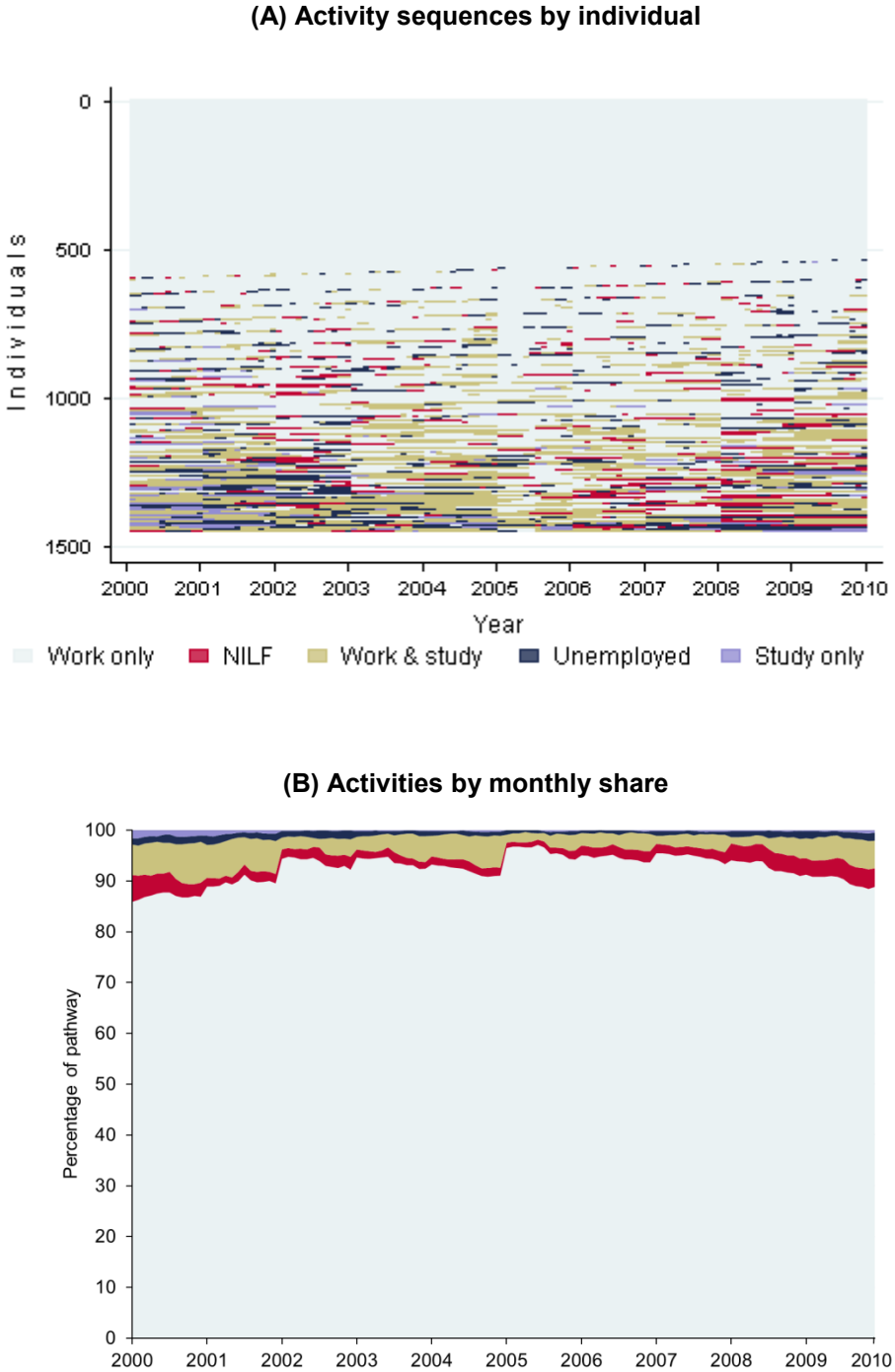
Broad education levels in the *Work* pathway do not change much over the ten years despite 41 per cent of individuals studying at some point, mostly while working. This is because 75 per cent of the individuals in this pathway who studied either completed a qualification at the same or lower level than their highest qualification, or began a higher level qualification that was either still being undertaken in 2010 or was abandoned earlier (see box 3.5 above). 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 3.4).[8] 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 3.7). This is much less than the two pathways characterised by significant NILF activity, where most young adults

[8] 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 3.5 below). In the *Prolonged NILF* pathway, 4–8 per cent of young adults are unemployed and 60–90 per cent are NILF (figure C.5).

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

Figure 3.4 **Activities in the *Work* pathway for young adults**



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

Table 3.6 **Selected characteristics of individuals in the *Work* pathway for young adults**

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 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 3.7 **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.

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 3.9 below). The average age in 2001 is 32.3 years (table 3.8 below). Women are predominant (81 per cent) and the pathway is characterised mostly by a progression from NILF to work (figure 3.5 below). 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 3.8).

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 3.8). 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 3.9). The NILF rate has reduced to 11 per cent (table 3.8) 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 3.8).

Although 37 per cent of individuals begin in work only (figure 3.5), by 2010 nearly all have has spells of work only, averaging 2.95 spells of 22 months per spell (table 3.9). In 2010, 85 per cent are employed (mostly without study) (figure 3.5). 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).

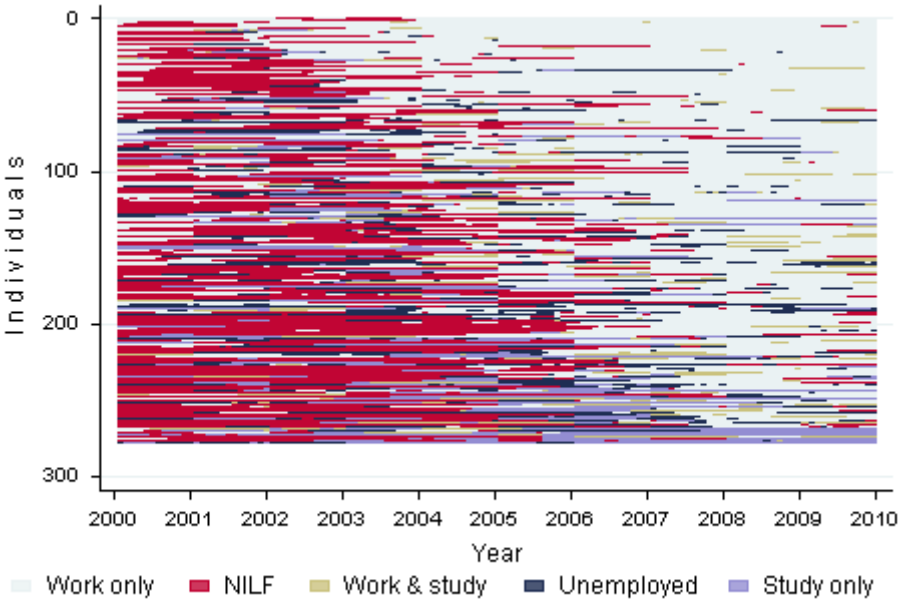
Unemployment is relatively common among the individuals in this pathway. In 2001, this pathway has the highest unemployment of young adults (9.3 per cent) (table 3.8). Over the next ten years, 61 per cent of individuals in this pathway average 2.4 spells of unemployment lasting seven months per spell (table 3.9).

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 3.8). About 60 per cent of individuals undertake some study (with or without work) during the ten-year period (not shown);[9] 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 3.9). The study activity leads to a relatively small increase in education levels (table 3.8). 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 (see box 3.5 above).

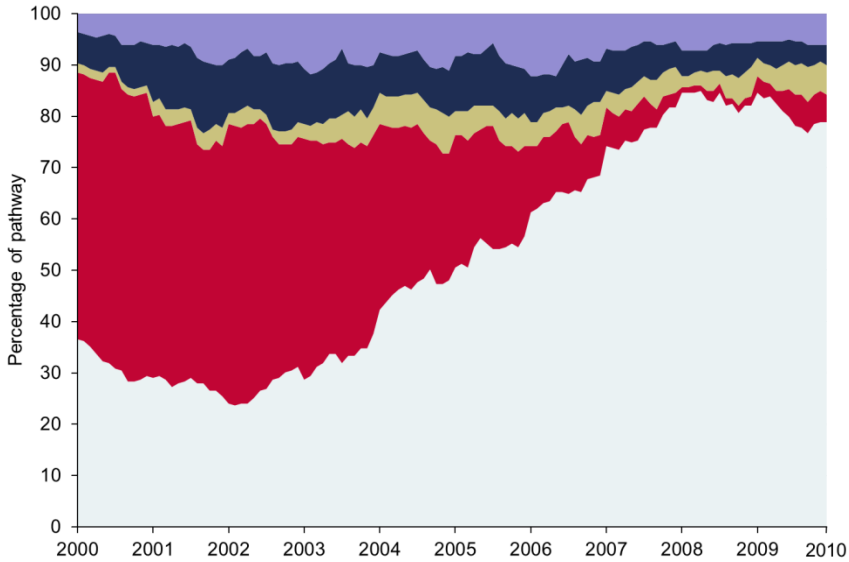
9 About 27 per cent of young adults in this pathway study with and without work.

Figure 3.5 **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 3.8 **Selected characteristics of individuals in the *NILF to Work* pathway for young adults**

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 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 3.9 **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 3.11 below). The *Work to NILF* pathway is the 'oldest' pathway for mature adults, with an average age of 49.1 years in 2001 (table 3.10 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 3.6 below).

In 2001, over 90 per cent of individuals are employed (figure 3.6). 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 3.10).

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 3.11). 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 3.6). 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.[10]

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 3.10).

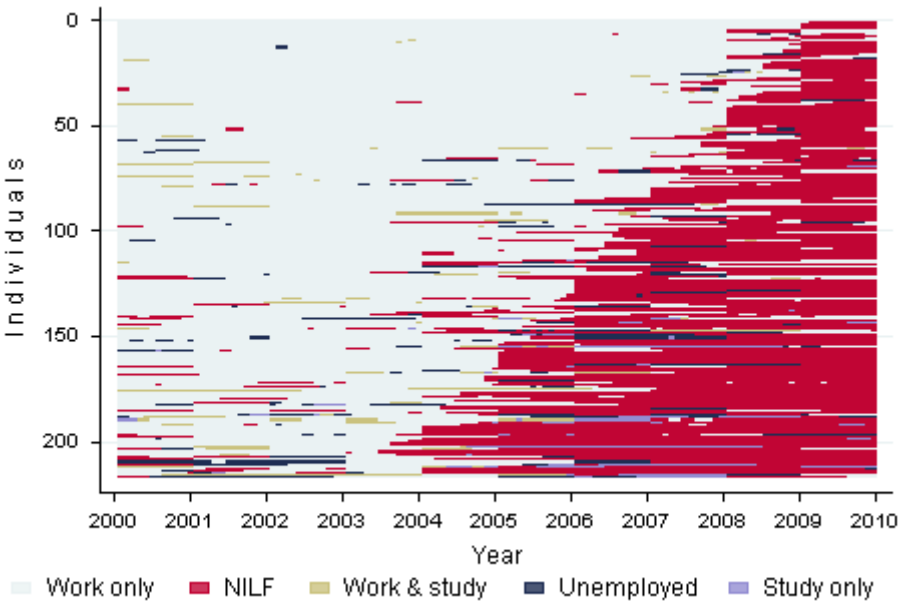
Spells of other activities are infrequent and tend to last about six months on average. The limited incidence and short duration of study means that education levels remain quite low. Most higher level qualifications require significantly more than six months to complete and are unlikely to be undertaken by individuals in a later stage of their life cycle.[11]

[10] 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 (appendix A).

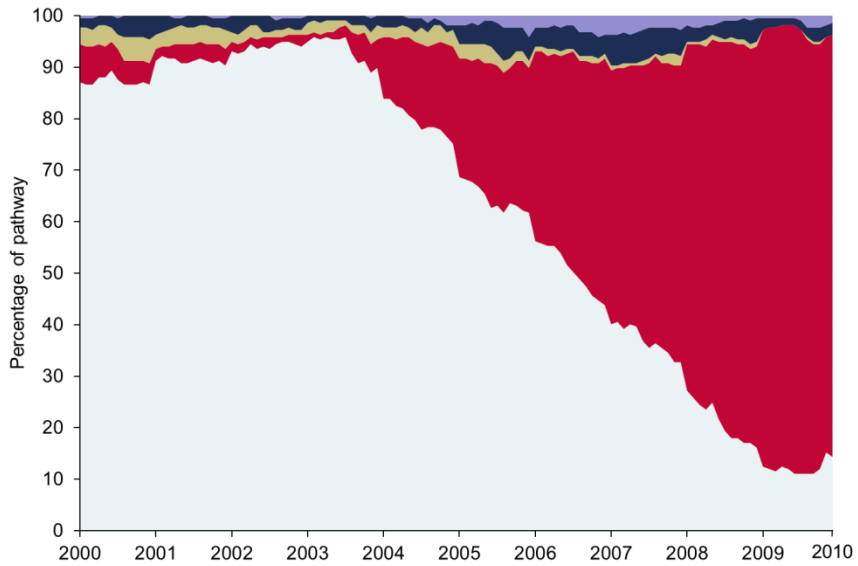
[11] However, of the 65 individuals in this pathway who studied, 6 spent two to four years continuously in study (with or without work) and acquired higher qualifications.

Figure 3.6 **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 3.10 **Selected characteristics of individuals in the *Work to NILF* pathway for mature adults**

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 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 3.11 **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.

Seniors in the *Work* pathway

Among seniors aged 55–64 in 2001, the *Work* pathway represents 21.5 per cent of individuals (table 3.13 below). The members of this pathway tend to be younger than those in the other senior pathways: 75 per cent are aged less than 60 years, whereas in the other senior pathways less than 50 per cent are aged less than 60 years (not shown). The *Work* pathway contains predominantly men (58 per cent) (table 3.12 below).

Unlike the other senior pathways, where activities are strongly associated with the transition to retirement and retirement itself, this pathway is dominated by work (figure 3.7 below). All individuals spend some time in work only and average 1.88 spells lasting 58 months (almost 5 years) per spell (table 3.13). More than one third of the individuals spend the entire ten years in work only (figure 3.7). Combining work with study is also common — one quarter of individuals average 1.55 spells of 10.4 months (table 3.13).

Education levels are higher in this pathway compared to the other senior pathways. About one quarter of seniors in the *Work* pathway has high education levels in 2001, and one third has medium-level education (table 3.12). Although 24 per cent of individuals combine study and work, there is no change in broad education levels by 2010, suggesting the study activity represents education at the same or lower levels, or part-qualification completions.

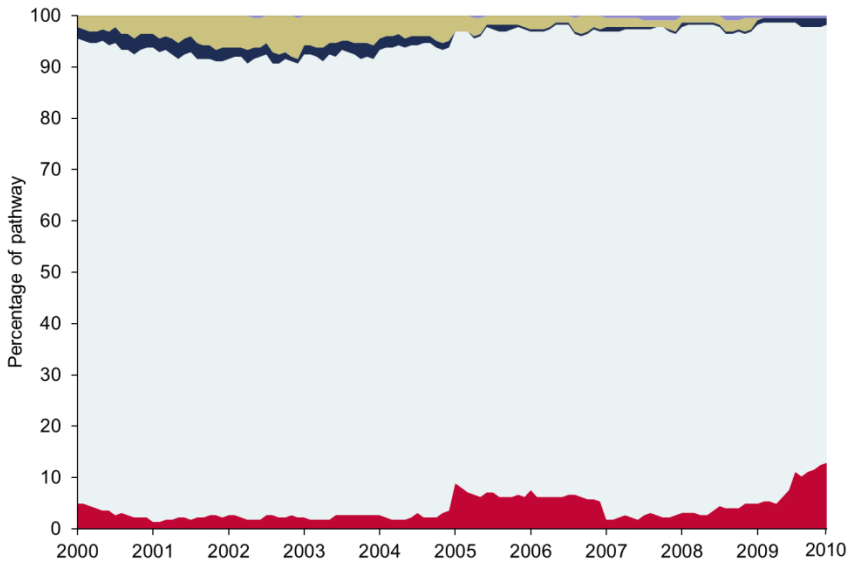
Few seniors experience unemployment, possibly because they tend to withdraw from the labour force upon leaving employment. However, for those individuals in this pathway who do become unemployed, the experience tends to be protracted (nine months) and recurrent (1.55 spells on average) (table 3.13).

This pathway displays the lowest rate of NILF among seniors in 2010, as most individuals in this pathway are yet to retire. In 2001, 4 per cent are NILF (table 3.12). During the ten-year period, only 37 per cent spend time NILF, averaging 1.75 spells of 7.5 months each (table 3.13). Monthly activity shares show a temporary, but small increase in NILF from July 2005 to June 2007 (figure 3.7). In 2010, 16 per cent of individuals are NILF and this is mostly due to the few individuals in this pathway who retire (table 3.12).

Figure 3.7 **Activities in the *Work* pathway for seniors**

(A) Activity sequences by individual

(B) Activities by monthly share



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

Table 3.12 **Selected characteristics of individuals in the *Work* pathway for seniors**

Characteristic	*Measure*	*W pathway*	*All senior pathways*
In 2001:			
age	years, average	57.8	59.2
gender	% female	41.78	53.25
locality (remoteness)	% major city	58.22	56.69
highest level of education[a]	% high	24.89	15.68
	% medium	37.33	34.9
	% low	37.77	49.42
unemployed	% U	2.22	2.20
NILF (incl. marginally attached)	% NILF	3.56	50.00
of which:			
home duties/childcare	% of NILF	25.00	17.78
study	% of NILF	0.00	0.76
marginally attached to labour force	% of NILF	12.50	0.38
other reasons[b]	% of NILF	62.50	81.07
In 2010:			
locality (remoteness)	% major city	56.00	53.92
highest level of education[a]	% high	25.33	15.87
	% medium	38.22	35.66
	% low	36.44	48.27
unemployed	% U	0.89	0.19
NILF (incl. marginally attached)	% NILF	16.44	80.31
of which:			
home duties/childcare	% of NILF	0.00	7.26
study	% of NILF	0.00	0.24
marginally attached to labour force	% of NILF	2.70	0.12
other reasons[b]	% of NILF	97.31	92.38

[a] Educational attainment has been classified as high (Bachelor 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 3.13 **Selected characteristics of activity patterns in the *Work* pathway (seniors)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	91.07	3.29	0.16	4.03	1.45
Share of path with at least one spell of the activity	100.00	24.44	3.56	36.89	12.89
Conditional on at least one spell of the activity:					
average number of spells	1.88	1.55	1.00	1.75	1.55
average length of spell (months)	58.00	10.44	5.50	7.50	8.71

[a] Pathway size 225 (21.5 per cent of seniors); average number of activities 1.78.

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

4 Future directions

This paper provides an analysis of transitions in Australia by describing the broad patterns (or pathways) in the monthly education and labour market activities of individuals over a ten-year period. Specifically, the paper used optimal matching and cluster analysis (OMCA) — two analytical techniques for summarising the different sequences of activities of individuals. These techniques advance the understanding of transitions in Australia. For example, a more informative characterisation of the factors that are associated with marginal attachment to the labour force is provided.

This chapter summarises the analytical approach used in this study and its results (section 4.1), outlines some additional analysis that could build on these results (section 4.2), and suggests other topical areas that may benefit from the application of OMCA techniques (section 4.3).

4.1 Summary

This study used the education and labour market calendar data from ten waves of the *Household, Income and Labour Dynamics in Australia* (HILDA) Survey to follow the sequences of activities undertaken by about 6500 individuals.

As has been demonstrated in this paper, the 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. Indeed, the usefulness of its design should inform the developers of other longitudinal surveys.

With so many individuals, activities and observations, it can be difficult to know where to begin an analysis. However, OMCA is ideal for exploring this type of (sequence) data. OMCA works by comparing individual activity sequences and forming groups of individuals based on the similarity of their activity patterns. Each group then represents a pathway. This process distils this complex information to a meaningful level and helps to focus the analysis.

Used in tandem, these two techniques are useful for summarising information encoded in sequences. This type of analysis can be very useful for policy makers — as an input to program evaluation and as a source of evidence for policy design.

In this study, four age-based segments were defined for individuals of working age: youths, young adults, mature adults and seniors. For each segment, defined according to age in 2001, OMCA identified four or five pathways based on similarities in individuals' activity patterns. Characteristics of individuals in each pathway were then explored.

For youths aged 15–24 in 2001, five pathways were identified: three were associated with education and the transition to work, and one was dominated by women's withdrawal from the labour force to raise children. One pathway was associated with work, although that pathway also showed extensive 'churn' as most individuals interspersed other activities with spells of work. Frequent churning into and out of the labour force can lead to tenuous labour force attachment. However, for some individuals, stable employment eventually results.

The activity sequences for young adults aged 25–39 were grouped into four pathways: two paths involved work (one with increasing education) and two were closely associated with women spending prolonged periods outside the labour force raising children (with one pathway showing their subsequent return to work).

Four pathways were identified for mature adults aged 40–54: work dominated one pathway, two others were driven by women spending time outside the labour force as they raised their children (and one of those pathways showed their return to work), and one pathway was associated with early retirement.

For seniors aged 55–64, four pathways were identified: one pathway was dominated by work and the other three were associated with retirement or transitions to retirement.

The tabulations and summary statistics presented in this paper provide some insight into transition pathways in Australia. However, further investigation and multivariate analysis is needed to disentangle the relationships between the characteristics of individuals and the pathways they follow.

4.2 Additional analysis

There remains considerable scope for additional analysis to build on the results in this paper.

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, an extension of the analysis in this paper could be to use the information from wave 11 of HILDA to represent 'outcomes', and then to examine econometrically the association between the 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'.

Another fruitful extension of the existing analysis is a more detailed investigation of study activity. This could be accomplished by tracking different types of education — for example school, vocational and higher education.[1]

4.3 Further potential uses of OMCA in areas of policy interest

The OMCA techniques used in this paper could be applied to many other topics, subject to the availability of suitable data — for example, understanding patterns in public housing tenure, housing arrangements or types of welfare payment.

Pathways relating to public housing tenancy could be investigated using the time of survey data in HILDA. Although limited to annual observations, these data could be sufficiently frequent for analysis as tenure in public housing tends to be much longer than one year (AIHW 2012).

In a similar vein, another data source is the accommodation 'calendar' in *Journeys Home* — a new longitudinal survey of factors affecting housing stability (for

[1] Some studies track apprenticeships separately from other types of study undertaken while working. Unfortunately, apprenticeships (and traineeships) cannot be identified reliably in HILDA.

details, see Wooden et al. 2012). These data could be used to characterise the housing or accommodation pathways of individuals over time according to patterns in housing situations, incorporating stable/secure housing through to various categories of homelessness. Applying OMCA to these data could provide insights into patterns of housing (in)stability that have not emerged in other analyses.

Administrative systems — such as those used to track welfare receipts by individuals — represent yet another source of data to which OMCA can be applied. If sufficient information is available, the analysis will reveal distinct patterns. This would provide a better understanding of whether (and which types of) welfare can become a trap. Such an analysis would be a valuable input into program evaluations.

The OMCA techniques could also be applied to the same topic of transitions using a different dataset, such as the *Longitudinal Surveys of Australian Youth* (LSAY). LSAY contains a calendar of labour market activity for a single age cohort of youths. In contrast to this study — in which the youth segment includes a mix of individuals who are initially aged 15–24 years and followed over time — LSAY begins with a group of individuals who are all 15 years old and tracks their activities until they are all about 25 years old. For examining youth transitions, an advantage of LSAY, compared to HILDA, is that the individuals are about the same age so that 'cohort effects' and age effects are not confounded. Despite these differences between the two surveys, the results (for the youth segment) could still be compared.

Whether sourced from HILDA or elsewhere, calendar-style data are particularly valuable for understanding individuals' transitions in the context of their longer-term experiences. OMCA techniques exploit the richness of such data to reveal overall similarities and differences in detailed patterns of activities.

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 in Stata are included.

A HILDA and the calendar data

The aim of this study is to gain a better understanding of the dynamic processes of transitions between labour market activities and education. This paper considers the transitions of youths as well as adults and seniors. The adult category is split between young adults (25–39 years old) and mature adults (40–54 years old), as it is expected that these two groups behave differently. Seniors (aged 55–64 years old) are included to analyse the pathways towards retirement.

Household, Income and Labour Dynamics in Australia (HILDA) is the only comprehensive longitudinal survey of Australians that covers all age groups of interest in this study. Ten years of data are used (from mid-2000 to mid-2010).[1] Using HILDA's calendar data it is possible to determine activity patterns within a year. Data collected at the time of the survey (for example, data on individuals' characteristics) can be matched to the calendar data to find possible explanations for the patterns.

This appendix describes the HILDA data used for the analysis. Section A.1 describes the calendar data, section A.2 describes related time of survey data, and section A.3 presents some caveats on the results.

A.1 Calendar data

The HILDA calendar is a rich source of information about labour market and education activities. The calendar has more detailed information about what has happened between annual interviews compared to the annual survey data. For example, questions are asked about any period of unemployment as part of the annual survey, but spells of unemployment between interviews are only captured in the calendar. Both sources of information are subject to recall error.

[1] Although wave 1 was conducted in late 2001, the calendar data collected at that time covered the period from mid-2000 until the time of the interview in 2001.

Figure A.1 HILDA education and labour market calendar, wave 1[a]

FG. CALENDAR

FG1 I now want you to think about the employment, job search and education related activities you have undertaken since 1st July 2000.

NOTE: THERE NEEDS TO BE A MARK AGAINST EVERY TIME PERIOD IN THE CALENDAR PRIOR TO THE DATE OF THE INTERVIEW.

		01 02 03	04 05 06	07 08 09	10 11 12	13 14 15	16 17 18	19 20 21	22 23 24	25 26 27	28 29 30	31 32 33	34 35 36	37 38 39	40 41 42	43 44 45	46 47 48	49 50 51
		2000						2001										
		Jul	Aug	Sept	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov
	Enrolled in school / educational course:																	
01	on FT basis																	
02	on PT basis																	
	Employed																	
03	Job..... FT / PT																	
04	Job..... FT / PT																	
05	Job..... FT / PT																	
06	Job..... FT / PT																	
07	Job..... FT / PT																	
08	Job..... FT / PT																	
09	Job..... FT / PT																	
10	Job..... FT / PT																	
11	Job..... FT / PT																	
12	Job..... FT / PT																	
13	Job..... FT / PT																	
14	Job..... FT / PT																	
15	**Not employed BUT looking for work**																	
16	**Neither employed nor looking for work**																	
		01 02 03	04 05 06	07 08 09	10 11 12	13 14 15	16 17 18	19 20 21	22 23 24	25 26 27	28 29 30	31 32 33	34 35 36	37 38 39	40 41 42	43 44 45	46 47 48	49 50 51

Interviewer Notes

(i) For each of the activities listed, put a cross in the part of the month that the activity began and finished, and then connect the crosses with a straight line. For activities that began before July 2000, put a cross in the very first box (July 2000). If an activity is going to continue after the interview date put a cross in last box prior to the interview date.

(ii) Only record data for the period up to, but not including, the month in which the interview takes place.

(iii) The months across the top are divided into three sections; (1) early in the month; (2) middle of the month; and (3) late in the month.

(iv) Number each job, and for each job, circle the appropriate code to indicate whether it was full-time (FT) (ie 35+ hrs/week) or part-time (PT) (ie <35 hrs/week). When entering any job check for any changes in hours during the period, and if they did switch from FT to PT or visa-versa, have two lines for that job number – one covering FT, the other PT.

(v) Holidays should not be treated as breaks in their usual activity (that is, work and study).

(vi) Job spells can be overlapping. Similarly, enrolled in school / education can overlap with all other activities.

(vii) Educational course or study refers to school or courses that lead to a qualification.

(viii) If a respondent is an apprentice and is unsure if they are studying as well as working, code as both.

[a] The design of the calendar was modified between waves 1 and 2. In wave 1, the calendar distinguishes between part- and full-time jobs but not in subsequent waves (Summerfield et al. 2011). In all waves except wave 1, the calendar asks respondents with more than one job to record the month and year they started each job.

Source: Reproduced, with permission of the Melbourne Institute of Applied Economic and Social Research and the Department of Families, Housing, Community Services and Indigenous Affairs, from the HILDA Survey Person Questionnaire, Wave 1 (http://www.melbourneinstitute.com/downloads/hilda/Questionnaires/PersonQuestionaireW1.pdf).

At the time of the annual interview, respondents record their status for the early, middle and latter parts of each month in the HILDA calendar. They record whether they are in full-time study, part-time study, employment (up to twelve jobs that may overlap), not employed but looking for work (in this study, this information measures unemployment) and neither employed nor looking for work (the measure of not in the labour force, NILF) (figure A.1).

Respondents are instructed to place a mark against every month third prior to the annual interview. Item non-response in the calendar would therefore be apparent.

Removing overlaps in the calendar data

For each wave, calendar data are collected back to the start of the previous financial year. This creates overlaps in the calendar data from one wave to the next. Without allowing for overlaps there would be gaps between the calendars because an individual is not necessarily interviewed in the same month each year.[2] In any given wave, interviews are conducted across a number of months, beginning in August and finishing as late as April the following year for some waves.[3]

For example, in wave 1, interviews began 24 August 2001 and finished 23 January 2002. For individuals interviewed in August for this wave, there are nearly 14 months of calendar data (collected in month thirds), but for individuals interviewed in January, there are nearly 19 months.

It is important to remove the overlaps from the data before they are merged into a single dataset. The calendar data for the financial year ending in the year of the survey are used to produce the dataset without overlaps (section A.3).

Figure A.2 illustrates how this is done for two hypothetical individuals in the first two periods of overlapping data. For the first individual, the wave 1 and wave 2 data overlap between July 2001 and September 2001; for the second individual the overlap extends for a further four months to January 2002. Irrespective of the length of overlap, the second wave is the source for the data for July 2001 to June 2002.

Similarly, the first individual is interviewed for wave 2 earlier than the second individual, but the calendar data used for July 2002 to June 2003 are from wave 3 for both individuals.

---

2 Another reason that there is at least two months of overlapping data in the calendar is that it allows consistency of information to be assessed (section A.3).

3 Interviews in the year following the survey year were infrequent across the ten waves, representing 4.3 per cent of all interviews at most (Summerfield et al. 2011).

Figure A.2 **An illustration of overlaps in the HILDA calendar[a]**

Waves 1 to 3

	Individual 1				Individual 2			
	Wave 1	Wave 2	Wave 3	…	Wave 1	Wave 2	Wave 3	…
Jul-00	X				X			
Aug-00	X				X			
Sep-00	X				X			
Oct-00	X				X			
Nov-00	X				X			
Dec-00	X				X			
Jan-01	X				X			
Feb-01	X				X			
Mar-01	X				X			
Apr-01	X				X			
May-01	X				X			
Jun-01	X				X			
Jul-01	X	X			X	X		
Aug-01	X	X			X	X		
Sep-01	X (interview)	X			X	X		
Oct-01		X			X	X		
Nov-01		X			X	X		
Dec-01		X			X	X		
Jan-02		X			X (interview)	X		
Feb-02		X				X		
Mar-02		X				X		
Apr-02		X				X		
May-02		X				X		
Jun-02		X				X		
Jul-02		X	X			X	X	
Aug-02		X	X			X	X	
Sep-02		X	X			X	X	
Oct-02		X (interview)	X			X	X	
Nov-02			X			X	X	
Dec-02			X			X	X	
Jan-03			X			X	X	
Feb-03			X			X (interview)	X	
Mar-03			X				X	
Apr-03			X				X	
May-03			X				X	
Jun-03			X				X	
.			.				.	
.			.				.	
.			.				.	

[a] X indicates a calendar entry corresponding to a particular month for a particular wave. Entries in red text indicate overlaps. Entries with green backgrounds indicate those selected for analysis.

The definition of education and labour market activities

The calendar in HILDA is unlike the calendars in other longitudinal surveys.[4] In the *British Household Panel Survey* (BHPS) and the *European Community Household Panel* (ECHP), respondents indicate their main activity, but HILDA's respondents indicate their labour market and education activities in separate parts of the calendar. The advantage of the approach used in HILDA is that individuals can record when multiple activities overlap (Watson 2009). Of particular interest are concurrent work and study activities.

Youths may combine work and study while at school or further education and during the transition to full time employment. Many individuals who are older than 25 years of age have completed their secondary and tertiary education, although some return to education at some point.

Work, study, and (concurrent) work and study categories are included in the activity variable to analyse the different ways individuals develop and use their human capital at different stages of life. The activity variable also includes values for 'unemployment' and 'NILF'. This results in a single activity variable with five mutually exclusive categories. This follows the approach taken by Albert Verdú and Davia (2010).[5]

While it is possible to disaggregate some of these categories further (for example, into full-time and part-time education), this would add significantly to the complexity of the data and the computational task for the subsequent analysis.

Deriving a monthly activity variable

For this analysis, the original calendar data are aggregated to monthly data. Aggregating the data retains most of its richness, and reduces it to a level that is practical for implementing optimal matching and cluster analysis techniques (appendix B).

In producing the monthly activity variable, three rules are applied to months that involve more than a single activity for each of the thirds (figure A.3). These rules define which activity is dominant in any month. They are often required for research based on labour market data (Quintini and Manfredi 2009).

---

4 For a comprehensive comparison of the HILDA survey with other longitudinal surveys see Watson (2009).

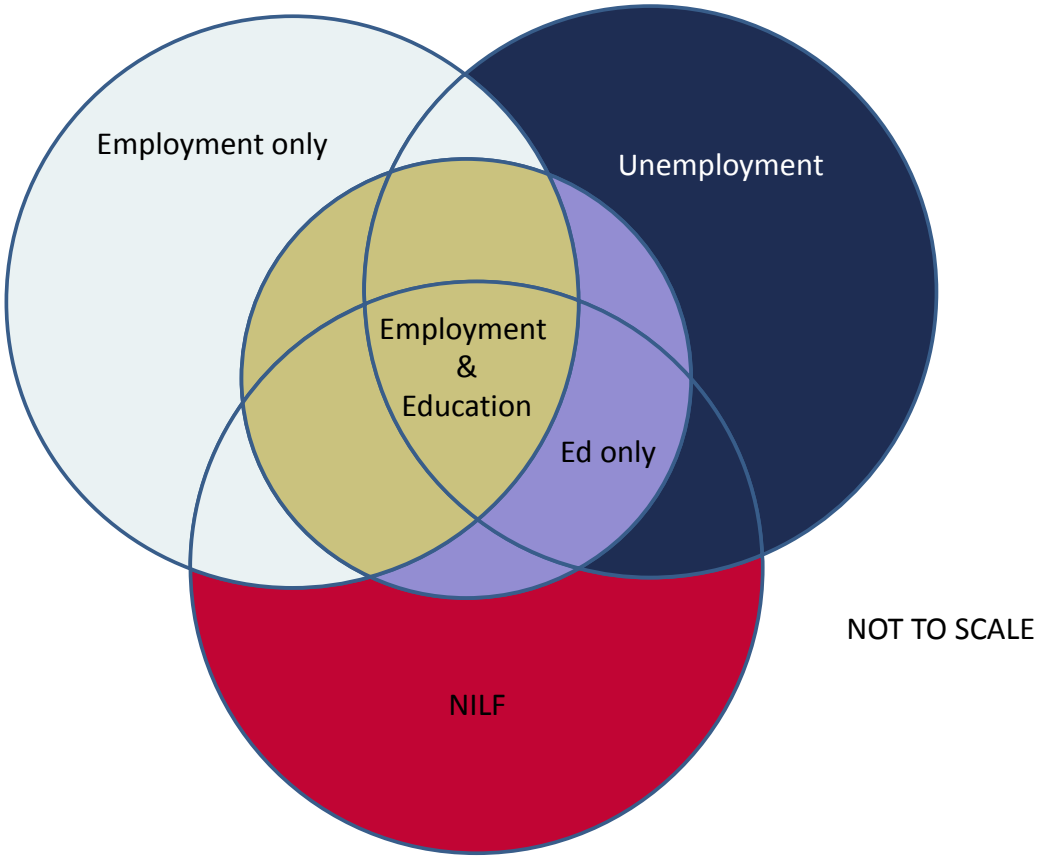
5 That paper uses 2005 data from a longitudinal survey of Spanish school leavers, called the education, training and labour market transition survey (ETEFIL).

Rule 1: Employment and study activities dominate unemployment and NILF

If any third of the month is spent in education only, then the activity for that month is defined as 'education only'. If another month third is spent in education and employment or in employment only, the activity for the month is defined as 'employment and education'.

Education activities are represented by the circle in the middle of figure A.3. Education overlapping with NILF or unemployment indicates that part of the month's activity was recorded as NILF or unemployment, but the activity for the whole month is categorised as 'education only' (purple). When education overlaps with employment, this is categorised as 'employment and education' (brown), regardless of whether one month third is spent in NILF or unemployment.

Figure A.3 **Aggregating activities over time — defining mutually exclusive and exhaustive activities for HILDA[a]**



[a] Based on the HILDA calendar, the labour market state of unemployment is defined as not being employed but looking for work.

Rule 2: Employment dominates unemployment or NILF

For months that do not involve any education, a month third spent in employment dominates any unemployment or NILF in the same month, and the value 'employment only' is allocated to the activity variable.

Rule 3: Unemployment dominates NILF

If part of the month is spent in unemployment and another in NILF, the month is allocated to unemployment. NILF is only recorded as the activity for a month if the whole month is spent outside the labour force.

The use of these rules has implications for the data. The data may indicate that employment is more stable than it actually is because it dominates other labour market states in the aggregation process. If unemployment or NILF lasted less than a month, spells of employment may appear longer than they actually are.

Similarly, the length of education spells could be overestimated. This could result in an underestimate of the occurrence of spells of unemployment and NILF lasting less than a month. It could also reduce the spell length of NILF and unemployment activities lasting more than a month.

A consequence of employment dominating unemployment is that unemployment rates could be underestimated and employment rates could be overestimated. With employment and unemployment dominating NILF, participation rates could be overestimated.[6]

Forming an activity calendar of 120 months

The dataset comprises individuals with full calendars only. That is, individuals must have participated in the survey each year, and be original HILDA sample members. The dataset therefore excludes those who did not respond to the survey in one or more years and individuals who were too young to be interviewed at the time of the first survey (including those who turned 15 years of age within the ten years). It also excludes new sample members (see Summerfield et al. (2011) for more information about the different sample member types in HILDA).

6 Activities using monthly and month-third data for each age segment were compared. Relative to monthly data, month-third data show: less study only and more unemployment for youths; more unemployment and less employment only for young adults and mature adults; and very little difference in the activities for seniors as most of their time is spent in employment (without study) and/or NILF.

As a result of these exclusions, for this paper, the original sample of 19 914 individuals who were enumerated in the first survey is reduced to a sub-sample of 6566 fully-responding individuals of working age in 2001 (this is discussed further in section A.3).

A.2 Time of survey data

The calendar data are used to identify activities over time and in grouping individuals with similar patterns of activities. Other information about individual characteristics, drawn from other parts of the annual survey, is labelled time of survey (TOS) data. TOS data serve several purposes in this paper:

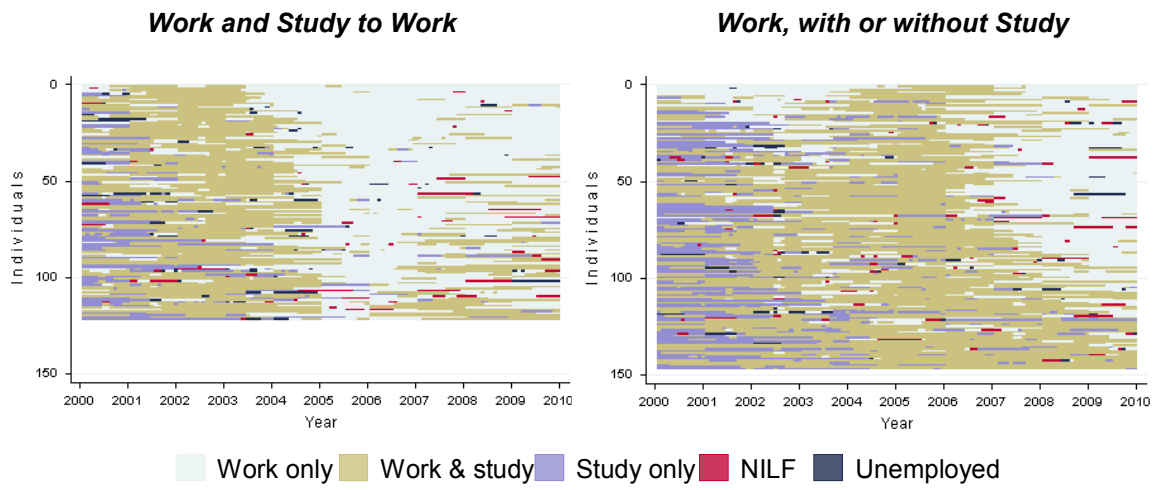
- as additional information about the activities (for example, occupation, level of study, full-time or part-time status for work and for study)
- to infer a reason for these activities (for example, illness/disability, caring responsibilities)
- to better understand the links between pathways and individuals' characteristics (for example, gender, location, and family background).

To illustrate the use of TOS data, consider the two youth pathways shown in figure A.4 — the *Work and Study to Work* pathway (left panel) and the *Work, with or without Study* pathway (right panel). The former involves more time in work only (light blue), and less time in work and study (brown) and study only (purple) than the latter.

TOS data reveal that youths in the pathway involving more time in study (right panel) are (on average) about a year younger and in 2001 have lower levels of education than youths in the other pathway (left panel). This suggests that youths in the *Work and Study to Work* pathway are not as advanced in their transitions to work as youths in the *Work, with or without Study* pathway.

In 2001, nearly 69 per cent of individuals in the *Work, with or without Study* pathway had a low level of education, compared to about 39 per cent of individuals in the *Work and Study to Work* pathway. By 2010, the percentage of individuals in these pathways with a low level of education was negligible — about 1 per cent in the *Work, with or without Study* pathway, and 2 per cent in the *Work and Study to Work* pathway. Relative to other youths, high percentages of youths in these pathways had high levels of education in 2010 — about 58 per cent in the *Work, with or without Study* pathway, and 51 per cent in the *Work and Study to Work* pathway.

Figure A.4 **A comparison of the *Work and Study to Work* and the *Work, with or without Study* pathways for youths[a]**



[a] For larger versions of the sequence index plots, see chapter 3 for the *Work and Study to Work* pathway and appendix C for the *Work, with or without Study* pathway.

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

A.3 Some caveats

Several caveats relate to the wave 1 data and the way in which the database is constructed for this analysis. The caveats relate to various potential sources of bias, which are discussed in this section.[7]

- The sample may not be representative of the population because of unequal probabilities of selection across individuals, resulting in sample selection bias.
- Longitudinal surveys are often subject to attrition bias, which occurs if the people who drop out of the sample after the first wave are a non-random selection of individuals from across the population.
- Recall bias is especially of concern for the calendar data in HILDA, since they are collected annually for a period exceeding a year.
- Two other potential sources of bias relate to: having a finite observation period and using time of survey data to make inferences about activities outside of the survey month.

[7] This appendix presents the potential sources of bias in the HILDA dataset that are most relevant to the analysis in the paper. See Watson and Wooden (2004) for a discussion of other potential sources of bias in the data, such as longitudinal inconsistencies in marital status, and transcription and data entry errors.

Sample selection bias

The reference population at the start of the HILDA survey in 2001, which was intended to be representative of the population, was primarily all Australians who were living in private dwellings. The survey did not (and does not) include, for example, individuals in hospitals or prisons as these are non-private dwellings. Individuals who live in remote areas were (and are) also not represented in the survey. See Wooden and Watson (2001) for further information about the reference population for HILDA.

Children aged 14 or younger are included in the sample (enumerated), but are only interviewed (that is, become respondents) once they turn 15 years of age. Data used in this paper do not include individuals who turned 15 after 30 June 2001.

Despite a low response rate in wave 1 — 39 per cent of individuals did not respond to the first survey[8] — Watson and Wooden (2012, p. 374) argue that the first wave of weighted data matched the 'broader population quite well'. Exceptions included Sydney residents and immigrants with a non-English speaking background (Wooden, Freidin and Watson 2002).

Attrition bias

According to Watson and Wooden (2012) the loss of sample members because of death or moving overseas (termed 'natural attrition') is not likely to cause the HILDA dataset to be unrepresentative of the Australian population. The authors state that other reasons for dropping out of the survey are more likely to contribute to attrition bias.[9]

Of the 13 969 people who were interviewed in the first wave, 7460 were also interviewed in waves 2 to 10, inclusive. Excluding the individuals lost through natural attrition, 87 per cent of those people interviewed in the first wave were re-interviewed in wave 2. The annual re-interview rate (which excludes natural attrition) rose to about 96 per cent by wave 10 (Watson and Wooden 2012).

The highest rate of attrition was from wave 1 to wave 2. After excluding 'natural attrition' resulting from death or moving overseas, the adjusted attrition rate was 13.2 per cent. This is comparable to the BHPS with an attrition rate of 12.4 per cent (Watson and Wooden 2004).

---

8 Those individuals are not included in the wave 1 sample of 19 914 (enumerated) individuals.

9 See Watson and Wooden (2004; 2006; 2012) for further information about attrition and the HILDA dataset.

Watson and Wooden (2004) compared characteristics of the individuals who dropped out of the sample (for reasons other than death or relocation overseas) to those who remained in the sample. Compared to individuals who participated in the first and second waves, individuals who dropped out of the survey before the second wave because they could not be contacted (often because they had changed address), or because they refused to participate further, were more likely to:

- live in Sydney
- be 15–24 years of age
- be single or in a de facto relationship
- be an Aboriginal or Torres Strait Islander, or born in a non-English speaking country
- be living in a flat, unit or apartment
- have a low level of education
- be unemployed.

Individuals with these characteristics will be under-represented in the sub-sample used in this paper, and the results presented may not reflect accurately the education and labour market activities of the underlying population. Henstridge observes that:

> It is likely that factors that lead to attrition — family disruption, employment changes and relocations — are variables of significant interest in the survey creating a situation where persons of greatest interest may well be the most difficult to collect longitudinal data for. (2001, p. 13)

The under-representation of unemployed people deserves further discussion. Although a relatively small proportion of the individuals in wave 1 were unemployed (about 4 per cent of the sample after excluding those who died or moved overseas between waves 1 and 2), nearly 19 per cent of these individuals dropped out of the sample before wave 2. This is high relative to those employed (full-time or part-time) or NILF, who have attrition rates for wave 2 at roughly 13 per cent (Watson and Wooden 2004).

The same authors modelled attrition based on pooled data from the first four waves. They found that individuals who had moved house were less likely to be contacted for the next survey. Homeowners were more likely than renters to be contacted, as were those who lived in a house rather than a flat or unit (Watson and Wooden 2006). Conditional on making contact, they found that the young and the elderly were less likely to respond to the survey, and that individuals with a Diploma or higher level of education were more likely to respond than someone whose highest level of education attained was Year 11 or below. Individuals who

work full-time were easier to contact, but less likely to respond compared to those who were NILF.

While finding that some characteristics were significant determinants of attrition in the HILDA survey, Watson and Wooden (2006, p. 9) also found that the overall explanatory power of their models of survey response was low, and concluded that, 'this is a desired outcome and presumably reflects the large random component in survey non-response'.

Use of weights to correct for attrition and sample selection bias

Weights are included in the HILDA time of survey data to account for unequal probabilities of selection in the initial sample and unequal probabilities of remaining in the sample. In other words, they are intended to counter the possible influence of sample selection bias and of attrition bias on statistical estimates based on the data. They are designed to re-align the sample at each wave to meet known population benchmarks, for example by age and gender subgroups. However, if there are unobserved characteristics that are linked to attrition, the weights will not account fully for the effects of attrition on the results of any data analysis (Watson and Wooden 2004).[10]

The results presented in this paper are unweighted because the techniques used do not permit the use of weights. Although weights could be applied to the results, they would not account for the fact that the majority of the nearly 14 000 individuals who responded to the first survey are excluded from the analysis.[11]

To select the sub-sample for analysis in this paper, individuals aged 65 and older in 2001 are excluded, accounting for about 2000 observations. About 5400 individuals who drop out of the HILDA sample after wave 1 are also removed. As a result of the way the HILDA calendar was designed, there are no gaps in the calendar data for the 6566 individuals who are left.[12]

Dropping almost 5400 observations at the second step introduces a source of bias that is not accounted for in the weights supplied with the HILDA data.

[10] For more detailed information about the weights in HILDA see Watson and Wooden (2012).

[11] In the wave 1 sample of 19 914 individuals, 13 969 were interviewed (that is, they were respondents). Of the remaining individuals who were enumerated but did not respond, the vast majority (about 4700) were younger than 15 on 30 June 2001, and were therefore too young to be interviewed.

[12] Instead of having missing data for month thirds where individuals were uncertain of their activity at a point in time, the respondent and interviewer used 'their "best guess" of what happened' (Watson 2009, p. 8).

Summerfield et al. (2011) report analysis of the 7460 individuals (aged 15 and older) who were re-interviewed in every wave up to and including wave 10. They indicate that 70 per cent of Indigenous people and 69 per cent of non-Indigenous people who were interviewed in wave 1 were interviewed in wave 10. However, only 45 per cent of Indigenous people were also interviewed in every intervening wave, which is low compared to 60 per cent of non-Indigenous people (Summerfield et al. 2011).

Applying person weights from the first wave for responding persons to the sub-sample used in the analysis gives an indication of whether it is representative of the population at the time of the first survey. The person weights are designed to sum to the sample size. Individuals with characteristics that are under-represented in the population are assigned a weight above 1; individuals with characteristics that are over-represented in the population are given a weight below 1. Therefore, if the sum of the weights is equal to 6566, it can be concluded that dropping individuals because they are non-responding or out of scope in one or more of the waves, or are over 64 years of age, will not bias the results of this analysis.

The sum of the weights is equal to 6250, roughly 300 (about 4.5 per cent) less than the sub-sample size.[13] This implies that the sub-sample includes more individuals who are over-represented, and fewer individuals who are under-represented, relative to the HILDA wave 1 sample. A comparison of the weighted and unweighted sample by age segment indicates that the under-represented are more likely to be in the youth and senior age groups than the young and mature adult age groups (not shown).

Benchmarking

In order to examine the effects of sample selection and attrition on the sub-sample used for the analysis more closely, key characteristics of the sub-sample are compared to those of the (broader) HILDA sample and to ABS data (table A.1). The HILDA statistics are based on wave 1 data, and the ABS data are matched as closely as possible to the time of the 2001 HILDA survey. Two sources of ABS data are used for this purpose. The age distributions are benchmarked against ABS historical population statistics and the labour force statistics against ABS *Labour Force Survey* (LFS) data.

[13] Twenty-four individuals are lost as a result of weighting because they have a weight of zero. This is because the benchmarks used in the weighting process were changed after wave 1 to be consistent with a change in the ABS definition of very remote areas (Watson 2012).

Table A.1 **A comparison of sub-sample with the full HILDA sample and ABS data, 2001**

Age distribution and labour force characteristics, unweighted[a]

	HILDA sub-sample 15–64 years	*HILDA broader sample 15–64 years*	*ABS*	
				Reference period (population)
Age distribution:[b]				
Youths (15–24 years)	13.4	18.9	20.4	30 June 2001 (15–64 years)
Young adults (25–39 years)	34.9	34.7	33.6	
Mature adults (40–54 years)	35.9	32.5	31.9	
Seniors (55–64 years)	15.9	14.0	14.1	
Labour force characteristics:				
Employment to population ratio	0.716	0.699	0.692	Oct 2001 (15–64 years)
Participation rate (%)	75.7	75.0	74.3	
Unemployment rate (%)	5.4	6.8	6.8	
Mean duration of unemployment (weeks)	35.8	31.4	48.6	Oct 2001 (15+ years)
Distribution of job duration:[c]				
Less than 1 year	18.3	20.9	22.9	Year ending Feb 2002 (15–69 years)
1 to under 5 years	34.5	35.8	36.2	
5 to under 10 years	19.0	17.6	17.0	
10 years or longer	28.3	25.7	24.0	

[a] Percentages may not add to 100 due to rounding errors. [b] Percentage of the working-age population (15–64 years) in each age segment. [c] Percentage of employed individuals in each job duration category.

Sources: ABS 2001 (*Labour Force*, Cat. no. 6203); ABS 2002 (*Labour Mobility*, Cat. no. 6209.0); ABS 2008 (*Australian Historical Population Statistics*, Cat. no. 3105.0); Authors' estimates based on HILDA waves 1–10; Watson and Wooden (2002).

In terms of the age distributions, the weighted HILDA sample (not shown) is very similar to the ABS data. However, the age distribution of the sub-sample, whether it is weighted (not shown) or unweighted, is different to the ABS data. The key difference is that there is a smaller proportion of youths (7 percentage points less) in the unweighted sub-sample compared to the ABS data. Weighting the sub-sample data brings it only part way towards this benchmark.

Relative to the ABS data, the sub-sample indicates:

- lower unemployment and slightly higher participation rates
- slightly larger employment to population ratio
- shorter average unemployment duration
- higher percentages of the working-age population who have been in their current job for five years or more relative to individuals with shorter tenure.

The employment to population ratios and the unemployment rates based on the (unweighted) HILDA sample are very close to the ABS statistics. The difference between the sub-sample and the ABS statistics may therefore be attributable to the exclusion of youths who are lost from HILDA after wave 1.

Other aspects of the HILDA sample do not reflect the ABS statistics as closely. The selection of individuals in the original HILDA sample has some flow-on effects for the sub-sample. The average duration of unemployment is shorter in HILDA and the sub-sample than the ABS statistics indicate. This divergence may be a result of the selection of the HILDA sample, leading to fewer people with longer unemployment durations in the samples.[14] Among the employed, tenure in current job appears longer in the HILDA sample and sub-sample than the ABS data — there is a higher proportion of individuals who have spent five years or longer in their current job. This is likely due to sample selection in HILDA, and weighting the figures (not shown) does not bring the sub-sample result much closer to the ABS statistics.

Watson and Wooden (2002, p. 24) also found average unemployment duration in the wave 1 HILDA sample to be below the ABS benchmark, but found other estimates (including job tenure) generally did meet their respective benchmarks:

> For the most part … the HILDA Survey data are generating estimates in line with ABS sources … estimates of key labour market indicators, such as the employment–population ratio and the unemployment rate, derived from the HILDA Survey data are quite close to ABS estimates from October 2001. Perhaps the most noticeable difference concerns unemployment duration …

With regard to the HILDA sub-sample used in this analysis, some of the statistics are fairly close to the ABS benchmarks, but there are others that are not close. As with the full HILDA wave 1 sample, the biggest difference between the two sets of statistics is for unemployment duration. Spells of unemployment should be longer than they appear to be in the HILDA sub-sample data.

The result of the benchmarking exercise indicates that sample selection and attrition may bias the results of this analysis.

The benchmarking also confirms that the weights would not re-align the sub-sample to the initial HILDA sample. (In the case of average unemployment duration, the weighted average is actually further away from the ABS average than the unweighted average is.) Not weighting the results is consistent with the approach taken by many studies in this area including Yu et al. (2012, p. 14), who argue that

[14] Comparing unweighted HILDA sample statistics with ABS statistics for the same age group (15 years and older) shows even larger gaps in average unemployment duration (not shown).

using unweighted data where there is potential for biased results is only an issue '*if* we seek to generalise our analysis to those of a representative population'.

Recall bias

As explained in section A.1, when individuals are surveyed, they are asked to recall their labour market status and enrolment in education during the previous 14–18 months. The further back someone is asked to recall, the less accurate their memory may be.

Studies have considered the evidence of inaccurate recall on calendar data (box A.1).

Box A.1 **Evidence of recall bias in longitudinal surveys**

Quintini and Manfredi (2009) analysed calendar and survey data from different sources and found vastly different measures of the average time taken by Spanish youths to find a job after leaving education. The calendar data indicated that, on average, Spanish youths take two years to find their first job after finishing education, whereas the survey data indicated that it takes six months.

The authors suggested that youths may be able to estimate how long the transition to the labour market took them for a survey, but find it difficult to give detailed information concerning the timing of each part-time or casual job they had prior to finishing their education and beginning their career for a calendar.

Another potential explanation is that they do not view this information as important as these jobs may be irrelevant to their career path. Alternatively, they are not motivated to remember or to provide this information.

In a study of transitions based on the British Household Panel Survey, Malo and Muñoz-Bullón (2003) indicated that recall bias would affect the accuracy of their results in theory, but also argued that previous research on this panel indicated that recall errors were usually random, with the exception of short-duration events, such as unemployment.

Watson (2009) includes a more extensive review of the types of recall errors that may lead to unreliable (biased) results, the reasons why they occur and factors that have been found to be linked to recall errors.

Recall errors in the HILDA calendar data

Overlaps in HILDA's calendar allow identification and analysis of differences in what individuals recall. An individual's activity spells are assumed to be recalled

perfectly if they can be matched in the overlapping periods exactly. In the first six overlapping periods in the HILDA calendar, exact matches were found for:

- 83 per cent of job spells
- 85 per cent of NILF spells
- 24 per cent of unemployment spells
- 66 per cent of full-time education spells and 40 per cent of part-time education spells (Watson 2009).

Spells that were not perfectly matched could either be: forgotten (not recorded in the later wave), remembered (not recorded in the earlier wave), or mistimed (beginnings and/or endings did not match in the overlaps). Most commonly, inconsistencies arose because an individual reported a spell in the first calendar, but 'forgot' in the second calendar. This implies that individuals are less likely to remember an activity spell as the time since that spell increases.

As mentioned earlier, the HILDA calendar was designed to avoid any gaps in the data. If respondents were uncertain of their activity at a point in time, they made a calculated guess with the help of their interviewer. This design feature may have contributed to inconsistencies in overlapping calendar data (Watson 2009).

Removing overlaps in the HILDA calendar data

The approach taken to remove overlapping calendar data by using data from 1 July in each wave uses data that is recalled back to the furthest point in time from the time of interview (see figure A.2). As discussed in Watson (2009), there are other methods to remove overlaps in the calendar data that would result in using data that are closer to the time of interview, and therefore less affected by recall errors. However, these approaches are considerably more difficult to implement as they depend on interview dates, and these dates differ across individuals as well as survey years.

The dataset used in this paper may therefore underestimate the number of spells (and overestimate spell duration) at the start of each financial year in particular. Spells of unemployment and part-time education are more likely to be biased as a result of recall errors during these months. For example, if an individual forgot to record a spell of unemployment that occurred more than a year earlier (during the overlap period), the number of spells in the corresponding sequence is underestimated and the length of an adjoining spell is overestimated.

Factors linked to recall errors in the HILDA calendar data

The probability of making different types of recall errors was found to be linked to spell length, the complexity of patterns of activities and to the characteristics of the individual (box A.2).

As these factors vary by pathway, these findings could have implications for the results of this study. First, pathways characterised by a complex sequence of labour market activities with spells of shorter duration are more likely to be affected by recall errors than pathways that are more stable. This implies that some of the less complicated pathways (for example the *Prolonged NILF* pathways for adults, and the *NILF*, *Early Work to NILF* and *Later Work to NILF* pathways for seniors), are less likely to be affected by recall errors than other pathways.

Adult and senior pathways that include spells of education are more likely to be affected by recall errors than youth pathways because education is a less common activity for the older age groups.

Pathways with a higher proportion of individuals with a high level of education (diploma or above) are less likely to be affected by errors recalling spells of full-time education than pathways with a higher proportion of individuals with relatively less education (Year 11 or below).

For youths and young adults, *Prolonged NILF* pathways are followed by individuals who have relatively lower levels of education compared to other pathways for their respective age segments. These pathways are therefore more likely to be affected by errors recalling spells of education. The *Churning with Work* pathway for youths also includes a small percentage of individuals with high levels of education in 2010 relative to the other three youth pathways. For young adults, this also applies to the *NILF to Work* pathway.

For mature adults, the *Prolonged NILF* and *Work to NILF* pathways are more likely to have errors in education spells, and for seniors the *NILF* and *Early Work to NILF* pathways are less likely to reflect full-time education spells accurately.

Box A.2 **Overlaps and recall errors in HILDA calendar data**

Watson (2009) studied patterns of recall errors — forgetting, remembering and mistiming of activity spells — in the overlapping sections of the HILDA calendar. In that study, the most important determinant of recall errors in labour force status was whether the individual was in the same activity at the time of the interview and the time being recalled. For example, individuals who were employed at the time of the interview had a mean predicted probability of forgetting a job spell of nearly 8 per cent, whereas individuals who were unemployed or NILF at the time of interview had a higher mean predicted probability of forgetting a job spell (about 30 per cent).

In general, recall errors for spells of employment, unemployment and NILF, were more likely the shorter the spell duration and the more complex the pattern of activity. However, unemployment and NILF spells were less likely to be forgotten when activity patterns were more complicated.

Errors in recalling shorter education spells were more likely than for spells of longer duration. However, recalling more complex patterns of activities was linked to a lower probability of forgetting or remembering a spell of education, but a higher probability of mistiming a spell of education.

Individuals with higher levels of education were less likely to forget or remember spells of education.

Individuals were more likely to make errors in reporting spells of any activity that were less common among people of their age group. For example, people aged 55 or over — who tend to be retired — reported spells of NILF more accurately than they reported spells of unemployment or employment. Similarly, people aged 15–24 years — for whom study is common — were less likely than older people to make mistakes recalling spells of (full-time) education.

Individuals tended to make consistent errors in forgetting or remembering spells across the waves.

It is important to consider that Watson's study considered recall errors within the overlaps. Recall errors are less likely in the non-overlapping months between surveys because there is less time between the survey and the period being recalled.

Source: Watson (2009).

Other potential sources of bias

The dataset used in this study includes ten years of data. It is not possible to see how long an individual has been in an activity prior to the start of the period, and likewise, it is not possible to know how long an individual will remain in an activity beyond the observation period. The data are, therefore, left and right censored, and this will tend to shorten the average duration of the activities affected, as the duration measures are not adjusted for censoring.

Another potential source of bias in the results of this analysis is the use of time of survey data to impute possible reasons why individuals are in particular activities between surveys. This may not be reasonable for individuals who have had changes in activities between surveys. However, it is unlikely that many of the characteristics of interest (the number of children, for example), are likely to be significantly out of step between two waves.

Similarly, it would be unreliable to impute annual income based on income reported at the time of the survey for someone who frequently cycles in and out of employment. On the other hand, for individuals who are in steady employment, it could be reasonable to impute income data in this way.

What are the possible impacts of bias?

The above discussion has revealed that there are some sources of bias that are likely to affect the results of this analysis. Table A.2 summarises these impacts where they differ by pathway within an age segment. It therefore does not show the impact of having a sub-sample with relatively too few youths and seniors compared to the HILDA sample in wave 1. In addition, it does not attempt to capture the impacts of imputing data based on TOS data, or the effect of censoring. TOS data do not affect the pathways observed, but do affect the explanations of the patterns of activities in the different pathways.

Censoring does not affect the size of the pathways, or the appearance of the pathways for the ten years considered. It affects the descriptive analysis of the pathways by censoring spells at the start and the end of the observation period. For pathways that are characterised by spells of short duration, this does not have a significant impact on average spell lengths. However, it is something to bear in mind when considering longer spells at the start or the end of the ten-year period, and especially those that span the whole ten years. This will mainly affect individuals in the *Work* pathways, and the *Prolonged NILF* and *NILF* pathways.

Table A.2 **Summary of potential impacts of bias on pathway characteristics**

By pathway

	Source of bias			
Pathway	*Data aggregation rules*[a]	*Sample selection*[b]	*Attrition*[c]	*Recall*[d]
Education to Work (Youths)	+Unemp/NILF & -study		✓	
Work and Study to Work (Youths)	+Unemp/NILF & -study		✓	
Churning with Work (Youths)				✓ more spells
Work, with or without Study				
Youths	+Unemp/NILF & -study			
Young adults	+Unemp/NILF & -study			
Work				
Young adults	+Unemp/NILF & -employed			✓ more spells
Mature adults	+Unemp/NILF & -employed			✓ more spells
Seniors	+Unemp/NILF & -employed			
Prolonged NILF				
Youths	+Unemp & -NILF	✓	✓	✓ study errors
Young adults	+Unemp & -NILF	✓	✓	✓ study errors
Mature adults	+Unemp & -NILF	✓	✓	✓ study errors
NILF (Seniors)	+Unemp & -NILF	✓	✓	✓ study errors
NILF to Work				
Young adults	+Unemp & -NILF	✓	✓	✓ more spells ✓ study errors
Mature adults	+Unemp & -NILF	✓	✓	✓ more spells
Work to NILF (Mature adults)	+Unemp & -employed		✓	✓ study errors
Early Work to NILF (Seniors)	+Unemp & -NILF			✓ study errors
Later Work to NILF (Seniors)	+Unemp & -employed			

[a] +/- indicates there should be more/less of this activity in the pathway. [b] A tick in this column indicates that the pathway should be bigger because it includes individuals with activity patterns that are assumed to be representative of people who are under-represented in the initial HILDA sample (people with long periods of unemployment), and people who are out of scope of the survey because they live in a non-private dwelling. It was not possible to determine the potential impact of selection bias due to non-coverage of remote areas. [c] A tick in this column indicates that the pathway should be bigger because it includes individuals who are more likely to leave the survey after the first wave. [d] A tick in this column indicates that the patterns of activities in the pathway are likely to have been biased by recall errors. People with low education are more likely to make errors recalling education spells. People with more complicated activity patterns are more likely to forget short spells so that their activities should contain more spells of shorter duration. For the relevant work pathways (including *Churning with Work*), this does not relate to the sub-group of individuals in work throughout the ten years.

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

Data aggregation rules

The rules applied to aggregate the month third to monthly data have some impacts on the results of this analysis. Short spells of less than a month in unemployment or NILF will not appear in the monthly data if some time in the month is spent in employment. This suggests that the pathways characterised by long spells of

employment — the *Work*, *Work to NILF* and *Later Work to NILF* pathways — would not show some short spells of unemployment or NILF, and the average duration of employment spells should be shorter.

Likewise, pathways characterised by significant time spent in study (with or without work) might overestimate the time spent in study, and underestimate the number of short spells of unemployment or NILF. This relates to the *Education to Work*, *Work and Study to Work*, and the *Work, with or without Study* pathways.

The *Education to Work* pathway for youths may reflect some short spells of work and study that do not represent concurrent work and study. For example, where an individual transitions from study only to work only within a month, this will be categorised as work and study. It is unlikely that this will impact significantly on spell length for education or work.

A consequence of unemployment dominating NILF is that the pathways characterised by long spells of NILF — mainly the *Prolonged NILF* and *NILF* pathways, but also the *NILF to Work*, *Work to NILF* and *Early Work to NILF* and *Later Work to NILF* pathways — reflect fewer short spells of unemployment, and longer spells of NILF than they would with the original month-third data.

Sample selection bias

As discussed above, sample selection bias may have some impact on the results presented in this paper. It is difficult to ascertain how the exclusion of individuals who live in non-private dwellings may impact the results of this analysis. Possibly, individuals living in non-private dwellings are doing so because they are in ill-health, have a disability or are in prison. If it is valid to assume that individuals live in non-private dwellings for these reasons, activity patterns are likely to involve long periods of NILF, at least in the early 2000s. This would suggest that these patterns should be more prevalent than they appear to be in the results presented in this paper.

The benchmarking exercise indicated that the HILDA wave 1 sample may not include enough individuals who have long periods of unemployment. This may have resulted in pathways with more unemployment appearing less prevalent than they should. Analysis of the HILDA sub-sample indicates that the pathways with more periods in unemployment are also those with long periods of NILF.

Attrition

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 drop out of the sample relative to pathways that involve a significant proportion of time working. This implies that these non-work pathways would otherwise be more prevalent than they appear to be.

Recall errors

Recall errors may have different impacts on the results, possibly changing the appearance of some of the plots. The plots that are dominated by NILF are not likely to reflect education spells accurately. Whether the differences are significant enough to cause bias in aggregate is difficult to determine as there may be some balancing out of dropped (forgotten) or added (remembered) spells, and of shortened or lengthened spells, for the pathway as a whole.

While table A.2 indicates that the results for all pathways are potentially affected by bias, the main concern would be that the non-work pathways are smaller than they should be. Other impacts are likely to be smaller, and some may cancel each other out in aggregate.

B Optimal matching and cluster analysis

In the social sciences, sequence analysis is frequently conducted using optimal matching (OM) to measure the similarity of individual sequences of activities and cluster analysis (CA) to identify groups of similar sequences. Together the techniques are referred to as OMCA.

OMCA is a tool to describe sequence data. It does not impose any assumptions about the data generating process (Martin and Wiggins 2011). According to Quintini and Manfredi (2009, p. 15), OMCA 'allows us to explore a dataset and discover (or confirm) some underlying patterns without any priors based on economic theory'.

Uncovering the same amount of information in the data using more typical descriptive statistics is difficult, if possible at all (Corrales-Herrero and Rodríguez-Prado 2012). The longer the sequences, the greater the number of possible activities and the larger the dataset, the greater the possible number of unique sequences. Analysis becomes too difficult, particularly where there are similarities in the order of activities across time, but differences in timing and persistence (Brzinsky-Fay 2007).

OMCA reduces the heterogeneity in the data to a level where patterns can be detected and sub-groups of individuals with similar sequences can be identified (Anyadike-Danes and McVicar 2010).

This appendix provides an overview of how the techniques work (sections B.1 and B.2) and their limitations (section B.3).[1,2]

1 Although there are alternative methods for sequence analysis, these are beyond the scope of this paper and will not be covered.

2 This paper uses Stata's SQ routines developed by Brzinsky-Fay, Kohler and Luniak (2006) for analysing sequence data. The routines do not support the use of survey weights.

B.1 Optimal matching

OM is used to compare sequences of events, activities or states across time (box B.1). The use of OM is a relatively recent development in the labour market transitions literature, compared to other fields such as biology.

Box B.1 **Origins and applications of OM**

OM originated in the field of computer science in the mid-1960s (Lesnard 2010). It was later adopted in other areas of science (for example, biologists used OM to compare deoxy-ribonucleic acid (DNA) sequences) (Brzinsky-Fay 2007). In the mid-1980s, the technique was first applied in the social sciences by Abbott and Forrest (Brzinsky-Fay 2007). By combining OM with cluster analysis, common patterns among sequences are more easily identified. A few examples are patterns of:

- types of housing (house, unit) and housing states (buying, renting)
- mental health patients' service use and hospitalisation
- steps in traditional dances
- sounds in bird calls (Martin and Wiggins 2011).

Although OM is the predominant approach to measuring dissimilarity between sequences, Martin and Wiggins indicated that the use of OM and cluster analysis have yet to meet widespread acceptance in the social science literature:

> OM … is not a brand new method any longer, but is still sufficiently maverick that most authors feel obliged to introduce its technicalities in some detail in every publication (rather as if one were to explain the idea of minimizing squared errors every time one published an OLS regression analysis). (2011, p. 386)

Over the past decade or more, the OM technique (and OMCA) has increasingly been used to better understand different career pathways and transitions in the labour market (Martin and Wiggins 2011), with the focus to date largely being on youths.

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 pair of sequences in the dataset is to each other, 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 transform it to be identical to another sequence.

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. Since an insertion in one sequence is equivalent to a deletion in the other, the two operations are grouped together and called 'indels'.

Each type of operation is given a 'cost', and the distance is the sum of these costs (box B.2).[3] 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 of the operations required to transform one sequence into another.

Box B.2 **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. 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.

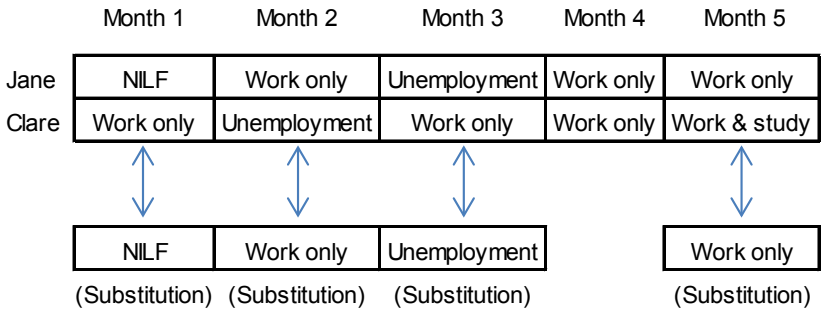
(Continued next page)

[3] The costs of substitutions can be allowed to differ according to the pair of activities concerned (discussed below).

Box B.2 (continued)

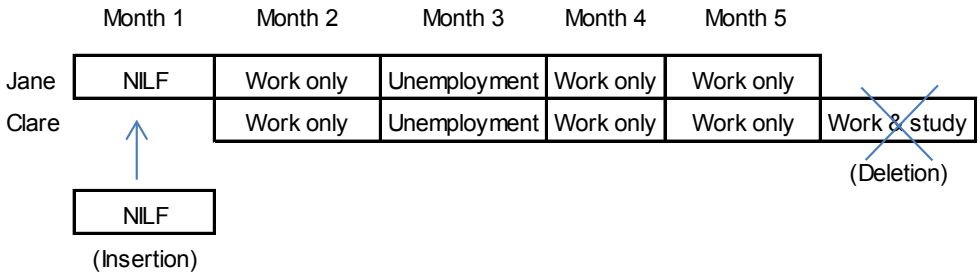
To keep things simple, no preference is given to a particular type of change. Each indel operation 'costs' 1 and each substitution 'costs' 2 (as a substitution is equivalent to an insertion at one point in a sequence and a deletion at the next point along the same sequence). So regardless of whether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

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 sequences for the other 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 with five months of work and a sequence with five months of NILF.

Source: Adapted from Quintini and Manfredi (2009).

Setting substitution and indel costs

The cost of substituting one activity for another must be specified before the OM stage of sequence analysis can be performed (Brzinsky-Fay and Kohler 2010). 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'.

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.[4] The other two approaches require substitution costs to be specified as a square matrix, in which the elements give the cost of substituting each type of activity for every other type. By using a matrix, the costs can differ according to the pair of activities concerned. However, as each cost represents a distance between two activities, the matrix must be symmetric — that is, the distance from activity A to activity B must be the same as that from activity B to activity A.[5]

In the theoretically-based approach, substitution costs are set in a subjective manner to show (for each activity) which of the activities ought to be 'closer' or 'more distant' according to theory (box B.3). 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 a more common choice in the literature and is used to set substitution costs for this analysis.

A second type of cost that must be chosen is the cost of indels. The setting of indel costs is rarely discussed in the literature (Albert Verdú and Davia 2010). As a substitution is equivalent to two indel operations (an insertion and a deletion), the relative magnitudes of the two types of cost can affect which types of operation are used to determine the distance between two sequences. Therefore, the researcher needs to decide in advance the relative importance of timing (indels) compared to similarity of sections of sequences (substitutions). For example, is it more important that two people have a spell of employment in the same period, or is it more important that they have the same sequence of employment, unemployment and

4 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.

5 Elements corresponding to the same two activities (the 'diagonal') will contain zeros to indicate that the distance between an activity and itself is zero.

NILF activities? The substitution costs used in this paper are all less than 2, and hence less than double the cost of an indel transformation, which is set to 1.

Although there is no agreement on how costs should be set, there does appear to be consensus on the importance of conducting sensitivity analysis with different costs. Several papers have considered the impact of different substitution costs on the results of OM (for example, Anyadike-Danes and McVicar 2003). Many papers reported that the results were robust to changing substitution costs (Corrales-Herrero and Rodríguez-Prado 2012). Some papers (including Brzinsky-Fay 2007) also tested the impact of varying the cost of indels, but found only slight differences in the results.

Box B.3 **Setting substitution costs on a theoretical basis**

Theory may suggest that the strength of attachment to the labour market can guide the setting of substitution costs. For example, Anyadike-Danes and McVicar (2010) set the cost of substituting between employment and unemployment to be lower than that between employment and NILF. However, theory does not indicate how 'close' or 'distant' any two activities should be, and so this decision is subjective (Hollister 2009).

The subjective nature of theoretically-based substitution cost matrices is reflected in the transitions literature. There are large differences in substitution costs, in both relative and absolute terms.

Many substitution costs used in the literature were not comparable to the substitution costs used in this paper. There were two reasons for this. First, the literature often relates to youths only, and this paper considers additional age groups, including three older groups that are likely to experience different transitions. Second, the activities in this analysis are different from those in most other papers. In particular, there are two activities in one category — work and study — so that the distance between work and study and either unemployment or NILF should be greater than the distance between work and study and either work only or study only.

For each age group, the substitution cost matrix has been determined using the probability of a transition from one activity to another in the dataset. That is, the substitution cost is inversely related to the average probability of transitions between activities A and B.[6] The substitution costs are the lowest for pairs of activities that have the greatest number of transitions between them. Box B.4 shows the substitution cost matrix for youths and compares it to the matrix for seniors.

6 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, averaged over time. Other data-driven options are to use the minimum or maximum probability for a substitution pair.

Box B.4 **Substitution cost matrices for youths and seniors**

The data-driven substitution cost matrices for youths and seniors are shown below. The diagonal elements are zero, as there is no change of activity. The off-diagonal elements represent the cost of substituting one activity with another. Therefore, the matrices are symmetric.

The costs are inversely related to the probability of a transition, and are less than (but very close to) 2 because the probability that an individual will transition to a different activity in any month is low. What is important to note in the tables is the relative differences in costs.

For youths, activities study only and work and study are relatively 'close'. By comparison, study only is 'further away' from work only. This is because there are more transitions in the data between study only and work and study than between study only and work only.

For seniors, study only is closer to NILF than any other activity. But the closest pair of activities is work only and NILF, since transitions from work to NILF are frequent in the data relative to any other type of transition, due to retirement.

Table 1: Youths

	Study only	Work & study	Work only	Unemployment	NILF
Study only	0.00000	1.98942	1.99820	1.99769	1.99759
Work & study	1.98942	0.00000	1.98142	1.99954	1.99964
Work only	1.99820	1.98142	0.00000	1.98981	1.99203
Unemployment	1.99769	1.99954	1.98981	0.00000	1.99761
NILF	1.99759	1.99964	1.99203	1.99761	0.00000

Table 2: Seniors

	Study only	Work & study	Work only	Unemployment	NILF
Study only	0.00000	1.99981	1.99995	1.99975	1.99890
Work & study	1.99981	0.00000	1.99793	1.99998	1.99994
Work only	1.99995	1.99793	0.00000	1.99823	1.98896
Unemployment	1.99975	1.99998	1.99823	0.00000	1.99871
NILF	1.99890	1.99994	1.98896	1.99871	0.00000

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

The substitution and indel costs used for this paper imply that greater importance is placed on the similarity of sections of sequences than the timing of activities. This is reasonable because the age segments used in this analysis include people of different ages and, therefore, different stages of life. In contrast, if a cohort of

individuals of the same age (as in the *Longitudinal Surveys of Australian Youth* (LSAY)) had been used, then they would be likely to make transitions at similar times, and so in that case, greater consideration would be given to the timing of activities.

After setting substitution and indel costs, for each pair of sequences the OM technique then determines the minimum total cost of the transformation. That total cost is then the 'distance' between those two sequences. Pairs of very different sequences have relatively larger distances than pairs of similar sequences (Corrales-Herrero and Rodríguez-Prado 2012). These distances are collected into a matrix and can be used to group similar sequences together.

B.2 Cluster analysis

Cluster analysis includes a variety of techniques that aim to identify groups in data. In the context of sequence analysis, cluster analysis identifies groups of individuals with similar activity sequences, based on the distances derived through OM.

In this paper, cluster analysis is hierarchical and agglomerative: it starts with all individuals in their own cluster (or group) and, in an iterative process, progressively combines the clusters until all individuals in the sample are in a single cluster.[7] Diagnostic information is then used to determine the point in this process at which the appropriate number of clusters is formed.

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.[8]

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

7 Instead of a hierarchical approach, the individuals could be partitioned into a pre-determined number of clusters.

8 There are several alternatives available, including single linkage, complete linkage, average linkage, centroid method and density linkage (see Lattin, Carroll and Green (2003) for further details).

clustering process the 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 tree diagram called a dendrogram (box B.5). 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, which is a matter of judgment.

Box B.5 **Interpreting dendrograms**

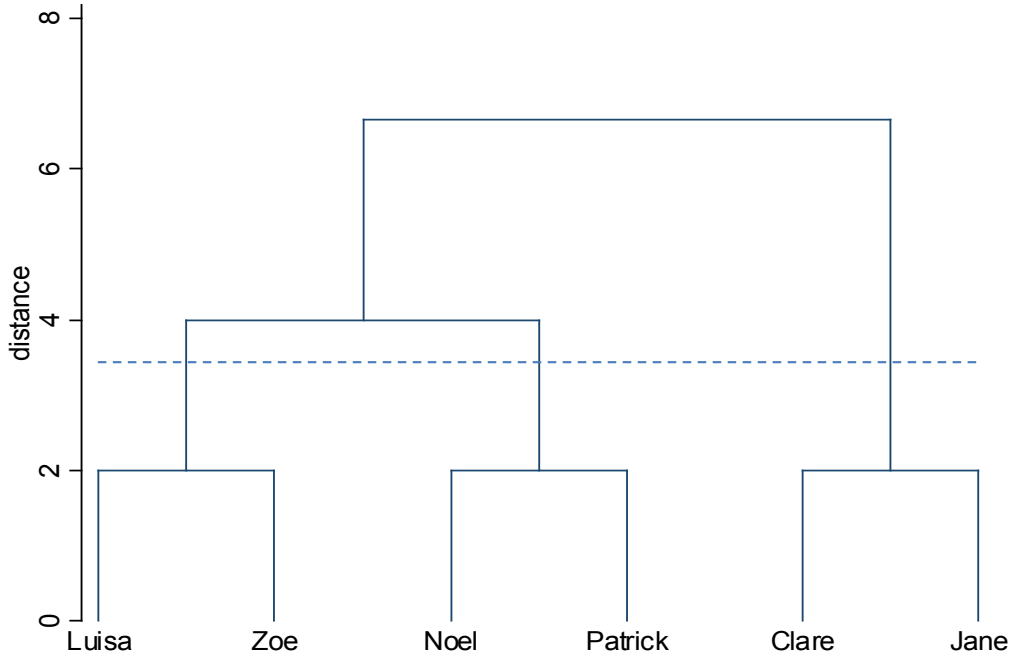
Consider four sequences in addition to Jane and Clare from box B.2. OM calculated that the minimum distance between Jane and Clare is 2 (2 indels). Patrick and Noel are also a distance of 2 apart (one substitution), but both are further than 2 away from Jane or Clare. Likewise, it would take 2 indels 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 B.5 (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.

Determining the number of clusters

There are several approaches to identifying the number of clusters. Most test statistics commonly used to validate cluster solutions cannot be applied to sequence data (Brzinsky-Fay 2007). The possible number of clusters ranges between one and the number of individuals included in the analysis. However, choosing either the upper limit or the lower limit of this range would be of no benefit as it would not reduce any of the variation in the data.

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 producing 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 B.5 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.

Some dendrograms are more difficult to interpret than others, making the number of clusters in the data less obvious to the researcher. Although not an issue in this analysis, the dendrogram may not show a distinct stage in the agglomerative process of clustering all individuals into one cluster where the cluster solution is less likely to change with increasing distance — the distances between clusters may appear to be very similar, without a large increase in distance at a particular stage.[9]

To reduce the degree of subjectivity involved 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 of the young adult, mature adult and senior age segments, the dendrograms suggested four clusters. These groups were found to be plausible for each age segment, and descriptive analysis readily provided a story for each cluster.

[9] Aisenbrey and Fasang (2010) suggest the use of cluster cut-off criteria in these instances to validate the number of clusters found in the data. Alternatively, the relative within and between cluster variation could be compared at each stage of the clustering process.

Labelling and describing the clusters

There are no set rules for labelling clusters. However, clusters are often labelled in a descriptive way. In the education and labour market transitions literature, the term pathway (which picks up on the sequencing of activities) is used instead of cluster. A pathway label may then derive from a commonly observed pattern of activities (for example, the *NILF to Work* pathway for young adults), or it may relate to the proportion of the time period spent in a particular activity or activities (for example, the *NILF* pathway for seniors and the *Prolonged NILF* pathways for the other age segments) (chapter 3). The labels are essentially arbitrary.

Sequence index plots are useful to see what 'most' individuals in the pathway do over time, and if there is a tendency for a particular activity to follow another activity (for example, a transition from working to retirement by older workers). The chronograph is another plot that is useful to see what the main activities are at various points in time.

Descriptive analysis is also valuable. Spell incidence and duration give additional information about what individuals are doing. Information on characteristics of individuals belonging to pathways indicates what individuals have in common with each other compared to individuals in other pathways.

The plots and descriptive analysis will indicate whether there is a reasonable level of within cluster variation and whether the pathways and their labels are meaningful (Corrales-Herrero and Rodríguez-Prado 2012).

Descriptive analysis of the pathways may also lead to forming a hypothesis about which factors are important for determining pathway membership. These can be tested using statistical models (box B.6). 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.

Box B.6 **Further utilisation of OMCA results**

The results of OMCA have been used in different ways in the literature on transitions. OM can be an input to cluster or regression analysis. Pathway (cluster) membership has also been used in regression analysis. The results of OM(CA) have been used as either an explanatory variable, or more typically, as the dependent variable in regressions.

For example, in Dorsett and Lucchino's (2012) analysis of British youth transitions, pathway membership was modelled as the dependent variable using a multinomial logit (MNL) model. The results indicated that gender, grades at school, housing tenure, parental qualifications, and the labour market status of family members were significantly associated with the future labour market outcomes of youths as represented by the pathway. In another study of British youths, Anyadike-Danes and McVicar (2003) conducted similar analysis using a MNL model. Corrales-Herrero and Rodríguez-Prado (2012) and Quintini and Manfredi (2009) applied MNL models to analyse the determinants of pathway membership for Spanish and US youths, respectively.

Instead of using pathway membership as the dependent variable, Malo and Muñoz-Bullón (2003) used the distance measure from OM as the dependent variable in a regression, where distance was measured from the median sequence for each age segment. However, it is unclear how to interpret the results of such a regression, as two very different sequences may have the same distance (McVicar and Anyadike-Danes 2010).

Alternatively, pathway membership may be viewed as an explanatory variable in an analysis of labour market and educational outcomes (Brzinsky-Fay 2007). For example, Fuller (2011) modelled wages with pathway membership as one of the explanatory variables. Han and Moen (1999) modelled characteristics of retirement and retirement plans using career pathway as an explanatory variable.

B.3 Limitations of OMCA techniques

Some limitations of OMCA techniques are discussed in the sequence analysis literature. This section presents the limitations that are relevant to this study and, where they exist, practical solutions to minimise their impact. First, it is difficult to account for sample design and attrition through the use of weights. Second, the substitution cost matrix must be symmetric, which in practice means that the cost of substituting one activity for another cannot be dependent on the direction of the transition. Third, as substitution costs are set in this analysis, the matrix does not change over time — so where a transition occurs in the sequence has no bearing on the cost of a substitution.

The difficulty in accommodating sample weights

The results in this paper are unweighted because the OMCA commands in Stata do not support weighting in either the OM or clustering processes. Also, the weights in the *Household, Income and Labour Dynamics in Australia* (HILDA) Survey data are inadequate in adjusting the sub-sample used in this analysis to be representative of the population (appendix A).

If appropriate weights were available, they could be used in estimating the transition probabilities and, therefore, in deriving the substitution cost matrix to be used in OM (Lesnard 2010). However, the clustering process would still not accommodate the weights.

The substitution cost matrix must be symmetric

Some researchers, for example Aisenbrey and Fasang (2010), argue that a symmetric substitution cost matrix is not reflective of reality. For example, the *Education to Work* pathway is characterised by a transition from study to work. As a result of the matrix structure, however, a substitution of work for study has the same cost as the substitution of study for work. This is considered to be a limitation by those who view substitutions in the OM process as representative of transitions in the sequence data.

However, Martin and Wiggins (2011) argue that this interpretation of the substitution cost matrix is a misunderstanding of the role the matrix plays in OM. It is important to bear in mind that the aim of a substitution cost matrix is not to mimic transitions, but to simply reflect how close different activities are to each other.

Substitution costs are not dynamic

In this paper, substitution costs are restricted to be the same across time. Lesnard (2010) states that assuming distances between activities are time invariant is a restrictive assumption. He argues that the distance between employment and unemployment, for example, should vary across time — they should be close to each other when the unemployment rate is high and distant when it is low.

Lesnard (2006) developed an extension to OM that allows substitution costs to vary across time.[10] According to Aisenbrey and Fasang (2010, p. 437), Lesnard's distance measure:

> … is particularly useful for applications … in which the exact timing of states within sequences is of particular theoretical importance, such as in the analysis of retirement trajectories …

However, dynamic substitution costs are not used in this study as the importance of the dynamic element has yet to be established empirically (Hollister 2009). Moreover, implementing dynamics would involve calculating 120 data-driven matrices instead of one. This would significantly add to the complexity of the computation task in OM.

[10] The approach, which has been named the dynamic Hamming model, is an extension to Hamming's distance, which is equal to the number of substitutions required to align two sequences. See Lesnard (2010) for further discussion about dynamic substitution costs.

C Results

Chapter 3 contains results and discussion for the following six pathways:

- youths in *Work and Study to Work* and *Churning with Work*
- young adults in *Work* and *NILF to Work*
- mature adults in *Work to NILF*
- seniors in *Work*.

This appendix contains results for the remaining 11 pathways. For each pathway, results are presented in a sequence index plot (labelled *(A) Activity sequences by individual*), a chronograph (labelled *(B) Activities by monthly share*), a table of selected characteristics of individuals in that pathway and for the age segment as a whole, and characteristics of the activity patterns in that pathway.

In each sequence index plot, every individual in that pathway is represented by a separate horizontal line. Each line shows the sequence of colour-coded activities, so that a change in colour from one period to the next represents a transition.[1] The chronograph displays the share of individuals in the pathway undertaking each activity for each month. In essence, while the sequence index plot captures the longitudinal nature of the sequence data, the chronograph does not.

C.1 Youth pathways

Results are presented for the following youth pathways:

- *Education to Work*
- *Work, with or without Study*
- *Prolonged NILF*.

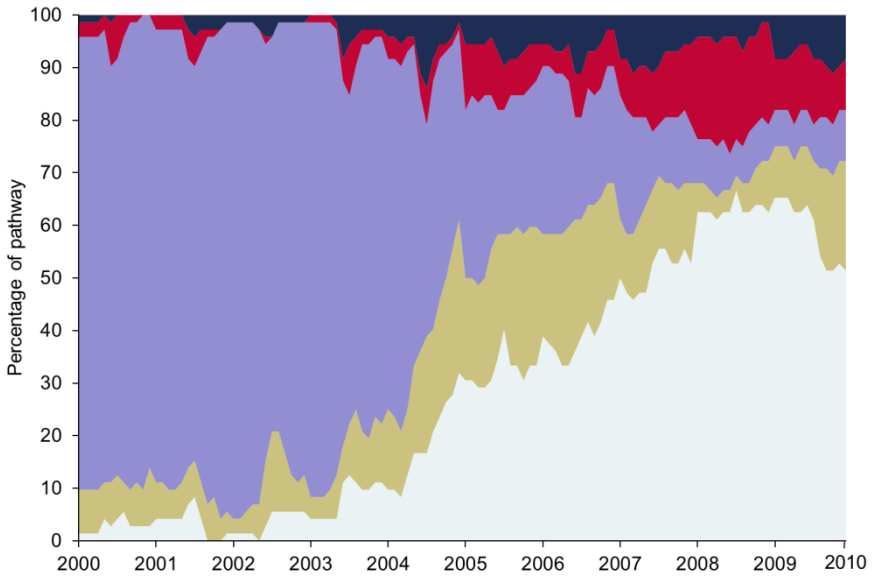
[1] Care is needed when examining sequence index plots. When the activities of many individuals are displayed for many time periods, these charts can be subject to overplotting that can distort the apparent incidence and duration of some activities (chapter 3).

Youths in the *Education to Work* pathway

Figure C.1 **Activities in the *Education 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 C.1 **Selected characteristics of individuals in the *Education to Work* pathway for youths**

Characteristic	*Measure*	*Ed–W pathway*	*All youth pathways*
In 2001:			
age	years, average	16.8	19.2
gender	% female	55.56	54.39
locality (remoteness)	% major city	63.89	61.35
highest level of education[a]	% high	1.39	9.23
	% medium	27.78	43.90
	% low	70.83	46.86
unemployed	% U	9.72	9.81
NILF (incl. marginally attached)	% NILF	75.00	27.37
of which:			
home duties/childcare	% of NILF	1.85	16.67
study	% of NILF	79.63	64.58
marginally attached to labour force	% of NILF	14.81	11.67
other reasons[b]	% of NILF	3.70	7.08
In 2010:			
locality (remoteness)	% major city	66.67	62.03
highest level of education[a]	% high	38.89	30.55
	% medium	52.78	55.41
	% low	8.33	14.02
unemployed	% U	8.33	4.45
NILF (incl. marginally attached)	% NILF	20.83	12.20
of which:			
home duties/childcare	% of NILF	40.00	66.36
study	% of NILF	33.33	13.08
marginally attached to labour force	% of NILF	0.00	5.61
other reasons[b]	% of NILF	26.67	14.95

[a] Educational attainment has been classified as high (Bachelor 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 C.2 **Selected characteristics of activity patterns in the *Education to Work* pathway (youths)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	28.08	12.96	47.47	6.86	4.63
Share of path with at least one spell of the activity	91.67	80.56	100.00	56.94	63.89
Conditional on at least one spell of the activity:					
average number of spells	2.41	2.90	2.88	1.71	1.80
average length of spell (months)	15.26	6.67	19.81	8.47	4.82

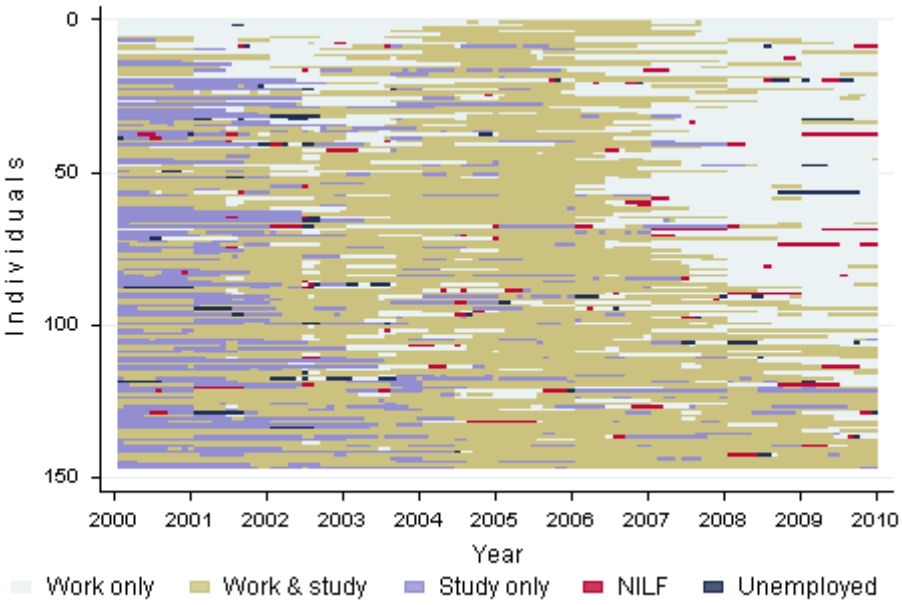
[a] Pathway size 72 (8.2 per cent of youths); average number of activities 3.93.

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

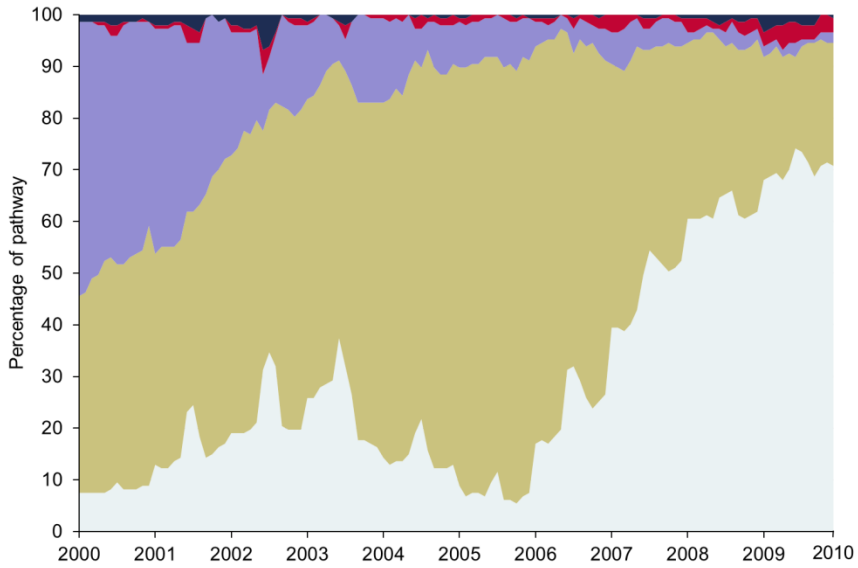
Youths in the *Work, with or without Study* pathway

Figure C.2 **Activities in the *Work, with or without Study* pathway for youths**

(A) Activity sequences by individual



(B) Activities by monthly share



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

Table C.3 **Selected characteristics of individuals in the *Work, with or without Study* pathway for youths**

Characteristic	*Measure*	*W +/- S pathway*	*All youth pathways*
In 2001:			
age	years, average	17.5	19.2
gender	% female	53.74	54.39
locality (remoteness)	% major city	68.71	61.35
highest level of education[a]	% high	6.12	9.23
	% medium	25.17	43.90
	% low	68.71	46.86
unemployed	% U	7.48	9.81
NILF (incl. marginally attached)	% NILF	35.37	27.37
of which:			
home duties/childcare	% of NILF	1.92	16.67
study	% of NILF	90.38	64.58
marginally attached to labour force	% of NILF	5.77	11.67
other reasons[b]	% of NILF	1.92	7.08
In 2010:			
locality (remoteness)	% major city	78.91	62.03
highest level of education[a]	% high	57.82	30.55
	% medium	40.82	55.41
	% low	1.36	14.02
unemployed	% U	3.40	4.45
NILF (incl. marginally attached)	% NILF	3.40	12.20
of which:			
home duties/childcare	% of NILF	80.00	66.36
study	% of NILF	20.00	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 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 C.4 **Selected characteristics of activity patterns in the *Work, with or without Study* pathway (youths)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	29.76	53.48	14.43	1.35	0.99
Share of path with at least one spell of the activity	97.28	100.00	80.27	36.05	25.85
Conditional on at least one spell of the activity:					
average number of spells	2.73	3.47	2.66	1.53	1.68
average length of spell (months)	13.42	18.50	8.11	2.94	2.72

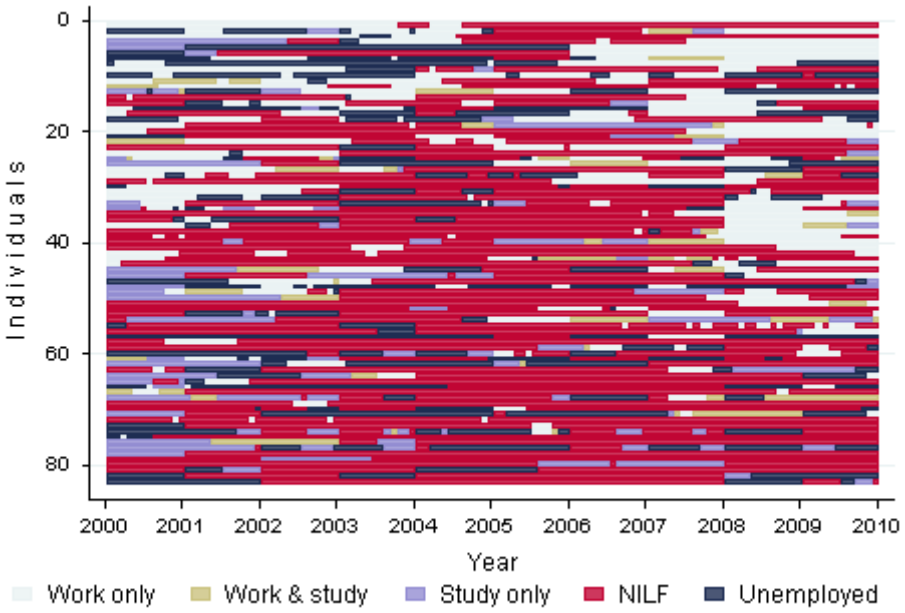
[a] Pathway size 147 (16.8 per cent of youths); average number of activities 3.39.

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

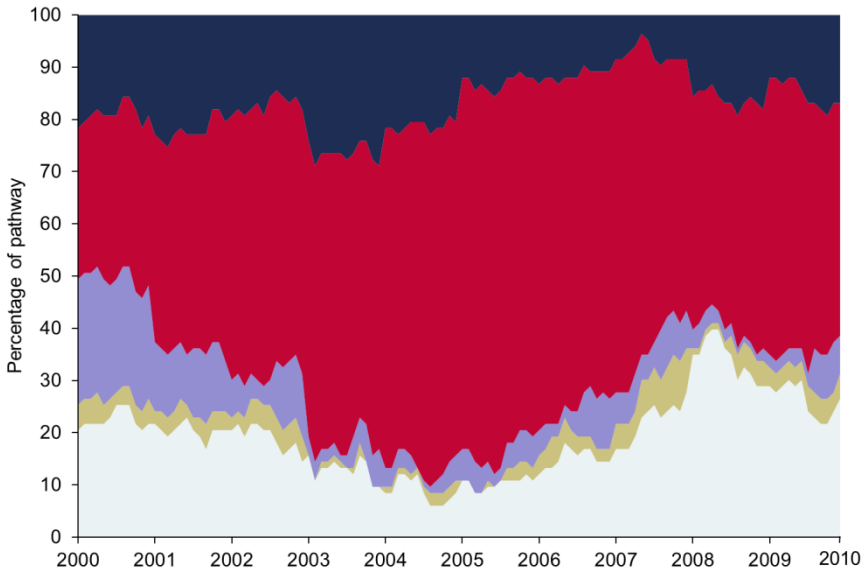
Youths in the *Prolonged NILF* pathway

Figure C.3 **Activities in the *Prolonged NILF* pathway for youths**

(A) Activity sequences by individual



(B) Activities by monthly share



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

Table C.5 **Selected characteristics of individuals in the *Prolonged NILF* pathway for youths**

Characteristic	*Measure*	*Prol NILF pathway*	*All youth pathways*
In 2001:			
age	years, average	20.3	19.2
gender	% female	80.72	54.39
locality (remoteness)	% major city	51.81	61.35
highest level of education[a]	% high	4.82	9.23
	% medium	33.74	43.90
	% low	61.44	46.86
unemployed	% U	26.51	9.81
NILF (incl. marginally attached)	% NILF	53.01	27.37
of which:			
home duties/childcare	% of NILF	61.36	16.67
study	% of NILF	20.45	64.58
marginally attached to labour force	% of NILF	6.82	11.67
other reasons[b]	% of NILF	11.36	7.08
In 2010:			
locality (remoteness)	% major city	39.76	62.03
highest level of education[a]	% high	4.82	30.55
	% medium	50.6	55.41
	% low	44.57	14.02
unemployed	% U	14.46	4.45
NILF (incl. marginally attached)	% NILF	49.40	12.20
of which:			
home duties/childcare	% of NILF	73.17	66.36
study	% of NILF	2.44	13.08
marginally attached to labour force	% of NILF	4.88	5.61
other reasons[b]	% of NILF	19.51	14.95

[a] Educational attainment has been classified as high (Bachelor 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 C.6 **Selected characteristics of activity patterns in the *Prolonged NILF* pathway (youths)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	19.19	3.30	7.56	53.00	16.95
Share of path with at least one spell of the activity	90.36	38.55	61.45	95.18	74.70
Conditional on at least one spell of the activity:					
average number of spells	2.72	1.56	1.80	2.72	2.66
average length of spell (months)	9.37	6.58	8.18	24.55	10.23

[a] Pathway size 83 (9.5 per cent of youths); average number of activities 3.60.

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

C.2 Young adult pathways

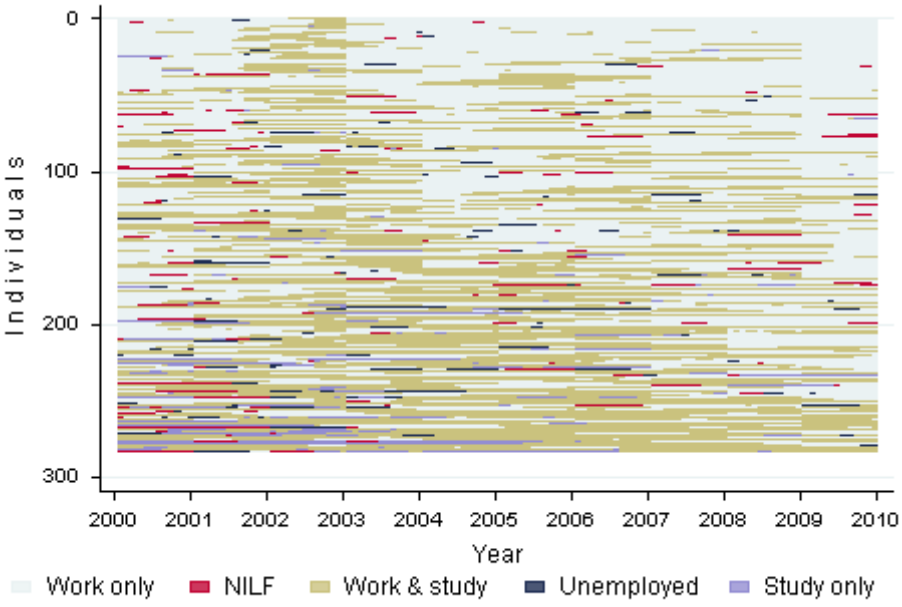
Results are presented for the following young adult pathways:

- *Work, with or without Study*
- *Prolonged NILF*.

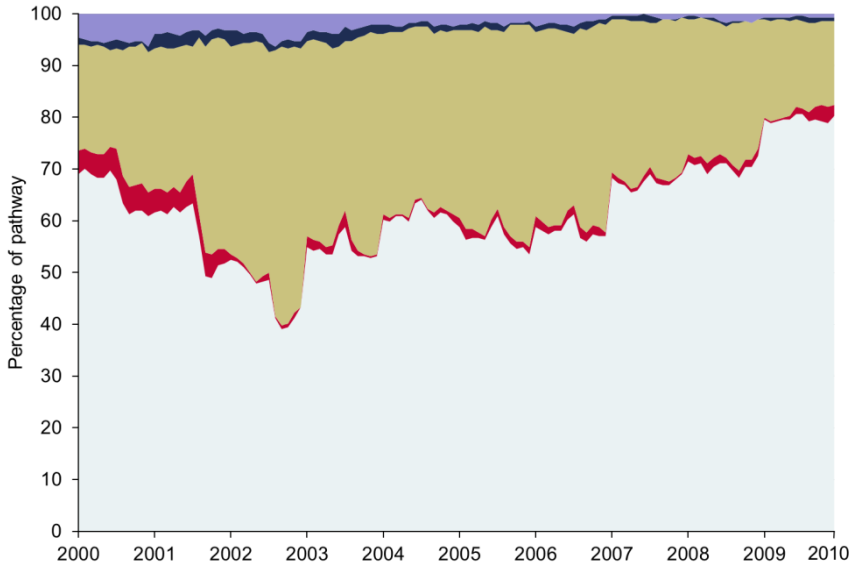
Young adults in the *Work, with or without Study* pathway

Figure C.4 **Activities in the *Work, with or without Study* 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 C.7 **Selected characteristics of individuals in the *Work, with or without Study* pathway for young adults**

Characteristic	*Measure*	*W +/- S pathway*	*All young adult pathways*
In 2001:			
age	years, average	31.8	32.7
gender	% female	56.69	54.78
locality (remoteness)	% major city	68.31	65.31
highest level of education[a]	% high	41.55	27.27
	% medium	48.59	46.3
	% low	9.86	26.42
unemployed	% U	2.11	3.23
NILF (incl. marginally attached)	% NILF	9.51	18.26
of which:			
home duties/childcare	% of NILF	62.96	75.60
study	% of NILF	29.63	8.13
marginally attached to labour force	% of NILF	3.70	2.87
other reasons[b]	% of NILF	3.70	13.40
In 2010:			
locality (remoteness)	% major city	64.79	62.12
highest level of education[a]	% high	55.99	30.62
	% medium	41.9	47.4
	% low	2.11	21.98
unemployed	% U	1.41	2.40
NILF (incl. marginally attached)	% NILF	2.82	12.54
of which:			
home duties/childcare	% of NILF	75.00	61.67
study	% of NILF	0.00	5.92
marginally attached to labour force	% of NILF	0.00	2.79
other reasons[b]	% of NILF	25.00	29.62

[a] Educational attainment has been classified as high (Bachelor 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 C.8 **Selected characteristics of activity patterns in the *Work, with or without Study* pathway (young adults)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	61.79	32.68	2.44	1.87	1.22
Share of path with at least one spell of the activity	100.00	100.00	26.76	33.10	25.70
Conditional on at least one spell of the activity:					
average number of spells	3.31	2.43	1.62	1.48	1.51
average length of spell (months)	22.43	16.14	6.76	4.59	3.79

[a] Pathway size 284 (12.4 per cent of young adults); average number of activities 2.86.

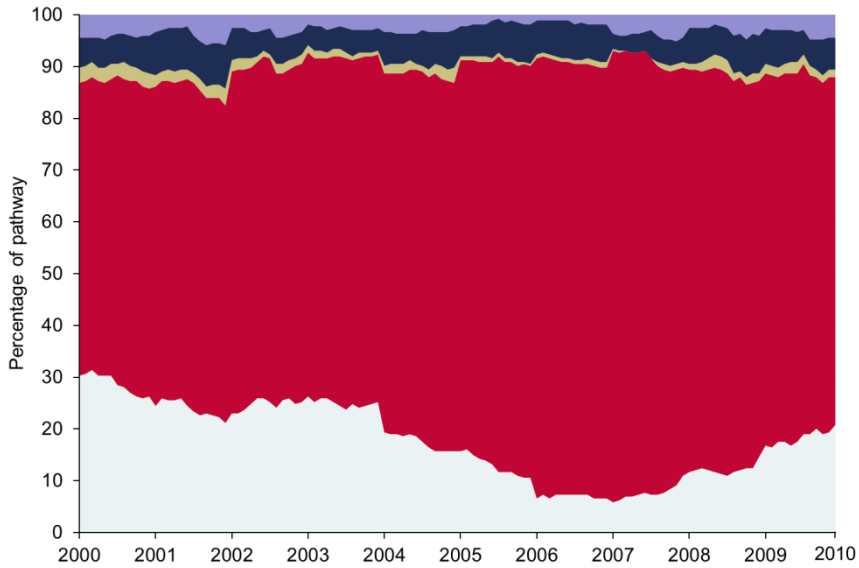
Source: Authors' estimates based on HILDA waves 1–10.

Young adults in the *Prolonged NILF* pathway

Figure C.5 **Activities in the *Prolonged NILF* 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 C.9 **Selected characteristics of individuals in the *Prolonged NILF* pathway for young adults**

Characteristic	*Measure*	*Prol NILF pathway*	*All young adult pathways*
In 2001:			
age	years, average	32.9	32.7
gender	% female	82.12	54.78
locality (remoteness)	% major city	61.68	65.31
highest level of education[a]	% high	14.95	27.27
	% medium	41.24	46.3
	% low	43.79	26.42
unemployed	% U	5.84	3.23
NILF (incl. marginally attached)	% NILF	66.79	18.26
of which:			
home duties/childcare	% of NILF	74.32	75.60
study	% of NILF	3.83	8.13
marginally attached to labour force	% of NILF	2.73	2.87
other reasons[b]	% of NILF	19.13	13.40
In 2010:			
locality (remoteness)	% major city	56.57	62.12
highest level of education[a]	% high	15.69	30.62
	% medium	43.06	47.4
	% low	41.23	21.98
unemployed	% U	6.20	2.40
NILF (incl. marginally attached)	% NILF	67.88	12.54
of which:			
home duties/childcare	% of NILF	65.05	61.67
study	% of NILF	2.15	5.92
marginally attached to labour force	% of NILF	1.08	2.79
other reasons[b]	% of NILF	31.72	29.62

[a] Educational attainment has been classified as high (Bachelor 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 C.10 **Selected characteristics of activity patterns in the *Prolonged NILF* pathway (young adults)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	17.79	1.53	3.11	71.51	6.06
Share of path with at least one spell of the activity	70.80	17.88	26.64	99.64	40.15
Conditional on at least one spell of the activity:					
average number of spells	2.10	1.51	1.51	2.07	2.08
average length of spell (months)	14.37	6.80	9.29	41.54	8.70

[a] Pathway size 274 (12.0 per cent of young adults); average number of activities 2.55.

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

C.3 Mature adult pathways

Results are presented for the following mature adult pathways:

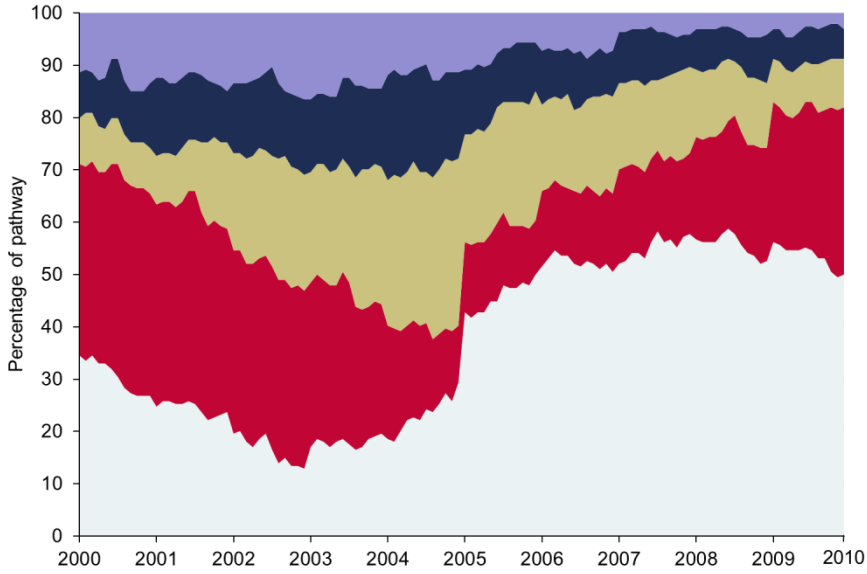
- *NILF to Work*
- *Prolonged NILF*
- *Work*.

Mature adults in the *NILF to Work* pathway

Figure C.6 **Activities in the *NILF to Work* 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 C.11 **Selected characteristics of individuals in the *NILF to Work* pathway for mature adults**

Characteristic	*Measure*	*N–W pathway*	*All mature adult pathways*
In 2001:			
age	years, average	45.6	46.7
gender	% female	64.43	53.82
locality (remoteness)	% major city	58.25	60.62
highest level of education[a]	% high	25.77	24.39
	% medium	39.68	40.06
	% low	34.54	35.56
unemployed	% U	14.43	3.53
NILF (incl. marginally attached)	% NILF	50.52	17.67
of which:			
home duties/childcare	% of NILF	51.02	43.51
study	% of NILF	10.20	4.33
marginally attached to labour force	% of NILF	5.10	3.37
other reasons[b]	% of NILF	33.67	48.80
In 2010:			
locality (remoteness)	% major city	58.25	58.33
highest level of education[a]	% high	33.5	25.53
	% medium	40.72	42.78
	% low	25.78	31.68
unemployed	% U	6.19	1.91
NILF (incl. marginally attached)	% NILF	34.54	26.89
of which:			
home duties/childcare	% of NILF	17.91	17.69
study	% of NILF	0.00	1.11
marginally attached to labour force	% of NILF	1.49	1.42
other reasons[b]	% of NILF	80.60	79.78

[a] Educational attainment has been classified as high (Bachelor 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 C.12 **Selected characteristics of activity patterns in the *NILF to Work* pathway (mature adults)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	37.53	17.29	9.02	24.82	11.34
Share of path with at least one spell of the activity	93.81	54.12	52.06	74.23	51.03
Conditional on at least one spell of the activity:					
average number of spells	2.84	2.30	2.01	2.31	2.82
average length of spell (months)	16.93	16.70	10.35	17.40	9.47

[a] Pathway size 194 (8.2 per cent of mature adults); average number of activities 3.25.

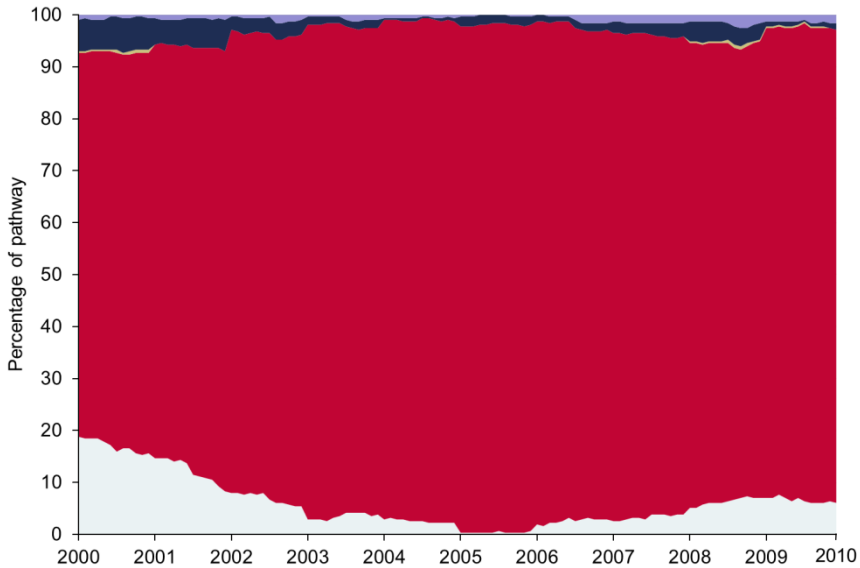
Source: Authors' estimates based on HILDA waves 1–10.

Mature adults in the *Prolonged NILF* pathway

Figure C.7 **Activities in the *Prolonged 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 C.13 **Selected characteristics of individuals in the *Prolonged NILF* pathway for mature adults**

Characteristic	*Measure*	*Prol NILF pathway*	*All mature adult pathways*
In 2001:			
age	years, average	48.9	46.7
hender	% female	71.97	53.82
locality (remoteness)	% major city	56.37	60.62
highest level of education[a]	% high	8.6	24.39
	% medium	30.57	40.06
	% low	60.82	35.56
unemployed	% U	3.82	3.53
NILF (incl. marginally attached)	% NILF	82.80	17.67
of which:			
home duties/childcare	% of NILF	40.77	43.51
study	% of NILF	0.00	4.33
marginally attached to labour force	% of NILF	0.77	3.37
other reasons[b]	% of NILF	58.46	48.80
In 2010:			
locality (remoteness)	% major city	53.50	58.33
highest level of education[a]	% high	8.6	25.53
	% medium	32.17	42.78
	% low	59.23	31.68
unemployed	% U	0.32	1.91
NILF (incl. marginally attached)	% NILF	92.99	26.89
of which:			
home duties/childcare	% of NILF	22.95	17.69
study	% of NILF	1.03	1.11
marginally attached to labour force	% of NILF	0.68	1.42
other reasons[b]	% of NILF	75.34	79.78

[a] Educational attainment has been classified as high (Bachelor 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 C.14 **Selected characteristics of activity patterns in the *Prolonged NILF* pathway (mature adults)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	6.14	0.11	0.91	90.22	2.61
Share of path with at least one spell of the activity	4.14	1.59	13.06	100.00	22.29
Conditional on at least one spell of the activity:					
average number of spells	1.35	1.20	1.27	1.56	1.54
average length of spell (months)	13.14	7.17	6.62	69.38	9.11

[a] Pathway size 314 (13.3 per cent of mature adults); average number of activities 1.78.

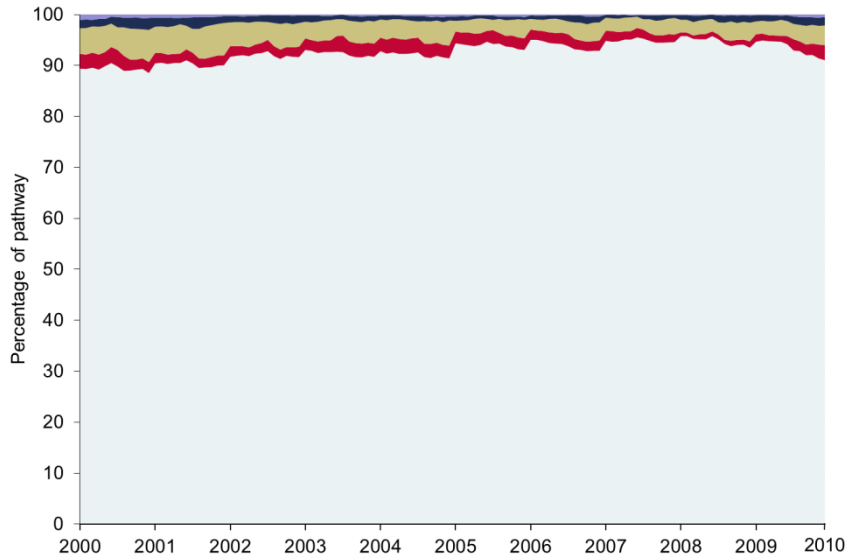
Source: Authors' estimates based on HILDA waves 1–10.

Mature adults in the *Work* pathway

Figure C.8 **Activities in the *Work* 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 C.15 **Selected characteristics of individuals in the *Work* pathway for mature adults**

Characteristic	*Measure*	*W pathway*	*All mature adult pathways*
In 2001:			
age	years, average	46.0	46.7
gender	% female	48.50	53.82
locality (remoteness)	% major city	62.00	60.62
highest level of education[a]	% high	27.63	24.39
	% medium	42.73	40.06
	% low	29.65	35.56
unemployed	% U	2.03	3.53
NILF (incl. marginally attached)	% NILF	3.01	17.67
of which:			
home duties/childcare	% of NILF	46.94	43.51
study	% of NILF	14.29	4.33
marginally attached to labour force	% of NILF	12.24	3.37
other reasons[b]	% of NILF	26.53	48.80
In 2010:			
locality (remoteness)	% major city	60.34	58.33
highest level of education[a]	% high	28.36	25.53
	% medium	45.74	42.78
	% low	25.91	31.68
unemployed	% U	1.66	1.91
NILF (incl. marginally attached)	% NILF	5.77	26.89
of which:			
home duties/childcare	% of NILF	12.77	17.69
study	% of NILF	3.19	1.11
marginally attached to labour force	% of NILF	4.26	1.42
other reasons[b]	% of NILF	79.79	79.78

[a] Educational attainment has been classified as high (Bachelor 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 C.16 **Selected characteristics of activity patterns in the *Work* pathway (mature adults)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	92.68	3.78	0.33	2.06	1.16
Share of path with at least one spell of the activity	100.00	34.93	7.00	25.23	19.09
Conditional on at least one spell of the activity:					
average number of spells	1.93	1.58	1.18	1.54	1.68
average length of spell (months)	57.75	8.21	4.76	6.36	4.34

[a] Pathway size 1 629 (69.2 per cent of mature adults); average number of activities 1.86.

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

C.4 Senior pathways

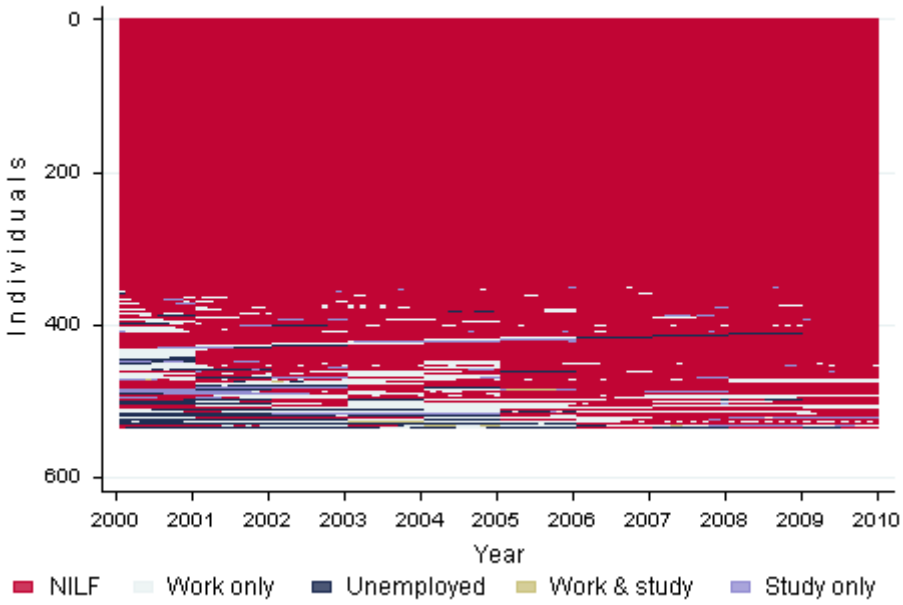
Results are presented for the following senior pathways:

- *NILF*
- *Early Work to NILF*
- *Later Work to NILF*.

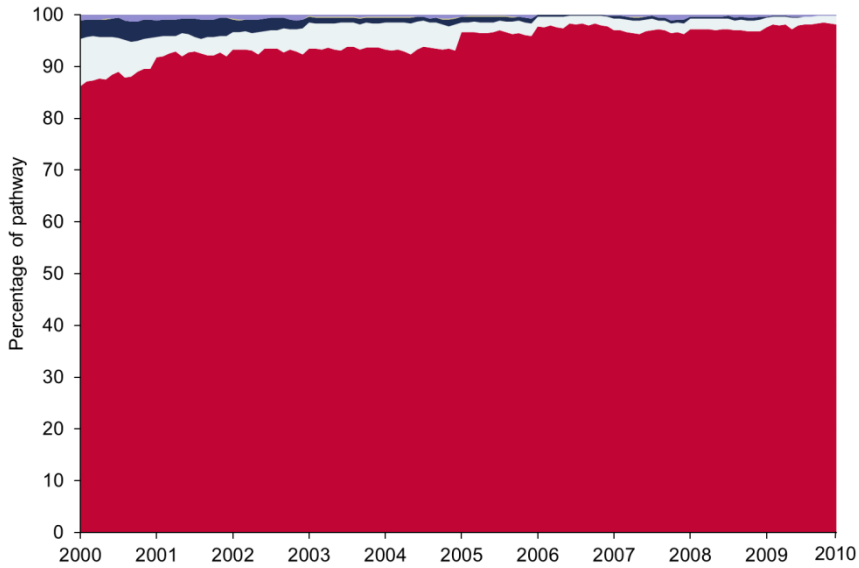
Seniors in the *NILF* pathway

Figure C.9 **Activities in the *NILF* pathway for seniors**

(A) Activity sequences by individual



(B) Activities by monthly share



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

Table C.17 **Selected characteristics of individuals in the *NILF* pathway for seniors**

Characteristic	*Measure*	*NILF pathway*	*All senior pathways*
In 2001:			
age	years, average	60.1	59.2
gender	% female	63.99	53.25
locality (remoteness)	% major city	54.66	56.69
highest level of education[a]	% high	9.88	15.68
	% medium	33.77	34.9
	% low	56.35	49.42
unemployed	% U	2.43	2.20
NILF (incl. marginally attached)	% NILF	92.91	50.00
of which:			
home duties/childcare	% of NILF	17.87	17.78
study	% of NILF	0.20	0.76
marginally attached to labour force	% of NILF	0.20	0.38
other reasons[b]	% of NILF	81.73	81.07
In 2010:			
locality (remoteness)	% major city	53.17	53.92
highest level of education[a]	% high	9.88	15.87
	% medium	33.96	35.66
	% low	56.15	48.27
unemployed	% U	0.00	0.19
NILF (incl. marginally attached)	% NILF	98.69	80.31
of which:			
home duties/childcare	% of NILF	9.64	7.26
study	% of NILF	0.19	0.24
marginally attached to labour force	% of NILF	0.00	0.12
other reasons[b]	% of NILF	90.17	92.38

[a] Educational attainment has been classified as high (Bachelor 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 C.18 **Selected characteristics of activity patterns in the *NILF* pathway (seniors)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	3.45	0.05	0.50	94.66	1.35
Share of path with at least one spell of the activity	27.05	1.12	6.53	100.00	9.14
Conditional on at least one spell of the activity:					
average number of spells	2.07	1.17	1.54	1.55	1.65
average length of spell (months)	7.39	4.71	5.96	73.44	10.69

[a] Pathway size 536 (51.2 per cent of seniors); average number of activities 1.44.

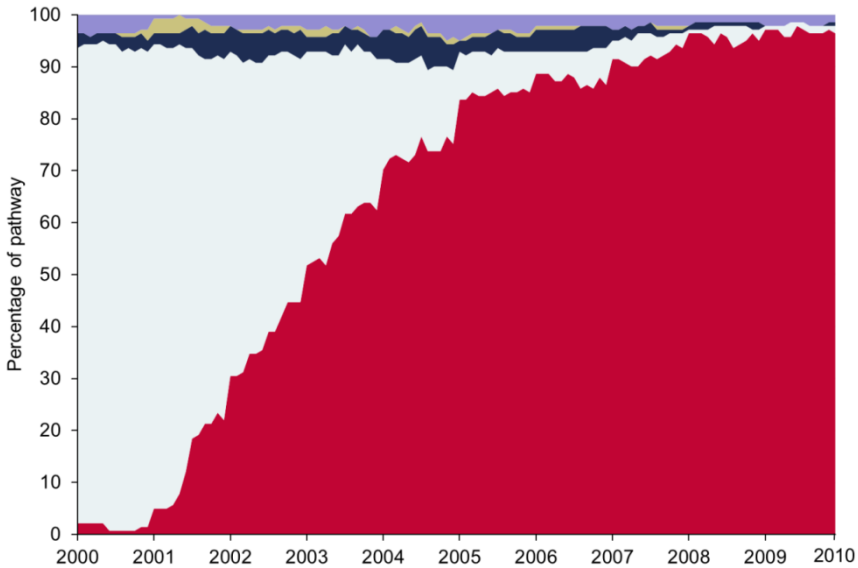
Source: Authors' estimates based on HILDA waves 1–10.

Seniors in the *Early Work to NILF* pathway

Figure C.10 **Activities in the *Early Work to NILF* pathway for seniors**

(A) Activity sequences by individual

(B) Activities by monthly share



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

Table C.19 **Selected characteristics of individuals in the *Early Work to NILF* pathway for seniors**

Characteristic	*Measure*	*E W–N pathway*	*All senior pathways*
In 2001:			
age	years, average	59.2	59.2
gender	% female	44.68	53.25
locality (remoteness)	% major city	56.74	56.69
highest level of education[a]	% high	14.19	15.68
	% medium	35.46	34.9
	% low	50.36	49.42
unemployed	% U	2.13	2.20
NILF (incl. marginally attached)	% NILF	7.80	50.00
of which:			
home duties/childcare	% of NILF	9.09	17.78
study	% of NILF	27.27	0.76
marginally attached to labour force	% of NILF	0.00	0.38
other reasons[b]	% of NILF	63.64	81.07
In 2010:			
locality (remoteness)	% major city	51.06	53.92
highest level of education[a]	% high	14.19	15.87
	% medium	38.29	35.66
	% low	47.52	48.27
unemployed	% U	0.00	0.19
NILF (incl. marginally attached)	% NILF	98.58	80.31
of which:			
home duties/childcare	% of NILF	2.16	7.26
study	% of NILF	0.72	0.24
marginally attached to labour force	% of NILF	0.00	0.12
other reasons[b]	% of NILF	97.12	92.38

[a] Educational attainment has been classified as high (Bachelor 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 C.20 **Selected characteristics of activity patterns in the *Early Work to NILF* pathway (seniors)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	29.88	0.60	2.43	64.10	2.99
Share of path with at least one spell of the activity	99.29	11.35	11.35	100.00	22.70
Conditional on at least one spell of the activity:					
average number of spells	2.03	1.31	2.50	2.05	2.06
average length of spell (months)	17.80	4.86	10.28	37.53	7.67

[a] Pathway size 141 (13.5 per cent of seniors); average number of activities 2.45.

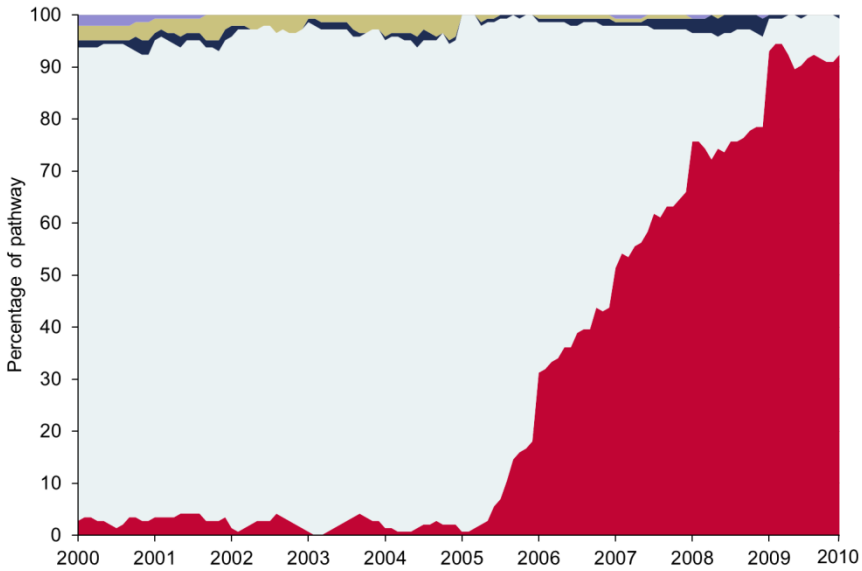
Source: Authors' estimates based on HILDA waves 1–10.

Seniors in the *Later Work to NILF* pathway

Figure C.11 **Activities in the *Later Work to NILF* pathway for seniors**

(A) Activity sequences by individual

(B) Activities by monthly share



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

Table C.21 **Selected characteristics of individuals in the *Later Work to NILF* pathway for seniors**

Characteristic	*Measure*	*L W–N pathway*	*All senior pathways*
In 2001:			
age	years, average	58.3	59.2
gender	% female	39.58	53.25
locality (remoteness)	% major city	61.81	56.69
highest level of education[a]	% high	24.31	15.68
	% medium	34.72	34.9
	% low	40.96	49.42
unemployed	% U	1.39	2.20
NILF (incl. marginally attached)	% NILF	4.17	50.00
of which:			
home duties/childcare	% of NILF	16.67	17.78
study	% of NILF	0.00	0.76
marginally attached to labour force	% of NILF	0.00	0.38
other reasons[b]	% of NILF	83.33	81.07
In 2010:			
locality (remoteness)	% major city	56.25	53.92
highest level of education[a]	% high	25.01	15.87
	% medium	35.41	35.66
	% low	39.57	48.27
unemployed	% U	0.00	0.19
NILF (incl. marginally attached)	% NILF	93.75	80.31
of which:			
home duties/childcare	% of NILF	5.19	7.26
study	% of NILF	0.00	0.24
marginally attached to labour force	% of NILF	0.00	0.12
other reasons[b]	% of NILF	94.81	92.38

[a] Educational attainment has been classified as high (Bachelor 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 C.22 **Selected characteristics of activity patterns in the *Later Work to NILF* pathway (seniors)[a]**

	Work only	*Work and study*	*Study only*	*NILF*	*Unemployment*
Average time in the activity (per cent)	68.47	1.63	0.29	28.47	1.15
Share of path with at least one spell of the activity	100.00	16.67	4.17	100.00	13.89
Conditional on at least one spell of the activity:					
average number of spells	1.97	1.58	1.00	1.59	1.65
average length of spell (months)	41.66	7.39	8.33	21.48	6.03

[a] Pathway size 144 (13.8 per cent of seniors); average number of activities 2.35.

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

References

Aisenbrey, S. and Fasang, A. 2010, 'New life for old ideas: the 'second wave' of sequence analysis bringing the 'course' back into the life course', *Sociological Methods & Research*, vol. 38, no. 3, pp. 420–62.

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.

AIHW (Australian Institute of Health and Welfare) 2012, *Housing Assistance in Australia 2012*, Cat. no. HOU 266, AIHW, Canberra.

Bachmann, R. and Burda M. 2007, *Sectoral Transformation, Turbulence, and Labour Market Dynamics in Germany*, Discussion paper no. 2007–008, SFB 649 (Collaborative Research Centre 649): Economic Risk, Humboldt University School of Business and Economics, Berlin, Germany.

Baxter, J. and Renda, J. 2011, 'Lone and couple mothers in the Australian labour market: differences in employment transitions', *Australian Journal of Labour Economics*, vol. 14, no. 2, pp. 103–22.

Black, D., Polidano, C. and Tseng, Y.-P. 2011, *The Re-engagement in Education of Early School Leavers*, Working Paper no. 13/11, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Borland, J. and Johnston, D. 2010, *How Does a Worker's Labour Market History Affect Job Duration?*, Working Paper no. 6/10, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

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.

——, —— and Luniak, M. 2006, 'Sequence analysis with Stata', *The Stata Journal*, vol. 6, no. 4, pp. 435–60.

Buddelmeyer, H. and Marks, G. 2010, *Annual Transitions Between Labour Market States for Young Australians*, National Vocational Education and Training Research and Evaluation (NVETRE) Program Report, NCVER, Adelaide.

Cai, L. 2010, *Work Choices of Married Women: drivers of change*, Visiting Researcher Paper, Productivity Commission, Canberra.

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.

Dandolopartners 2012, Interim Evaluation of the National Partnership on Youth Attainment and Transitions, Report for the Department of Education, Employment and Workplace Relations, Dandolopartners, Melbourne.

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.

Fabrizi, E. and Mussida, C. 2009, 'The determinants of labour market transitions', *Giornale degli Economisti e Annali di Economia* (Journal of Economists and the Annals of Economics), vol. 68, no. 2, pp. 233–65.

Fasang, A. 2008, Family biographies and retirement processes: a comparative analysis of West Germany and the United Kingdom, PhD thesis, Jacobs Center on Lifelong Learning and Institutional Development, Jacobs University Bremen, Germany, unpublished.

—— 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).

Fitzpatrick, D., Lester, L., Mavromaras, K., Richardson, S. And Sun, Y. 2011, *From Education to Employment: How Long Does it Take?*, Research Report, NCVER, Adelaide.

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.

Gilfillan, G. and Andrews, L. 2010, *Labour Force Participation of Women Over 45*, Productivity Commission Staff Working Paper, Canberra.

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.

Henstridge, J. 2001, *The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Weighting and Imputation*, HILDA Project Discussion Paper No. 3/01, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

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.

—— and McKenzie, P. 2001, *Patterns of Success and Failure in the Transition from School to Work in Australia*, Longitudinal Surveys of Australian Youth (LSAY) Research Report no. 18, ACER, Melbourne.

Lattimore, R. 2007, *Men Not at Work: An Analysis of Men Outside the Labour Force*, Productivity Commission Staff Working Paper, Canberra, January.

Lattin, J., Carroll, J. and Green, P. 2003, *Analyzing Multivariate Data*, Thomson Learning, Ontario, Canada.

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.

—— and —— 2010, *Does Optimal Matching Really Give Us Anything Extra for the Analysis of Careers Data? An Application to British Criminal Careers*, Working Paper, Economics and Finance Research Group, Queen's University, Belfast.

Magnac, T. 2000, 'Subsidised training and youth employment: distinguishing unobserved heterogeneity from state dependence in labour market histories', *The Economic Journal*, vol. 110, October, pp. 805–37.

Malo, M. and Muñoz-Bullón, F. 2003, 'Employment status mobility from a life-cycle perspective: a sequence analysis of work-histories in the BHPS', *Demographic Research*, vol. 9, article 7, pp. 119–62.

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.

Mroz, T. and Savage, T. 2006, 'The long-term effects of youth unemployment', *Journal of Human Resources*, vol. XLI, no. 2, pp. 259–93.

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.

Polidano, C. and Mavromaras, K. 2010, *The Role of Vocational Education and Training in the Labour Market Outcomes of People with Disabilities*, National Vocational Education and Training Research and Evaluation (NVETRE) Program Report, NCVER, Adelaide.

Productivity Commission 2005, *Economic Implications of an Ageing Australia*, Research Report, Canberra.

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.

Sane, R. 2011, Transitions out of part-time work: evidence from Australia, Working Paper, April, http://ssrn.com/abstract=1803008 (accessed 16 July 2013).

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

Schoon, I., McCulloch, A., Joshi, H. E., Wiggins, R. D. and Bynner, J. 2001, 'Transitions from school to work in a changing social context', *Young*, vol. 9, no. 1, pp. 4–22.

Stovel, K. and Bolan, M. 2004, 'Residential trajectories : using optimal alignment to reveal the structure of residential mobility', *Sociological Methods & Research*, vol. 32, no. 4, pp. 559–98.

Summerfield, M., Dunn, R., Freidin, S., Hahn, M., Ittak, P., Kecmanovic, M., Li, N., Macalalad, N., Watson, N., Wilkins, R. and Wooden, M. 2011, *HILDA User Manual — Release 10*, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Watson, N. 2009, Disentangling overlapping seams: the experience of the HILDA survey, *HILDA Conference Proceedings 2009*, Melbourne, 16–17 July.

—— 2012, *Longitudinal and Cross-sectional Weighting Methodology for the HILDA Survey*, HILDA Project Technical Paper no. 2/12, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

—— and Wooden, M. 2002, *Assessing the Quality of the HILDA Survey Wave 1 Data*, HILDA Project Technical Paper no. 4/02, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

—— and —— 2004, 'Sample Attrition in the HILDA Survey', *Australian Journal of Labour Economics*, vol. 7, no. 2, June, pp. 293–308.

—— and —— 2006, *Modelling longitudinal survey response: the experience of the HILDA survey*, HILDA Project Discussion Paper no. 2/06, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

—— and —— 2012, 'The HILDA survey: a case study in the design and development of a successful household panel survey', *Longitudinal and Life Course Studies*, vol. 3, no. 3, pp. 369–81.

Wooden, M., Bevitt, A., Chigavazira, A., Greer, N., Johnson, G., Killackey, E., Moschion, J., Scutella, R., Tseng, Y.-P., and Watson, N. 2012, 'Introducing 'Journeys Home'', *Australian Economic Review*, vol. 45, no. 3, pp. 368–78.

——, Freidin, S. and Watson, N. 2002, 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1', *Australian Economic Review*, vol. 35, no. 3, pp. 339–48.

—— and Watson, N. 2001, *The Household, Income and Labour Dynamics in Australia (HILDA) Survey: An Introduction to the Proposed Survey Design and Plan*, HILDA Project Discussion Paper no. 1/00, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

Youth Connections National Network 2013, The Space In-Between — Future Policy Directions for Youth at Risk, www.youthconnect.com.au/wp-content/uploads/2013/03/Space-in-between-4web.pdf (accessed 6 June 2013).

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.