Producing Working-Life Histories in the BHPS and UKHLS

Liam Wright

Department of Epidemiology and Public Health, University College London

April 1, 2020

Abstract

The document describes a Stata algorithm for producing working-life histories for participants in the British Household Panel Survey (BHPS) and the United Kingdom Household Longitudinal Study (UKHLS). It also describes in detail questionnaire items related to working lives from the two studies.

Contents

1. Introduction

In this paper, I describe a Stata algorithm to produce working-life histories for participants in the British Household Panel Survey (BHPS) and the United Kingdom Longitudinal Study (UKHLS; also known as Understanding Society). To my knowledge, this is the first published algorithm for deriving working-life histories in both surveys. Methods using BHPS data alone have been previously published (Halpin, 1997, 2000; Maré, 2006; Oskrochi & Crouchley, 2000; Paull, 2002) and have influenced the method used here. Simply scaling these up to the UKHLS is not possible, however, as questions related to working-life histories differ between the two surveys. Further, there have been developments in imputing missing sequence data since the previous methods were published, which arguably mean fewer strong assumptions need to be made when cleaning the data (Halpin, 2012, 2013, 2016).

Deriving work-life histories in the two surveys is a time intensive task, involving many decisions which are arbitrary and perhaps contentious. I have two hopes in publishing this paper and the accompanying code. The first is that researchers will not need to create their own code from scratch, which I suspect has precluded some studies. The second is that I will get feedback on the work. My original purpose in deriving the histories was to calculate months unemployment during participants' youths (Wright, 2019). Decisions in the data cleaning have been made with this in mind. This may mean the code is not as widely useful as it could be, but I would like to improve the work through time, based on readers' comments. If you would like to make a suggestion, or if you spot an error, please create an issue on GitHub (project link here).

The structure of this paper is as follows. In Section 2, I provide some background information on the BHPS and UKHLS. In Section 3, I describe questions relating to working-life histories in the two surveys. In Section 4, I describe my own method, including discussion of how to use the code published alongside this paper.

This work was carried out during my PhD. I'd like to thank my supervisors, Professor Jenny Head and Dr Stephen Jivraj, for their guidance and advice, and to thank my PhD colleagues who are a joy to work with. Thanks also to David Maré, whose work deriving work-life histories in the BHPS made this task far easier, and to the Institute for Social and Economic Research, who maintain excellent documentation of the two surveys.

2. Background

The BHPS and its successor the UKHLS are two yearly panel surveys of households from across the UK. Both are managed by the Institute for Social and Economic Research (ISER) at the University of Essex and are available through the UK Data Service. In this section, I provide some basic information on the design of these surveys before detailing the questions related to working-life histories they collect.

First, a note on the data. In this paper, I use Stata files from the harmonized the BHPS and UKHLS dataset (Institute for Social and Economic Research, 2019).[1] One of the many positive characteristics of this data is the consistency in how folders, files and variables are named. Files from wave `i` of the BHPS are contained in the folder `bhps_wi` and those from wave `i` of the UKHLS are contained in folder `ukhls_wi`. Cross-wave files are contained in folders `bhps_wx` and `ukhls_wx`, respectively.[2]

Yearly files in the BHPS are prefixed with '`bw_`', where `w` is a letter corresponding to the position the wave is in the alphabet (`a`=1, `b`=2,..., etc.). UKHLS files are pre-fixed with '`w_`' and follow the same rule. The main individual adult interview (`indresp`) files for Wave 13 of the BHPS and Wave 6 of the UKHLS are therefore called `bm_indresp` and `f_indresp`, respectively, and stored in the folders `bhps_w13` and `ukhls_w6`.

This rule is also used for naming wave-specific variables. (Cross-wave variables are not given a prefix.) ISER names variables consistently across the BHPS and UKHLS, where possible. For instance, the variable for main economic activity, `jbstat`, is named `bw_jbstat` in the BHPS and `w_jbstat` in the UKHLS. Variables which differ slightly between the two surveys are appended with '`_bh`' in the BHPS version.[3] Some variables are also recoded for harmonization across the two surveys. I report values in the dataset, rather than the original survey, where this is the case.

In this document, I refer to variables or files from specific waves using the appropriate prefix (e.g. `c_jbstat` for the `jbstat` variable from Wave 3 of the UKHLS). When referring to a variable or file in general, I use the non-specific prefix '`bw_`' or '`w_`' (e.g. `w_jbstat`).

Related variables often share a common stub. For instance, in the BHPS, variables for start day, month and year of a participant's current job are called `bw_jggbd`, `bw_jggbm`,

[1] The code works with both the End User License and Special License versions of the dataset. The Special License dataset includes a variable for birth month, which is used in the algorithm if available. If not, birth month is assumed to be June. Readers should use the Special License dataset where possible.

[2] Some earlier versions of the dataset used the form `us_wi` for the UKHLS folders. I include a folder stub macro in the do files which can be changed if the naming convention is amended (or earlier files are used).

[3] ISER also provides non-harmonized BHPS files through the UK Data Service which do not have these properties.

`bw_jggby4`. For brevity, I report such variables using the rule, `bw_stub/abc` (i.e. `bw_jggb/dmy4` for the start date).

Finally, when discussing survey questions, I refer to the 'universe' of participants who were asked them. To define the universe, I sometimes use logical operators and functions using Stata syntax. For those new to Stata, important operators and functions are described in Table 1.

Table 1: Stata's Logical Operators and Functions[4]

Operator/Function	Description
&	AND
|	OR
!	NOT
==	EQUALS TO
!=	NOT EQUALS TO
>=	GREATER THAN OR EQUALS TO
<=	LESS THAN OR EQUALS TO
inlist(x,y,z,...)	x==y | x==z | ....
inrange(x,y,z)	x>=y & x<=z
x[_n-1]	x from previous row

2.1 BHPS

The BHPS started in 1991 and ended after 18 waves in 2008. Fieldwork began in September each year and lasted 4-9 months. Households were initially recruited from residences in mainland Britain, south of the Caledonian Canal, and followed across the UK thereafter. Original sample members (OSMs) were followed into new households, and new household members were eligible for inclusion in the study for as long as they resided with an OSM.[5] Newborn children of OSMs were followed independently of their parents, if they moved household.

Five thousand households were initially recruited to the study, with the sample designed to be representative of the general population. In Wave 9, a further 3,000 households were recruited from Scotland and Wales to enable subgroup analysis in these countries. In Wave 11, 2,000 households were introduced from Northern Ireland. The same rule for following these participants applied as to participants recruited in Wave 1. The BHPS also incorporated a sub-sample of low income households from the UK European Community Household Panel in Wave 7 of the survey, though these were only followed to Wave 11 when funding expired.

From Waves 1-15, the survey was recorded on paper. In Wave 16, computer assisted interviewing was introduced, and dependent interviewing was used with the wording of several

[4] For brevity, I sometimes use 'OR' instead of '|' and '=' instead of '=='.

[5] These household members are termed Temporary Sample Members (TSMs).

questions adapted according to participant's answers in previous waves. The survey each wave consisted of several questionnaires, some asked annually and others intermittently. Each wave included an adult interview, answered by household members aged 16 or over on December 1st of the sample year.[6] The adult interview contained a self-completion module, but most questions were asked via interviewer. The majority of adult interviews were conducted face-to-face, though, from Wave 3 onward, a small number were answered by telephone (2.41% of all interviews). Interviews could also be completed by proxy household member, which occurred in 2.44% of adult interviews.

Some questions in the face-to-face interview were not asked in the proxy or telephone interviews, including a number of questions related to working-life histories. For instance, an Employment History module collecting information on economic activities since the previous wave was consistently asked of face-to-face respondents only. Face-to-face and telephone interviews were identical, however - with a few important exceptions - from Wave 15 onwards. For brevity, in what follows, I refer to the face-to-face interview and Wave 15-18 telephone interviews as the 'full' interview.[7]

A lifetime "Employment Status History" questionnaire, collecting information on activities since first leaving full-time education, was given to full-interviewees in Waves 2, 11, and 12. In Wave 3, a job history questionnaire was given to collect further information on spells of employment gathered in the employment status history questionnaire from the previous wave. The method to derive working-life histories described here uses data from the recurring adult interview and the employment status history questionnaires only. The data from these are stored in the files `bw_indresp` (Waves 1-18), `bw_jobhist` (Waves 1-15), `bw_jobhstd` (Waves 16-18), and `bw_lifemst` (Waves 2, 11, and 12).

2.2 UKHLS

The UKHLS started in 2009 and continues to the present. Nine waves of data are available, at the time of writing (01 April 2020). Fieldwork takes approximately 24 months, with some overlap between waves. Participants are interviewed every 12 months, on average.

Households in the UKHLS were originally recruited from across the UK, with individuals from ethnic minority backgrounds over-sampled. 30,000 households were recruited in Wave 1. Remaining BHPS sample members were able to join the UKHLS in Wave 2 (6,700 households), and in Wave 6, a sample boost of 2,500 immigrant and ethnic minority households was added.

[6] As fieldwork began in September, this means a small number of 15 year olds were interviewed in the adult interview.

[7] Type of interview is stored in the variable `bw_ivfio`. 1=Face-to-face, 2=Proxy, 3=Telephone.

Similar to the BHPS, original sample members are followed across households, with new household members eligible for inclusion as long as they reside with an OSM.[8] Unlike the BHPS, in the UKHLS, only newborn children of OSM *mothers* are followed independently. Fathers of newborn OSM's are also independently followed.[9]

The UKHLS includes an adult interview every wave, conducted by face-to-face, telephone or web interview or via proxy (latter comprises 7.12% of interviews).[10] A number of questions related to employment histories are not included in the proxy interview, but the face-to-face, telephone and CAPI interviews are identical. I accordingly refer to these as 'full' interviews. The UKHLS is recorded on computer and dependent interviewing is used extensively.

Waves 1 and 5 included a lifetime activity history module for full interviewees. Unlike the BHPS, the UKHLS does not include a job history questionnaire to collect further information on spells of employment elicited in the lifetime activity history questionnaire. The method described here for deriving working-life histories uses data from the adult interview and employment status history questionnaires only. The relevant data are stored in the `w_indresp` (Waves 1-9) and `w_empstat` (Waves 1 and 5) files.

The UKHLS also includes a separate 'Innovation Panel' used to test new survey questions and methodologies. I do not use data from this panel here, given the differences in questions asked, and the purpose of, this group.[11]

[8] In ethnic minority household sub-samples, only household members who belong to an ethnic minority are treated as OSMs.

[9] Fathers of newborn OSMS are identified as Permanent Sample Members (PSMs) in the data (`w_sampst`)

[10] Web interview introduced in Wave 8. Type of interview and interview mode are storied in the variables `w_ivfio` and `w_indmode`, respectively.

[11] Dataset available here.

3. Activity Data in the BHPS and UKHLS

3.1 BHPS

The BHPS contains many questions on working-life histories, several of which overlap in the time period being asked about. In this section, I introduce the relevant data in each of the `bw_indresp`, `bw_jobhist`, `bw_jobhstd`, and `bw_lifemst` files in turn.

3.1.1 bw_indresp

Main Activity

The `bw_indresp` files contain several questions relating to current working status and economic activity since the previous wave. There are differences in question wording and routing across waves.

Each wave contains a question related to current main activity, `bw_jbstat`, which is shown in the box below. In Wave 1, the question was asked only of face-to-face respondents, placed later in the interview than in subsequent waves, and had a different ordering for the response categories. The variable for Wave 1 is recorded in `ba_jbstat_bh`, though ISER also supply a recoded version, `ba_jbstat`, with the harmonized dataset.[12] At the time of writing (01 April 2020), there is an issue with how these variables in Wave 1 have been cleaned.[13]

bw_jbstat: Which of the following best describes your current situation?

Universe: All Participants

1. Self employed
2. In paid employment (full or part-time)
3. Unemployed
4. Retired from paid work altogether
5. On maternity leave
6. Looking after family or home
7. Full-time student/at school
8. Long term sick or disabled
9. On a government training scheme
97. Something Else

Current Employment

[12] Proxy respondents were asked a separate main activity question in Wave 1. As far as I can tell, the variable is not supplied with the harmonized dataset, nor do ISER file `ba_jbstat` or `ba_jbstat_bh` with the responses.

[13] See this issue in the Understanding Society User Forum.

A separate set of questions are asked each wave on whether a participant is currently employed. All respondents are asked whether the participant was in paid work last week (`bw_jbhas`), and if not, whether they were away from a job (`bw_jboff`). If the respondent answers yes to either of these questions, they are asked a series of questions on job characteristics, including whether the participant is self-employed or an employee (`bw_jbsemp`), the start date of the job (`bw_jbbg/dmy4`) and on working hours. ISER derive a variable for full-respondents which specifies whether the participant is employed in full-time or part-time work (`bw_jbft_dv`). Telephone (Waves 3-14) and proxy respondents (all waves) are directly asked whether the participant is employed in full or part-time work (`bw_prjbft`). If the respondent does not respond yes to either `bw_jbhas` or `bw_jboff`, further questions related to job-search activity are asked. From Wave 6 onwards, for full interviewees, these conform with the International Labour Organization's (2013) definition of unemployment.

Further description of the main variables is provided in Table 2.[14] A routing diagram for the face-to-face interview from Waves 1-15 is displayed in Figure 1. Following the introduction of dependent interviewing in Wave 16, a number of alternative questions were added to the survey for returning full interviewees. ISER fill in the original variables as if these were asked, however, and routing is substantively similar in these waves. A routing diagram for (non-full) telephone and proxy interviews is displayed in Figure 2.

Answers to the main activity and current employment questions are not always consistent. For instance, a small number of participants state being in employment but not having a job when asked directly, and vice versa. Cross-tabulations of the variables are displayed in Table 15.

Education

The `bw_indresp` files also contains a number of questions on whether the participant is attending, or has attended, school and/or a place of further education, and at what age they left. These questions are asked to all new respondents in the full adult interviews, including rising 16 year olds, and to all proxy respondents. There was a small change to the routing of these questions in Wave 2, when respondents whose main activity was full-time education were not asked the age at which they had left school. Telephone respondents in Waves 3-14 were not asked any of these questions. Further description of these variables is displayed in Table 3. A routing diagram is provided in Figure 3.

For returning participants interviewed face-to-face, questions on spells of full-time education since the previous wave were asked. In Waves 2-7, participants who did not report full-time student as their main activity were asked whether they had attended any educational institution full-time since September 1st 1989+Wave (`bw_edlyr`). (The September 1st 1989+Wave rule ensures that the period overlaps with the start of fieldwork for the previous

[14] Question wording in this table is from Waves 1-15 face-to-face or telephone interviews in the first instance.

wave.) Start dates were not asked for and only one end date was elicited (bw_edend/my4). Participants presumably answer about the most recent spell, though this was not specified in the question. In Waves 8-18, face-to-face respondents were asked questions about *all* spells of full-time education since September 1st 1989+Wave. Both start and end dates were collected in this case. I do not use responses from Waves 2-7 given the limited date information.

Corresponding information was not elicited from proxy or telephone respondents at any point.[15] Education history variables from Waves 2-18 are described in Table 4 and a routing diagram is displayed in Figure 4. Variables from Waves 8-18 are provided in the bw_indresp files in wide format, with spell number, i, appended to the variable name.

A small proportion of individuals who state being currently in school or further education according to the education variables do not report full-time student as their main activity. Cross-tabulations are displayed in Table 16.

Annual Employment History

Besides information on current activities, the full adult interviews in each wave of the BHPS include an 'Employment History' module, collecting information on changes in activity since the previous wave or interview. The design of this module changed in Waves 1, 2 and 16. Variables from this module are stored in the files bw_indresp (Waves 1-18), bw_jobhist (Waves 1-15), and bw_jobhstd (Waves 16-18).[16]

In Waves 1-15, the procedure in the Employment History module was to first assess whether the participant's current job or other activity began after September 1st 1989+Wave, and if so, elicit information on previous activities until the earliest spell began before or on that date. In Waves 16-18, with the introduction of dependent interviewing, the procedure was changed to check whether the participant had ended the activity from the previous interview/wave, and if so, elicit information on spells up to the current period. In other words, in Waves 1-15, the module worked backwards to the previous wave, while in Waves 16-18, questions worked forwards to the current period.

In all waves, information on the 'index' spell is stored in the bw_indresp file, and information on other spells contained in the bw_jobhist or bw_jobhstd files. In the rest of this sub-section, I discuss the variables I use from the Employment History module which are contained in bw_indresp.

In Wave 1, current activity was first defined by whether an individual was self-employed or in paid employment based on responses to ba_jbsemp. If the participant was not in employment, ba_jbstat_bh was used to define current activity instead. For those in employment, responses to the questions ba_jsbg/dmy4 (self-employed) and ba_jbbg/dmy4

[15] Respondents in the proxy interview were asked if the participant had any spells of full-time education since September 1st 1989+Wave, but no information on start or end dates was elicited.

[16] bw_jobhist files are also available for Waves 16-18 but these are based on post-processing by ISER to create consistency across waves.

(paid-employment) were used to fill another variable, `ba_cjsbg/dmy4`, which specified the start date of the current job. Respondents who were not in employment were asked for the start date of their current activity directly, with responses recorded in `ba_cjsbg/dmy4`. This variable was then used by the interviewer to check whether the start date was after September 1st 1990, with participants asked directly if unclear (i.e. in case of missing dates). Whether the activity began before, on, or after September 1st 1990 is stored in the indicator variable `ba_cjsbggly`.

In Waves 2-15, an amendment was made to take account of the possibility that full-time students may also be employed. In these waves, `bw_jbsemp` was only used to define the participant's current activity if the participant did not report full-time education as their main activity (i.e. `bw_jbstat`!=7). If the participant was in full-time education or otherwise not in employment, another question on current activity was asked, `bw_nemst`. This was identical to `bw_jbstat`, except for removing categories 1 and 2 regarding self-employment or paid work. The question is displayed below.

bw_nemst: Please look at this card and tell me which best describes your current situation.

Universe: `bw_ivfio=1 & (bw_jbstat=7 | !inlist(bw_jbsemp,1,2))`

3. Unemployed
4. Retired from paid work altogether
5. On maternity leave
6. Looking after family or home
7. Full-time student/at school
8. Long term sick or disabled
9. On a government training scheme
10. Something else

Participants asked `bw_nemst` were then asked the start date of this spell (`bw_cjsbg/dmy4`), with interviewers checking whether this was after September 1st 1989 + Wave (`bw_cjsbggly`). Similar to Wave 1, for non-students who were self-employed or in paid employment, responses to `bw_jsbg/dmy4` or `bw_jbbg/dmy4` were used by the interview to fill the `bw_cjsbg/dmy4` and `bw_cjsbggly` variables. Further detail on the variables from the Employment History module from Waves 1-15 contained in the `bw_indresp` files are displayed in Table 5. Routing diagrams for Waves 1 and 2-15 are displayed in Figure 5 and Figure 6, respectively. Responses to `bw_nemst` and `bw_jbstat` are not always consistent (see Table 17 for cross-tabulations). Notably, a small number of participants report being full-time students in `bw_jbstat` but change their answer in `bw_nemst`.

The routing of the Employment History module in Waves 16-18 is more complex. In these waves, participants were only routed into the module if they were either:

1. Currently in employment (`bw_jbhas = 1 | bw_jboff = 1`), and:
   (a) newly employed or with changed job characteristics, if personally interviewed in previous two waves, or
   (b) in a job which began after September 1st 1989+Wave, otherwise.
2. Not currently employed (`bw_jbhas != 1 & bw_jboff != 1`), and:
   (a) not retired in both current and previous interviews, if personally interviewed in previous two waves, or
   (b) in an activity (`bw_jbstat`) which began after September 1st 1989+Wave (`bw_cjsbg/dmy4`; `bw_cjsbggly`), otherwise.

If either of these conditions were met, participants were then asked what they were doing at the time of the previous interview, if interviewed in last two waves, or on September 1st 1989+Wave, otherwise. Participants interviewed within the last two waves were asked whether the information contained in `bw_jbstat` from the previous interview was correct (`bw_cjsck3`), and if not correct, what they were doing at this date instead (variable `bw_cjsstly`; shown below).[17] Participants not interviewed in the past two waves were asked `bw_cjsstly` directly. The date at which this activity ended (`bw_cjse/dmy4`) or whether this was the current status (`bw_cjscjs`) was then elicited, with the module ending if this was the participant's current activity. The reason for ending the index activity was also elicited if the participant was in a job which had now ended (`bw_jhstpy`).

The main Employment History module variables from Waves 16-18 contained in the `bw_indresp` files are described in Table 6. Routing diagrams are displayed in Figure 7. As

bw__cjsstly: Please look at this card and tell me which best describes your situation on (September 1st 1989+Wave/<INTDATE>)?

Universe: `bw  cjsck2 = 2 | (bw cjsck2 = 1 & bw  cjsck3> 1)`

1. Self employed
2. In paid employment (full or part-time)
3. Unemployed
4. Retired from paid work altogether
5. On maternity leave
6. Looking after family or home
7. Full-time student/at school
8. Long term sick or disabled
9. On a government training scheme
10. Something else

[17] `bw_cjsstly` is filled for those who state `bw_jbstat` from the previous wave was correct.

elsewhere, responses to `bw_jbstat` and `bw_cjsstly` (for current spells) are not always consistent. Cross-tabulations are displayed in Table 18.

Telephone respondents were able to complete the full Employment History module in Waves 15-18. In Waves 3-14, telephone respondents answered a separate set of questions on the date they began their current activity. Respondents who did not give full-time student as their main activity were asked for the start date of their main activity (`bw_presbg/my4`), and if unclear, whether this was after September 1st 1989+Wave. Respondents who stated being in paid work in either the `bw_jbhas` or `bw_jboff` questions were also asked for the start date of this job (`bw_prjbbg/my4`), and if unclear, whether this was after September 1st 1989+Wave (`bw_prjbly`). Proxy respondents answered this same set of questions in Waves 3-18. No start date information was collected from telephone or proxy respondents in Waves 1 or 2. A routing diagram is given in Figure 2.

Lifetime Status History

Waves 2, 11, and 12 included an 'Employment Status History' questionnaire as part of the face-to-face adult interview. This collected information on changes in status from first leaving full-time education (FTE) to the present day. In Wave 2, all participants were eligible to answer the questionnaire, while in Wave 11 (12), only members of the Scotland and Wales (Northern Ireland) boost samples were eligible. The `bw_indresp` files contain three variables from this questionnaire: `bw_ledendm`, `bw_ledeny4`, and `bw_lednow`. These contain responses from the first question of the interview, which asked "When was the first time that you left full-time education?". `bw_ledendm` and `bw_ledeny4` contain the month and year of the end date, while `bw_lednow` is an indicator for if the participant has never left (0. Never left). (In Waves 11 and 12, participant were given the extra prompt to ignore gaps between school and further education of less than one year.) Other data from the questionnaire is placed in the `bw_lifemst` files (see Section 3.1.4). A routing diagram is displayed in Figure 10

3.1.2 bw_jobhist

The `bw_jobhist` files contain information from the Employment History module for activities prior to the current spell. The question wording and routing for these spells did not change across Waves 1-15. These files have also been constructed by ISER post-hoc for Waves 16-18, but I do not use these in my code and discuss files from Waves 1-15 only here.

Participants in Waves 1-15 whose current activity started after September 1st 1989+Wave were asked which of several descriptions was closest to the what they were doing before the current spell (`bw_jhstat`; described below) and what the start dates were (`bw_jhbg/dmy4`), with questions repeated until the earliest start date was prior to September 1st 1989+Wave.

Once all activities were enumerated, the interviewer inquired further about episodes of employment (`inlist(bw_jhstat,1,2)`), including whether the participant was self-employed or in part-time or full-time employment (`bw_jhsemp`, asked if `bw_jhstat==2`) and the reason

for leaving the job (bw_jhstpy). Individuals currently in employment whose prior period of employment ended due to finding a better job were asked in further detail about what attracted them to their current job (bw_jblky). The main variables from the questionnaire are described in further detail in Table 7. Question routing is displayed in Figure 8.

bw_jhstat: Can you look at this card please and tell me which of the descriptions comes closest to what you were doing immediately before then?

Universe: bw_cjsbly == 2

1. Doing a different job for the same employer
2. Working for a different employer
2. In paid employment (not self employed)
2. Working for myself (self employed)
3. Unemployed
4. Retired from paid work altogether
5. On maternity leave
6. Looking after family or home
7. Full-time student/at school
8. Long term sick or disabled
9. On a government training scheme
10. Something else

Responses in bw_jobhist display "seam effects" with some inconsistencies between responses for the overlapping period in consecutive waves (i.e. between September 1st 1989+Wave and the interview date of the prior wave). The nature of these inconsistencies is not random – for instance, individuals are more likely to retrospectively reclassify periods of unemployment as inactivity. See Paull (2002) for more details.

3.1.3 bw_jobhstd

The bw_jobhstd files contain information from the Employment History module in Waves 16-18 for activities starting after the previous interview or September 1st 1989+Wave. Participants whose index spell had ended by the current interview were asked the same bw_jhstat question as in Waves 1-15 - though in this case, for spells occurring *after* the index spell, and for spell end dates (bw_jhend/dmy4) - with questions repeated until the current activity was reached (bw_jhcjs). Job characteristics for spells of employment prior to the current activity were also elicited, including whether the participant was self-employed or in part-time or full-time employment (bw_jhsemp, where bw_jhstat==2) and the reason for leaving the job (bw_jhstpy). Participants who had moved into their current job from another job and who stated leaving because of a better job were asked to give further detail on what attracted them to their current position (bw_jblky). The main variables are described further in Table 8 and questionnaire routing is displayed in Figure 9. Cross-tabulations between

(current spell) `bw_jhstat` and `bw_jbstat`, again occasionally inconsistent, are displayed in Table 19.

Newly interviewed individuals in all waves of the Employment History questionnaire who did not report a spell of work since the previous wave were also asked if they have ever had a paid job (`bw_jbhad`) and, if so, for job characteristics, including the year in which they left (`bw_jlendy4`). Responses to these questions are stored in the `bw_indresp` file. I do not use these here, however, as individuals may have changed statuses in the interim - for example, moving between unemployment and economic inactivity.

3.1.4 bw_lifemst

Participants in the Economic Status History questionnaire who stated having left full-time education were asked about all status changes to the time of the interview (`bw_leshst`; shown below), with changes elicited from this first leaving education to the present date. Participants were asked the date at which their status changed (`bw_leshe/my4`) or whether it was the current status (`bw_leshne`). The main variables are described further in Table 9. Note, a sequence of consecutive jobs could be recorded as a single status in this questionnaire.

bw_leshst: Which description on this card comes closest to what you first did after leaving full-time education?/Which of the descriptions on the card best describes what you did next, even if it was only for a month or so?

Universe: `bw_lednow != 0`

1. Self employed
2. Full-time paid employment
3. Part-time paid employment
4. Unemployed
5. Retired from paid work altogether
6. On maternity leave
7. Looking after family or home
8. Full-time student/at school
9. Long term sick or disabled
10. On a government training scheme
11. National Service/War Service
12. Something else

Given the long recall period, individuals were given the extra option to provide seasons rather than months for the timing of status changes. In Wave 2, Winter was not split into Winter at the start or end of the calendar year, whilst in Waves 11 and 12 options for "Winter (Jan/Feb)" and "Winter (Dec)" were provided. A routing diagram for the entire questionnaire

is shown in Figure 10. ISER also provide start date variables (`bw_leshs/my4`) in the data files, which are taken from the end dates for the previous spell. Data are provided in long format with the variable `bw_leshno` indicating the spell. Responses to the `bw_leshst` and `bw_jbstat` questions show a small number of inconsistencies. Cross-tabulation are given in Table 20.

3.2 UKHLS

Many of the questions asked in the BHPS are also asked in the UKHLS. However, several differences exist between the two surveys.

3.2.1 w_indresp

Main Activity

As in the BHPS, the main current economic activity status variable in the UKHLS is `w_jbstat`, which is asked of both full and proxy respondents. The UKHLS version of this question add extra categories for 'unpaid worker in family business' and (from Wave 3) 'working in an apprenticeship'. More detail is displayed in the box below. Full interviewees who give full-time student as their main activity are also asked what type of educational institution they attend (`w_edtype`).

w_jbstat: Which of these best describes your current situation?

Universe: All Participants

1. Self employed
2. In paid employment (full or part-time)
3. Unemployed
4. Retired from paid work altogether
5. On maternity leave
6. Looking after family or home
7. Full-time student/at school
8. Long term sick or disabled
9. On a government training scheme
10. Unpaid worker in family business
11. Working in an apprenticeship
97. Something else

Current Employment

Questions on current employment are elicited similarly to the BHPS. `w_jbhas` and `w_jboff` elicit whether the participant had paid employment in the last week or was otherwise off from work. Those in work are asked for job characteristics, including on whether they are self-

employed or an employee (w_jbsemp), and on start dates (new interviewees only) and job hours. ISER provide a derived variable for whether the respondent is in full- or part-time employment for full-interview respondents (w_jbft_dv). Proxy respondents are separately asked whether the participant is in full- or part-time work if they are unable to give specific working hours (w_pjbptft). For full-respondents not in work, questions on job search activity consistent with the International Labour Organization's (2013) definition of unemployment are asked.

Full-respondents who have not been interviewed previously and who are not in work are also asked whether they have ever had a job (w_jbhad), and for the end date of their last job if so (w_jlend/my). As with the BHPS, I do not use this information as participants may have moved between non-employment activities in the interim. In Wave 1, currently retired individuals (a_jbstat=4) who are not in work (a_jboff=2) were also asked for the date they retired (a_retdate/my).

Further description of the current employment variables is given in Table 10. Routing diagrams for full- and proxy respondents are displayed in Figure 11 and Figure 12, respectively. Again, as in the BHPS, answers to the main activity and current employment questions are not always consistent. Cross-tabulations are displayed in Table 21.

Full-Time Education

The wording and routing of questions on education differ between the BHPS and UKHLS. In the full interview of the UKHLS, new respondents are again asked whether they are attending, or have attended, school (w_school). Participants who did attend are asked to give the age at which they left (w_scend), and are then asked whether they are currently attending, or have attended, a place of further education (w_fenow), and for the age at which they left (w_feend). The same questions are asked of all proxy respondents in Waves 1-3 and, from Wave 4 onwards, of proxy respondents answering on behalf of participants who have never been interviewed personally, or on behalf of, before. Description of the relevant variables is provided in Table 11. A routing diagram is displayed in Figure 13.

From Wave 2 onwards, an Annual Event History module was included the adult interview to be completed by returning full participants. One section of this module related to spells of fulltime education since the previous (full) interview. Participants who state full-time student as their main activity in the prior interview (w_ff_jbstat=7)[18] are asked whether they have been continuously in full-time education since then, and if not, for the spell end date. Participants whose spell of education had ended are then asked, along with other returning participants, for the start and end dates of any other spells of education since the previous interview, including whether any of the spells is still continuing. Unlike in the BHPS, the type of educational institution attended is not elicited in these questions, though for

[18] _ff_ is used to name variables which have been 'fed-forward' from the previous interview. There is sometimes a discrepancy between the fed-forward variable and the variable in the previous wave. This is because fed-forward values are not cleaned by ISER but the main variables are.

current students this may be taken from w_edtype. Description of the relevant variables and a routing diagram are displayed in Table 12 and Figure 14. A cross-tabulations for current status and current education is displayed in Table 22.

Annual Employment History

Another section of the Annual Event History module is on activities since the previous interview. The routing of these questions is complex, and depends on values contained in w_ff_jbstat and w_ff_emplw.[19] There are several routes through the module. Participants who were not employed at the time of the previous interview according to w_ff_jbstat (w_ff_jbstat>2) are asked whether they have continuously been in the same activity since that date (w_notempchk). Participants who were employed according to either w_ff_jbstat or w_ff_emplw are asked whether they have continuously been in paid work since the prior interview (w_empchk). These questions are not mutually exclusive as a small number of participants give a non-employment activity as their main activity whilst also state having a job.

Individuals whose job or non-employment activity had ended are asked when this occurred (w_empstend/dmy) and what they did next (w_nxtst). Those whose job ended are also asked for the reason why (w_stendreas).[20] Participants who state being continuously employed since the previous wave who were not self-employed according to w_ff_jbstat are asked whether they have moved employer (w_jbsamr) or their job has changed within their firm (w_samejob) and, if so, on what date (w_jbend/dmy). Those who moved employer are further asked the reason why (w_jbendreas). Participants whose job changed were then asked whether this was their current job (w_cjob).

This question was also asked of those in non-employment whose next activity was paid work and those who stated not continuously being employed but whose next activity was paid work (w_nxtst==1). If it was their current activity, what attracted the participant to the job is elicited (w_cjbatt).[21] If it wasn't their current activity, characteristics of the role, such as working hours, whether self-employed or employee, end reason(s), and end dates are elicited.[22] Participants whose next activity after the index spell wasn't paid work (w_nxtst==2) are instead asked whether this is their current activity (w_cstat), and if not, what the end date was (w_nxtend/dmy).

Participants whose activity after their index activity had also ended (w_cjob==2 or w_cstat==1) were then asked a series of questions on following activities with questions looped until their current status was reached. For spells of employment, job hours, reason(s) for

[19] w_ff_emplw is a fed-forward value for employment derived from answers to jbhas and jboff.

[20] From Wave 6, participants have been able to give multiple reasons why a job ended. Prior to this only the main reason was elicited.

[21] Again, from Wave 6, participants were able to give multiple reasons.

[22] A question on end reasons was only introduced in Wave 5. Presumably, this was mistakenly omitted in previous waves.

leaving, whether employed or self-employed and attraction to current job are additionally asked.

The description and routing of the questions on economic activity statuses since respondents were last interviewed are displayed in Table 13 and Figure 15. Note, for individuals asked both `w_empchk` and `w_notempchk`, ISER state that question wording referred to activities via the `w_empchk` route (link).

Seam effects are observed in this questionnaire. Table 23 compares responses for current activities in this module with `w_jbstat`.

Lifetime Employment Status History

In Waves 1 and 5, a lifetime 'Employment Status History' questionnaire was given to full interviewees. In Wave 1, this questionnaire was answered by individuals issued in the first six months of the survey. Study co-ordinators cut the questionnaire after the first six months of Wave 1 as they found the survey was taking too long to complete. The Wave 5 questionnaire was answered by non-former BHPS participants who did not answer the questionnaire in Wave 1, due to either being issued they survey after month 6 or entering in a later wave.[23]

The `w_indresp` files contain one question from the Employment Status History module which asked full-time students in further or higher education whether they had been in fulltime education continuously since school (`w_lgaped`). Those who had had a break were routed further into the module. How students were identified differed across the two waves. In Wave 1, students were identified using `a_fenow` which was asked of all participants who had left school (`a_school=1`). In Wave 5, students were identified by main activity (`e_jbstat=7`) and type of institution attending (`inlist(e_edtype,3,4,5)`). A routing diagram is displayed in Figure 16.

3.2.2 w_empstat

Further/higher education students who had not been continuously in full-time education since school and other participants who were not currently in education were asked for all status changes (`w_leshst`) and their start dates (`w_lesh/emsy4`) from first leaving full-time education. As in the BHPS, participants were allowed to give seasons instead of months, with options given for Winter at start or end of the calendar year.

Participants were also asked a number of questions on the characteristics of their first job. I do not use this information here. Relevant variables from the questionnaire are described further in Table 14 and routing diagrams are displayed in Figure 16. Data are supplied in long format in the `w_empstat` file, with the variable `w_spellno` identifying the spell. Answers to

[23] Households are randomly assigned to interview months, which means allocation to the life history module in Wave 1 is a natural experiment. However, attrition rates are higher among those who were asked to complete the Lifetime Employment Status History module in Wave 1, so the questions that could be answered using this data (e.g. on recall bias) are limited.

`w_jbstat` and for the current spell in `w_leshst` differ in a small number of cases (see Table 24 for cross-tabulations below).

w_leshst: Which description on this card [comes closest to what you first did after leaving full-time education?]{first spell}/[best describes what you did next, even if it was only for a month?]{second spell onwards}

Universe: Non-student | `w_lgaped = 1`

0. Current status reached, no further changes
1. Self employed
2. Full-time employed
3. Full-time employed
4. Unemployed
5. Retired
6. Maternity leave
7. Looking after family or home
8. Full-time student/at school
9. Long-term sick or disabled
10. On a government training scheme
11. National Service/War Service
12. Something else

4. Deriving Working-Life Histories in the BHPS and UKHLS

In this section, I discuss the working-life history code accompanying this paper. I begin by discussing the logic of the code and then describe the procedure for using the do files.

4.1 Underlying Logic

The over-arching aim of the code is to produce a single dataset containing a sequence of single, "main" activities for each individual in the BHPS and UKHLS, ideally uninterrupted from date of birth to last interview. Insisting upon a single activity is, of course, a simplification. Activities need not be mutually exclusive. For instance, individuals may work alongside full-time education or have caring responsibilities whilst being retired. However, more is not possible with the data.

The code proceeds by cleaning data from each of the separate working-life modules within the two surveys. It then merges this data together, giving precedence to particular modules over others and using lower precedence modules only where gaps remain. The pre-merged files are saved separately, in case the user wants the data from one or a few modules only.

Defining A Spell

To derive an accurate sequence of "main" activities for a participant, one must know when an activity began and when it ended. However, identifying correct start and end dates in the BHPS and UKHLS is not straightforward. This is for three reasons. First, only start or end dates are collected in many of the activity history modules, and most do not explicitly ask for dates when an activity was a main activity. Second, date information is often missing or incomplete. Third, activities may be mis-remembered and responses are often inconsistent across survey modules or waves. There is no way of cleaning the data that ensures complete accuracy. At most, we can only hope inaccuracies are kept within acceptable bounds.

To expand upon the first issue, in the Annual Activity History module in Waves 1-15 of the BHPS, participants are asked to give the start date of their current activity, the start date of the activity carried out immediately before then, and the start date of the activity immediately before that, and so on until a specified date is reached (September 1st in the year before fieldwork commenced). The wording of the questions does not require participants to state when an activity was a main activity or when the activity ended at all. To create a sequence from these data, an assumption must be made about these dates.

Here, I assume that, where start dates are elicited, the end date of an activity is the date the subsequent activity began and, where end dates are elicited, the start date is the date the prior activity ended. I also assume that a given activity is the main activity in the time between two adjacent start or end dates. This assumption is clearly inaccurate where activities

overlap and where the main activity alternates through time. The assumption is likely to hold for most individuals - Table 15 and Table 21 show the majority of individuals give similar answers to different questions about current activity - but unlikely to hold for all. Yet, it is required to create the activity histories.

No assumption is required to obtain start and end dates in the Annual Education History modules as these are asked directly.[24] When merging the files, I assume that spells of full-time education are the main activity. I do this because I am primarily interested in unemployment. Job-seeking to support education is likely to be qualitatively different than job-seeking at a sole activity: education may provide many of the 'latent' benefits associated with employment, such as an identity, time structure, and social network (Jahoda, 1982). Further, unemployment alongside education is less likely to give a negative signal to prospective employers.[25]

Imputing Missing Dates

I derive activity start and end dates at the month-year level. Day dates are available in some modules, but I discard these given the high levels of missingness and likely inaccuracies in the non-missing data – can you remember the exact day you started your previous job?

I work with three date variables: month, season and year. Any or all of these may be missing (though a season is implied by a month). In this section, I outline the process followed in the code for imputing missing dates. I adopt a more conservative approach for imputing dates than Maré (2006). I do not impute dates for activities where the range of feasible values is (in my opinion) too wide, choosing to leave gaps instead. These gaps may be filled by the reader using principled methods, such as multiple imputation, which are preferable to the rudimentary approach used here (in essence, taking the midpoint from the possible range of dates). I do not use multiple imputation here because (a) methods are developing and (b) imputation models and substantive models (the final model of interest) should be "compatible" - that is, described by the same joint distribution (van Buuren, 2018). Compatibility cannot be ensured in advance.

I use a sequential method for imputing dates. In the first step, I create upper and lower bounds for the date to be imputed based on the existing information available in the date itself (e.g. year observed but not month), the dates observed from prior and subsequent spells, and dates implied by the data collection process (e.g. the interview date and the previous interview, where dependent interviewing is used). For instance, if a respondent states an activity started in 2008 but does not specify the month, then I infer that the spell started sometime between

[24] BHPS Waves 8-18 and UKHLS Waves 2-9. As mentioned, the Annual Education History module in Waves 2-7 of the BHPS did not include start and end dates so is not used here.

[25] When using this dataset, the analyst should consider ignore activities between spells of education – for instance, during summer holidays – as again these may be qualitatively different to other spells of unemployment. See Furlong (2006).

January and December 2008, inclusive. If a later spell began in September 2008, then I can adjust the upper bound downwards further. In some cases, the upper and lower bound are identical for month-year, season-year or year variables. Where this is the case, the missing date is overwritten with this value as it is implied by the other data.

Sometimes, season but not month is available. In this case, I calculate the upper and lower bounds (at a maximum) using the following rule:

Season	**Lower Bound**	**Upper Bound**
(Early-Year) Winter	January	February
Spring	March	May
Summer	June	August
Autumn	September	November
(Late-Year) Winter	December	December

Note, in the Lifetime Employment History module in Wave 2 of the BHPS, winter is not split into early or late year winter. In this case, I first use the lower and upper bound dates to check whether timing can be inferred from the data - which I can if, for instance, an activity is followed by another beginning in April of the same year. If possible, I set the season and follow the procedure above. Otherwise, I change season to missing.

Where calculating the bounds uses information from subsequent or prior spells, it is assumed that the participants records activities chronologically. This should be true given how questions are worded, but in practice, some individuals record spells non-chronologically. I discard a participant's data (for that module in that wave) where responses can be identified as ordered non-chronologically and where it matters for imputing missing values (i.e. where they have a missing date that could be before or after another recorded spell).

Following this step, I then impute month-years for spells which have season-year information available. If a given activity is the only spell in a particular season, I set the month using the following rule: January if early-year winter; April if spring; July if summer; September if autumn; December if late-year winter. If there are several spells in a particular season-year, I divide the gap between the upper and lower bounds equally between the number of spells with the same upper and lower bounds. This procedure may create non-integer month values, which I leave rather than round. This works where the data are stored in spell format but needs accounting for if the user would like data to store the data by month (for instance, if carrying out Sequence Analysis).[26]

The next step is to impute dates for activities which have month and season and/or year missing. (Spells which have month or season but not year missing are treated as fully

[26] In this case, just use the Stata functions `ceiling()` or `floor()` to amend the dates.

missing as I assume the observed data to be uninformative.) First, the lower and upper bounds are recalculated based on the newly imputed data. Then, dates are imputed for any activity where the gap between the lower and upper bound does not exceed a threshold value. By default, I set the threshold to six months, though users are able to change this (see Section 4.2 on using the do files). The imputation procedure is the same as above where the gap is divided equally between spells which share the same upper and lower bound values.

This is carried out first for spells with month missing only, then for spells with missing month and year. (Upper and lower bounds are recalculated between these two steps.) The reason for doing this sequentially, rather than at the simultaneously, is that, where a spell with missing month is adjacent to a spell with missing month and year, the upper and lower bounds for the spells will only partially overlap. In some cases, the simultaneous approach can lead to spells losing their chronological order. More frequently, it can make the time between spells more compressed. See the table below.

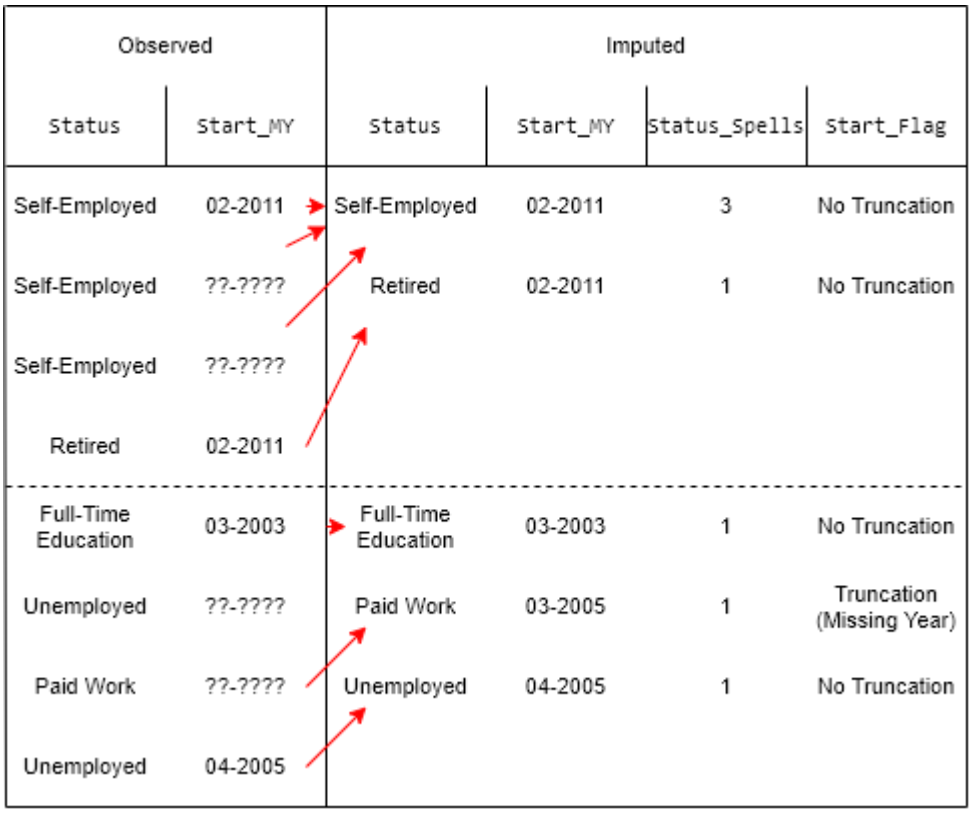
Observed Date	**Sequential Imputation**		**Simultaneous Imputation**
	Step 1	**Step 2**	
08-2009	08-2009	08-2009	08-2009
??-2009 Upper Bound: 08-2009 Upper Bound: 12-2009	10-2009	10-2009	10-2009
??-???? Upper Bound: 08-2009 Upper Bound: 02-2010	??-???? Upper Bound: 10-2009 Upper Bound: 02-2010	12-2009	11-2009
02-2010	02-2010	02-2010	02-2010

I choose a threshold of six months as, where there is only one spell in a specific gap, the imputed date can be out by at most three months. This may be too inaccurate for some readers, in which case the threshold can be reduced. It may be too conservative for others, in which case the threshold can be increased. In my experience, though, this threshold fills a high number of the gaps in the data without making unreasonable assumptions about the length of particular activities. Note, if the threshold value is set to zero, the step in which month is imputed from season is still carried out.

At this point, there may still be some activities with missing dates. Rather than estimate these dates, the aim now is to use available information to define a minimum period in which the activity could have been carried out. For instance, if one activity which has month missing but is observed to have started in 2008 and is followed by an activity that began in 2010, the latest that activity could have started is December 2008. If month and year are missing, the activity is assumed to have started in the month prior to the subsequent spell, at the latest. Similar reasoning is used where end dates are estimated instead: if an activity with year observed only is preceded by an activity which ended in an earlier year, the earliest in could have ended is in January of the observed year. If year and month are missing - or the

two activities are from the same year - the activity is assumed to have ended in the month after the prior spell ended, at the earliest.

This logic generates dates which are truncated and leaves gaps in sequences. If multiple adjacent activities still have missing dates after the imputation steps, minimum period information is only inferable for the first activity in the sequence, if adjusting end dates, or for the last activity in the sequence, if adjusting start dates. In general, this means the other activities with missing dates have to be dropped, but where the adjacent activities are of the same type (identified by status and, for employment, job hours), I collapse these spells to bridge the gap (see below). I record the number of reported spells an activity is made up of in a variable `Status_Spells`. The dataset also includes status end reason and job attraction information, which contain the count of spells with the specific end reason or job attraction. Whether the date have been truncated or imputed is recorded in two flag variables, `Start_Flag` and `End_Flag`.



Truncating the dates rather than throwing the data away entirely has two advantages. First, it provides long sequences which - in my view - only depend on weak assumptions. Second, it retains information on the activities which are likely to extend across the gap. This is useful if the reader wants to impute the missing sequence data using, for instance, the Halpin (2012, 2013, 2016) method, which uses the statuses either side of a gap to impute the statuses within the gap.

The procedure just described for imputing dates is carried out in each of the Annual Employment History and Employment Status History modules, separately by module and by

wave. The process for imputing missing Annual Education History dates differs slightly in that the dates from other spells are not used to create upper and lower bounds as spells of education do not need to be consecutive.

Once start dates have been imputed, the end dates are then calculated from information in other spells or from the interview procedure itself (or vice versa) – for instance, the end dates for a current activity is set as the date of the interview. Following this, the algorithm then collapses sequences within a module across waves. I use a simple rule to do this: where spells overlap, the start date of the spell elicited in the later wave is truncated to the month the spell from the earlier wave ended. This rule creates seam effects, meaning the timing and number of status transitions is likely to be inaccurate, but this should reduce error from recall bias.

The cleaned data from each module are saved separately in the files:

- BHPS Life History.dta
- UKHLS Life History.dta
- BHPS Education History.dta
- UKHLS Education History.dta
- BHPS 1-15 Annual History.dta
- BHPS 16-18 Annual History.dta
- UKHLS 16-18 Annual History.dta

Note, for the UKHLS Annual History module, I prepend an initial employment or retirement spell which is collected during a participants first interview.

Creating Initial Spells of Full-Time Education

In the code, I also create an initial spell of education, which, where possible, spans from birth to the point at leaving full-time education for the first time (or to last follow-up where the participant remains in education). The date an individual first left full-time education is only asked directly in the Employment Status History module. Where available, I use this to create a single spell of education which starts at birth and ends at the date provided. Unlike in other modules where I am more conservative, I assume the education ended in June, if not observed: June is around the time academic courses typically finish in the UK.

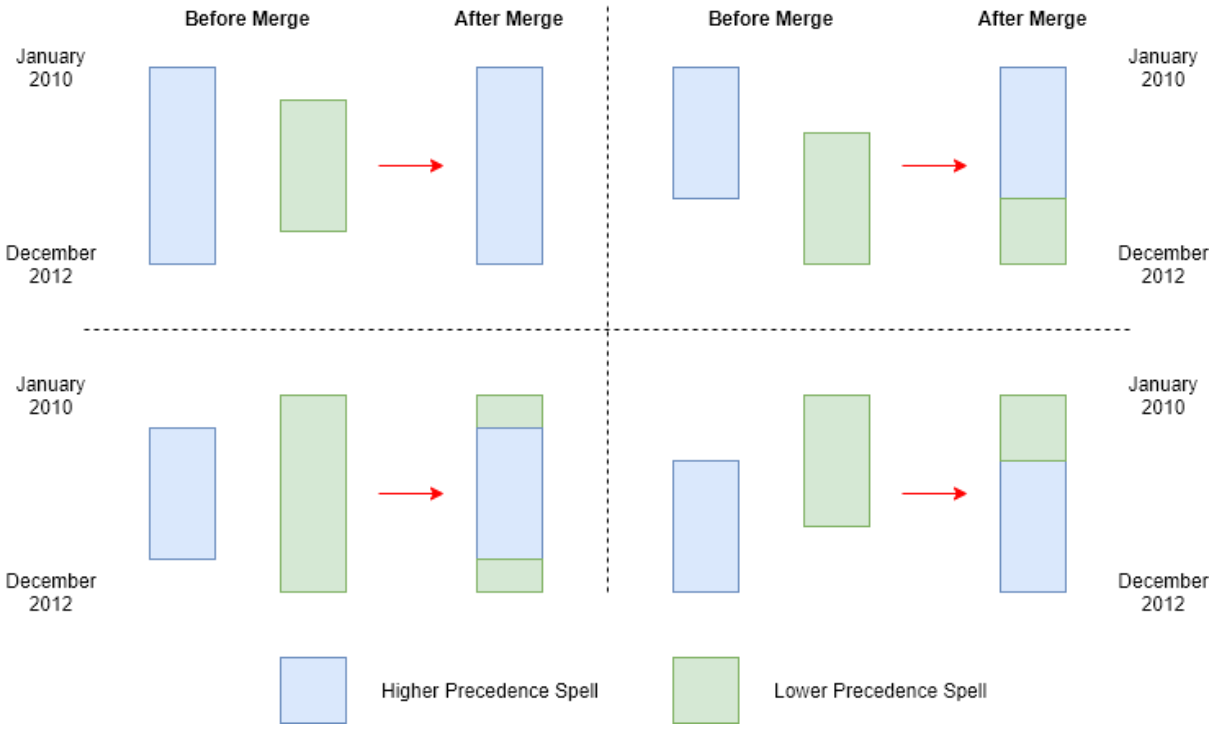
For participants who do not have information from the Employment Status History module, I use responses to questions on what age the participant (a) left school and (b) left further education. I assume that the end month of these spells was the June of or after their birthday.[27] If the individual left school before age 20 and - if applicable - further education before age 24, I assume there was no gap in the participant's education. However, if the participant left at a later age, I allow that there may have been a gap. If the participant left

[27] Where birth month is not available – i.e. is missing or not included with the dataset (as with the End User Licence dataset) – I assume the participant is born in June.

school before age 20 but left full time education at age 24 or later, I create a spell from birth to the date of leaving school, and record this as a potential ongoing school. As individuals may also state they are still in education, rather than that they have left, if these responses are taken at an age which suggest no gap in education (school before age 20, further education before age 24), I also create a spell of education from birth to the interview date which is identified as ongoing. The derived education dates are used in the final dataset and also stored in a dataset, "Education Variables – Cleaned.dta" created by the code.

Merging Activity History Files

Once the data from the separate modules are cleaned, the algorithm then merges the datasets together. The data are first collapsed across waves for similar modules (e.g. the Annual History modules) using the same rule for individual modules described above. Then, the modules are brought together. I order the modules lexicographically where there is overlap, with Annual Education History data taking precedence over Annual History data taking precedence over the Employment Status History data. My reasoning is that, where an individual is in full-time education, this should take precedence over other activities, and that Annual History data should take precedence over Employment Status History data because of the shorter recall period and higher detail of questions asked. The following graphic shows how overlapping spells are adjusted. Once the Annual History, Annual Education History and Lifetime Employment Status history modules have been merged, the initial education spell is prepended.

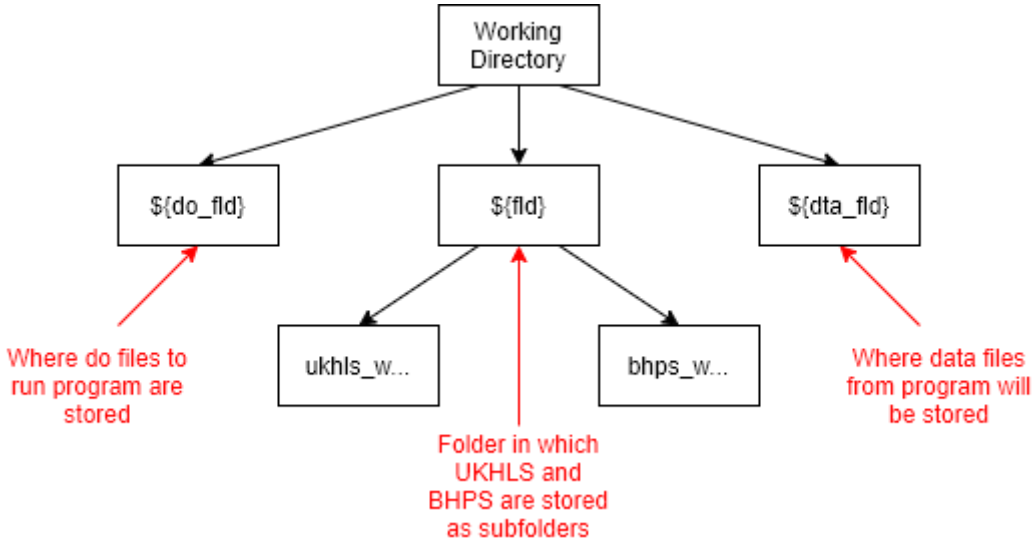


4.2 Using the Do Files

The algorithm is written as a set of Stata .do files. Stata is proprietary and requires a (rather expensive) licence. I now realise that I should have used an open source language, such as R. I apologise for this.

The do files are straightforward to run but require a few inputs from the user. Instructions are below. You will need to download the UKHLS and BHPS harmonized dataset to run the files.[28] I cannot provide the working-life history data directly given licencing restrictions.

1. Open "Launch Programme.do".
2. Change the working directory and global macros in Lines 22-31 to reflect your folder structure.[29] The code will assume you have the folders structured as below. Please use forward slashes rather than back-slashes in file paths as the back-slash is an escape character in Stata.



3. Next – if necessary – change the macros in Lines 33-38 which contain the folder stubs for the BHPS and UKHLS files and the file suffix which is added to Special Licence Access files.[30]
4. Change the macros in Lines 40-47 which contain the number of UKHLS and BHPS waves you want to use and the wave numbers for BHPS and UKHLS life history waves (included for future-proofing.)
5. Change the maximum gap length you want to impute missing dates for in Line 50.

[28] Links here, here and here.

[29] In Stata, a global macro is a piece of text associated with a name. Macros can be used to create commands. For instance, if you have a global macro `x` which stores the value 12, the command `generate newvar = ${x}` will create a new variable, `newvar`, equal to 12. (`${x}` tells Stata that you want to use the macro `x`.) Macros are very useful. For more on macros, see the Stata manual.

[30] Before Wave 8, the UKHLS folder stub was `us_` not `ukhls_`, so included for future proofing.

6. Finally, change the global macro `run_full` to "YES" and press the Execute button to run the full program.
    - Alternatively, the do files for each module can be run individually by selecting relevant lines from Lines 77-94. Lines 58-66 will need to be run before doing this.

The code should take 10-20 minutes to run. It has been tested with the End User Licence and Special Licence BHPS Waves 1-18 and UKHLS Waves 1-9 harmonized datasets in Stata 16 and Stata 17. It will work with either Stata IC or Stata SE. To my knowledge, it should also work with earlier versions of Stata and the BHPS and UKHLS data, so long as harmonised surveys are used. The data files produced will be over 800mb in size.

I hope you find the working-life history files useful. Again, if you have any comments, please direct them to the GitHub page here.

All the best,

Liam

References

Furlong, A. (2006). Not a very NEET solution. *Work, Employment and Society*, *20*(3), 553–569. https://doi.org/10.1177/0950017006067001

Halpin, B. (1997). *Unified BHPS Work-Life Histories: Combining Multiple Sources into a User-Friendly Format*. http://doc.ukdataservice.ac.uk/doc/3954/mrdoc/pdf/techpap.pdf

Halpin, B. (2000). *BHPS Work-life History Files, Version 2*. http://doc.ukdataservice.ac.uk/doc/3954/mrdoc/pdf/newman.pdf

Halpin, B. (2012). *Multiple Imputation for Life-Course Sequence Data*. http://teaching.sociology.ul.ie/seqanal/shortptex.pdf

Halpin, B. (2013). *Imputing Sequence Data: Extensions to initial and terminal gaps, Stata's mi Imputing Sequence Data: Extensions to initial and terminal gaps, Stata's mi **. http://www.ul.ie/sociology/pubs/wp2013-01.pdf

Halpin, B. (2016). Multiple Imputation for Categorical Time-series. *Stata Journal*, *16*(3), 590–612.

Institute for Social and Economic Research. (2019). *Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009: Special Licence Access*. UK Data Service. http://doi.org/10.5255/UKDA-SN-6931-10

International Labour Organization. (2013). *Resolution I: Resolution concerning statistics of work, employment and labour underutilization*. https://www.ilo.org/wcmsp5/groups/public/—dgreports/—stat/documents/normativeinstrument/wcms_230304.pdf

Jahoda, M. (1982). *Employment and Unemployment: A Social-Psychological Analysis*. Cambridge University Press.

Maré, D. C. (2006). *Constructing Consistent Work-life Histories: A guide for users of the British Household Panel Survey*. https://www.iser.essex.ac.uk/files/iser_working_papers/2006-39.pdf

Oskrochi, G., & Crouchley, R. (2000). *Using SPSS and Gauss to Unify the BHPS Work History Data*.

Paull, G. (2002). *Biases in the Reporting of Labour Market Dynamics* (pp. 1–65). Institue for Fiscal Studies. https://www.ifs.org.uk/wps/wp0210.pdf

van Buuren, S. (2018). *Flexible imputation of missing data* (Second edition). CRC Press, Taylor & Francis Group.

Wright, L. (2019). *The 'Scarring' Effect of Youth Unemployment on Mental Health and Subjective Wellbeing* [Report for Transfer from MPhil to PhD, University College London]. https://bit.ly/2OGAnmQ

Appendix

A.1. Routing Diagrams

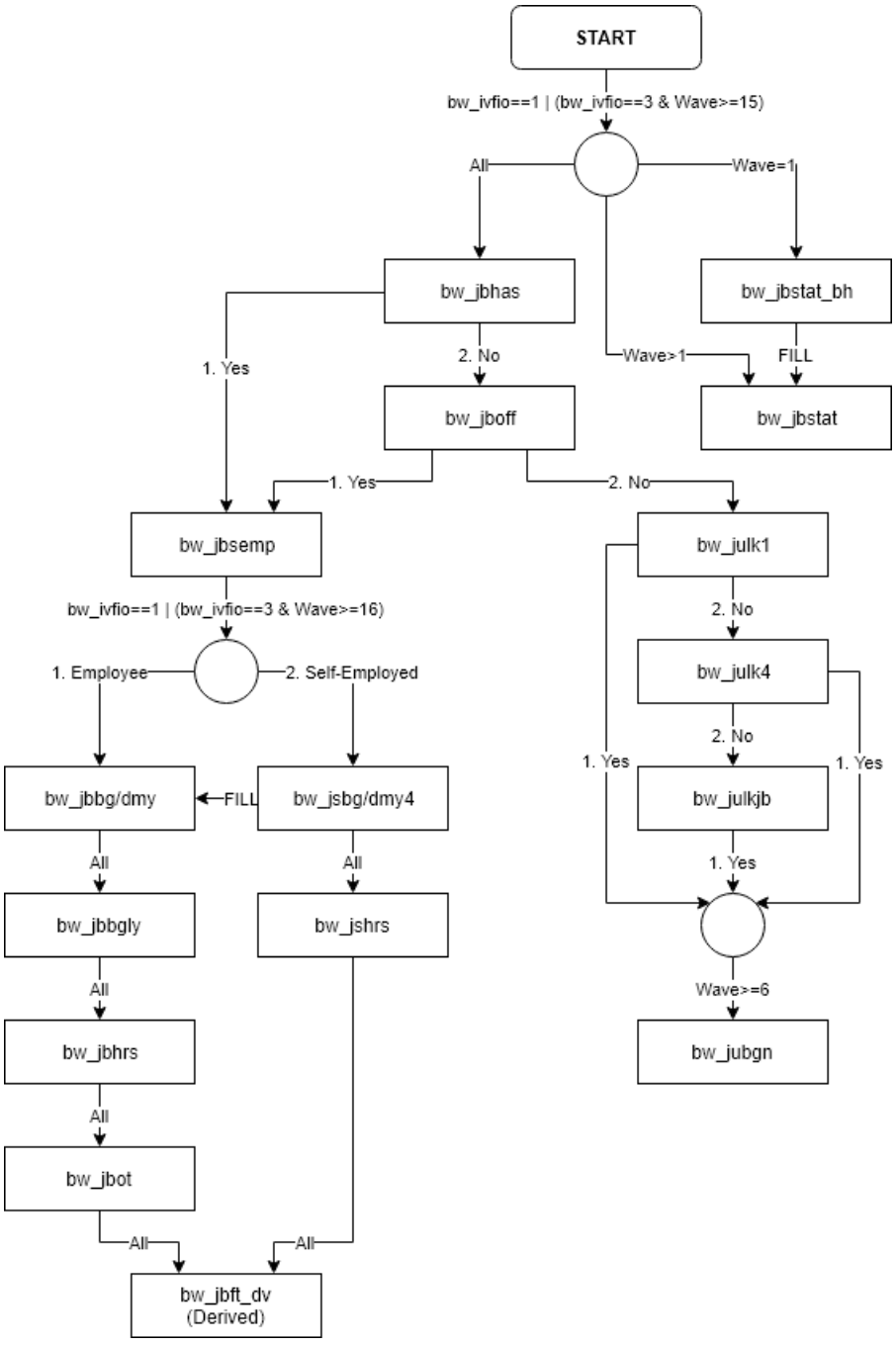


Figure 1: Routing of Current Activity and Job Questions, Full Interview, Waves 1-15 BHPS

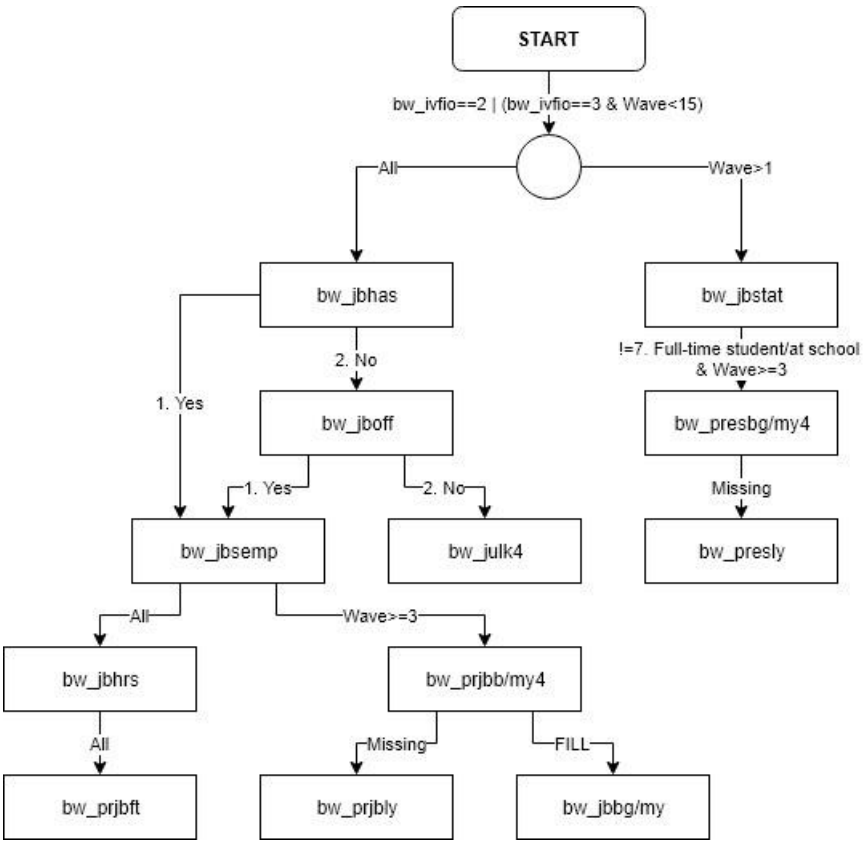


Figure 2 : Routing of Current Activity and Job Questions, Non-Full Interview, BHPS

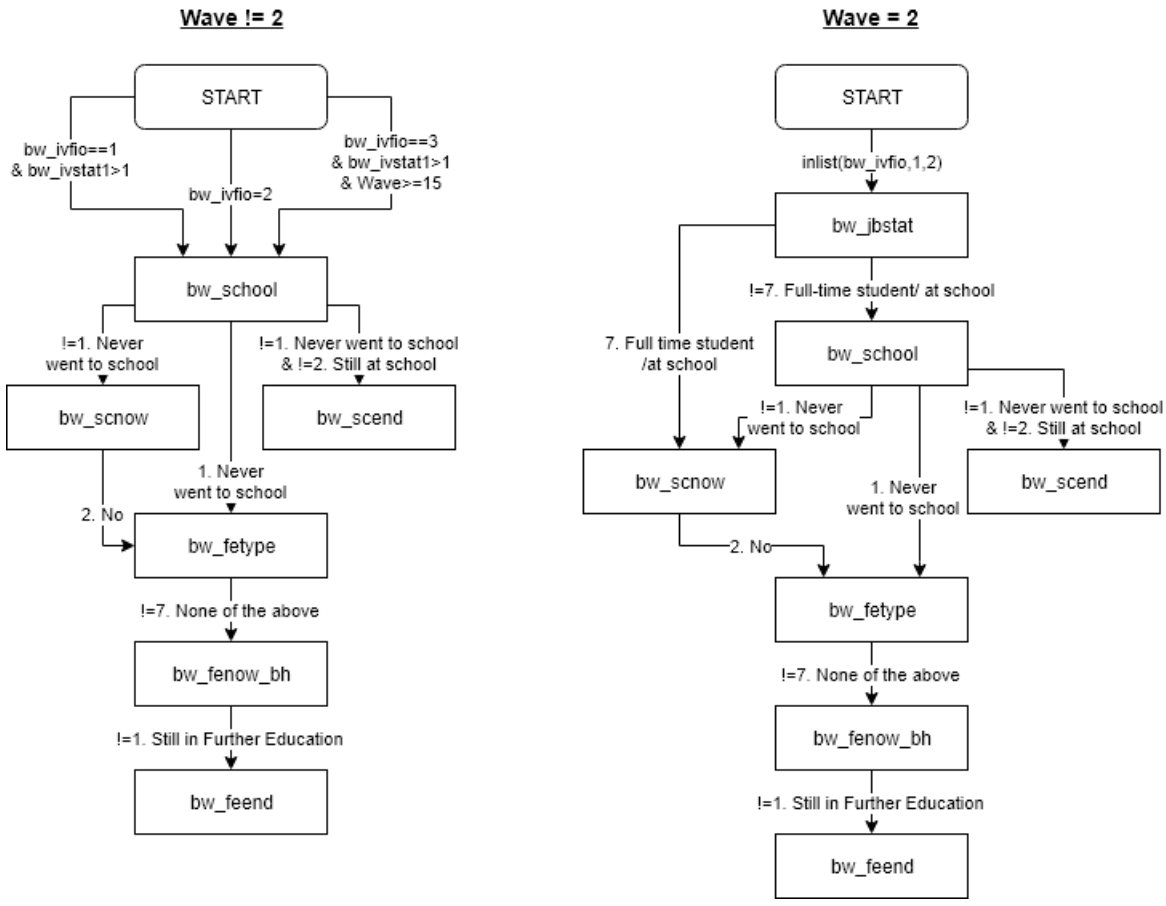


Figure 3: Routing of School and Further Education Leaving Age Questions, BHPS.

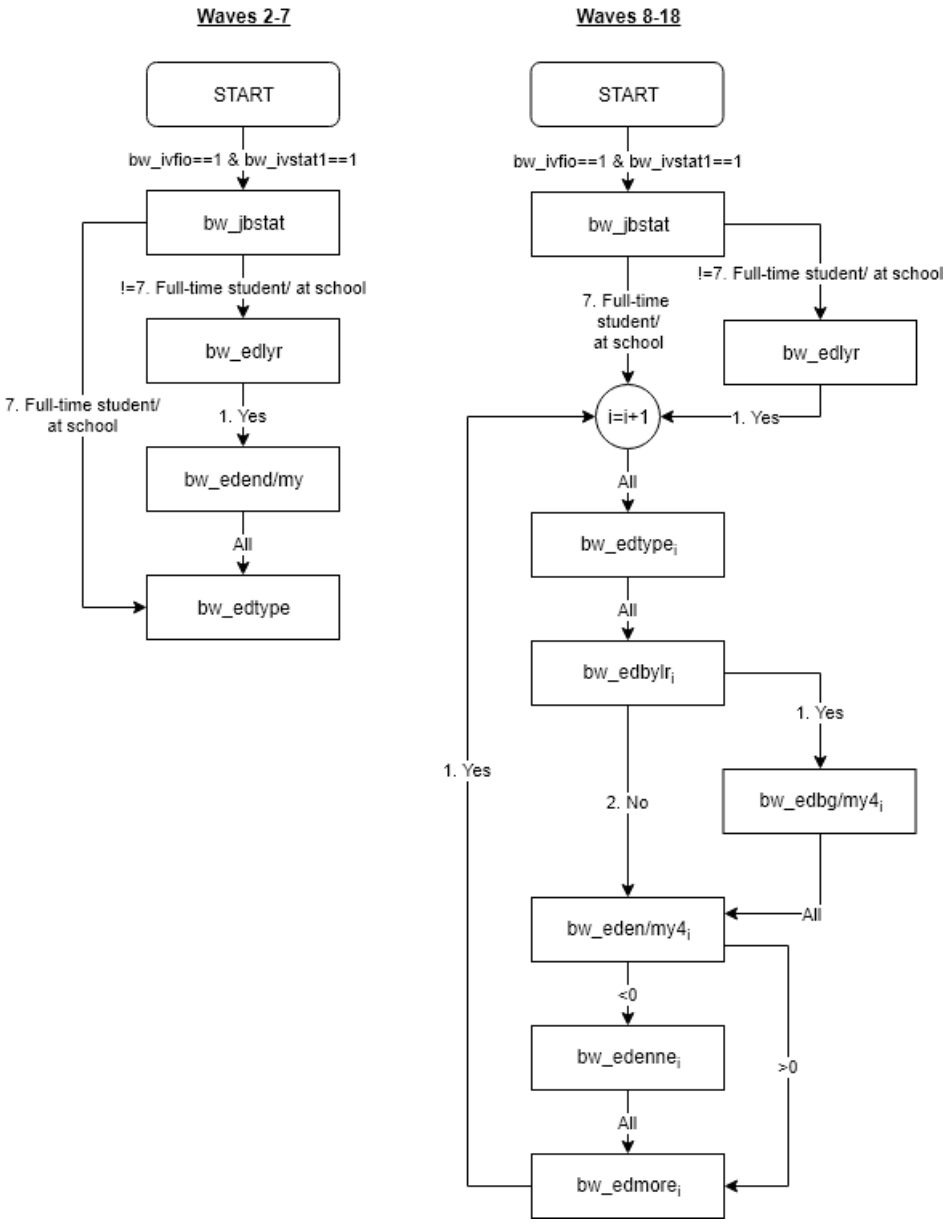


Figure 4: Full-time education since last wave, Waves 2-18, BHPS.

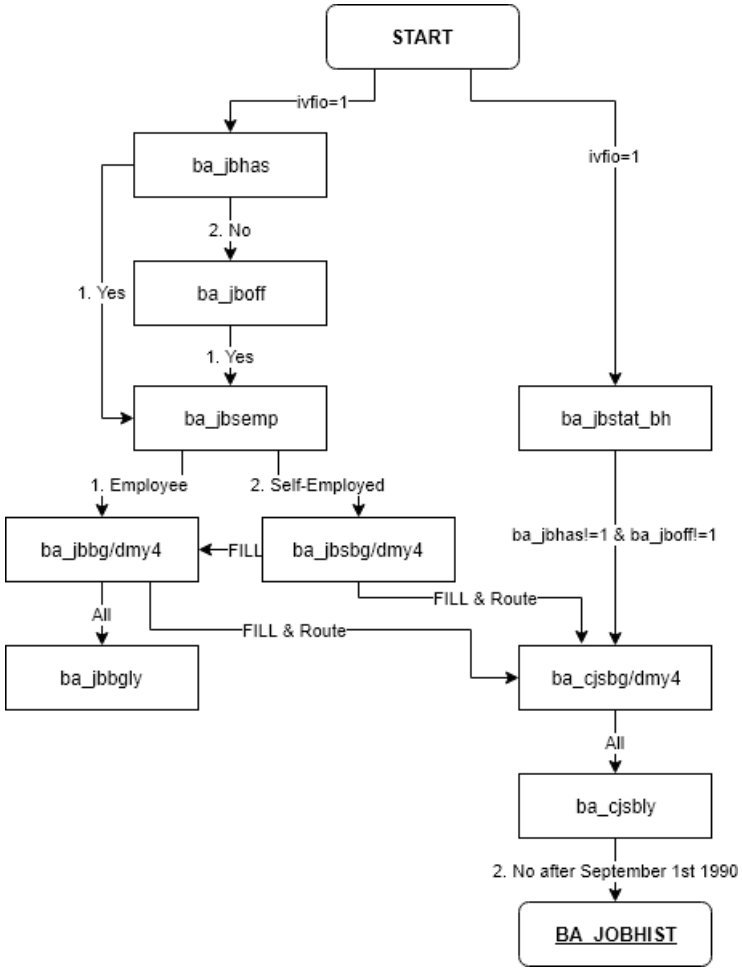


Figure 5: `ba_indresp` *Annual Employment History, Wave 1, BHPS.*

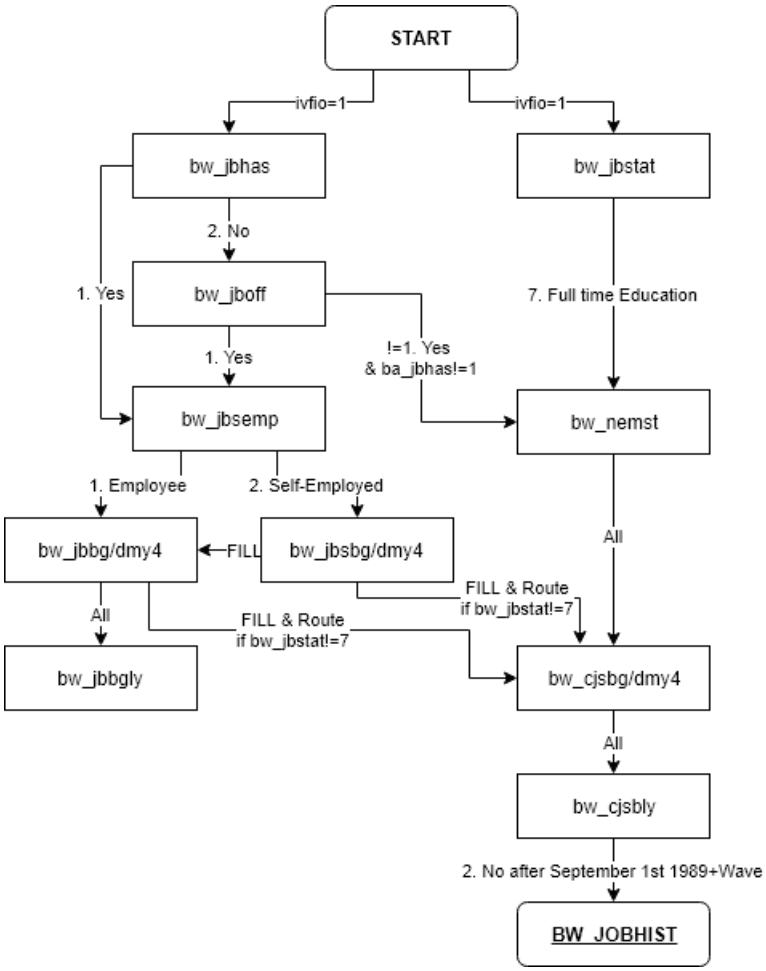


Figure 6: **`bw_indresp`** *Annual Employment History, Waves 2-15, BHPS.*

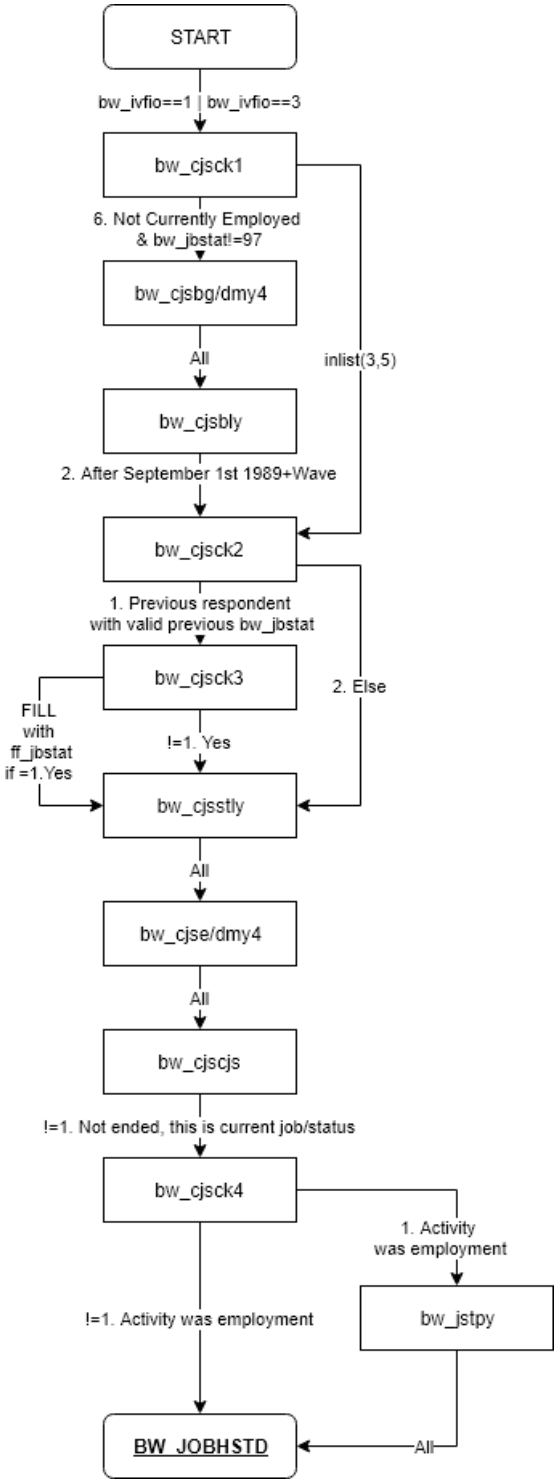


Figure 7: **`bw_indresp`** *Annual Employment History, Waves 16-18, BHPS.*

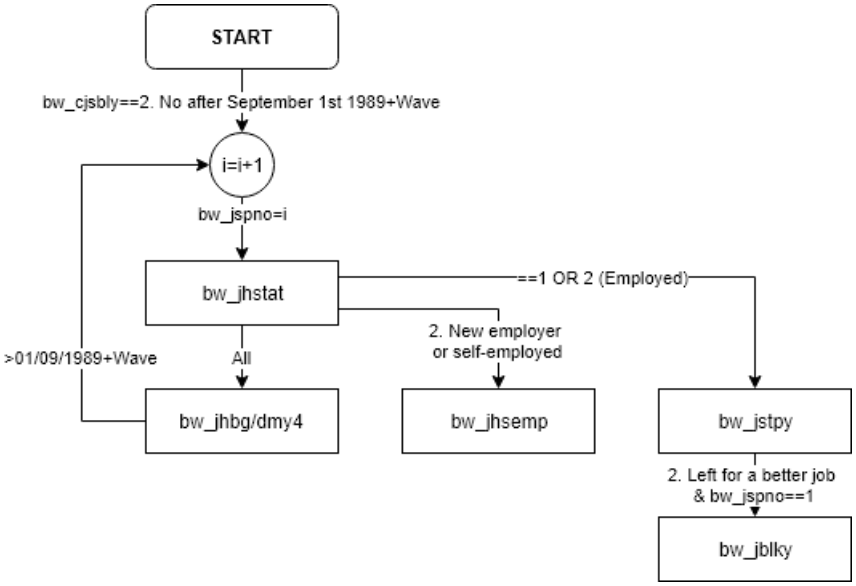


Figure 8: `bw_jobhist` *Annual Employment History, Waves 1-15, BHPS.*

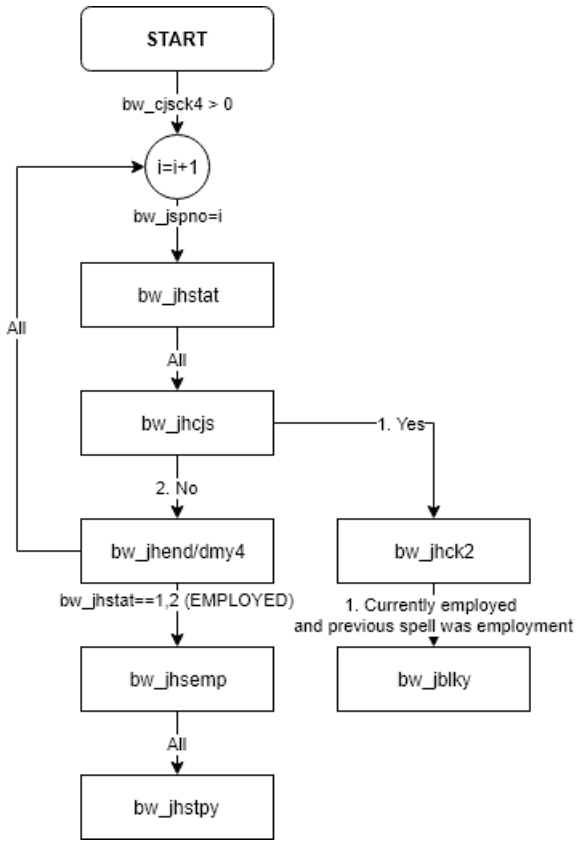


Figure 9: **`bw_jobhstd`** *Annual Employment History, Waves 16-18, BHPS.*

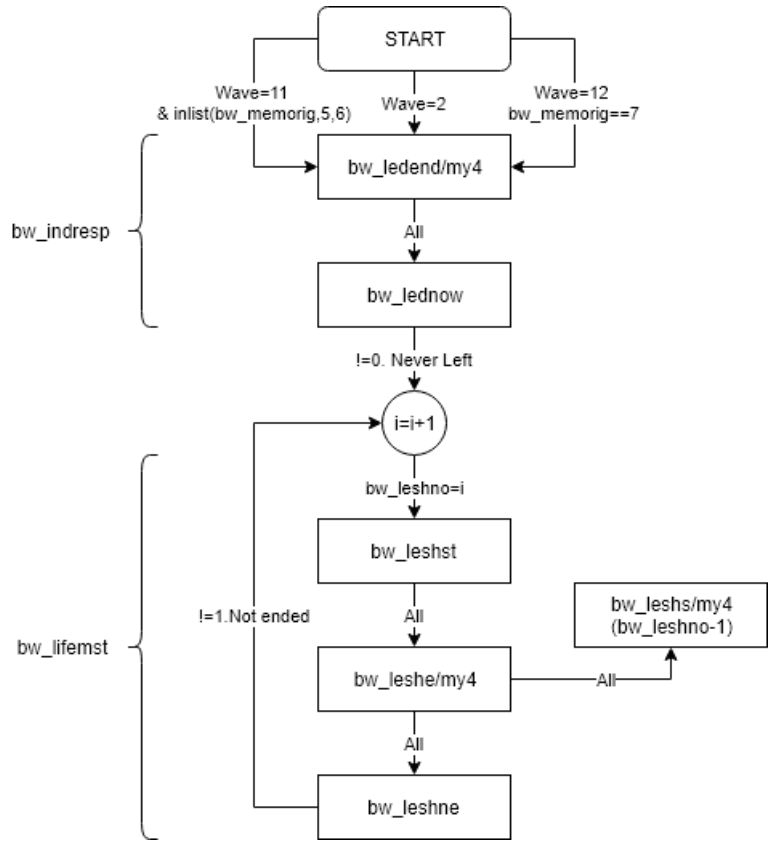


Figure 10: Lifetime Employment Status History, Waves 2, 11 and 12, BHPS.

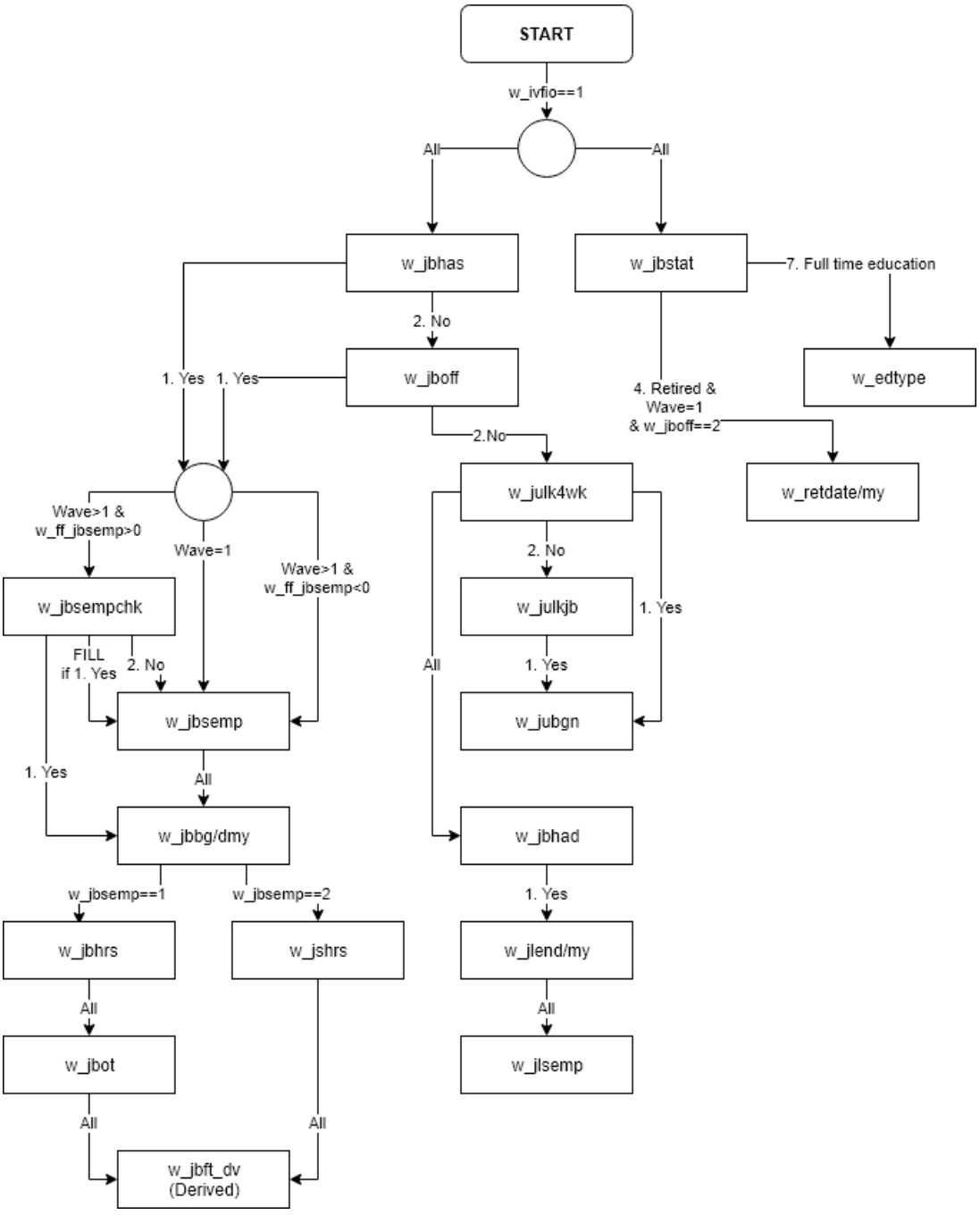


Figure 11: Current Employment, Full Interview, UKHLS.

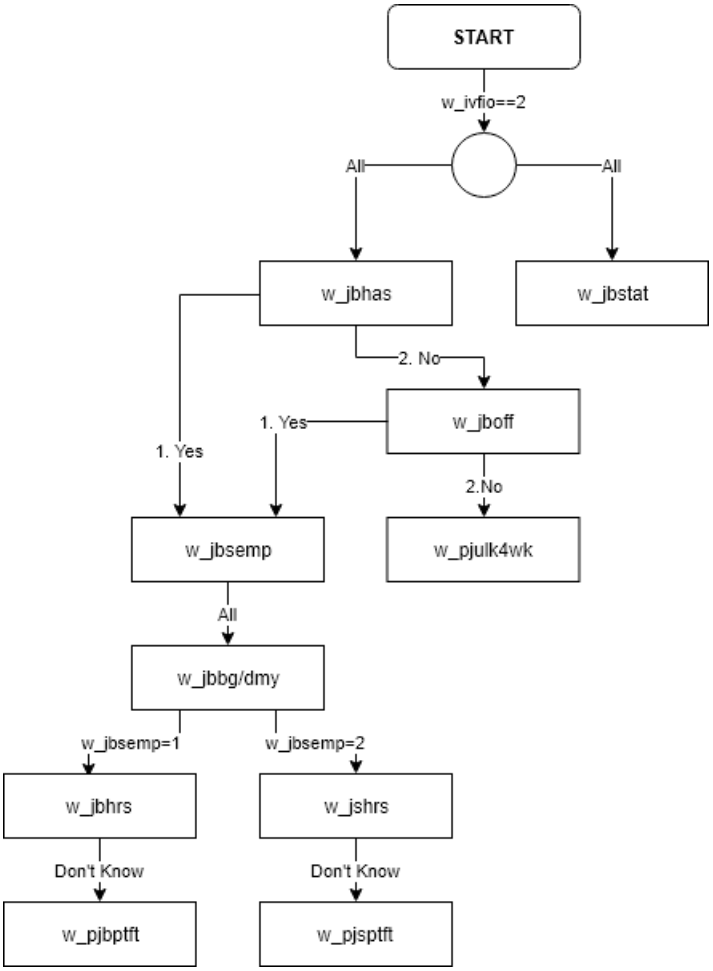


Figure 12: Current Employment, Proxy Interview, UKHLS.

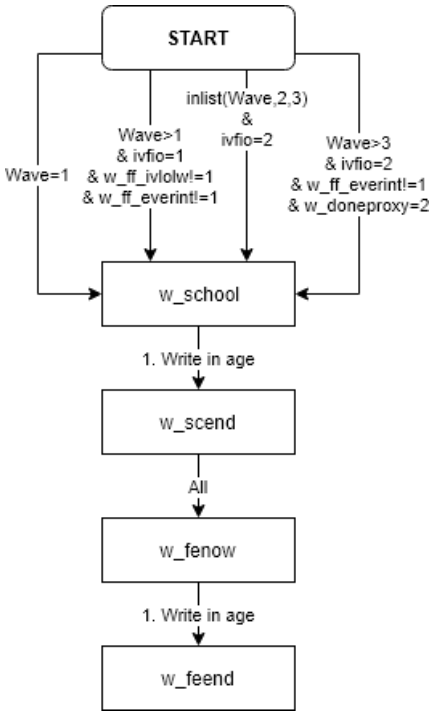


Figure 13: School and Further Education Leaving Ages, UKHLS.

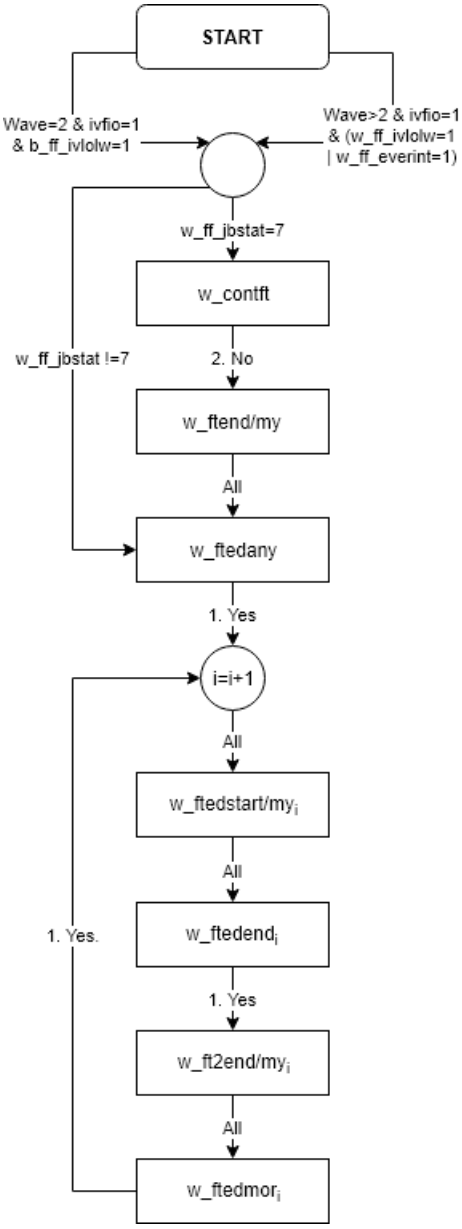


Figure 14: Full-Time Education since previous interview, Waves 2-9, UKHLS.

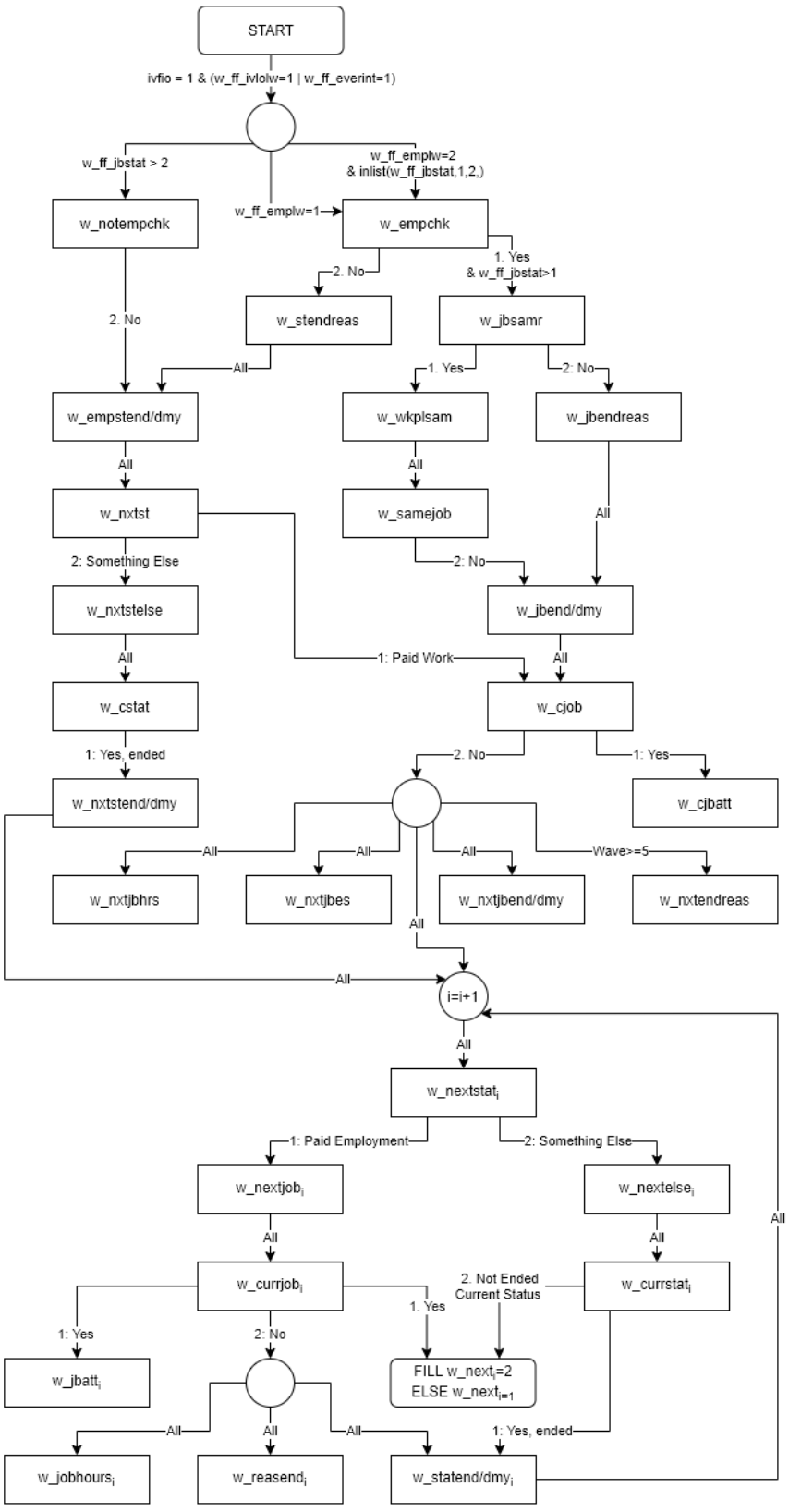


Figure 15: Employment history since previous interview, Waves 2-8, UKHLS.

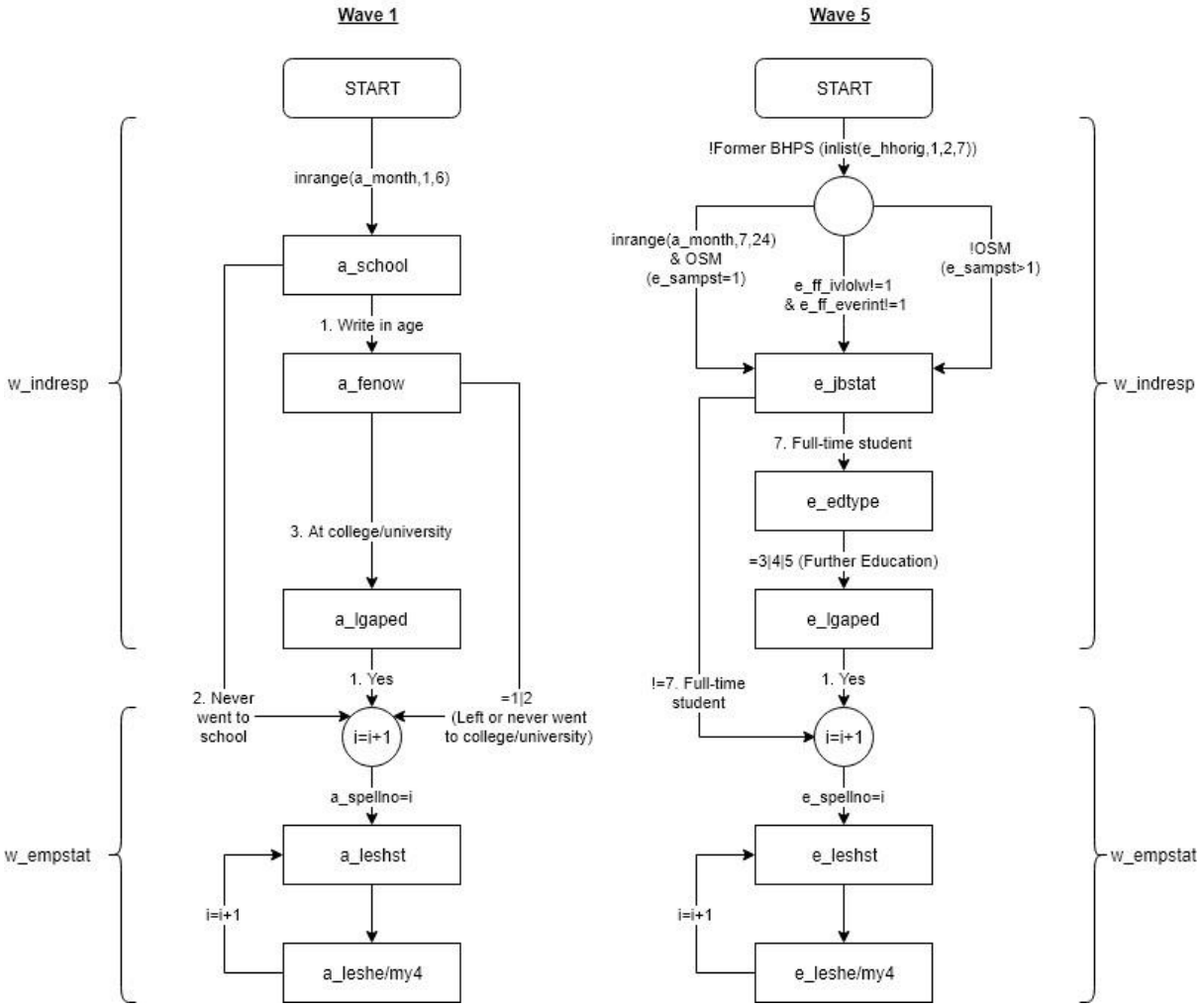


Figure 16: Employment Status History, Waves 1 and 5, UKHLS.

A.2. Question Wording

Table 2: Current Employment Questions, Waves 1-18 BHPS

Variable	Question	Responses	Universe
bw_jbhas	Did you do any paid work last week...?	1. Yes 2. No	All
bw_jboff	...did you have a job that you were away from last week?	1. Yes 2. No 3. Waiting to take up job	bw_jbhas=2
bw_jbsemp	Are you an employee or self-employed?	1. Employee 2. Self-employed	bw_jbhas=1 OR bw_jboff=1
bw_jbhrs	How many hours, excluding overtime and meal breaks, are you expected to work in a normal week?	≥ 0	bw_jbsemp=1 OR (bw_jbsemp=2 & bw_ivfio>1)
bw_jbot	And how many hours overtime do you usually work in a normal week?	≥ 0	bw_jbsemp=1 & bw_ivfio=1
bw_jshrs	How many hours in total do you usually work a week in your job?	≥ 0	bw_jbsemp=2
bw_jbft_dv	Is participant's job part- or full-time?	1. Full-time 2. Part-time	bw_jbhas=1 OR bw_jboff=1
bw_prjbft	Would you say (her/his) current job is part-time or full-time?	1. Part time 2. Full time	(bw_jbhas=1 OR bw_jboff=1) & bw_ivfio> 1
bw_jbbg/dmy4	What was the date you started working in your present position...?	DMY	bw_jbsemp=1 & bw_ivfio=1
bw_jbbgly	Is bw_jbbg/dmy4 before September 1st 1989+Wave?	1. Yes 2. No	bw_jbsemp=1 & bw_ivfio=1
bw_jsbg/dmy4	On what date did you start doing your present job...?	DMY	bw_jbsemp=2 & bw_ivfio=1
bw_prjbbg/my4	When did you start working in this job?	DMY	bw_jbsemp=2 & bw_ivfio> 1
bw_prjbly	Was it after September 1st 1989+Wave?	1. Yes 2. No	bw_prjbbg/my4< 0 & bw_ivfio> 1
bw_julk1	Have you looked for any kind of paid work or government training scheme in the last week...?	1. Yes 2. No	bw_jboff=2 & bw_ivfio=1
bw_julk4	Have you looked for any kind of paid work or government training scheme in the last four weeks?	1. Yes 2. No	bw_jboff=2 & bw_julk1!=1
bw_julkjb	...would you like to have a regular paid job even if only for a few hours a week?	1. Yes 2. No	bw_julk4=2 & bw_ivfio=1
bw_jubgn	If a job or a place on a government training scheme had been available in the week ending last Sunday, would you have been able to start within two weeks?	1. Yes 2. No	bw_julk1=1 OR bw_julk4=1 OR bw_julkjb=1 & bw_ivfio=1

Table 3: School and Further Education Leaving Ages, BHPS.

Variable	Question	Responses	Universe
bw_school	How old were you when you left school?	1. Never went to school 2. Still at school	bw_ivfio=3 OR (bw_ivfio=1 & bw_ivstat1!=1)
bw_scend	Write in age	> 0	bw_school$<$ 1
bw_scnow	Is respondent still at school?	1. Yes 2. No	bw_school!=1
bw_fetype	tell me which, if any, of these further education institutions have you attended full-time or are attending?	1. Nursing school/Teaching Hospital 2. College of futher/higher education 3. Other College or training establishment 4. Polytechnic/Scottish Central Institutions 5. University 7. None of above	bw_school=1 OR bw_scnow=2
bw_fenow	How older were you when you left there, or when you finished or stopped your course?	1. Still in further education	bw_fetype<7
bw_feend	Write in age	> 0	bw_fenow!=1

Table 4: Education Spells since Last Wave, BHPS 8-18

Variable	Question	Responses	Universe
bw_edlyr	Have you attended any educational institution full-time since September 1st 1989+Wave?	1. Yes 2. No	bw_jbstat!=7
bw_edend/my4	When did you leave this education institution?	-3. Not left >0 (if left)	bw_ivfio=1 & bw_edlyr=1
bw_edtype	Could youtell me what type of education institution (you are attending/you attended last)?	Various	bw_jbstat=7 OR bw_edlyr=1
bw_edlyr$_i$	Have you attended any educational institution full-time since September 1st 1989+Wave?	1. Yes 2. No	bw_jbstat!=7
bw_edtype$_i$	Could youtell me what type of education institution (you are attending/you attended last)?	Various	bw_ivfio=1 & bw_edmore$_{i-1}=1$
bw_edbylr$_i$	Did this period of full-time education start after September 1st 1989+Wave?	1. Yes 2. No	bw_ivfio=1 & bw_edmore$_{i-1}=1$
bw_edbg/my4$_i$	On what date did you start this period of education?	>0	bw_edbylr$_i=1$
bw_eden/my4$_i$	When did you leave this education institution?	>0 (if left)	bw_ivfio=1 & bw edmore$_{i-1}=1$
bw_edenne$_i$	When did you leave this education institution?	1. Not ended	bw_eden/my4$_i<0$
bw_edmore$_i$	Have you attend any other education institution full-time since September 1st 1989+Wave?	1. Yes 2. No	bw_ivfio=1 & bw_edmore$_{i-1}=1$

Table 5: ***bw_indresp*** *Employment History Questions, Waves 1-15, BHPS*

Variable	Question	Responses	Universe
bw_nemst	Please look at this card and tell me which best describes your current situation?	Various	Wave>=2 & (bw_jbstat=7 OR bw_jboff>1)
bw_cjsbg/dmy4	On what day did your present spell of [bw_nemst] begin?	DMY	All
bw_cjsbly	Is [bw_cjsbg/dmy4] after September 1st 1989+Wave or before?	1. Yes, September 1st 1989+Waveor before 2. No, after September 1st 1989+Wave	All

Table 6: ***bw_indresp*** *Employment History Questions, Waves 16-18, BHPS*

Variable	Question	Responses	Universe
bw_cjsck1	CHECK: Individual changed status/job since last interview/wave	Various	All
bw_cjsbg/dmy4	You told me earlier that you were [bw_jbstat]. On what date did your present spell of [bw_jbstat] begin?	DMY	bw_cjsck1=6 & bw_jbstat!=97
bw_cjsbly	Is date at [bw_cjsbg/dmy4] before <INTDATE>/ September 1st 1989+Wave?	1. Yes, before <INTDATE>/ September 1st 1989+Wave. 1. No, <INTDATE>/September 1st 1989+Wave or after.	bw_cjsck1=6 & bw_jbstat!=97
bw_cjsck2	CHECK: Respondent has valid bw_jbstat from previous interview	1. Yes 2. No	bw_jbstat=2
bw_cjsck3	When we interviewed you on <INTDATE>, you were [ff_jbstat]. Is that correct?	1. Yes 2. No	bw_cjsck2=1
bw_cjsstly	Please look at this card and tell me which best describes your situation on September 1st 1989+Wave/ <INTDATE>.	Various	bw_cjsck1=3 OR bw_cjsck1=5 OR bw_cjsck3=2
bw_cjse/dmy4	On what date did you stop doing that job/activity?	DMY	bw_cjsck1=3 OR bw_cjsck1=5 OR bw_cjsck3=2
bw_cjscjs	On what date did you stop doing that job/activity?	1. Not ended, this is current job/status	bw_cjsck1=3 OR bw_cjsck1=5 OR bw_cjsck3=2
bw_cjsck4	CHECK: bw_cjse/dmy4>0 & activity was employment	1. Yes 2. No	bw_cjscjs!=1
bw_jstpy	Would you...tell me which of these statements best describes why you stopped doing that job?	Various	bw_cjsck4=1

Table 7: ***bw_jobhist*** *Employment History Questions, Waves 1-15, BHPS*

Variable	Question	Responses	Universe
bw_jhstat	Can you look at this card please and tell me which of the descriptions comes closest to what you were doing immediately before then?	Various	bw_cjsbly=2
bw_jhbg/dmy4	And on what date did you start doing that?	DMY	bw_cjsbly=2
bw_jhsemp	Were you a full-time employee, a part-time employee, or self-employed?	1. Full-time employee 2. Part-time employee 3. Self-Employed	bw_jhstat=2
bw_jstpy	Would you...tell me which of these statements best describes why you stopped doing that job?	Various	bw_jhstat<=2
bw_jblky	What was the main thing about your present job that attracted you to it?	Various	bw_jhstat<=2 & bw_jspno=1

Table 8: ***bw_jobhist*** *Employment History Questions, Waves 16-18, BHPS*

Variable	Question	Responses	Universe
bw_jhstat	...can you tell me which of the descriptions comes closest to what you were doing immediately after...?	Various	bw_cjsck4!=0
bw_jhcjs	...is this your current job/activity?	1. Yes 2. No	bw_cjsck4!=0
bw_jhck2	CHECK: Current and previous spells were employment	1. Yes 2. No	bw_jhcjs=1
bw_jblky	What was the main thing about your present job that attracted you to it?	Various	bw_jhck2=1
bw_jhend/dmy4	And on what date did you stop doing that and start your next job or other activity?	DMY	bw_jhck2=2
bw_jhsemp	Were you a full-time employee, a part-time employee, or self-employed?	Various	bw_jhstat<=2 & bw_jhck2=2
bw_jstpy	Would you...tell me which of these statements best describes why you stopped doing that job?	Various	bw_jhstat<=2 & bw_jhck2=2

Table 9: Lifetime Employment Status History Questions, Waves 2, 11 and 12, BHPS

Variable	Question	Responses	Universe
bw_ledend/my4	When was the first time you left full-time education?	MY	See Figure 10
bw_lednow	When was the first time you left full-time education?	0. Never Left 1. Never went to school	See Figure 10
bw_leshst	Which description...comes closest to what you first did after leaving full-time education/did next...?	Various	bw_leshne!=1
bw_leshe/my4	When was the next time your situation changed...?	MY	bw_leshne!=1
bw_leshne	When was the next time your situation changed...?	1. Not Ended	bw_leshne!=1

Table 10: Current Employment Questions, Waves 1-9 UKHLS

Variable	Question	Responses	Universe
w_jbhas	…did you do any paid work last week - that is in the seven days ending last Sunday - either as an employee or self-employed?	1. Yes 2. No	All
w_jboff	...did you have a job that you were away from last week?	1. Yes 2. No 3. Waiting to take up job	bw_jbhas=2
w_jbsempchk	And are you still an {if ff_JBSEMP = 1} [ff_JBSEMP] ?	1. Yes 2. No	(w_jbhas =1 OR w_jboff=1) & inlist(ff_jbsemp,1,2)
w_jbsemp	Are you an employee or self-employed?	1. Employee 2. Self-employed	bw_jbhas=1 OR bw_jboff=1
w_jbbg/dmy	And on what date did you start working in your present job? If you have been promoted or changed grades, please give me the date of that change. Otherwise please give me the date when you started doing the job you are doing now for your present employer or working self-employed.	MY	(w_jbhas =1 OR w_jboff=1) & (missing(ff_jbstat) OR missing(ff_emplw))
w_jbhrs	Thinking about your (main) job, how many hours, excluding overtime and meal breaks, are you expected to work in a normal week?	≥ 0	bw_jbsemp=1 OR (bw_jbsemp=2 & bw_ivfio>1)
w_jbot	And how many hours overtime do you usually work in a normal week?	≥ 0	bw_jbsemp=1 & bw_ivfio=1
w_jshrs	How many hours in total do you usually work a week in your job?	≥ 0	bw_jbsemp=2
w_jbft_dv	Is participant's job part- or full-time?	1. Full-time 2. Part-time	bw_jbhas=1 OR bw_jboff=1
w_pjbptft	Would you say that [NAME] 's current job is part-time or full-time?	1. Part time 2. Full time	w_jbsemp=1 & w_jbhrs=DK & w_ivfio> 1
w_pjsptft	Would you say that [NAME] 's current job is part-time or full-time?	1. Part time 2. Full time	w_jbsemp=2 & w_jshrs=DK & w_ivfio> 1
w_julk4wk	Have you looked for any kind of paid work or government training scheme in the last four weeks?	1. Yes 2. No	w_jboff=2

Variable	Question	Responses	Universe
w_julkjb	...would you like to have a regular paid job even if only for a few hours a week?	1. Yes 2. No	w_julk4wk=2 & w_ivfio=1
w_jubgn	If a job or a place on a government training scheme had been available in the week ending last Sunday, would you have been able to start within two weeks?	1. Yes 2. No	w_julk4wk=1 OR w_julkjb=1
w_j1none	Since leaving full-time education, have you ever had a paid job?	1. SPONTANEOUS Still in full-time education 2. Never had a paid job 3. Had a paid job	inlist(ff_ivlolw=2,3,MIS) & ff_everint!=1 & Wave>1
w_jbhad	Have you ever had a paid job at all, apart from any casual or holiday work?	1. Yes 2. No	w_jboff=2 & ff_ivlolw!=1 & ff_everint!=1
w_jlend/my	In what month and year did you leave your last paid job?	MY	w_jbhad=1 & ff_ivlow!=1 & ff_everint!=1
w_jlsemp	Were you working as an employee or were you self-employed?	1. Employee 2. Self-Employed	w_jbhad=1 & ff_ivlow!=1 & ff_everint!=1
w_retdate/my	Can I just check, in what month and year did you retire?	MY	W_jboff=2 & w_jbstat=4 & Wave=1
w_edtype	Are you...	1. At School 2. At Sixth Form College 3. At FE College 4. At HE College 5. Or at University?	w_jbstat=7

Table 11: School and Further Education Leaving Ages, UKHLS.

Variable	Question	Responses	Universe
w_school	How old were you when you left school?	1. Write in age 2. Never went to school 3. Still at school	ff_ivlolw!=1 & ff_everint!=1
w_scend	WRITE IN AGE LEFT SCHOOL	Age (years)	w_school=1
w_fenow	And how old were you when you left college or university, or when you finished or stopped your course?	1. Write in age 2. Never went to college or university 3. At college/university	w_school=1
w_feend	WRITE IN AGE LEFT COLLEGE/UNIVERSITY	Age (years)	w_fenow=1

Table 12: Education Spells since Last Wave, UKHLS

Variable	Question	Responses	Universe
w_contft	Last time we interviewed you, you were in full-time education. Have you been in continuous full-time education since [ff_IntDate]? Being on holiday from school or between school and University counts as being in full-time education even if you had a job at that time.	1. Yes 2. No	See Figure 14
w_ftend/my4	Can you tell me the month and year you finished that period of full-time education?	MY	See Figure 14
w_ftedany	Have you had any periods of full-time education since [ff_IntDate] ? {if ff_JBSTAT not equal to 7} / Have you had any other periods of full-time education since then? {if ContFT = 2}	1. Yes 2. No	See Figure 14
w_ftedstart/my4$_i$	Can you tell me the month and year you started that/your next period of full-time education?	MY	w_ftedany = 1 OR w_ftedmore$_{i-1}$=1
w_ftedend$_i$	Has that period of full-time education ended?	1. Yes, ended 2. No, not ended	w_ftedany = 1 OR w_ftedmore$_{i-1}$=1
w_ft2end/my4$_i$	Can you tell me the month and year you finished that period of full-time education?	MY	w_ftedend$_i$=1
w_ftedmor$_i$	Have you had any other periods of full-time education since then?	1. Yes 2. No	w_ftedend$_i$=1

Table 13: Annual Employment History Questions, Waves 2-9, UKHLS

Variable	Question	Responses	Universe
w_notempchk	Last time we interviewed you, you were [ff_JBSTAT] . Have you been continuously [ff_JBSTAT] since [ff_IntDate] ?	1. Yes 2. No	See Figure 15
w_empchk	Last time we interviewed you, you were [ff_JBSTAT] . Have you been continuously [ff_JBSTAT] since [ff_IntDate] ? {if ff_empLW = 1 & ff_JBSTAT less than 3} / Last time we interviewed you, you were doing some paid work. Have you been continuously in paid work since [ff_IntDate] ? {if (ff_empLW = 1 & ff_JBSTAT greater than 3) OR (ff_empLW = 2 & ff_JBSTAT less than 3)}	1. Yes 2. No	See Figure 15
w_stendreas	Can you tell me why you stopped doing that job?	Various	w_empchk=2
w_empstend/dmy4	On what date did you stop being [ff_JBSTAT] {if NotEmpChk = 2} / working in the job you were doing on [ff_IntDate] {if EmpChk = 2} ?	MY	w_notempchk=2 OR w_empchk=2
w_jbsamr	And have you worked continuously for the same employer since [ff_IntDate] ?	1. Yes 2. No	w_empchk=1 & ff_jbstat!=1
w_jbendreas	Can you tell me why you stopped doing that job?	VARIOUS	w_jbsamr=2
w_wkplsam	Are you still working at the same workplace as the job you had on [ff_IntDate] ?	1. Yes 2. No	w_jbsamr=1
w_samejob	Have you been working continuously in the same job since [ff_IntDate] ?	1. Yes 2. No	w_jbsamr=1

Variable	Question	Responses	Universe
w_jbend/dmy	On what date did you stop working in the job you were doing on [ff_IntDate] ?	MY	w_jbsamr=2 OR w_samejob=2
w_nxtst	Immediately following that period of [ff_JBSTAT] {if NotEmpChk = 2} / job {if EmpChk = 2} , did you have a period of paid work or did you do something else?	1. Paid Work 2. Something Else	w_notempchk=2 OR w_empchk=2
w_cjob	Was that {if NxtSt = 1} / your {if JbSamR = 2 | SameJob = 2} next job your current job?	1. Yes 2. No	w_nxtst=1 OR w_jbsamr=2 OR w_samejob=2
w_cjbatt	What was the main thing about your current job that attracted you to it?	Various	w_cjob=1
w_nxtjbhrs	Was that job full-time or part-time, where part-time is working less than 30 hours a week?	1. Full-time 2. Part-time	w_cjob=2
w_nxtjbes	In that job were you employed or self-employed?	1. Employed 2. Self-Employed	w_cjob=2
w_nxtjbend/dmy4	On what date did you end that job?	MY	w_cjob=2
w_nxtendreas	Why did you stop doing that job?	Various	w_cjob=2
w_nxtstelse	What were you doing? Were you...	1. Unemployed/looking for work 2. Retired 3. On maternity/paternity leave 4. Look after the family or home 5. In full-time education/student 6. Long-term sick or disabled 7. On a government training scheme 8. Something else	w_nxtst=2
w_cstat	Has this period of being [NxtStElse] ended or is this what you are doing now?	1. Yes, ended 2. Not ended, current status	w_nxtst=2
w_nxtstend/dmy4	On what date did you stop being [NxtStElse] ?	MY	w_nxtst=2
w_nextstat$_i$	Immediately following that period of [NxtStElse]] {if NxtSt = 2} / period of [NextStat(i-1)] {if NxtSt = 2 and 2nd or subsequent loop} / job {if NxtSt = 1} , did you have a period of paid employment or did you do something else?	1. Paid employment 2. Something else	See Figure 15

Variable	Question	Responses	Universe
w_nextjob$_i$	Was this next period of employment...	1. Doing a different job for the same employer 2. Working for a different employer 3. Working for yourself/self-employed 4. Or did you return to the same job with the same employer?	w_nextstat$_i$=1
w_currjob$_i$	Is this your current job?	1. Yes 2. No	w_nextstat$_i$=1
w_jbatt$_i$	What was the main thing about your current job that attracted you to it?	Various	w_currjob$_i$=1 & w_nextjob!=4
w_jobhours$_i$	Was this job full-time or part-time, where part-time is working less than 30 hours a week?	1. Full-time 2. Part-time	w_currjob$_i$=2
w_reasend$_i$	Can you tell me why you stopped doing that job?	Various	w_currjob$_i$=2
w_nextelse$_i$	What were you doing? Were you...	1. Unemployed/looking for work 2. Retired 3. On maternity/paternity leave 4. Look after the family or home 5. In full-time education/student 6. Long-term sick or disabled 7. On a government training scheme 8. Something else	w_nextstat$_i$=2
w_currstat$_i$	Has this period of being [NextElse] ended or is this what you are doing now?	1. Yes, ended 2. Not ended, current status	w_nextstat$_i$=2
w_statend/dmy$_i$	On what date did you stop doing that and start your next job or other activity?	MY	w_currstat$_i$=1 OR w_currjob$_i$=2

Table 14: Lifetime Employment Status History Questions, Waves 1 and 5, UKHLS

Variable	Question	Responses	Universe
w_lgaped	Can I just check, apart from taking a holiday or a gap year, did you leave full-time education between finishing at school and starting at college or university?	1. Yes 2. No	See Figure 16
w_leshst	Which description on this card comes closest to what you first did after leaving full-time education / best describes what you did next, even if it was only for a month?	Various	w_leshst[_n-1]!=0
w_leshe/my4	When did you start this period of being [w_leshst]?	MY	bw_leshne!=1

A.3. Cross Tabulations

Table 15: Cross-Tabulation of Main Activity and Current Employment Questions, BHPS

bw_jbstat	**Employee**	**Self-Employed**	**Not Employed**	**Waiting to take up job**
Self-Employed	454	15,936	249	28
In Paid Employment	117,446	1,358	369	106
Unemployed	392	67	8,453	164
Retired	231	103	47,700	45
On Maternity Leave	808	21	114	2
Looking after Family or Home	383	106	17,693	56
Full-Time Student	5,392	114	8,693	52
Long-Term Sick/Disabled	302	30	10,006	9
Government Training Scheme	281	2	300	7
Something Else	205	47	732	38

Table 16: Cross-Tabulation of Main Activity and Current Education, BHPS

bw jbstat	bw_scnow=1	bw_fenow_bh=1	bw_edend/my4=-3	bw_edenne=1
1. Self-Employed	4	11	4	21
2. In Paid Employment	59	317	61	439
3. Unemployed	12	23	11	42
4. Retired	5	4	3	12
5. On Maternity Leave	1	2	0	4
6. Looking after Family or Home	5	24	4	44
7. Full-Time Student	2,913	2,403	0	5,823
8. Long-term sick/disabled	17	18	5	21
9. Government Training Scheme	1	44	29	34
97. Something Else	12	19	1	30

Table 17: Cross-Tabulation of **bw_nemst** *and* **bw_jbstat**, *Waves 2-15, BHPS*

bw_jbstat	bw_nemst							
	3	**4**	**5**	**6**	**7**	**8**	**9**	**97**
1. Self-Employed	33	24	1	28	2	11	1	104
2. In Paid Employment	63	84	1	50	15	15	2	146
3. Unemployed	5,880	41	4	235	18	243	18	64
4. Retired	38	34,144	3	584	4	373	1	61
5. On Maternity Leave	3	7	55	10	0	2	0	0
6. Looking after Family or Home	108	263	7	12,975	20	202	1	29
7. Full-Time Student	13	0	1	9	10,404	6	12	60
8. Long-term sick/disabled	24	68	1	43	4	7,056	2	13
9. Government Training Scheme	11	2	0	2	11	7	193	3
97. Something Else	20	18	1	47	13	18	2	384

Table 18: Cross-Tabulation of **bw_cjsstly** *and* **bw_jbstat**, *Waves 16-18, BHPS*

bw jbstat	bw cjsstly									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**97**
1. Self-Employed	144	38	1	3	0	7	0	0	0	0
2. In Paid Employment	32	1,105	8	13	0	26	49	6	7	5
3. Unemployed	3	4	40	0	0	5	4	2	2	0
4. Retired	3	9	2	23	0	7	0	4	0	0
5. On Maternity Leave	0	21	0	0	1	2	0	0	0	0
6. Looking after Family or Home	2	6	3	0	0	39	0	0	0	1
7. Full-Time Student	0	14	2	0	0	5	793	3	1	0
8. Long-term sick/disabled	0	8	4	2	0	3	1	17	0	0
9. Government Training Scheme	0	1	1	0	0	1	0	0	4	1
97. Something Else	5	9	10	15	0	20	12	8	0	26

Table 19: Cross-Tabulation of `bw_jhstat` *and* `bw_jbstat`*, Waves 16-18, BHPS*

`bw_jbstat`	`bw_jhstat`									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**97**
1. Self-Employed	321	36	5	1	0	4	1	1	0	38
2. In Paid Employment	46	4,335	15	4	3	9	8	3	3	142
3. Unemployed	6	19	592	2	5	27	12	21	2	27
4. Retired	3	8	3	342	0	12	0	3	0	6
5. On Maternity Leave	1	11	1	0	39	1	0	0	0	0
6. Looking after Family or Home	1	10	6	1	4	199	0	3	0	8
7. Full-Time Student	3	48	1	1	1	3	317	0	1	16
8. Long-term sick/disabled	0	3	0	2	0	1	0	103	0	3
9. Government Training Scheme	0	4	0	0	0	0	0	0	36	1
97. Something Else	0	17	9	5	2	6	4	5	1	45

Table 20: Cross-Tabulation of **bw_leshst** *and* **bw_jbstat**, *Waves 2, 11 and 12, BHPS*

bw jbstat	bw_leshst											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. Self-Employed	1,086	32	8	3	1	2	2	7	2	0	0	2
2. In Paid Employment	29	6,319	1,698	3	5	6	19	57	3	7	1	5
3. Unemployed	1	1	2	746	2	2	58	5	32	2	0	8
4. Retired	2	12	11	17	2,807	0	327	7	105	0	0	11
5. On Maternity Leave	0	8	1	0	0	37	0	0	0	0	0	1
6. Looking after Family or Home	5	1	34	31	70	5	1,488	9	28	0	0	4
7. Full-Time Student	0	1	1	1	0	0	0	991	0	3	0	2
8. Long-term sick/disabled	1	7	3	15	13	0	42	4	668	1	0	3
9. Government Training Scheme	0	1	1	2	1	0	0	4	0	66	0	0
97. Something Else	2	2	3	4	0	0	7	8	1	0	0	25

Table 21: Cross-Tabulation of Main Activity and Current Employment Questions, UKHLS Waves 1-9

w_jbstat	**Employee**	**Self-Employed**	**Not Employed**	**Waiting to take up job**
Self-Employed	2,127	28,171	1,107	134
In Paid Employment	186,959	3,040	1,435	269
Unemployed	516	86	19,677	386
Retired	841	693	91,771	80
On Maternity Leave	1,820	52	363	10
Looking after Family or Home	317	130	22,553	85
Full-Time Student	6,857	190	21,105	251
Long-Term Sick/Disabled	301	23	14,374	16
Government Training Scheme	79	4	272	6
Unpaid, Family Business	21	19	213	4
On Apprenticeship	322	3	48	6
Something Else	236	85	1,725	76

Table 22: Cross-Tabulation of Main Activity and Current Education, UKHLS Waves 1-9

w_jbstat	w_school=**1**	w_fenow=**1**	inlist(1,w_contft, w_ftend$_i$)
1. Self-Employed	10	173	94
2. In Paid Employment	209	1,469	2,192
3. Unemployed	86	235	747
4. Retired	7	285	21
5. On Maternity Leave	1	13	20
6. Looking after Family or Home	6	143	110
7. Full-Time Student	5,408	5,628	12,303
8. Long-term sick/disabled	35	85	48
9. Government Training Scheme	1	28	23
10. Unpaid, Family Business	0	4	6
11. On Apprenticeship	2	27	38
97. Something Else	29	48	112

Table 23: Cross-Tabulation of **`w_jbstat`** *and current status derived from Annual History, Waves 2-9, UKHLS*

`w_jbstat`	Current Status derived from Annual History											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**97**
1. Self-Employed	20,040	2,678	114	179	2	270	29	52	2	10	0	287
2. In Paid Employment	2,378	142,807	317	212	28	641	834	102	23	9	15	190
3. Unemployed	65	342	9,600	196	23	1,704	485	928	25	9	3	414
4. Retired	525	672	446	74,244	1	1,185	0	1,060	1	31	0	204
5. On Maternity Leave	42	1,448	13	11	152	32	7	0	0	0	0	2
6. Looking after Family or Home	112	213	1,514	619	69	14,047	45	430	1	21	0	284
7. Full-Time Student	72	2,644	305	9	3	121	11,196	27	13	0	5	113
8. Long-term sick/disabled	18	187	1,097	593	0	505	19	8,531	2	2	0	53
9. Government Training Scheme	4	19	61	2	0	13	18	6	56	0	0	5
10. Unpaid, Family Business	16	15	20	26	0	39	2	3	0	36	0	39
11. On Apprenticeship	4	217	8	1	0	4	15	1	12	0	10	22
97. Something Else	48	133	225	96	3	286	92	59	5	9	2	458

Table 24: Cross-Tabulation of `w_leshst` *and* `w_jbstat`*, Waves 1 and 5, UKHLS*

w_jbstat	w_leshst											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1. Self-Employed	1,842	443	110	31	15	10	27	10	4	1	1	20
2. In Paid Employment	225	11,574	3,511	199	24	92	140	118	27	20	5	89
3. Unemployed	15	134	43	1,131	14	16	159	18	97	6	0	32
4. Retired	68	684	198	56	6,158	15	481	8	162	3	2	35
5. On Maternity Leave	1	71	21	2	2	128	3	1	0	0	0	1
6. Looking after Family or Home	18	92	54	89	43	53	1,508	12	40	2	1	30
7. Full-Time Student	1	11	14	8	0	0	5	227	0	0	0	6
8. Long-term sick/disabled	12	70	17	53	23	6	84	9	923	3	0	14
9. Government Training Scheme	0	0	0	4	0	0	0	3	2	24	0	0
10. Unpaid, Family Business	5	3	0	3	1	0	4	0	0	0	0	6
11. On Apprenticeship	0	12	1	1	0	0	1	1	0	2	0	3
97. Something Else	7	17	15	13	10	0	40	4	8	1	0	67