Australian Government

Department of Social Services

Australian Institute of Family Studies

Growing Up in Australia:
The Longitudinal Study of Australian Children (LSAC)
LSAC Technical Paper No. 31

The Longitudinal Study of Australian Children

Event history calendar

Jennifer Prattley, Neha Swami, Paul Hockey, Kristen Power, Bosco Rowland and Lisa Mundy

April 2024

Acknowledgements

This report makes use of data from *Growing Up in Australia*: The Longitudinal Study of Australian Children (LSAC). LSAC is conducted in a partnership between the Department of Social Services (DSS), the Australian Institute of Family Studies (AIFS) and fieldwork providers (the Australian Bureau of Statistics (ABS) up to and including Wave 9C, and Roy Morgan, from Wave 10 onwards). Advice is provided by an advisory group of leading researchers.

The authors wish to thank the wider AIFS LSAC team for their contributions to the paper and project. The authors would especially like to thank LSAC participants for their ongoing contribution to the study.

Contents

List of tables

List of figures

Glossary

Term	Description
LSAC	*Growing Up in Australia*: The Longitudinal Study of Australian Children
EHC	Event history calendar
Episode	Period of time with a known start and end date in which the individual had a constant employment status
Long form	A long form dataset has multiple rows per individual

Introduction

Growing Up in Australia: The Longitudinal Study of Australian Children (LSAC) is a major study following the development of 10,000 young people and their families. The study began in 2003 with a representative sample of children from urban and rural areas of all states and territories in Australia. It follows 2 cohorts of children: the 'B' (baby) cohort who were born between March 2003 and February 2004 and the 'K' (kindergarten) cohort who were born between March 1999 and February 2000 (see Mohal et al., 2023 for details).

This technical report provides information about the event history calendar (EHC) data collected in LSAC. These data capture the employment, study and residential history of LSAC respondents, referred to collectively as 'life history data' or 'life episodes'. Episode or life history data are obtained by observing individuals over time, focusing on the timings and types of episodes occurring for the individual under observation (Andersen & Keiding, 2001). These data can be used to study retrospective information on work and careers, relationship patterns and education pathways (Eerola & Helske, 2016).

In this report we focus on the employment data as an example, collected as part of the EHC in Waves 7 and 8 for the K cohort. EHC data were not collected for the B cohort in Waves 7 and 8 as they had not yet reached the age at which the EHC was first implemented for the K cohort (16 years). Moving forward, LSAC will collect EHC data for both the B and K cohorts in Wave 10.[1] Researchers can use EHC data on employment to analyse when individuals transition in and out of work and the length of time spent in employment, as well as to examine relationships between employment trajectories and factors of interest (such as socio-economic status or education). The EHC employment data can be complex to prepare for analysis. Hence, in this paper we provide information about these data and how they can be used, with a focus on how to combine the data files for analysis.

Aims

The aims of this paper are to:

- describe the EHC employment data and how they were collected and stored
- describe a method for preparing a consolidated dataset for analysis
- provide examples of how to interpret and understand the information available in each wave as well as the consolidated file.

EHC employment data

The EHC was administered to each K-cohort Young Person (previously known as Study Child) in Wave 7, in 2016, when they were 16–17 years old, and again in Wave 8, in 2018, when they were 18–19 years old. Each Young Person was asked to complete the EHC interview, which involved the guided completion of a calendar-based data collection instrument. In contrast to the standard LSAC interview, the EHC interview didn't follow a set series of questions. The interview was instead structured by the design of the calendar instrument, with the interviewer asking questions as necessary to obtain sufficient information to complete it.

The calendar provides a visual summary of the Young Person's 'life episodes' for home, study and work for the 2 years prior to their LSAC interview (see below for a precise description of reporting periods). In this paper, we focus on the WORK (employment) EHC module. EHC employment data are available for 95% of the 3,089 Wave 7 K-cohort respondents (*N* = 2,931) and 88% of the 3,037 Wave 8 K-cohort respondents (*N* = 2,673).

What is an episode in the employment module?

The focus of the employment EHC calendar is to record changes in employment, employer, full-time or part-time status, reasons for working and job search activity.

An episode is a period of time in which the respondent had no changes in employment status. The following situations are regarded as a change in employment status:

---

1 Note that in Wave 9C1 and 9C2 both B- and K-cohort children were 16 years+; however, no EHC was collected due to COVID limitations.

- stopping or starting a job (if 2 weeks or more in duration)
- changing from 'working' to 'not working'
- changing employer
- changing full-time or part-time work status.

Employment includes:

- working for pay, profit, commission or payments in kind, in a job or business or on a farm (comprising employees and owner managers)
- working without pay in a family business or on a farm (i.e. contributing family workers).

Employment does not include:

- unpaid domestic work
- work experience/internships
- volunteer community services.

Methods and data collection

During the interview, the young person entered responses using a calendar instrument (see Figures 1 & 2 below). Data were entered episode by episode, with fields to define work/non-work, employer, full-time or part-time status, start and end month and reasons for change. The instrument then populated the visual calendar with the new episode.[2]

Figure 1 shows an example of a completed work episode for a young person who was working part-time at McDonalds and left the job in February to focus on study. Figure 2 shows how this young person's work history appears in the calendar (shaded green). The young person worked at McDonalds from March 2016 to January 2017, was not working between February and May 2017, then worked at Sassy Hair from June 2017 until the time of the interview.

Figure 1: Example of a completed work episode

Figure 2: Example of the completed employment domain time period

---

2 In a few instances, additional relevant data were collected via interviewer-recorded comments which were used during data processing to add or edit episodes. This occurred for *N* = 16 respondents in Wave 7 and *N* = 3 in Wave 8. Item i**d40u2** in the main survey datasets indicates instances where this is the case for a respondent.

The following is paraphrased from interviewer instructions on filling in the employment calendar:

- The calendar provides a visual summary of work–life episodes for the last 2 years/since last interview.
- Each row represents an episode, shaded different colours within the relevant time period to represent periods of work/non-work.
- Each episode or change requires a new row.
- Episodes must be added in chronological order, with the next episode not starting before the start date of any earlier episode.
- Time periods for different episodes can overlap (e.g. working in 3 jobs at one time).
- In Wave 8, some episodes from Wave 7 are prefilled in the calendar and should be edited to indicate if the episode was still current or had ended.
- All months of the calendar must have an episode filled.
- All calendars must end with an episode ending at the current period (e.g. not working, looking for work).

Data files

EHC employment data collected at Waves 7 and 8 are stored in 2 supplementary files (**ehcegrk16** and **ehcegrk18** respectively) separate from the main survey files (**lsacgrk16** and **lsacgrk18**). However, main datasets contain some variables relevant to EHC employment data; for example, number of employment episodes in EHC, consent question, etc.

Both the Wave 7 and 8 EHC datasets are in long form, where one row represents an episode of either work or non-work for an individual. Each individual can have multiple rows, depending on how many episodes they had. The unique identifier **hicid** can be used to identify all episodes for a given person. The Wave 7 EHC file has 3,796 rows that contain employment data for 2,931 respondents. The Wave 8 EHC file has 6,392 rows that contain employment data for 2,673 respondents.

Reporting period for EHC employment data

Figure 3 shows the retrospective reporting period for EHC employment data for Waves 7 and 8. The Wave 7 reporting period (shown in blue) extended from the date of the Young Person's Wave 7 interview back to either the date of their Wave 6 interview or 2 years prior, if they had not taken part in Wave 6. The data from the Wave 7 EHC cover the period from when respondents were aged 14–15 to 16–17.

The reporting period for Wave 8 (shown in green) was from the date of the Wave 8 interview back to either the date of the Wave 7 interview or 2 years prior, for those who had not taken part in Wave 7. The data from the Wave 8 EHC cover the period from when respondents were aged 16–17 to 18–19. Hence, the total reporting period covered by EHC data for Waves 7 and 8 combined is from 2014 (14–15 age) to 2018 (18–19 age).

Figure 3: Reporting period for event history calendar employment data
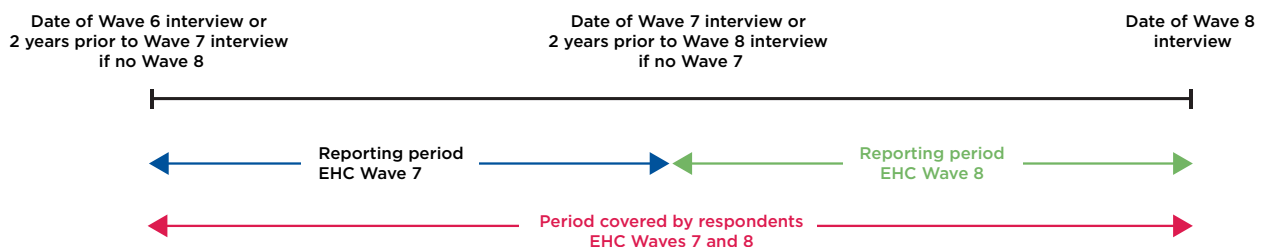
The response rates for the EHC data for Waves 7 and 8 are shown in Table 1. A total of 3,241 individuals completed the EHC in at least one of Waves 7 or 8. Of these, 2,363 (72.9%) completed the EHC in both waves, 568 (17.5%) in Wave 7 only, and 310 (9.6%) in Wave 8 only.

Table 1: Response rates for the EHC data, Waves 7 and 8

Responding Individuals	Obs.	%
Wave 7 only	568	17.5
Wave 8 only	310	9.6
Both waves	2,363	72.9
Total	3,241	100.0

Notes: These numbers refer to employment data collected via the EHC only (numbers may not add up to exactly 100.0% due of rounding). Employment data collected in the main surveys at Waves 7, 8 and later may be available for some or all of these individuals. For 310 individuals, in green in Figure 3, with EHC only available for Wave 8, items relating to employment collected in main surveys at Wave 7 and earlier may still be available.

Source: LSAC employment EHC data. Total *N* = 3,241.

Content

Table 2 shows the descriptions of the variables available in the Wave 7 and Wave 8 EHC employment data files. Further details are available in the LSAC data dictionary. The dictionary also lists variables relevant to the EHC that are stored in main survey Wave 7 and Wave 8 data files.

EHC employment variables can be grouped into 4 themes:

- identifiers and consent variables
- change and characteristics of episodes (e.g. timing, duration, type)
- reasons for stopping work
- job search activity (Wave 8 only).

A single-letter prefix is used in variable names in the EHC and main survey data files. This prefix indicates the age of the study child at the time data were collected, taking the value **i** for age 16–17 years (Wave 7 for K cohort) and j for age 18–19 years (Wave 8 for K cohort). For more on variable naming protocols see the LSAC Data User Guide. In this document we replace the age prefix with * when referring to variables in multiple waves.

Each of the Wave 7 and Wave 8 EHC employment and main survey datasets contains the unique Young Person identifier **hicid**, which can be used to merge the different datasets. The variable **id40u2** indicates if manual editing was required post-data collection in data processing and, if yes, whether extra episodes were created. In the main Wave 7 survey dataset, variables ***id40s1** and ***id40s2** indicated whether consent for EHC data collection was given by the parent and the Young Person, respectively. Consent from both individuals was required for EHC participation. In Wave 8, given respondents were 18–19 years old, specific consent for the EHC was no longer asked of parents or the Young Person, and consent was implied if Young Person completed the EHC Calander.

Table 2: Items relating to EHC employment collected for K cohort

Variable	Question/item	Values	Main W7	Main W8	EHC W7	EHC W8	Notes
Identifier and consent variables							
hicid	Unique Young Person identifier	Number	✓	✓	✓	✓	
*id40s1	If parent provided consent to proceed with EHC	1 Yes; 2 No	✓				
*id40s2	If young person provided consent to proceed with EHC	1 Yes; 2 No	✓				
*id40s7	Whether EHC employment data added to file	1 Yes; 2 No	✓	✓			
*id40u2	EHC editing	1 No editing required; 2 Manual entry from interviewer comments – data not entered in interview using EHC; 3 Editing required – extra employ episodes created	✓	✓	✓	✓	
*datint	Date of wave interview	Date	✓	✓	✓	✓	
Variables related to timing, duration and nature of episodes							
*id40s4	Number of employment episodes in EHC	Number	✓	✓			
*empcount	Number of episodes	Number			✓	✓	
loopindx	Order of repeating items (Starting 1 to 10)	1…10			✓	✓	1 is the oldest episode. Loopindx is unique to the dataset they are in (Wave 7 or 8). The max value of loopindx in Wave 7 is 6.
*ehcesmon	Start month of episode	Number			✓	✓	
*ehcesyea	Start year of episode	Number (Year)			✓	✓	
*ehceemon	End month of episode	Number			✓	✓	
*ehceeyea	End year of episode	Number (Year)			✓	✓	
*ehcelength	Length of episode in months	Number			✓	✓	

Variable	Question/item	Values	Main W7	Main W8	EHC W7	EHC W8	Notes
*ehcenowk	No work	0 No; 1 Yes			✓	✓	
*ehcenofurwk	No further work	0 No; 1 Yes			✓		This indicates non-work episodes where respondents had previous work episodes in the calendar but no further work.
*ehcecurr	If episode related to current employer	0 No; 1 Yes			✓	✓	
*ehcelabl	Employer index	String			✓	✓	Employer indexes are unique to the dataset they are in.
*ehce06	Full-time/part-time status	1 Full time; 2 Part-time			✓	✓	
Variables related to reasons for stopping work							
*ehcerswa	Reasons stopped work – Left/change of employer	0 No; 1 Yes			✓	✓	
*ehcerswb	Reasons stopped work – Changed full-time/part-time	0 No; 1 Yes			✓	✓	
*ehce08a	Reasons stopped work – Physical health problem	0 No; 1 Yes			✓	✓	
*ehce08b	Reasons stopped work – Mental health problem	0 No; 1 Yes			✓	✓	
*ehce08c	Reasons stopped work – Did not like the job	0 No; 1 Yes			✓	✓	
*ehce08d	Reasons stopped work – Problems with others/being bullied	0 No; 1 Yes			✓	✓	
*ehce08e	Because of discrimination	0 No; 1 Yes			✓	✓	
*ehce08f	Did not need the money	0 No; 1 Yes			✓	✓	
*ehce08g	Pregnancy/caring for own children	0 No; 1 Yes			✓	✓	Confidentialised. These variables were converted to -9 missing for general release. ehce08g was only selected 4 times in both waves.
*ehce08h	Other caring responsibilities	0 No; 1 Yes			✓	✓	

Variable	Question/item	Values	Main W7	Main W8	EHC W7	EHC W8	Notes
*ehce08i	Child care too expensive, unsuitable	0 No; 1 Yes			✓	✓	Confidentialised as per ehce08g above
*ehce08j	Studying, returning to study	0 No; 1 Yes			✓	✓	
*ehce08k	Lost job	0 No; 1 Yes			✓	✓	
*ehce08l	Job ended/temporary/ seasonal	0 No; 1 Yes			✓	✓	
*ehce08m	Unsatisfactory work arrangements	0 No; 1 Yes			✓	✓	
*ehce08n	Self-employed business closed down	0 No; 1 Yes			✓	✓	
*ehce08o	Study child moved to another location	0 No; 1 Yes			✓	✓	
*ehce08p	Other (specify)	0 No; 1 Yes			✓	✓	No specify option was available in Wave 7. The description in Wave 7 was simply 'Other'.
*jothrea	Enter other reason stopped work	String				✓	This variable is only available in Restricted Release datasets.
Variables regarding job search activity							
*look	Actively looking for work	1 Yes; 2 No				✓	
*mthslk	Months actively looking for work	Number				✓	
*mthsst	Whether could have started work immediately if found job	1 Yes, all months; 2 Yes, less than all months; 5 No				✓	

Understanding EHC data with an example

Tables 3a and 3b provide a glimpse of parts of EHC data available for the Young Person with hicid = 51101070 and explain how to interpret the data.

Table 3a: Example: Person with HICID 51101070 Wave 7 data file

hicid	iempcount	iid40u2	idatint	loopindx	iehcenowk	iehcenofurwk	iehcecurr	iehcelabl	iehcesmon	iehcesyea	iehceemon	iehceeyea	iehcelength	iehcerswa	iehcerswb	iehce06
51101070	2	No editing required	9-May-16	1	No	No	No	Employer 1	2	2015	9	2015	7	Yes	No	Part-time
51101070	2	No editing required	9-May-16	2	No	No	Yes	Employer 2	12	2015	-9	-9	-9	-9	-9	Part-time
		This Young Person had 2 episodes in the Wave 7 dataset, which were reported at the time of their Wave 7 interview in May 2016.						The first episode with Employer 1 began in February 2015 and ended in September 2015, giving a duration of 7 months.			The second episode with Employer 2 began in December 2015 and because it was ongoing, had no end date.					Both employment episodes were part-time work.

Table 3a shows the Wave 7 EHC employment dataset for the respondent with HICID 51101070. The respondent had 2 episodes (**iempcount**) recorded on the date of their Wave 7 interview, 9 May 2016 (**idatint**). We have sorted the data by **loopindx** so that the earliest episode appears first and the most recent appears last.

The first episode was of work (**iehcenowork = No**), [equivalent statement about iehcenofurwork], and was not for the respondent's current employer (**iehcecurr = No**). The month and year variables tell us the respondent started that work episode in February **(iehcesmon = 2)** 2015 (iehcesyea) and ended in September (**iehceemon = 9**) 2015 (**iehceeyea**). That is a total duration of 7 months (iehcelength). The respondent stopped this work episode because they changed employer (**iehcerswa = Yes**) but didn't change from part-time to full-time (**iehcerswb = No**). This employment episode was part-time (**iehce06**). The second episode was also of work for the respondent's current employer. It began in December 2015 and was for part-time work. Since at the time of interview the work episode was ongoing, there is no end date and **iehceemon**, **iehceeyea** and **iehcelength** are all set to -9. Since the respondent is currently in this work episode, there is no reason for stopping it, so **iehcerswa** and **iehcerswb** are also set to -9.

Table 3b: Example: Person with HICID 51101070 Wave 8 data file

hicid	jempcount	jid40u2	jdatint	loopindx	jehcenowk	jehcecurr	jehcelabl	jehcesmon	jehcesyea	jehceemon	jehceeyea	jehcelength	jehcerswa	jehcerswb	jehce06
51101070	2	No editing required	13-Aug-18	1	No	No	Employer 1	5	2016	5	2017	12	Yes	No	Part-time
51101070	2	No editing required	13-Aug-18	2	No	Yes	Employer 2	6	2017	-9	-9	-9	-9	-9	Full-time
		This Young Person had 2 episodes in the Wave 8 dataset, which were reported at the Wave 8 interview in August 2018.						The first episode was part-time work with a start date of May 2016 and an end date 12 months later in May 2017.				The Young Person moved on to a new episode in June 2017 with a new employer, it was a full-time position that they were still in at the time of their interview.			

Table 3b shows the Wave 8 EHC employment dataset for the respondent with HICID 51101070. The respondent had 2 episodes (**iempcount**) recorded on the date of their Wave 8 interview, 13 August 2018 (**idatint**). Again, data is sorted by **loopindx**, so the earliest episode appears first and the most recent appears last. The first episode was of work (**iehcenowork = No**) and was not for the respondent's current employer (**iehcecurr = No**). The month and year variables tell us the respondent started that work episode in May (**iehcesmon = 5**) 2016 (**iehcesyea**) and ended in May (**iehceemon = 5**) 2017 (**iehceeyea**). That is a total duration of 12 months (**iehcelength**). The respondent stopped this work episode because they changed employer (**iehcerswa = Yes**) but didn't change from part-time to full-time (**iehcerswb = No**). This employment episode was part-time (**iehce06**). The second episode was also of work for the respondent's current employer. It began in June 2017 and was for full-time work. Since at the time of interview the work episode was ongoing, there is no end date and thus, **iehceemon**, **iehceeyea** and **iehcelength** are all set to -9. Since the respondent is currently in this work episode, there is no reason for stopping it, so **iehcerswa** and **iehcerswb** are also set to -9.

Constructing a consolidated dataset

Although the 2 employment data files can be analysed as discrete datasets, they can also be combined, enabling users to construct individual employment histories that span a longer period. The resulting dataset would contain reconciled episodes from the 2 original files with 'seam effects' removed, where possible. Seam effects are a form of measurement error that can arise where data are linked wave by wave. The possibility of recall error and combining data from repeated panel observations to create continuous histories can lead to a concentration of transitions or abnormal heaping of data at the seam between 2 waves of a panel. In the context of the LSAC EHC, seam effects could occur where an episode of employment was reported in both Wave 7 and Wave 8 but with conflicting information. Thus, the possibility of measurement error increases when 2 waves of data are joined, as there may be measurement errors in the first wave and the second wave (Lynn, 2009; O'Muircheartaigh, 1996).

In theory, there should be no gaps or overlapping episodes occurring at the join between Waves 7 and 8, because individuals at Wave 8 were asked to retrospectively report from the time of the Wave 7 interview up until the time of the Wave 8 interview. However, recall error could lead to erroneous reporting of start and/or end dates of episodes, employer or full-/part-time status, giving rise to inconsistencies between Waves 7 and 8 interviews. For example, a small number of respondents with data in Wave 7 and Wave 8 (<3%) have their first Wave 8 episode beginning earlier or later than their Wave 7 interview date. Data users should give careful consideration on how to address these issues if they want to combine data files across waves.

In selecting the appropriate approach for combining EHC data from both waves, data users should consider the number and nature of episodes, the relationship between successive episodes, dependence structure and research questions (Amorim & Cai, 2015). Methodological challenges arise when combining complex sequences of episode history data, such as incomplete information in the form of censoring (e.g. some trajectories can be right-censored, being observed only for a sub-period) and truncation (e.g. left truncation where start times of some episodes are unknown). Data users should understand the assumptions they make to model transitions carefully, because including or excluding partially observed trajectories impact the interpretation of the results.

This section provides an example approach for combining the Wave 7 and Wave 8 EHC employment datasets (**ehcegrk16** and **ehcegrk18**) to form a single consolidated dataset of employment episodes that data users can use to analyse employment transitions. The approach involves appending episodes stored in the Wave 8 dataset to those in Wave 7, identifying instances where 2 or more episodes could be collapsed into a single episode, addressing inconsistencies in overlapping episodes and handling missing date information. New values are generated for items such as employer index, number and duration of episodes in this dataset. The process is explained in detail below.

Step 1: Combining Wave 7 and Wave 8 datasets

A single combined EHC employment dataset could be created by appending all records in the Wave 8 dataset (*Total episodes* = 6,392) to the Wave 7 dataset (*Total episodes* = 3,796). An initial unedited version of the dataset would have *N* = 10,188 records, representing 3,241 individuals, grouped using the **hicid** identifier. Before appending, the age prefix ('i' for Wave 7 and 'j' for Wave 8) should be removed from all variable names, except **hicid** and **loopindx**, which don't include the prefix.

Step 2: Editing the combined dataset

Actions taken to edit and clean the combined dataset include:

- addressing unknown start and end date information
- determining whether any episodes reported in Wave 8 could be collapsed or merged with any reported at Wave 7
- checking for inconsistencies in overlapping episodes
- generating new values for some variables where required.

Formatting and cleaning date information

Start and end dates of episodes are available at the year and month level. Information in ***ehcesmon** (start month of episode) and ***ehcesyea** (start year of episode) can be combined to generate a single indicator of when an episode began, and ***ehceemon** and ***ehceeyea** can be combined for a single indicator of when an episode ended.

Table 4: Start and end dates availability in the initial combined EHC employment dataset

	Wave 7 EHC	Wave 8 EHC
Both start dates and end dates available	1,020	4,084
Period of work	735	1,972
No work	285	2,112
Only start date available	1,485	2,308
Period of work	1,485	2,308
Only end date available	34	
Period of work	34	
No start and end dates available	1,257	
Period of work	112	
No work	1,144*	
Not determinable	1	
Total episodes	3,796	6,392

Notes: *In Wave 7, both start dates and end dates are missing for 1,144 episodes because those respondents were not in employment at all during the EHC period (i.e. they have a single non-work episode in the calendar).
Source: LSAC employment EHC data

As shown in Table 4, in the Wave 7 dataset, both start and end dates are available for 1,020 of the 3,796 episodes. Of these 1,020 episodes, 735 are for the periods of work and 285 are of no work. There are 1,485 episodes that only have start dates available and all are for periods of work. Thirty-four episodes have only end date available and all are periods of work. There are 1,257 episodes in the Wave 7 EHC employment dataset that do not have both end or start dates and the majority of them are periods of non-work (1,144). For one episode with an unknown start and end date, it is not possible to determine whether it was for a period of non-work or work (***ehcenowk** unknown).

Both start and end dates are available for 4,084 out of all of the 6,392 episodes in the Wave 8 dataset. The 2,308 episodes with unknown end dates are periods of work that were ongoing at the time of the Wave 8 interview (***ehcenowk = No**, ***ehcecurr = Yes**).

For non-work episodes, the discrepancy in unknown start and end dates in Wave 7 compared to Wave 8 is due to differences in how non-work was recorded. In Wave 7, when respondents did not work at all in the EHC period, this was recorded as a single episode of 'no work' (***ehcenowk = Yes**) with no start or end date. Non-work episodes for respondents who did work within the EHC period were instead recorded as 'no further work' (***ehcenofurwk = Yes**) with start and end dates recorded for all episodes (*n* = 285). In Wave 8, all non-work episodes were recorded as 'no work' (***ehcenowk = Yes**) regardless of whether a respondent worked in the EHC period, with start and end dates recorded for all episodes.

How each of the episodes with unknown start or end dates are treated should depend on whether the episode is for a period of non-work or work, and the Young Person's participation in the relevant waves. The next section describes in detail the treatment of these episodes.

Combining episodes

Once data users know the extent of total missing information on start and end dates for both Wave 7 and Wave 8 episodes in the combined employment EHC dataset, the next step is to address that missingness by editing episodes to create a continued employment history that covers the entire reporting period from 2014 (14–15 age) to 2018 (18–19 age). In this example, we suggest some methods that could be adopted for various scenarios observed in the combined dataset:

- Episodes that are fully contained within the reporting period for each wave (i.e. they started and ended prior to the date of interview) can retain the original start and end dates as they do not overlap the reporting period for another wave.

- If an episode of non-work reported in Wave 7 is contiguous to one reported in Wave 8, the 2 could be collapsed to form one single episode of non-work in the consolidated dataset. In such cases, the start date of the single episode would be the earliest of the start dates (the Wave 7 start date) and the end date would be the latest of the end dates (the Wave 8 end date).

- There are individuals in work at the time of their Wave 7 interview who have no end date for that episode in the Wave 7 EHC file. In this case, the 2 scenarios are: (a) if they did not take part in Wave 8, then the end date in the consolidated file is the date of the Wave 7 interview or (b) if they did take part in Wave 8, then episodes of work reported at the Wave 8 interview are assessed to determine if they are a continuation of the episode of work that was ongoing at the time of the Wave 7 interview. For our example, we will designate a Wave 8 work episode as a continuation of one reported in Wave 7 if all of the following criteria are met:

  a. The Wave 7 episode represented work that the respondent was in at the time of the Wave 7 interview (**iehcecurr = Yes**).

  b. The start month (**jehcesmon**) and year (**jehcesyea**) of the Wave 8 episode were either earlier than, the same as, or immediately following the month (**iehcesmon**) and year (**iehcesyea**) of the Wave 7 interview.

  c. For the Wave 8 episode, neither the indicator of change in employment (**jehcerswa**) nor the indictor of change in full-time or part-time status (**jehcerswb**) had value 1 (Yes).

  d. The Wave 7 episode and Wave 8 episode have the same full-time/part-time status value (***ehce06**).

If the above criteria for continuation are met, the Wave 7 and Wave 8 episodes in question can be merged or combined as one single episode that spanned the Wave 7 and Wave 8 reporting periods. The start date is taken as that reported in the Wave 7 interview and end date as reported in Wave 8. Only one episode from Wave 7 and one episode from Wave 8 for the respondent meet the above criteria.[3]

Episodes that cannot be combined

Episodes of work in Wave 7 that were ongoing at time of interview (**iehcecurr = Yes**) but cannot be combined with an episode reported at Wave 8, remain separate, distinct episodes. These can be allocated an end month (**iehceemon**) and year (**iehceeyea**) of the respondent's Wave 7 interview date and have the length of episode (**iehcelength**) calculated accordingly.

Similarly, episodes of work in Wave 8 that are overlapping or consecutive with those reported in Wave 7 but not identified as a continuation of a Wave 7 episode, such as where multiple jobs were held concurrently, are retained as separate distinct episodes.

Considerations for combining contiguous episodes

When combining contiguous Wave 7 and Wave 8 episodes, consideration should be given to the following:

- The Waves 7 and 8 episodes will have different dates of interview (**datint**). Whether to retain either, both or exclude this variable will depend on the research question. If both dates are considered relevant, we recommend retaining the age prefix for ***datint** in Step 1 to include both **idatint** and **jdatint** in the file.

---

3 The employer index (**ehcelabl**) in each of the Wave 7 and Wave 8 datasets is not useful in the process of identifying collapsible episodes. Index numbers refer to the order of employers within each wave and were not carried across waves. Consequently, an episode of work for Employer 1 in Wave 8 would not necessarily be a continuation of an earlier episode reported in Wave 7 that was also labelled Employer 1.

- Although EHC editing (**id40u2**) for most episodes will be coded 1 ('No editing required'), there may be cases where one of the two combined episodes will have a different value (2 or 3). If that information is important to the analysis, we recommend using the value for the combined episode that indicates the episode was edited.
- Combined episodes should retain the 'current employer' (**ehcecurr**; yes/no) value from the Wave 8 episode.
- Length of completed episode in months (**ehcelength**) must be recalculated for all combined episodes where the Wave 8 episode has a non-missing value for this variable (i.e. the episode was completed).

Count and index variables

The number of episodes (***empcount**), episode index (**loopindx**) and employer index (***ehcelabl**) values are specific to the original data files (i.e. 2-year time period corresponding to each wave). These values can remain unchanged for Young People who only took part in one of the two waves. However, in a combined file they would no longer be meaningful for Young People who participated in both Wave 7 and Wave 8. For example, when the datasets are appended, the first episode in Wave 7 and first episode in Wave 8 will both have an episode index (**loopindx**) value of 1, which is incorrect as the Wave 8 episode is no longer the first chronologically. The description below shows methods of regenerating these variables for Young People who participated in Wave 7 and Wave 8.

- The value for number of episodes (**empcount**) can be recalculated as the number of episodes for each respondent in the combined dataset. This value will be the same for all episodes for a particular Young Person.
- The values for the episode index (**loopindx**) can be regenerated in the combined data file. Because the start date of each episode cannot be before the start date of the previous episode (see Methods and data collection), episode index values can be re-assigned sequentially in order of the start month (**jehcesmon**) and year (**jehcesyea**) of each episode for a Young Person. The episode with the earliest start date for the Young Person will have an episode index (**loopindx**) of 1. The episode with the latest start date will have an episode index (**loopindx**) of the total number of episodes for that Young Person (this is the same value as **empcount** above).
- The values for the employer index (**ehcelabl**) can be regenerated to some extent in the combined dataset with some caveats as described below. This can be done by retaining the values for Wave 7 employers, then reassigning index numbers to the Wave 8 employers starting after the highest employer index number in Wave 7. For example, if a Young Person had 2 different employers in each wave with no overlapping episodes – Employers 1 and 2 in Wave 7, and Employers 1 and 2 in Wave 8 – the Wave 8 employers would become Employers 3 and 4 in the combined dataset. Any matching employers within a wave should retain matching employer index values. If any Wave 7 and Wave 8 episodes are combined, any employers in Wave 7 with the same employer as the Wave 7 combined episode and any employers in Wave 8 with the same employer as the Wave 8 combined episode would be given the same employer index value.

  A hypothetical example with the original and regenerated episode (**loopindx**) and employer indexes (**ehcelabl**) for a Young Person can be seen in Figure 4. In Wave 7, this Young Person changed employer after episode 1 and recorded 2 episodes with the second employer. In Wave 8, they began at one employer, moved to a second employer, then returned to the first employer. The last episode in Wave 7 and first in Wave 8 were able to be combined (combined episode 3).

Figure 4: An example of combining Wave 7 and Wave 8 loop index (loopindx) and employer index (ehcelabl) in a consolidated file
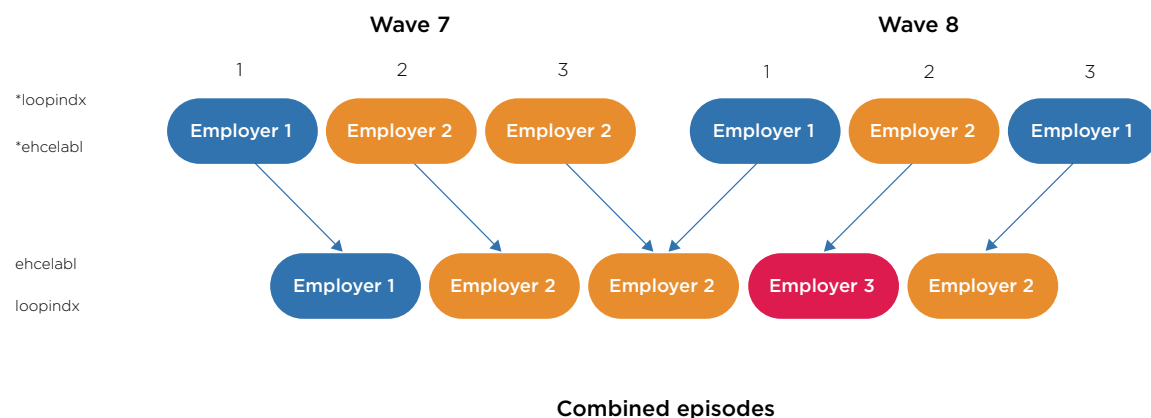
Because 2 episodes were combined, we can assume Employer 2 in Wave 7 and Employer 1 in Wave 8 are the same employer (combined Employer 2). We can therefore identify that episodes 2 and 3 in Wave 7 and episodes 1 and 3 in Wave 8 are all with the same employer. These can be given the same employer index number in the combined dataset (combined Employer 2). The remaining employer in Wave 8 is then given the next available index number (combined Employer 3).

Caution must be taken in interpreting regenerated employer indexes (**ehcelabl**). As noted earlier, employer index (***ehcelabl**) values are not retained from one wave to the next. It is impossible to identify whether an employer in one wave is the same as in another unless that employer is in episodes in both waves that are contiguous and can be combined, as in the above example. When a Young Person has EHC data in both waves, regenerated employer indexes will not accurately differentiate employers between files in the following cases:

a. No episodes can be combined between the waves. In this case, changes of employer between Wave 7 and Wave 8 also cannot be identified.

b. Two episodes can be combined between the waves but there is more than one different employer in both waves. In this case, all changes of employer can still be identified.

An illustration of case B can be seen in Figure 4. In this example, we can identify that Employer 2 in Wave 7 and Employer 1 in Wave 8 are the same employer. For the other employers in Wave 7 (Employer 1) and Wave 8 (Employer 2), we can identify that they are different to the employer in the combined episode but not whether they are the same or different employers to each other.

Understanding consolidated EHC dataset through an example

Example: Person with HICID 51101070

For this example, we continue with person HICID 51101070, whose Waves 7 and 8 datasets we discussed separately in Understanding EHC data with an example. This individual has the following recorded employment episodes:

For Wave 7, 2 episodes were reported at the time of their Wave 7 interview in May 2016, as shown in Table 2a previously:

● The first episode of part-time work with employer 1 began in February 2015 and ended in September 2015, giving a duration of 7 months.

● The second episode of part-time work with employer 2 began in December 2015 and, because it was ongoing, had no end date.

For Wave 8, 2 episodes were reported at the Wave 8 interview in August 2018, as shown in Table 2b previously:

● The first episode of work was part-time, started in May 2016, and ended 12 months later in May 2017.

● The second episode of work with a new employer was full-time, started in June 2017 and, because it was ongoing at the time of the interview, had no end date.

First, we simply create an appended dataset. Table 5a shows the appended dataset for the person with HICID 51101070 where the age prefix ('i' for Wave 7 and 'j' for Wave 8) has been removed from all variable names and only variables common to both datasets have been retained for brevity.

Table 5b shows the combined data for the person with HICID 51101070 where the middle 2 episodes have been combined and **empcount**, **loopindx**, **ehcelabl**, **ehcelength** have been edited following the rules discussed earlier:

● one episode of part-time work from February 2015 to September 2015, recorded in the Wave 7 dataset

● one episode of part-time work from December 2015 to May 2017.

This is derived from combining the ongoing work episode in the Wave 7 dataset starting in December 2015 with the work episode recorded as starting in May 2016 (Wave 7 interview, i.e. a proxy for any work date on or before the Wave 7 interview, hence likely ongoing) and ending in May 2017. The amalgamation of 2 episodes could occur because the relevant:

● Wave 7 episode was ongoing at the time of that interview, both were for part-time work, and the start date of the Wave 8 episode was the same as the Wave 7 interview date.[4]

● There was one episode of full-time work from June 2017 to at least August 2018.

---

4 There is the possibility that these episodes were in fact for 2 different employers and should not be amalgamated, with cleaning and editing of life history data rarely error free. In that case, the respondent would retain 4 episodes and the end date of the second and start date of the third would both be May 2016.

Table 5a: Example: Appended dataset from both Waves 7 and 8 for person with HICID 51101070

	hicid	empcount	id40u2	datint	loopindx	ehcenowk	ehcecurr	ehcelabl	ehcesmon	ehcesyea	ehceemon	ehceeyea	ehcelength	ehcerswa	ehcerswb	ehce06	
Wave 7	51101070	2	No editing required	9-May-16	1	No	No	Employer 1	2	2015	9	2015	7	Yes	No	Part-time	
	51101070	2	No editing required	9-May-16	2	No	Yes	Employer 2	12	2015	-9	-9	-9	-9	-9	Part-time	To be combined
Wave 8	51101070	2	No editing required	13-Aug-18	1	No	No	Employer 1	5	2016	5	2017	12	Yes	No	Part-time	
	51101070	2	No editing required	13-Aug-18	2	No	Yes	Employer 2	6	2017	-9	-9	-9	-9	-9	Full-time	

Table 5b: Example: Person with HICID 51101070 combined data file

	hicid	empcount	id40u2	datint	loopindx	ehcenowk	ehcecurr	ehcelabl	ehcesmon	Ehcesyea	ehceemon	ehceeyea	ehcelength	ehcerswa	ehcerswb	ehce06
Combined	51101070	3	No editing required	9-May-16	1	No	No	Employer 1	2	2015	9	2015	7	Yes	No	Part-time
	51101070	3	No editing required	9-May-16	2	No	No	Employer 2	12	2015	5	2017	18	Yes	No	Part-time
	51101070	3	No editing required	13-Aug-18	3	No	Yes	Employer 3	6	2017	-9	-9	-9	-9	-9	Full-time

Notes: We can't tell whether Employers 1 and 3 are the same or different employers for reasons explained earlier. For the combined episode we retain date of Wave 7 interview, given it is assumed that episode is a continuation from Wave 7. Green = taken from Wave 7 episode; Purple = taken from Wave 8 episode and Orange = regenerated/recalculated.

Source: LSAC employment EHC data

Summary

This technical report presents detailed information on the EHC, with a focus on employment data, to aid researchers in understanding the method of collection and the key variables. The report also provides an example of one of the ways in which data users can combine EHC employment data from across waves to form a consolidated dataset. Examples are provided on how to interpret and understand the EHC employment information available in each wave, as well as information on how to create a consolidated EHC employment data file.

There are some caveats that users must keep in mind when using the EHC employment data. The most common drawback of retrospectively collected employment data is the potential for recall error – the inability of respondents to accurately recall circumstances from their past (Solga, 2001) and that respondents might oversimplify their work careers, and under-report or over-report certain episodes (Manzoni et al., 2010). All these issues may be compounded when consolidating multiple waves of EHC employment data.

Some useful resources on combining event history data and how to model them in the analysis are:

- Amorim, L. D., & Cai, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology*, 44(1), 324–333.
- Annual Reviews, Event History Analysis: www.annualreviews.org/doi/10.1146/annurev-statistics-022513-115558.
- Gash, V. (2010). What is Event History Analysis?: hummedia.manchester.ac.uk/institutes/methods-manchester/docs/eha.pdf.

References

Amorim, L. D., & Cai, J. (2015). Modelling recurrent events: A tutorial for analysis in epidemiology. *International Journal of Epidemiology*, *44*(1), 324–333.

Andersen, P. K., & Keiding, N. (2001). Event-history analysis in continuous time. In (Eds.) N. J. Smelser & P. B. Baltes, *International encyclopedia of the social & behavioral sciences* (pp. 4946–4956). Pergamon. doi.org/10.1016/B0-08-043076-7/02103-3

Eerola, M., & Helske, S. (2016). Statistical analysis of life history calendar data. *Statistical Methods in Medical Research*, *25*(2), 571–597.

Lynn, P. (2009). Methods for longitudinal surveys. *Methodology of Longitudinal Surveys*, 1–19.

Manzoni, A., Vermunt, J. K., Luijkx, R., & Muffels, R. (2010). 2. Memory bias in retrospectively collected employment careers: A model-based approach to correct for measurement error. *Sociological Methodology*, 40(1), 39–73.

Mohal, J., Lansangan, C., Gasser, C., Howell, L., Hockey, P., Duffy, J. et al. (2023). Growing Up in Australia: *The Longitudinal Study of Australian Children – Data User Guide, Release 9.1C2*, April 2023. Melbourne: Australian Institute of Family Studies. doi:10.26193/QR4L6Q

O'Muircheartaigh, C. (1996). Measurement errors in panel surveys: Implications for survey design and for survey instruments. *Società Italiana di Statistica, Atti della XXXVIII Riunione Scientifica, Rimini: Maggioli Editore*, 1, 219–230.

Solga, H. (2001). Longitudinal surveys and the study of occupational mobility: Panel and retrospective design in comparison. *Quality and Quantity*, *35*, 291–309.