The *Journeys Home* Accommodation Event-History Files

Abraham Chigavazira and David Ribar[1]
Melbourne Institute of Applied Economic and Social Research
December 2016

Introduction

The *Journeys Home* (JH) survey is a longitudinal, interview-administered survey of 1,682 disadvantaged Australians who were initially either homeless or at high risk of homelessness. JH followed these subjects through six waves of interviews which were spaced approximately six months apart. A unique feature of JH is an Accommodation Calendar that asked people to report each place that they resided at and between interviews and that is available to licensed users as a stand-alone file (JH_ACCCAL) with the different releases of the JH data (see Bevitt et al. 2014 for more information). For each place that people stayed, the calendar includes measures of the type of accommodation (e.g., living in own place, living with parents, sleeping rough), the date that the person began living in that accommodation, and meta data indicating the wave and date in which the information was first recorded and indicating possible problems in recording the data. In principle, the Accommodation Calendar can be used for event-history analyses of the durations of accommodation spells and transitions between accommodations. However, such analyses require several additional measures that were not included in the calendar file. They also require researchers to address some problems in the Accommodation Calendar, such as missing start dates or accommodation types, breaks in the calendar, and inconsistent entries.

These issues, which are moderately complex and require considerable understanding of the data, will be common to all researchers who wish to conduct event-history analyses (see, e.g., Cobb-Clark et al. 2016). Accordingly, we have written programs in STATA (jh_acc_eh.do) and SAS (jh_acc_eh.sas) that apply consistent rules to the Accommodation Calendar and other information available from the JH general files and output three event-history files that can be used in alternative types of analyses:

- ACCSPELL, a file with spell-level observations that can be used in descriptive event-history analyses or in multivariate event-history analyses with fixed (non-time-varying) covariates;
- ACCWSPELL, a file with wave-within-spell-level observations that can be used in continuous-time event-history analyses with covariates that are fixed within waves but might vary across waves; and
- ACC3SPELL, a file with third-of-a-month-within-spell-level observations that can be used in discrete-time event-history analyses with covariates that might vary by third-of-a month (10-11-day) intervals.

This document describes the programs and how the event-history data are created from the JH Accommodation Calendar. It describes the Accommodation Calendar, discusses some of the problems with the data, outlines the strategies that the programs take to address these problems and create event-history measures, and documents the contents of the resulting files.

---

[1] Corresponding author: david.ribar@unimelb.edu.au. The authors thank Andrew Bevitt, Nicolas Herault and Yi-Ping Tseng for helpful advice and comments.

Journeys Home **Accommodation Calendar Description and Issues**

The JH Accommodation Calendar collected information on each subject's accommodation type at the time of the first (Wave 1) interview and housing changes from Waves 2 to 6. For these subsequent interviews, respondents answered questions on housing changes between their last interview and their current interview (the period between interviews is called the "reference period"). The calendar asked respondents to give information of the accommodation type they moved to, and when they moved into that accommodation as shown in Figure 1. The maximum number of changes in accommodation that could be collected per reference period was 21.

Figure 1: Accommodation Calendar Data Structure

Accommodation 0	Accommodation type 0	Start date 0
Accommodation 1	Accommodation type 1	Start date 1
Accommodation 2	Accommodation type 2	Start date 2

The Accommodation Calendar follows a "wide format" with a single record for each individual, with arrays of variables for each spell capturing accommodation type, start date, which wave spell was recorded, and any recoding done on spell start date (see Bevitt et al. 2014 for full documentation). There were a total of 7,234 spells in the Accommodation Calendar.
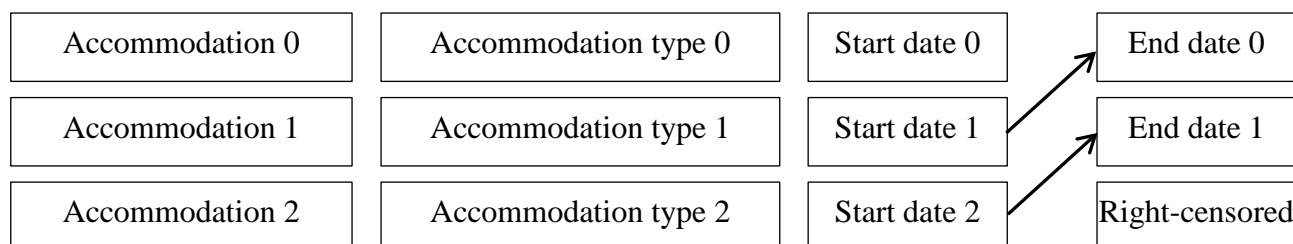
Dates are not recorded in a fine continuous time measure. Respondents were asked to state whether the accommodation spell started at the beginning, middle or end of the month. Thus dates are recorded in thirds of the month were the first third is the 1st of the month, 2nd third is the 11th of the month and the last third is the 21st of the month. The nature of date collection creates a number of issues addressed by the program.

These data can potentially be used in event-history analyses of the durations spent in a given accommodation and transitions between accommodations. To conduct an event-history analysis, however, analysts need additional information,
- the date when they first observe the person in the spell,
- an indicator for whether this date is the actual beginning of the spell (an indicator for whether the spell is "left-censored"),
- the date when they last observe the person in the spell, and
- an indicator for whether this date is the actual end of the spell (an indicator for whether the spell ends in an exit or is "right-censored").

In theory, these additional measures are easily derived from the available information. Consider the example from Figure 1 and suppose that all of the start dates are available. We can take the start dates of spells 1 and 2 as the end dates of spells 0 and 1, as shown in Figure 2. Because there is no spell (and hence no start date) following spell 2, we would mark its end date as missing or censored.

Figure 2: Transforming Accommodation Calendar to Event-History Data

Accommodation 0	Accommodation type 0	Start date 0	End date 0
Accommodation 1	Accommodation type 1	Start date 1	End date 1
Accommodation 2	Accommodation type 2	Start date 2	Right-censored

In practice, however, there are issues that arise in the Accommodation Calendar that make the transformation more difficult. Also, there are other types of measures that we would want to include, such as measures for the duration of the spell, the type of the preceding spell, and the type of the following spell.

A number of issues are addressed by program to create the spell data. These issues include assigning dates for right and left censored spells, addressing breaks in spells, treatment of recoded dates and calculation of duration. All these issues are addressed when the ACCSPELL file is created and are also incorporated into the ACCWSPELL and ACC3SPELL files.

Left-censored spells. 310 Accommodation Calendar spells have missing start date (MCWHEN#) values and are left-censored. For these spells, the programs set SPLCENS to one and insert the earliest observed date as the observed start date (SPSTART), if one is available. We applied two rules. If the spell was on-going as of the interview date (MCINT# was one), we used the interview date as the observed start date. If the spell began and ended between interviews (MCINT# was zero), we used the start date of the next spell as the observed start date if that date was non-missing and left the start date as missing if the start date of the next spell was missing.

Right-censored spells. Right-censoring occurs when an accommodation is the last that is observed for someone, when the start date for the next spell is missing, or when there is a break between spells. For these spells, the programs set SPEXIT to zero and insert the latest observed date as the observed end date (SPEND). If a spell was on-going as of an interview date (MCINT# was one), we use the HCHANGE variable taken from the JHGEN_LONG (available in the JHLRD_LONG, JHLIM_LONG or JHOS_LONG file) file to identify the subsequent waves when the person was still in the same accommodation and use the date of the last such interview as the observed end date. If the spell began and ended between waves, we used the spell's start date, if one was available, as the end date.

Breaks in spells. The Accommodation Calendar has a variable, MCALL#, that indicates whether all the spells during a reference period were reported (conversely, whether there were "breaks" in the spells). The indicator is needed because the calendar sometimes stopped collecting information during a reference period. This could occur if the person refused to continue the calendar, failed to provide a start date AND accommodation type for a spell, or provided more than 20 spells during a reference period. In these cases, the Accommodation Calendar contains the spells within the reference period (if any) that preceded the break and the spell that overlapped the interview. To address breaks, we inserted an artificial left-censored spell at the position of the break, applied the rules given in the previous paragraph for addressing right-censored spells, and then removed the artificial spell.

3

Other issues. Some spell start dates in the Accommodation Calendar were re-coded by the JH team to correct for overlapping seams, implausible dates and missing spells (see Bevitt et al. 2014 for more information). In the calendar data the variable that indicates recoded start dates is MCDTR and is named SPDTR in the spell data. We treated these recoded dates as accurate recodes so no treatment was done to the data.

Description of the event-history files

ACCSPELL. The file ACCSPELL contains one observation per spell (7,234 observations total), thus multiple observations for most respondents. Each observation is uniquely identified by a cross-wave person identifier, XWAVEID, and a spell number, SPNUM. Each observation also contains the information from the Accommodation Calendar, including the accommodation type, SPACC, the reported start date, SPWHSD, and meta data for the survey wave in which the spell was first reported, JHWAVE, the date of the interview when the spell was reported, WIDT, whether the spell overlapped with the interview date when it was reported, SPINT, whether the spell date was recoded, SPDTR, whether all spells in the reference period were reported, WCMPL, and whether the spells appear to have been reported in consecutive order during the reference period, WORDR. As mentioned, the observations also contain measures for the spell start and end dates, SPSTART and SPEND, left-censoring, SPLCENS, and a completed spell, SPEXIT. All date variables are stored in standard SAS or Stata date formats.

Additionally, the file includes the spell duration, SPTDUR, which is calculated as SPEND – SPSTART + 1; thus, all spells have a duration of at least one. The observation also has the accommodation type of the previous spell, SPLSTACC, and of the following spell, SPNXTACC. Table 1 lists the variables and the summary statistics.

Some more notes and descriptions of the measures:
- XWAVEID and SPNUM identify spells
- XWAVEID and JHWAVE can link the data back to JHGEN_LONG (or JHLRD_LONG, JHLIM_LONG, or JHOS_LONG)
- SPNUM: spell number. Gives the sequence of the accommodation spells. In the calendar data the first accommodation spell is spell zero.
- SPEND: spell end date is the start date of the next accommodation type. Spell end date is imputed to be the interview date if that is the final wave that person responded in the Journeys Home survey. Spell end date is set to spell start date for right censored spells were the next spell start date is not recorded in the calendar data.
- SPSTART: spell start date from the calendar data. Left censored spells have an imputed start date. Those whose spell is left censored and that spell coincides with being the current accommodation at the interview date we impose the start date to be the interview date. For left censored spells that do not coincide with the interview then the start date is start date of the next spell (if the date is recorded).
- SPTDUR: total spell duration. Spell duration is calculated as SPEND minus SPSTART plus one.
- SPLCENS: spell is left censored if it has missing start date. An approximate start date is then imputed for all possible cases.
- SPEXIT: spell ended with exit. Spells coded as did not end as an exit are right censored spells. The end date is imputed for those spells as described in the SPEND variable description.
- SPNXTACC: next accommodation type. This derived from taking the next spell's

accommodation type. The missing value codes are -98 for unknown and -99 for unknown because spell information is not complete (MCALL==0 and SPINT==1)

- SPLSTACC: the final accommodation spell observed in the data. This is taken from the last observed spell. If accommodation type in last spell is missing then SPLSTACC is missing and set to -99.

Table 1: ACCSPELL variable list and descriptive statistics

Variable	Variable label	Mean	Min	Max	N
XWAVEID	Journeys Home XWAVEID	10898.74	10001	21571	7,234
SPNUM	Spell number	3.143	0	33	7,234
SPACC	Current spell accommodation type	5.053	-2	97	7,234
SPDTR	Spell start date recode	-0.605	-3	5	7,234
SPINT	Spell coincides with interview	0.668	0	1	7,234
JHWAVE	Wave when spell reported	3.163	1	6	7,234
SPWHSD	Spell start date (date FMT)	6-Apr-12	14-Sep-69	1-May-14	6,924
WCMPL	All spells reported in wave	-1.276	-9	2	7,234
WORDR	Spells reported in chron. order	-1.344	-9	1	7,234
WIDT	Int. date when spell reported (date FMT)	31-Oct-12	2-Sep-11	12-May-14	7,234
SPNXTACC	Next spell accommodation type [A]	-19.842	-99	97	7,234
SPLCENS	Spell is left-censored [A]	0.0429	0	1	7,234
SPEXIT	Spell ends with exit [A]	0.7190	0	1	7,234
SPSTART	Spell start date (date FMT) [A]	14-Apr-12	14-Sep-69	10-May-14	7,145
SPEND	Spell end date (date FMT) [A]	15-Feb-13	9-Sep-11	12-May-14	7,145
SPTDUR	Spell total duration [A]	308.156	1	16252	7,145
SPLSTACC	Last spell accommodation type [A]	-19.823	-99	97	7,234

[A] signifies spell-level measure added to ACCSPELL file.

ACCWSPELL. The file ACCWSPELL contains information of the spell data divided into segments for each reference period. The program uses ACCSPELL as the input file in SAS and JHGEN_LONG, ACCSPELL and JH_ACCCAL in STATA as input files. The dataset can contain multiple observations per spell, as some spells extend across multiple waves. The principal reason for organising the data this way is to facilitate multivariate continuous-time hazard and survival analyses in which some or all of the covariates change across survey waves.

As with ACCSPELL, each spell is uniquely identified by the cross-wave person identifier, XWAVEID, and the spell number, SPNUM. Each observation (segment) is uniquely identified by XWAVEID, SPNUM, and the spell segment number, SPSEGNUM. To link to data the JH survey data use the XWAVEID and JHWAVE, where the JHWAVE identifies the wave at the end of the segment. ACCWSPELL contains all of the spell-specific measures from ACCSPELL but adds several other variables that are specific to each segment, such as an indicator for whether the segment ends with an exit, SPSEGEXT, the segment

start date, SPSEGST, the segment end date, SPSEGEND, and the cumulative spell duration through the end of the segment, SPSEGCDR. Table 2 lists the variables and the summary statistics.

Some more notes and descriptions of the measures:
- XWAVEID and SPNUM identify spells
- XWAVEID, SPNUM, and SPSEGNUM identify spell segments (equivalently XWAVEID, SPNUM, and JHWAVE identify spell segments)
- XWAVEID and JHWAVE can link the data back to JHGEN_LONG (or JHLRD_LONG, JHLIM_LONG, or JHOS_LONG)
- SPSEGST: spell segment start date. The start date of the spell within the reference period. The variable SPSTART reflects the actual beginning of when the spell started.
- SDPSEGEND: spell segment end date. The end date of the spell segment within the reference period. The variable spend reflects the actual end date of the spell.
- SPSEGEXT: spell segment exit. Whether the spell segment ended within the reference period.
- SPSEGCDR: spell segment duration cumulative. Cumulative of the total spell computed as SPSEGEND minus SPSTART plus one. This variable does not reflect duration of the segment, data user has to compute duration of the segment.

Table 2: ACCWSPELL dataset variable list and descriptive statistics

Variable	Variable label	Mean	Min	Max	Obs
XWAVEID	Journeys Home XWAVEID	10897.45	10,001	21571	14,147
SPNUM	Spell number	2.293914	0	33	14,147
SPACC	Current spell accommodation type	4.325511	-2	97	14,147
SPDTR	Spell start date recode	-1.02163	-3	5	14,147
SPINT	Spell coincides with interview	0.8301407	0	1	14,147
JHWAVE	Wave of spell segment [W]	3.552485	1	6	14,147
SPWHSD	Spell start date (date FMT)	6-Oct-11	14-Sep-69	1-May-14	13,677
WCMPL	All spells reported in wave	-2.633491	-9	2	14,147
WORDR	Spells reported in chron. order	-2.680992	-9	1	14,147
SPNXTACC	Next spell accommodation type [A]	-26.75634	-99	97	14,147
SPLCENS	Spell is left-censored [A]	0.0332226	0	1	14,147
WIDT	Int. date when spell reported (date FMT)	27-Jul-12	2-Sep-11	12-May-14	14,147
SPEXIT	Spell ends with exit [A]	0.670955	0	1	14,147
SPSTART	Spell start date (date FMT) [A]	17-Oct-11	14-Sep-69	10-May-14	14,058
SPEND	Spell end date (date FMT) [A]	13-Apr-13	9-Sep-11	12-May-14	14,058
SPTDUR	Spell total duration [A]	545.0124	1	16252	14,058
SPLSTACC	Last spell accommodation type [A]	-34.03096	-99	97	14,147
SPSEGNUM	Spell segment number [W]	1.887185	1	6	14,147
SPSEGEND	spell segment end date [W]	30-Nov-12	2-Sep-11	12-May-14	14,058

6

SPSEGEXT	spell segment exit [W]	0.3676398	0	1	14,147
SPSEGCDR	spell segment duration cumulative [W]	411.5461	1	16252	14,058
SPSEGST	spell segment start date [W]	27-Jun-12	14-Sep-69	10-May-14	14,058

[A] signifies spell-level measure.
[W] signifies segment-level (wave-level) measure added to or changed in ACCWSPELL file.

ACC3SPELL. The file ACC3SPELL has spell data in blocks of thirds of the month, reflecting the way the calendar data were collected. The program uses ACCSPELL as the input file in SAS and JHGEN_LONG, ACCSPELL and JH_ACCCAL in STATA as input files. The program calculates the number of 10-11 day blocks within the spell. The dataset can contain multiple observations per spell, as some spells extend across multiple blocks. The principal reason for organising the data this way is to facilitate multivariate discrete-time hazard and survival analyses in which some or all of the covariates change over time.

ACC3SPELL takes the spells from ACCSPELL as input and then creates one observation for each block that the spell touches. Because of this, blocks can be repeated across spells, and the number of blocks in a month can exceed three if there are multiple spells that begin or end in the same block. For example, a spell that was recorded to end on 21 November 2013, and a spell that began 21 November 2013 would both produce blocks for the last third of November 2013.

As with ACCSPELL, each spell is uniquely identified by the cross-wave person identifier, XWAVEID, and the spell number, SPNUM. Each observation (block segment) is uniquely identified by XWAVEID, SPNUM, and the block segment number, SPDISNUM. To link to data the JH survey data use the XWAVEID and JHWAVE, where the JHWAVE identifies the wave at the end of the segment. ACC3SPELL contains all of the spell-specific measures from ACCSPELL but adds several other variables that are specific to each segment, such as the month, SPMONTH, and year, SPYEAR. For each spell, ACC3SPELL also has a variable for the total number of segments, SPDISSNM. The last segment in a spell can be identified when SPDISNUM equals SPDISSNM, and a spell exit can be identified as the last segment in a spell with SPEXIT equal to one. Table 3 lists the variables and the summary statistics.

Some more notes and descriptions of the measures:
- XWAVEID and SPNUM identify spells
- XWAVEID, SPNUM, and SPDISNUM identify spell segments (equivalently XWAVEID, SPNUM, SPMONTH, and SPYEAR identify spell segments)
- XWAVEID & JHWAVE can link the data back to JHGEN_LONG (or JHLRD_LONG, JHLIM_LONG, or JHOS_LONG)
- SPYEAR: spell year. The year of the block
- SPMONTH: spell month. The month of the block. There can be more than 3 blocks per month as the program creates a duplicate block for spells close to the edges of interview dates and creates a separate block for each new spell within the same 10 day window.
- SPDISSNM: spell total 10-11 day blocks segments
- SPDISNUM: spell 10-11 day block segment number

Table 3: ACC3SPELL dataset variable list and descriptive statistics

Variable	Variable label	Mean	Min	Max	Obs
XWAVEID	Journeys Home XWAVEID	10915.42	10001	21571	223,665
SPNUM	Spell number	1.180167	0	33	223,665
SPACC	Current spell accommodation type	3.745995	-2	97	223,665
SPDTR	Spell start date recode	-1.792793	-3	5	223,665
SPINT	Spell coincides with interview	0.9403617	0	1	223,665
JHWAVE	Wave at end of spell segment [3]	2.891825	1	6	223,665
SPWHSD	Spell start date (date FMT)	13-Feb-08	14-Sep-69	1-May-14	221,423
WCMPL	All spells reported in wave	-5.113643	-9	2	223,665
WORDR	Spells reported in chron. order	-5.129971	-9	1	223,665
WIDT	Int. date when spell recorded (date FMT)	2-Apr-12	2-Sep-11	12-May-14	223,665
SPNXTACC	Next spell accommodation type [A]	-35.96557	-99	97	223,665
SPLCENS	Spell is left-censored [A]	0.0100239	0	1	223,665
SPEXIT	Spell ends with exit [A]	0.5854604	0	1	223,665
SPSTART	Spell start date (date FMT) [A]	1-Mar-08	14-Sep-69	10-May-14	223,576
SPEND	Spell end data (date FMT) [A]	5-May-13	9-Sep-11	12-May-14	223,576
SPTDUR	Spell total duration [A]	1892.239	1	16252	223,576
SPLSTACC	Last spell accommodation type [A]	-59.44496	-99	97	223,665
SPDISSNM	Spell total 10-day segments [A]	187.4682	1	1603	223,665
SPDISNUM	Spell 10-day segment number [3]	94.23408	1	1603	223,665
SPYEAR	Spell year [3]	2010.263	1969	2014	223,576
SPMONTH	Spell month [3]	6.501373	1	12	223576

[A] signifies spell-level measure.
[3] signifies segment-level (wave-level) measure added to or changed in ACC3SPELL file.

Running the programs

The STATA version of the program is jh_acc_eh.do, and the SAS version is jh_acc_eh.sas. To run the programs you need to modify the input and output directory names at the beginning of the SAS (the libname) and STATA program file (the global macros). The input datasets are JH_ACCCAL (the calendar data) and JHGEN_LONG file. Those with JHLIM_LONG, JHLRD_LONG or JHOS_LONG replace the MYDATA macro just below the directory macros with the alternate longitudinal dataset. When the program has finished running it produces the ACCSPELL, ACCWSPELL and the ACC3SPELL file, which are placed in the output directory specified at the beginning of the program.

References

Bevitt, Andrew, Abraham Chigavazira, Rosanna Scutella, Yi-Ping Tseng, and Nicole Watson. 2014. *Journeys Home User Manual: User Manual Version: 6.0.2.* Melbourne: Melbourne Institute of Applied Economic and Social Research.

Cobb-Clark, Deborah A., Nicolas Herault, Rosanna Scutella, and Yi-Ping Tseng. 2016. "A Journey Home: What Drives How Long People Are Homeless?" *Journal of Urban Economics* 91:57-72.