Constructing Consistent Work-life Histories:
A guide for users of the British Household Panel Survey

David C Maré

(ISER University of Essex & Motu Research)

ISER Working Paper
Number 2006-39

Institute for Social and Economic Research

The Institute for Social and Economic Research (ISER) specialises in the production and analysis of longitudinal data. ISER incorporates the following centres:

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.

- ESRC UK Longitudinal Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It was responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study

- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national datasets from all over Europe.

The British Household Panel Survey is one of the main instruments for measuring social change in Britain. The BHPS comprises a nationally representative sample of around 5,500 households and over 10,000 individuals who are reinterviewed each year. The questionnaire includes a constant core of items accompanied by a variable component in order to provide for the collection of initial conditions data and to allow for the subsequent inclusion of emerging research and policy concerns.

Among the main projects in ISER's research programme are: the labour market and the division of domestic responsibilities; changes in families and households; modelling households' labour force behaviour; wealth, well-being and socio-economic structure; resource distribution in the household; and modelling techniques and survey methodology.

BHPS data provide the academic community, policymakers and private sector with a unique national resource and allow for comparative research with similar studies in Europe, the United States and Canada.

BHPS data are available from the Data Archive at the University of Essex
http://www.data-archive.ac.uk

Further information about the BHPS and other longitudinal surveys can be obtained by telephoning +44 (0) 1206 873543.

The support of both the Economic and Social Research Council (ESRC) and the University of Essex is gratefully acknowledged. The work reported in this paper is part of the scientific programme of the Institute for Social and Economic Research.

Acknowledgement:
This paper was started while I was a Visiting fellow at the Institute for Social and Economic Research (ISER) at the University of Essex in 2002. I am grateful to ISER for their hospitality and support, and especially to Stephen Jenkins for his support and encouragement during the paper's long gestation. Thanks also to Gillian Paull, Peter Lynn, Elena Bardasi, and Mark Bryan for helpful comments. The contents of this paper do not, however, necessarily reflect the views of ISER.

Author contact: dave.mare@motu.org.nz.

This research was undertaken using data from the British Household Panel Survey; Waves 1-13, 1991-2004 (ESRC Research Centre on Micro-social Change (2005))

Readers wishing to cite this document are asked to use the following form of words:

> **Maré, David C (August 2006) 'Constructing Consistent Work-life Histories: A guide for users of the British Household Panel Survey', ISER Working Paper 2006-39. Colchester: University of Essex.**

For an on-line version of this working paper and others in the series, please visit the Institute's website at: http://www.iser.essex.ac.uk/pubs/workpaps/

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ UK
Telephone: +44 (0) 1206 872957
Fax: +44 (0) 1206 873151
E-mail: iser@essex.ac.uk
Website: http://www.iser.essex.ac.uk

ABSTRACT

This paper assembles existing documentation about the data available from the British Household Panel Survey (BHPS) that can be used to create consistent and complete work-life histories for respondents. It discusses some of the practical and conceptual challenges in deriving consistent work histories from the BHPS data, and documents aspects of the data and its manipulation. It brings together summaries of how other researchers have dealt with the challenges. It also serves to document a newly derived dataset of consistent work life histories.

The material in this paper should be seen as a complement to the BHPS Work-Life History files available though the UK Data Archive. It provides background information on the many issues that were dealt with in the creation of those files. The current paper is intended to assist users of the BHPS to understand the inevitable tradeoffs involved in deriving consistent work-life histories, and to make a well-informed judgement about strengths and weaknesses of the Data Archive files for any specific form of analysis. It also provides guidance for enthusiastic users who wish to derive their own versions of the files.

NON-TECHNICAL SUMMARY

The British Household Panel Survey collects extensive information about respondents' work-life histories. Two retrospective accounts of histories are supplemented annually by further spell information and details of activity as at the date of interview. Unfortunately, respondents' accounts of their histories are not always consistent with information collected earlier. This paper documents the challenges in building an internally-consistent picture of respondents' histories that is also as consistent as possible with their previous reports. The paper thus updates and extends earlier work by Halpin (1997), Halpin (2000), Paull (2002) and Oskrochi and Crouchley (2000).

The paper also documents a newly-derived version of the BHPS work-history files that will be made available through the UK Data Archive, and provides guidance on the use of the files.

Table of Contents

List of Tables

List of Figures

List of Appendix Tables

1. Background

I never meant to embark on this exercise. In July 2002, I arrived as a Visiting Fellow at the ISER University of Essex with the intention of using the BHPS work-history files, which are available through the UK Data Archive. I wanted to use these files as the building blocks for creating lifetime occupational histories. In the course of preliminary analyses, I struck some puzzles in the data that I could not readily resolve. Before I knew it, I was buried in the complexities of the BHPS, which I had not used before. Gaining a sound understanding of the dataset one is using is always a good idea, and the excellent BHPS online documentation made the task relatively painless.

However, I continued to be puzzled by some aspects of the work-life data. (At least some of the apparent anomalies that I found were due to my own lack of understanding of the available datasets, and the lack of comprehensive documentation for the work history files.) I soon discovered that I was not alone - several other researchers had embarked on similar exercises. At least some had returned to the base data to re-derive a work-history file rather than rely on the available processed files. Those I talked to said that they had spent several weeks (I suspect longer) deriving a file that was suitable for their own purposes, but which would need more work if it were to be of use to other researchers. With one major exception, these efforts have produced user-specific versions of histories that have not been made generally available.

After a few weeks of trying to 'tweak' the processed files, I decided that I too should tackle the task of creating my own version of the files, with the hope that my efforts could also enable comparability and easier replication of results, and reduce the costs faced by others who might follow. To get an idea of the range of uses that others had made of the available datasets, and to learn about the efforts that they had made, I circulated some informal questions to BHPS users. The 10 responses are summarised in Appendix A.

The current paper evolved out of notes that I was making for myself from various existing sources. Inevitably, I would forget some bits of information as I absorbed other features of the BHPS, or of the way that others had dealt with particular issues in particular files. I therefore found it useful to assemble relevant pieces of information in a single location. On the assumption that this would be equally useful for others who embark on a similar course, I have spent a little longer making my notes tidier, and hopefully accessible to other researchers.

Parts of this paper, in particular the description of the features of the BHPS data, draw heavily on existing sources, including:

- the excellent BHPS documentation [Taylor, Marcia Freed (ed). with John Brice, Nick Buck and Elaine Prentice-Lane (2002) British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices. Colchester: University of Essex. available online at http://www.iser.essex.ac.uk/bhps/doc]
- The documentation of the BHPS work-life history files, by Halpin ((1997), (2000))
- Published papers documenting others' attempts [Paull (2002) ; Oskrochi and Crouchley (2000)]

I have benefited from seeing programs created by fellow researchers, which in many places are themselves clearly the result of collaboration and code-sharing. I have drawn on programs from:

- Brendan Halpin/ Adrian Birch: SPSS code (which appears to incorporate previous programming by Nick Buck, NJ Davey, J Gershuny, Mark Taylor, and possibly many others)
- Stata code from ISER colleagues Elena Bardasi and Mark Bryan.

The paper in fact goes further than just summarising existing documentation. It also documents my own attempts to build a 'clean' work history file. In the course of doing this, I have brought together in a comparative way some of the documentation of other attempts to derive work history files. Hopefully this will serve to highlight the choices that have been made by other researchers, and the range of possible solutions to the various challenges of creating consistent work life history files.

1.1. Work-Life History Data - What are they?

The BHPS Work-Life History data (hereafter WLH data) provide a window into people's life-time patterns of activity. There are three main forms of lifetime retrospective data available from the BHPS.

- **Employment Status and Job History** [Record Types *BLIFEMST*, *CLIFEJOB*, *wJOBHIST*, as well as relevant information from *wINDRESP*]
- **Marital & Cohabitation History** [Record Types *BMARRIAG*, *BCOHABIT*, as well as relevant information from *wINDRESP*]
- **Fertility & Child History** [Record Types *BCHILDAD*, *BCHILNT*, as well as relevant information from *wINDRESP*]

My focus in this paper is solely on the first of these three. The employment status and job histories summarise each respondent's work experience, as well as spells of non-labour market activity, ideally from the time of first leaving full-time education until the latest interview date.

While in principle the BHPS data contains all that is needed to build these life histories, in practice it is a considerable challenge to derive consistent histories.

1.2. What is a ‘consistent’ work-life history file?

When I refer to 'consistent' work-life histories, I have in mind two quite different meanings. The difference between these meanings is at the heart of the problems of extracting histories from the BHPS data. The first meaning is that the resulting history should be consistent with the responses given by respondents. The second meaning is that the resulting history should be internally consistent, meaning that it is a non-overlapping sequence of spells that accounts for all of the respondent's experience since leaving full-time education.

Unfortunately, the recorded information about the timing and sequence of spells is often not internally consistent. Achieving internal consistency is therefore achieved only at the loss of some consistency with information provided by the respondent. Internal consistency, as described above, is itself a somewhat unrealistic objective, since people's reported work-life histories may well involve overlapping spells (eg: if the respondent has more than one job). Furthermore, histories may also be reported or recorded with error, creating overlaps or gaps where none exist. The standards for internal consistency will to some extent also depend on the use that is to be made of the data.

Reducing the complexities of working life to a series of sequentially coded spells will inevitably do some violence to reality. The objective in creating a consistent work life history file is to obtain a sequential-spell history that approximates people's working lives, and to minimise the biases introduced into whatever analysis is subsequently undertaken.

Whether particular ways of deriving histories introduce biases depends in part on the nature of the analysis to be undertaken. For instance, a single long spell of full-time employment may be split into three because a different reported (or recorded) occupational code in one wave makes the spell look like a different spell. For calculating lifetime work experience, this splitting does not matter. If the objective were to analyse spell length, the splitting of the spell into three would make the derived dataset much less suitable.

My primary objective in creating a consistent work-life history was to obtain job information (and in particular, industry and occupation information) that spanned each respondent's working life. I wanted to obtain measures of accumulated experience in different sorts of jobs, as well as in spells of unemployment or types of 'not-in-labour-force' activities. For this purpose, I wanted to derive a non-overlapping sequence of spells that accounted for all of a respondent's working life, and which included as much information as possible about jobs held.

2. The Base data from the BHPS - Summary by source

This section summarises selected information about each of the data sources that contain information that is relevant for the construction of consistent histories. The material is presented by data source. Appendix B contains copies of relevant questionnaire items and variable names, taken from the BHPS documentation [Taylor et al (2005)].

The BHPS WLH data are contained in four types of file:

- *wINDRESP* - the main individual questionnaire responses from wave *w*
- *wJOBHIST* - employment (job) history information from wave *w*
- *BLIFEMST* - retrospective employment status history, collected in wave 2 (*B*)
- *CLIFEJOB* - retrospective employer spell history, collected in wave 3 (*C*)

We follow the convention of using a lower-case '*w*' to refer to the wave prefix. Thus, *AINDRESP* refers to the *INDRESP* file for wave 1 (A). We also extend the notation to refer to relative wave positions. For instance, in referring to the overlap between one wave's *JOBHIST* and the previous wave's *INDRESP* files, we may refer to *wJOBHIST* and *(w-1)INDRESP*.

2.1. wINDRESP: Main annual survey response

Employment Status

There are three measures of current employment status: that on the Household Composition Form; that arising from the direct status question J2 (wJBSTAT); and that arising from the sequence of questions about whether the respondent did any paid work in the last week, whether away from a job and whether seeking work. The first of these is likely to be reported by someone else, and therefore one of the others is to be preferred if available. In a minority of cases, there may be inconsistencies (e.g. where full-time students have a part-time job, but define themselves as students). Where the interest is in self-defined status, wJBSTAT should be used, but the routing of the questionnaire, and hence the availability of data at various points, depends on the paid work questions. Thus, in order to select out all those in employment, the combination of wJBHAS and wJBOFF should be used.

Source: BHPS User Manual (Taylor et al (2005))

2.1.1 Timing of job characteristics

Occupation, industry, sector, hours of work, etc. are measured as at the interview date (ie: at the right censored end of the spell)

2.1.2 Questionnaire Routing

Current Employment Status

Within the *wINDRESP* questionnaire, there are two separate sources for deriving current employment status.

The first is a question that everyone answers, early in the interview, which asks

"Please look at this card and tell me which best describes your current situation?" [Variable JBSTAT]

Self employed	Family care
In paid employ	FT student
Unemployed	Long term sick/disab
Retired	Govt trng scheme
On matern leave	Something else

This is sometimes referred to as 'main activity'.

The second source of employment status is collected independently of the first, and is obtained from the Employment section of the individual questionnaire. All respondents are asked

> *Can I just check, did you do any paid work last week - that is in the seven days ending last Sunday - either as an employee or self employed? [Variable wJBHAS]*

and if they have not done any paid work in the previous week, they are asked:

> *Even though you weren't working did you have a job that you were away from last week? [Variable wJBOFF]*

Details about the respondent's current (main) job are collected for everyone who answers yes to one of the these two questions. Respondents answering no to both questions are asked, for a second time, about their current situation (except for Wave A - see below).

Note that the two reports of status need not be the same. Some differences will be due to the differing nature of the questions. Some will be the result of respondents giving inconsistent answers. Paull (2002) provides a good discussion of the differences across the first nine waves, with comparisons across demographic groups.

Over all of the *wINDRESP* observations, 3.0% (0.31+1.24+1.44) of non-proxy, non-missing responses were inconsistent.[1] Surprisingly, about 10% (0.31/3.0) of the inconsistent cases related to respondents whose main activity (*JBSTAT*) was reported as either employment or self-employment, but who did not have a job (*JBHAS*=1 or (*JBHAS*=2 & *JBOFF*=1)).

Table 1: Employment Status:
Comparison of Two measures from Pooled *wINDRESP* files

	wNEMST Employed (coded 'inapplicable')	*Not Employed (Same category)*	*Not Employed (diff category)*	*Total*
wJBSTAT				
Employed	87,909	n/a	482	88,391
	(56.36%)		(0.31%)	(56.66%)
Not employed	1,929	63,418	2,253	67,600
	(1.24%)	(40.65%)	(1.44%)	(43.34%)
Total	89,838	63,418	2,735	155,991
	(57.59%)	(40.65%)	(1.75%)	(100%)

Note: Excludes proxy responses, missing and wild responses. Shaded cells are inconsistent.

Full-time / Part time

The *wINDRESP* measure of whether a job is a full-time or part-time job is a derived variable. It is based on total hours, i.e. including both normal and overtime hours. It is computed for both employees and the self employed. It uses *wJBHAS* and *wJBOFF* to identify those with jobs, and combines responses to the hours questions *wJBHRS* and *wJBOT*.

Permanent/ Temporary/ Casual

The question(s) about whether the job is permanent, temporary, casual, seasonal, etc. relates to what the respondent identifies as their "current job", and precedes the clarification about what constitutes a 'main job' (ie: most hours, or if equal hours, then highest paid). In cases of multiple job-holding, it is therefore possible that the response relates to a job other than the main job. The mismatch is presumed to be small.

Spell-start dates for Full-time Students

Information about the date at which the current employment status spell started is coded differently for full-time students with jobs, than for others with jobs. The routing is summarised in the following table.

Respondents with jobs have the start date of their current job recorded as *wJBBG**. The question that they are asked relates to a particular job:

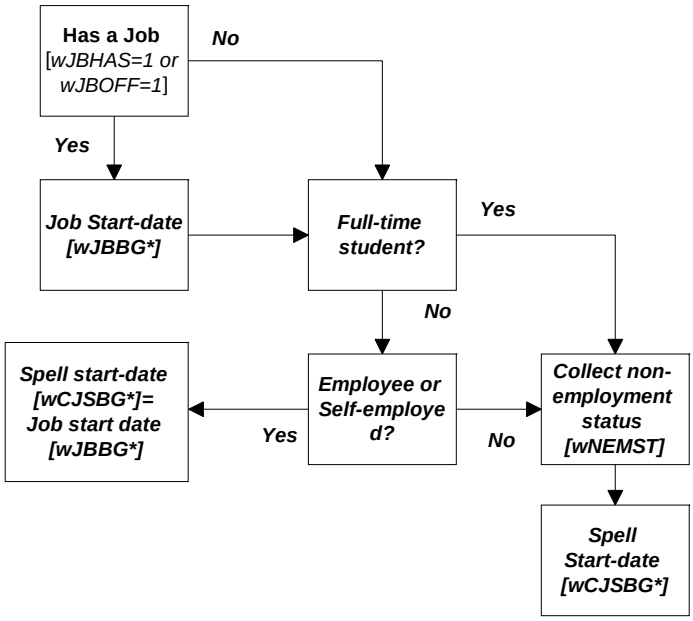
> *What was the date you started working in your present position? If you have been promoted or changed grades, please give me the date of that change. Otherwise please give me the date when you started doing the job you are doing now for your present employer. [Coded as wJBBG*]*

For most of these respondents with jobs, this information is transferred to the variable *wCJSBG**. For those without jobs, the information that is entered into variables *wCJSBG** comes from a question asking about the start date of the non-employment spell identified in *wNEMST*. Full-time students (as identified in the earlier 'main activity' question recorded as

[1] Paull (2002, P8, footnote 35) reports that inconsistency arises in 1.2 percent of cases. This is lower than the 3.0% that I find. The difference is due to different sample selection and a different definition of inconsistency. Paull confines her attention to the first 9 waves, to individuals giving a full interview (*wIVFIO*==1), and to original sample members from Wave G onwards (memorig==1). I include all observations. Both Paull and I exclude missing and proxy responses from our analyses but I count as inconsistent observations where *wNEMST*=-8 and *wJBSTAT*>2 (the cell with 1.24% in the table). I thank Gillian Paul for her help in identifying these differences.

wJBSTAT) who hold a job are asked for the start date of their current non-employment spell (presumably as a full-time student) and it is this date that is entered as *wCJSBG**, not the start date of their job spell.

Figure 1: Routing of Full-time students to capture current spell start [wCJSBG*]



Length of spell

wINDRESP contains a derived measure of spell length. *wCJSTEN* measures the length of time in the current labour market spell, whether, employee, self employed or not employed in number of days. Where day is missing, this is assumed to be one. For years before 199LY[2], where month is missing, this is assumed to be July. This measure is based on reported start date and recorded interview date. It does not contain independent information. If spells are going to be merged from separate files, it is preferable to re-derive these variables based on possibly adjusted dates.

2.1.3 Proxy responses & Telephone interviews

Each year, between 2.8 and 6.0 percent of responses are collected by proxy or (starting in wave 3) telephone interview. The proxy and telephone questionnaires collect less information than does the main questionnaire. The proxy and telephone questionnaires do, however, attempt to collect a good deal of the relevant employment information in *wINDRESP*.

Exceptions include:

- Employment Status: asked only once (as *wJBSTAT*). *wNEMST* is not available from Proxy responses.
- Earnings data: collected in bands rather than in pounds.
- Dates: collected as month and year (no day recorded), and stored under different variable names (eg: *wPRCJSBG** rather than *wCJSBG**)
- Hours of work: A question on full-time v part-time is asked directly, rather than being derived from responses to hours of work questions.

[2] The notation '199LY'is used in the BHPS documentation to indicate 'last year' - the period 12 months prior to the start of fieldwork for the wave (e.g.1 September 1990 - 31 August 1991 for Wave One)

Table 2: Percent of responses accounted for by proxy or phone (by Wave)

WAVE	1	2	3	4	5	6	7	8	9	10	11	12	13
Percent of responses	3.4	3.9	6.0	4.4	4.6	3.2	3.3	3.3	2.8	3.3	4.2	5.5	5.5

2.1.4 Imputation

Certain income and housing variables contain imputed values. Imputation methods vary and are documented in Section V of the BHPS user manual. Broadly, imputation is a combination of hotdeck and regression (predictive mean matching) methods, with some cross-wave imputation.

2.1.5 Wave-specific differences

Date of interview

In all waves except wave one, there are variables recording the year of interview [*wDOIY* and *wDOIY4*, for 2-digit and 4-digit year variables respectively]. There is no such variable in Wave One. However, all interviews were undertaken in 1991, so the gap is easily rectified.

Employment Status

The coding schedule for information on employment status changed between Wave one and Wave Two. Only the ordering, and not the range of possible responses was affected. The variable affected is *wJBSTAT*. A table summarising the coding schedule and changes is included in Appendix Table C1 (with the codings used by Halpin summarised in Table C2).

Employment Status for non-employed (wNEMST)

As noted above, respondents who reported not having a job are asked about their main activity twice. The exception is in Wave A, where the early question about main activity (coded as *wJBSTAT*) is the only source of this information.

Permanent v Temporary jobs

The *wINDRESP* question asking whether a respondent's current job was permanent or temporary was split into two from Wave 9 onwards. A table summarising the coding schedule and changes is included in Appendix Table C3.

2.2. wJOBHIST: Annual employment history update

The Employment History: Record Types wJOBHIST and variables wCJSBLY, wNJBS

. . . the employment history is routed in terms of current status as defined by the job-holding questions. In a limited number of cases, respondent or interviewer confusion led to erroneous ordering or overlapping periods. The majority of these have been cleaned, but a few unresolvable inconsistencies remain. Where the last employment spell given started after 1st September 199LY, or the current spell started after this date, and no information about previous spells was given, a single wJOBHIST record with missing data was created.

Record Type wJOBHIST

This record contains information from the employment history over the period from 1st September 199[y-1] to the date of interview. There is one record for each spell identified at questions J12-J14, with job characteristic information from questions [J16 to J31] appended where relevant.

These records will only exist for respondents whose current labour force spell began after 1.9.199[y-1]. The additional key wJSPNO, identifies the sequence of job spell, with the most recent first.

Derived variables are those from wJHENDD onwards. Where a job is with the same employer as previously mentioned and/or at the same workplace (wJHSTAT=1), then the values for wJHSIC wJHSECT and wJHSIZE are copied from the relevant record.

Source: BHPS User Manual (Taylor et al (2005))

2.2.1 Identifying job spells

wJOBHIST collects information about all types of spells. Information is collected in two stages. First, a complete sequence of spells is determined, and start-dates recorded. Distinct spells are defined as a change in type of activity, as chosen from the following list:

SHOWCARD J2 [recorded as variable wJHSTAT]

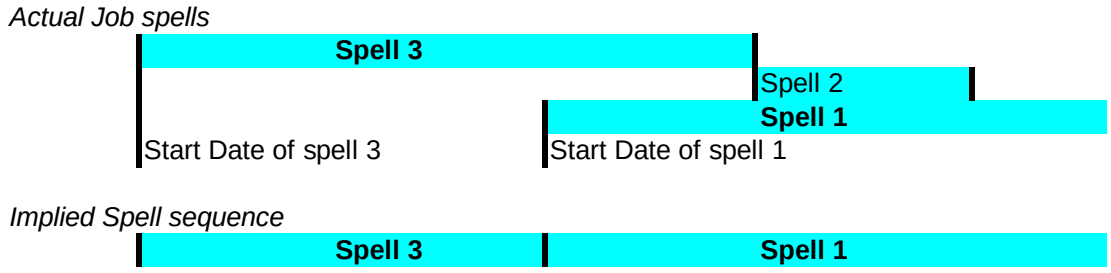
01	Doing a different job for same employer	05	On maternity leave
02	Working for a different employer/ In paid employment (not self employed)/ Working for myself (self employed)	06	Looking after family or home
		07	In full-time education/student
		08	Long term sick or disabled
		09	On a government training scheme
		10	Something else (PLEASE GIVE DETAILS)
03	Unemployed/looking for work		
04	Retired from paid work altogether		

Note that spell start dates only are collected, and not end-dates. It is implicitly assumed that the start date of a spell coincides with the end-date of the prior spell.

For those with jobs, the current spell is a spell in their current main job, which may be immediately preceded with another job spell. Spell type may therefore be the same for two consecutive employment spells where the respondent changes jobs, whereas the type of non-employment spell must change between consecutive *wJOBHIST* spells.

Information is collected on the basis of consecutive main jobs. 'Main job' is defined as the job with the most hours, or if hours are equal in two or more jobs, the job with the highest rate of pay. Spells are consecutive in that respondents are asked for consecutive start dates. The following figure illustrates the way that overlapping spells are captured in *wJOBHIST*. Only the most recent spell (Spell 1) is captured fully. Subspells such as Spell 2 are not captured at all, and previous overlapping spells (Spell 3) are shortened.

Figure 2: Treatment of overlapping spells in wJOBHIST



In the second stage of the *wJOBHIST* questions, further information is obtained for each spell of paid employment [*wJHSTAT* = 1 or 2]. This information includes hours of work, contract type, industry, occupations, etc (although note the comments below under the heading of questionnaire routing).

2.2.2 Order of spells

Information about *wJOBHIST* spells are collected in reverse order. Questions that ask whether a job was with the *same* employer [*wJHSTAT*] or at the *same* workplace [*wJHPLDF*] therefore compare the spell in question to the chronologically *subsequent* spell.

2.2.3 Timing of job characteristics

Occupation, full-time status, etc. relates to the occupation as at the start of the spell. The exception is information about pay, which is collected as at 1 September of the previous year, or at the start of the spell if that is more recent.

2.2.4 Questionnaire Routing

Employment Status

Job and employment status history information is collected for the period between 1 September in the year before the interview year, until the interview. It is thus collected only for respondents whose current spell does not entirely cover this window - i.e: for those whose current spell started after 1 September in the prior year.

Interviews occur between August and May. The 'window' that is covered by the *wJOBHIST* records is thus between 11 months (for respondents interviewed in August) and 20 months (for respondents interviewed in May). Table 3 shows the distribution of interviews across interview months, for the first 13 waves

Full-time Employee / Part-time Employee / Self employed

Hours of work for *wJOBHIST* spells are collected as a full-time v part-time indicator, as part of a question that also identifies whether, for job spells, the respondent is an employee or self-employed [*wJHSEMP*]. A full-time / part-time distinction is made only for employees, and not for the self-employed. As noted below, no full-time / part-time distinction is made in Wave A.

This information is not, however, collected for respondents whose job is with the same employer as the subsequent job [*wJHSTAT*=1]. This is the case for 16.3 percent of wJOBHIST spells from the first ten waves.

Firm Size, Industry, and Sector

For some *wJOBHIST* non-employment spells, coding for firm size, industry, and sector has been carried from the (chronologically) next spell if it is an employment spell. Thus, firm variables exist even non-employment spells. This is presumably an error, and these codes need to be changed to missing.

Table 3: Month of Interview - by wave

Interview	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*	*Wave*

month	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*
	(A)	*(B)*	*(C)*	*(D)*	*(E)*	*(F)*	*(G)*	*(H)*	*(I)*	*(J)*	*(K)*	*(L)*	*(M)*
Aug	0	0	0	0	0	14	7	5	0	0	0	0	0
Sep	2006	3774	3504	2759	1766	4796	5114	5734	5626	5445	5482	5573	6933
Oct	5378	4405	3833	4157	4187	3333	3682	3424	3742	3789	4630	4363	5169
Nov	2568	1263	1367	1683	2315	834	1711	1071	1789	929	3536	2564	2335
Dec	291	153	239	406	432	226	387	329	514	147	1655	1484	755
Jan	0	51	223	245	256	135	153	174	1019	155	1964	1439	615
Feb	0	69	227	99	127	52	85	86	1526	568	744	749	219
Mar	0	127	133	75	112	33	41	50	630	1799	541	243	140
Apr	0	3	74	52	46	15	12	33	387	1464	311	173	68
May	0	0	0	5	8	0	1	0	390	1309	4	9	4
Missing	21	0	0	0	0	0	0	0	0	0	0	0	0
Total	10264	9845	9600	9481	9249	9438	11193	10906	15623	15605	18867	16597	16238

Permanent/ temporary/ casual

No information is collected on whether job spells captured in *wJOBHIST* are permanent, temporary or causal. Some relevant information is contained in reports of why respondents *left* jobs[*wJHSTPY*]. One of the options is "temporary job ended". (The full range of responses is shown in Table C4.) It seems plausible to infer that a job that was left because it was temporary was in fact a temporary job. It is not as clear that respondents in temporary jobs would necessarily report only this response. This indirect measure of permanent/ temporary status is likely to understate the extent of temporary job holding.

Table 4: *wJHSTPY* - Reasons for leaving a job

Possible responses	
Promoted	Left to have baby
Left for better job	Children/home care
Made redundant	Care of other person
Dismissed or sacked	Other reason
Temporary job ended	Missing or wild
Took retirement	Inapplicable
Stopped health reas	

2.2.5 Wave-specific differences

Employment Status

The change in ordering of employment status codes that was noted above for the variable *wJBSTAT* in *wINDRESP* also applies to the Employment Status as collected for the Employment History [recorded as *wJHSTAT*]. The coding schedule is summarised in Table C1.

Full-time Employee / Part-time Employee / Self employed

The *wJOBHIST* question in Wave 1 provides no information about whether jobs were full-time or part-time. From Wave 2 onwards, this distinction is made for employees, but not for the self-employed. The affected variable is *wJHSEMP*. The coding schedule is summarised in Table C1..

2.2.6 Proxy Respondents

No employment history information is collected from telephone or proxy interviews.

2.3. BLIFEMST: Retrospective Employment Status history (collected in Wave B)

Record Type BLIFEMST
This record contains information about employment status spells in the period since the respondent first left full time education. There is one record for each spell reported in answer to questions L50, L51 and L52. The record contains end date for each spell except the final spell which should be recorded as not ended. The start date for each spell and the spell length in months are included as derived variables.
While data collected here may have been compared with the single year job history information to resolve internal ambiguities, there has been no attempt to enforce consistency between data collected here and that contained in the records AJOBHIST, BJOBHIST etc. Season codes were used when the respondent could not remember exact month. In the calculation of spell length it was assumed, by convention, that winter would be coded as January, spring as April, summer as July and Autumn as October. However where these season codes are used in a spell which starts and finishes in the same calendar year, the length is set to -3, indicating less than 12 months but exact length indeterminate.

Source: BHPS User Manual (Taylor et al (2005))

2.3.1 Identifying job spells

Spell information is collected chronologically, starting with the first spell following full-time education, and proceeding up to (and including) the spell that was current at the time of the Wave 2 interview. Distinct spells are defined as a change in type of activity, as chosen from the following list:

SHOWCARD 18 [recorded as variable BLESHST]

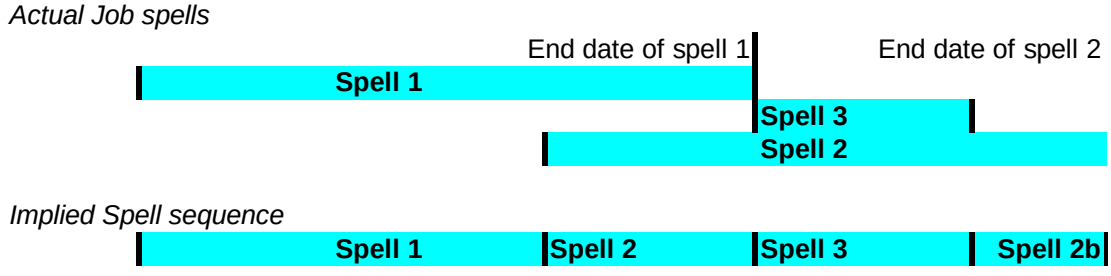
01	Self-employed	07	Looking after family or home
02	Full-time paid employment	08	Full-time Student/At School
03	Part-time paid employment	09	Long term sick or disabled
04	Unemployed	10	On a government training scheme
05	Retired from paid working altogether	11	National Service/War Service
06	Maternity leave	12	Something else (please give details)

Note that the date information is collected for each change of status. The recorded end date of one spell is thus assigned to the start date of the following spell.

Information is recorded on consecutive spells of main activity. The spells are delineated by the date at which the respondent's 'situation changed'.

Figure 3 shows an example of how overlapping spells are (probably) recorded in *BLIFEMST*. Spell 1 could be a period of family care, which overlaps with a period of part-time work (Spell 2), and is followed by a brief full-time spell (Spell 3). The recorded sequence depends on what the respondent regards as a change in their situation. The implied sequence shown in the figure is one possible interpretation, under which only spell 3 is captured fully. Spell 2 in fact appears as two short spells.

Figure 3: Treatment of overlapping spells in *BLIFEMST*



2.3.2 Questionnaire Routing

All respondents are asked when they first left full-time education. On the basis of the response to this question, all respondents who are still in full-time education should be excluded from the life history questions. In practice, there are a number of respondents whose *BLIFEMST* histories commence with one or more spells of full-time education.

2.3.3 Proxy Respondents

No employment status history information is collected from telephone or proxy interviews.

2.4. CLIFEJOB: Retrospective Job (Employer) history (collected in wave C)

Record Type CLIFEJOB
This record contains information about jobs held in employment spells in the period since the respondent first left full time education up to the beginning of data collection in the main panel - i.e. 1st September 1990. There is one record for each spell reported in answer to questions L5 to L13. The definition of a job spell in this record type is different from that used in wINDRESP and wJOBHIST: here a job spell corresponds to a continuous spell with a single employer, while in the main panel data spells may also be defined by changes in occupation or promotions while working with the same employer. An additional spell from the responses to questions L14 to L17 is generated, if the respondent started work with their current employer before 1.9.90. Note that, for this spell the date and occupational information may be different from that collected in the employment section about the current job, since the information there relates to start of the current job, and respondents may have done more than one job in a single spell with the same employer.

The final record of this type for each individual may be one of three different kinds: a) a generated present employer record as indicated above; b) a completed record where this was the last job the respondent has had to date, and this began before 1.9.90 and has finished; c) a record containing only status and start date, where the job began after 1.9.90. In this last case further information about this job and any subsequent jobs held by the respondent will be contained in the main panel record types (e.g. AINDRESP or AJOBHIST). The type of last record is indicated by the variable CLJENST. This information is also contained in the variable CLJRST on record type CINDRESP.

The CLIFEJOB records have been checked to ensure that the spell sequence (indicated by CLJSEQ) is in ascending order of start dates. However there is some multiple job holding reported, so job spells may overlap, and end dates may not be consistently in order.

Source: BHPS User Manual (Taylor et al (2005))

2.4.1 Identifying job spells

Spell information is collected chronologically, starting with the first employer following full-time education.

The job spells identified in *CLIFEJOB* are spells with a single *employer*. If the respondent has several jobs with the same employer, they are recorded as a single spell. The questionnaire asks for information on the sequence of spells with different employers (which may overlap).

Spell start date and spell end date are both collected for each employer. Although the information on job title was asked only for the *main* job (ie: the one with the most hours, or else the highest paid), the routing of the questionnaire allows respondents to report overlapping employer spells.

Having reported details of the main job with an employer, and the date that they left that employer, respondents are asked "have you had any more paid jobs since then?". It is ambiguous whether this means since leaving that employer, or since starting with that employer. If they answered yes, they were subsequently asked about the start date of the spell with the next employer. It is therefore valid to report overlapping spells.

CLIFEJOB contains no explicit information on non-employment spells (except as reflected by gaps between periods covered by jobs). Some information is available from the stated reason that the employer spell ended (*CLJYLFT*).

The collection of spell information stops when the next spell start date occurs after the date covered by *AJOBHIST* (Sept 1 1990) or refers to a spell with the same employer as for the current (Wave 3) job. Spell information is collected for the spell with the current employer if the spell started prior to Sept 1 1990.

2.4.2 Timing of job characteristics

Job characteristics are measured as at the beginning of the employer spell.

2.4.3 Questionnaire Routing

The life-time job history questions are asked only of respondents who were interviewed in Wave 2 and who had had at least one job of a month or more prior to the job that they might have held at the time of the Wave 3 interview.

Respondents were then shown the calendar derived from the *BLIFEMST* and other retrospective questions in Wave 2, and asked whether the calendar was correct. In only 49% of cases (3461/7014 valid responses) was the earlier calendar confirmed as correct.

2.4.4 Proxy Respondents

No employer history information is collected from telephone or proxy interviews.

3. Resolving issues in the base data

The previous section summarised some of the features of the different datasets containing information that is needed to construct work and life histories. This section discusses some of the data issues that need to be dealt with. A particular focus is on missing and inconsistent date information, which will cause difficulties for deriving consistent work histories. This section also comments on options for imputing missing data. Discussion of issues that arise when combining the different datasets is deferred until the next section.

Where possible, I summarise the ways that different researchers have dealt with the issues, inevitably oversimplifying some issues. One of the difficulties in documenting the derivation of WLH files is the number and complexity of adjustments that are made to the base data. I focus on the approaches of Halpin (1997), Halpin (2000), Oskrochi and Crouchley (2000), and Paull (2002), since these are published sources. I also summarise my own approach, and make reference in places to alternative approaches of which I am aware. Differences in approach are in part due to the differing objectives of the researchers.

Halpin's objective was to create consistent WLH files, combining information from all of the datasets. He has derived several different datasets, based on combining and reconciling different subsets of the available information. The datasets also differ in the detail with which spells are defined (job spells; employer spells; employment status spells). The different datasets are described in Appendix D below, and in Halpin (2000).[3] Halpin has also prepared versions of the information with two different files structures - a calendar-structure, with each (individual, month, status) observation capturing the status of an individual for a particular month; and an episode structure, with each (individual, start, end, status) observation summarising a spell. Halpin (2000) contains a fuller summary of his approach and methods. (See also Section 5.3.3 and Appendix E.)

Oskrochi and Crouchley (2000) document the derivation of a consistent WLH dataset, based on spells of different employment status. They create a monthly panel (calendar format) summarising labour market status (defined as in *wJBSTAT*) at each point. They combine information from all available data sources. They use their derived dataset for spell duration analysis as well as for deriving cumulative experience in each labour market state.

Paull (2002) derives WLH files in order to examine the importance of reporting biases in the measurement of labour market dynamics. She undertakes two separate exercises. First, she creates a dataset that captures information about periods that are reported on twice, in consecutive interviews. Second, she creates a lifetime WLH file of employment status episodes, with periods of employment divided into employer spells. Paull reports on four different methods of deriving histories. It is her 'reconciliation' method that is described in the sections that follow.

My own objective was to create a complete consistent WLH file, summarising spells in different 'jobs' (occupations), as well as in different employment status spells. For this, I needed to combine information from all the datasets. I wished to derive measures of accumulated experience, and was not intending to use the derived dataset to analyse spell lengths.

My accounts of the way that researchers have dealt with various issues are not necessarily comprehensive. The treatments that I report will in many cases not deal entirely with all of the issues that arise. The researchers themselves note in their published work that the range of adjustments is too complex to report fully. The presentations here should therefore be seen as indicating the range of possible approaches to different issues. For a fuller understanding

[3] I have used the version of Halpin's files that was available in 2002. A more updated version of the files is now available from the UK Data Archive as study 3954.

of a particular researcher's approach, their published works in most cases provide a more detailed guide.

3.1. Consistent Date sequences

Consistent histories require sequences of spell start and end-dates that generate consecutive, non-overlapping spells accounting for the respondent's entire history since leaving full-time education. There are two types of difficulties that arise in obtaining such sequences from the recorded data. First, date information may be partially or totally missing. Second, although date information may be present, it may generate inconsistent histories, due to negative durations, overlapping spells, or gaps between spells. This section summarises and discusses the extent of these problems in the various BHPS datasets.

My general approach is to keep track of the feasible range for each start and end date. Thus, for each start (and end) date, I retain the reported value as it appears in the raw data, and also the minimum and maximum of the feasible date range. If possible, the reported date is used. Where this is not possible, I try to set the start and end dates within feasible ranges: A spell cannot start after its maximum feasible end-date, or before the minimum feasible end-date of the previous spell. Similarly, a spell cannot end before its minimum feasible start date, or after the maximum feasible start date of the subsequent spell. The following sections provide more information on how this general approach is applied to reconciling different sorts of inconsistencies that occur in the data.

It was the presence of negative durations, overlaps and gaps in the Halpin files that had initially caused me concern. (These are detailed below in Appendix Section E2.) I had started using Halpin's *newpan* file, in order initially to avoid the complications of the long term retrospective (*BLIFEMST* and *CLIFEJOB*) files. I had planned to use only the five most recent years of experience, and restrict my sample to the 5^{th} to 10^{th} waves. The features of the Halpin files that first caused me concern were that: a) a number of spell durations were negative; and b) the spell histories did not always account for the entire period between the respondent leaving full-time education and the interview date (or else accounted for more months than was possible).

Table 5 summarises the date issues in each of the BHPS base datasets, and in the Halpin files. The criteria for identifying negative durations, gaps and overlaps differ for the main files and the Halpin files. This is due to the different conventions used in recording spell timing. In the main files, the end of a spell is concurrent with the start of the next spell. In the Halpin files, the end date of a spell and the start date of the next spell are consecutive rather than concurrent. Gaps therefore arise if the difference between one spell ending and the next one starting is greater than 1. Similarly, concurrent end and subsequent start dates are treated as overlaps in the Halpin files.

Table 5: Issues with Dates and Timing - Summary

	Missing Dates				*Inconsistent dates*	*Dubious sequences*	
	Totally Missing		*Partly Missing*		*Negative duration [end<beg]*	*Gap occurs before [beg>end[-1]+1)*[b]	*Overlaps previous [beg<end[-1]+1]*[b]
Type of discrepancy	*Endpoints*	*Other*	*Year only known*	*Year+Seas known*	*xx instances (xx indivs;xx obs*	*xx instances (xx indivs;xx obs)*	*xx instances (xx indivs;xx obs)*
Data Source							
Obs / Individuals							
BLIFEMST	S: 37	S: 77	S: 2,228	S: 4,245	0	0	0
35,474 / 9,026	E: 9,026	E: 73	E: 1,764	E: 3,847			
CLIFEJOB	S: 12	S: 230	S: 2,406	S: 6,249	0	1,582	113
32,773 / 7,074	E: 4,498	E: 209	E: 2,266	E: 5,767		(1,014; 6,719)	(103; 616)
AJOBHIST	S: 93	S: 3	S: 209	S: 0	0	0	0
3,382 / 2,362	E: 17	E: 3	E: 4	E: 0			
AINDRESP	S: 590		S: 1,072	S: 0	2	0	0
10,264	E: 0		E: 21	E: 0			
BJOBHIST	S: 34	S: 2	S: 135	S: 32	0	0	0
2,918 / 2,060	E: 2	E: 3	E: 0	E: 4			
BINDRESP	S: 406		S: 726	S: 287	2	0	0
9,845	E: 0		E: 0	E: 0			
CJOBHIST	S: 28	S: 5	S: 136	S: 11	0	0	0
2,918 / 2,102	E: 4	E: 5	E: 7	E: 1			
CINDRESP	S: 623		S: 623	S: 84	3	0	0
9,600	E: 0		E: 0	E: 0			
DJOBHIST	S: 75	S: 0	S: 250	S: 1	0	0	0
3,011 / 2,107	E: 7	E: 0	E: 7	E: 1			
DINDRESP	S: 480		S: 1,209	S: 7	5	0	0
9,481	E: 0		E: 0	E: 0			
EJOBHIST	S: 101	S: 2	S: 252	S: 2	0	0	0
3,112 / 2,162	E: 3	E: 2	E: 13	E: 1			
EINDRESP	S: 497		S: 1,245	S: 1	1	0	0
9,249	E: 0		E: 0	E: 0			
FJOBHIST	S: 72	S: 2	S: 245	S: 0	0	0	0
3,019 / 2,095	E: 2	E:2	E: 7	E: 0			
FINDRESP	S: 380		S: 1222	S: 0	0	0	0
9,438	E: 0		E: 0	E: 0			
GJOBHIST	S: 126	S: 1	S: 312	S: 0	0	0	0
3,941 / 2,692	E: 17	E: 2	E: 12	E: 0			
GINDRESP	S: 483		S: 1,516	S: 0	4	0	0
11,193	E: 0		E: 0	E: 0			
HJOBHIST	S: 83	S: 2	S: 301	S: 0	0	0	0
3,505 / 2,476	E: 9	E: 3	E: 8	E: 0			
HINDRESP	S: 514		S: 1,434	S: 0	2	0	0
10,906	E: 0		E: 0	E: 0			
IJOBHIST	S: 103	S: 94	S: 509	S: 0	0	0	0
5,179 / 3,682	E: 21	E: 97	E: 48	E: 0			
IINDRESP	S: 622		S: 2,352	S: 0	0	0	0
15,625	E: 0		E: 0	E: 0			
JJOBHIST	S: 10	S: 72	S: 587	S: 0	0	0	0
5,451 / 3,790	E: 3	E: 74	E: 19	E: 0			
JINDRESP	S: 650		S: 2,506	S: 0	3	0	0
15,605	E: 0		E: 0	E: 0			
KJOBHIST	S: 121	S: 68	S: 542	S: 0	0	0	0
5,973 / 4,387	E: 29	E: 69	E: 33	E: 0			
KINDRESP	S: 937		S: 2,358	S: 0		0	0
18,867	E: 0		E: 0	E: 0			
LJOBHIST	S: 61	S: 46	S: 459	S: 0	0	0	0
5,043 / 3,721	E: 45	E: 47	E: 18	E: 0			
LINDRESP	S: 1,088		S: 2,268	S: 0		0	0
16,597	E: 0		E: 0	E: 0			
MJOBHIST	S: 14	S: 57	S: 524	S: 0	0	0	0
4,803 / 3,588	E: 18	E: 57	E: 28	E: 0			
MINDRESP	S: 1,065		S: 2,397	S: 0		0	0
16,238	E: 0		E: 0	E: 0			

Table 5 (continued)

	Missing Dates				Inconsistent dates	Dubious sequences	
	Totally Missing		Partly Missing		Negative duration [end<beg]	Gap occurs before [beg>end[-1]+1)[b]	Overlaps previous [beg<end[-1]+1][b]
Type of discrepancy	*Endpoints*	*Other*	*Year only known*	*Year+Seas known*	*xx instances (xx indivs;xx obs*	*xx instances (xx indivs;xx obs)*	*xx instances (xx indivs;xx obs)*
Data Source							
Obs / Individuals							
DERIVED DATASETS [a]							
NEWPAN	S: 15	S: 49	S: 0	S: 0	2,010[b]	2,825	653[e]
122,742 / 21,817	E: 5	E: 49	E: 0	E: 0	(1,613; 19,950)	(1,659; 12,333)	(584; 6,201)
XLEMPE	S: 20	S: 38	S: 0	S: 0	986[c]	931	132[f]
77,094 / 21,815	E: 0	E: 38	E: 0	E: 0	(871; 7,088)	(825; 4,575)	(127; 1,134)
XLJOBE	S: 20	S: 41	S: 0	S: 0	1,406[d]	1,071	136[g]
85,641 / 21,815	E: 0	E: 41	E: 0	E: 0	(1,249; 11,575)	(944; 5,426)	(130; 1,309)
LJEMPE	S: 0	S: 0	S: 0	S: 0	0	0	0
84,011 / 21,815	E: 0	E: 0	E: 0	E: 0			

a) For derived datasets, gaps occur if start dates and prior end-dates are separated by a difference of more than 1; overlaps occur if the difference is less than one
b) 1,965/2,010 negative durations are of duration -1.
c) 986/986 negative durations are of duration -1
d) 1,406/1,406 negative durations are of duration -1
e) 653/653 overlaps are of length one. ie: start date and prior end-date coincide
f) 112/132 overlaps are of length one. ie: start date and prior end-date coincide
g) 114/136 overlaps are of length one. ie: start date and prior end-date coincide
h) End-date endpoints for wINDRESP files, shown in column one, are interview dates

3.1.1 Partially missing dates

In all of the original BHPS datasets, dates are recorded in three parts - day, month, and year. Dates are partially missing if any one or two of these is missing. In some cases where month information is missing, the season (spring, summer, autumn, winter) was recorded. For the long-term retrospective files *BLIFEMST* and *CLIFEJOB*, respondents could indicate the *season* when a spell started or finished, when they were unable to recall the actual month. Although season responses were not an option on the questionnaires for *wINDRESP* or *wJOBHIST*, seasons were also reported and coded for spell dates in waves 2 to 5. The number of cases in *wINDRESP* and *wJOBHIST* where months are recorded as seasons is relatively small, peaking at 287 in *BINDRESP*.

The following table summarises the types of missing date data that arise.

Table 6: Categories of missing date information

DD/MM/YY	Day, month and year known
??/MM/YY	Complete information
??/SS/YY	Partially missing - Season and Year known
??/**/YY	Partially missing - Year alone is known
??/MM/**	Partially missing - Month alone is known
??/**/**	Totally Missing

*KEY: DD=known date; MM=known month; YY=known year; SS=known season; **=Missing value; ??=Any value*

For the purposes of creating histories, researchers have generally coded spell start and end-dates as month and year. ie: day information is not used. For this reason, ??/MM/YY is treated as 'Complete information'. Bryan uses day information to determine which month a spell date should be allocated to. If the date falls on or before the 15th of the month, it is allocated to the prior month.

Season and Year known

As shown in Table 5, some dates are reported as seasons in all files from Wave 2 (B) to Wave 5 (E). The greatest number of cases occur in the long term retrospective files *BLIFEMST* and *CLIFEJOB*. This is not surprising, since it is only in these two datasets that the questionnaire

contained an option to report seasons rather than dates. In the other files, respondents or interviewers must have chosen to 'write in' seasons when months were not remembered. Just over 10 percent of the dates reported in *BLIFEMST* are reported as seasons, and just over 1/3 of dates in *CLIFEJOB* are reported as seasons.

In order to allocate dates when only season and year are known, most researchers initially allocate season codes to a particular month, in the middle of the season. (code 13 = winter =January; code 14 = spring = April; code 15 = summer = July; code 16 = autumn = October). The exception is Crouchley and Oskrochi (2000), who appear to deal with these cases as they do cases where only year is known - see below.

A potential problem arises in the case of winters. A spell reported as having starting in the winter of a year could have started at the beginning of the year or at the end of the year. The allocation of winter to January could thus be incorrect. Halpin, Maré and Paull therefore subsequently adjust some allocated winter seasonal dates, but using different rules.

Halpin reallocates winter dates only for records in *BLIFEMST*. The reallocation is based on the relationship between the allocated date and the dates that precede or follow it. The algorithm is shown in Table 7.

Table 7: Re-assigning the winter season

Researcher	*Algorithm*
Halpin w2cal.sps	For BLIFEMST, adjust winter using the following rule: $Y_{t-1}<Y_t<Y_{t+1}$ Winter=January $Y_{t-1}=Y_t<Y_{t+1}$: Winter=December $Y_{t-1}<Y_t=Y_{t+1}$: Winter=January $Y_{t-1}<Y_t$; spell incomplete: Winter=January $Y_{t-1}=Y_t$; spell incomplete: $M_{t-1}<3$ or $M_{t-1}>9$: Winter=$S_{t-1}+1$ $3<=M_{t-1}<=9$: Winter=December $Y_{t-1}=Y_t=Y_{t+1}$: $M_{t-1}<3$: Winter=$S_{t-1}+1$ $M_{t-1}>=3$: Winter=November M_t = Month; Y_t = Year; S_t = start date = M_t: Y_t

Maré	• Winter initially allocated to January
	• Where the assignment of winter to January causes a date conflict, assign winter to December (as long as that doesn't itself cause a conflict)
	  • Where $S_{max,t}<E_{min,t-1}$ and $S_{min,t}+11<=E_{max,t}$ then S_t = December
	  • Where $E_{min,t}<S_{max,t}$ and $E_{max,t}+11<=S_{max,t+1}$ then E_t = December.
	  • S_t = start date $\in[S_{min,t}, S_{max,t}]$; E_t = end date $\in[E_{min,t}, E_{max,t}]$
	• Retain (slightly overlapping) minimum and maximum possible values 2 months either side of the assigned month.

Oskrochi & Crouchley	• Season-only dates treated in the same way as missing dates (see below)
Paull (p 40)	• Winter initially allocated to January • Inconsistent months originally imputed from a reported season were adjusted (within season) to match the reported month from another interview

Maré does not reallocate dates in a deterministic way, and applies the same rules to all data sources (or when combining data sources). Each date is recorded as a triple, being the 'best guess', and the minimum and maximum of the feasible range. One requirement of taking this

approach is that these minima and maxima must be recalculated whenever a date change is made.

For seasonal dates, the 'best guess' is the seasonal mid-point. The minimum and maximum dates are 2 months either side of the assigned month. eg: 13 = Winter is assigned provisionally to January but could be anywhere between November and March. Winters are reallocated from around January to around December only if their allocation to the midpoint causes a conflict that cannot be resolved by an intra-seasonal reallocation. Even then, the reallocation to December is only done if it does not create a further date conflict. The feasible range once the month has been allocated to December is October to February.

Year alone is known

Where only the year of a spell start or end date is known, the default treatment for most researchers is to allocate the date to mid-year (either June or July). Some researchers use more complicated algorithms, as shown in Table 8. The essence of all the approaches is to allocate the month to the mid-point of the feasible range.

Month alone is known

Where the year is unknown, the date is treated the same as a totally missing date. It would be feasible to combine the known month with date information from preceding and subsequent spells to determine or impute the year, although it appears that no researchers have done so.

3.1.2 Totally missing

In some cases where date information is totally missing, it can be imputed by relying on information from prior or subsequent spells. The first two columns of Table 5 show the extent of totally missing dates in the various datasets.

End-points

The first column shows the number of missing dates that occur at end-points (first start date or last end-date). These are shown separately because there is limited scope to impute these dates based on neighbouring spells. The final end-date is missing for all individuals in *BLIFEMST*, since the *BLIFEMST* file contains a record for the spell that was current at the date of interview in Wave 2 (B). The incomplete final spells all have missing end dates. For the same reason, all of the end dates in the *wINDRESP* file should be missing. The data in the table therefore show instead the occurrence of missing interview dates. Missing interview dates arise only in Wave 1 (A), and because interview year is known (all interviews took place in 1991), the 21 missing interview dates show up in the third column (year known).

Some of the missing initial start dates in *BLIFEMST* (37), and *CLIFEJOB* (12) can be filled in by using information on the date at which the respondent completed full-time education. There are actually This information was collected as variables *AFEEND* and *ASCEND* in Wave A for most respondents, and updated in waves B and C for those who left full-time education between the Wave A interview and the Wave C interview. The *wFEEND* and *wSCEND* questions are not, however, identical to the *BLEDEND(M/Y)* question that is the routing question for the *CLIFEJOB* life history questions. *BLEDEND(M/Y)* asks for the month and year when the respondent *first* left full-time education, whereas the *wFEEND* and *wSCEND* questions are preceded by questions about the sort of school or further educational institution that the respondent has attended, and ask about the date of leaving that institution. If the respondent had a job between school and further education, BLEDEND will be earlier than *wFEEND*.

Comparing *BLEDEND* with the maximum of the leaving ages (*AFEEND*, *ASCEND*, *BFEEND*, *BFEEND*) shows that the two are equal for 52 percent of the 9026 respondents, and within a year for a further 23 percent. Of the remaining 25 percent, 7 percent are missing, 13 percent have *BLEDEND* showing an earlier date, and in 5 percent of cases *BLEDEND* is later than leaving age. There are 35 of the 37 cases where *BLEDEND* is missing for which non-missing values of leaving age can be used. (In only one of these cases, the reported end-date is earlier than the reported leaving age.

Table 8: Allocating partially missing dates - only year known

Researcher	*Algorithm*
Halpin * *w2cal.sps*	For *BLIFEMST*, allocate Missing months as follows: $Y_{t-1}<Y_t<Y_{t+1}$ Month=July $Y_{t-1}<Y_t=Y_{t+1}$: Month=$M_{t+1}/2$ $Y_{t-1}=Y_t<Y_{t+1}$: Month=$(M_{t-1}+12)/2$ $Y_{t-1}=Y_t=Y_{t+1}$: Month=$(M_{t-1}+M_{t+1})/2$ $Y_{t-1}<Y_t$; spell incomplete: Month=July $Y_{t-1}=Y_t$; spell incomplete: Month=$(M_{t-1}+12)/2$ $Y_t<Y_{t+1}$; M_{t+1} missing: Month=July (M_{t+1}=July) $Y_{t-1}<Y_t=Y_{t+1}$; M_{t+1} missing: Month=April (M_{t+1}=August) $Y_{t-1}=Y_t=Y_{t+1}$; M_{t+1} missing: Month=$M_{t-1}+(12-M_{t-1})/3$ $M_{t+1}=M_{t-1}+2*(12-M_{t-1})/3$ M_t = Month; Y_t = Year; S_t = start date = M_t: Y_t
* *save3_5a.sps*	For *CLIFEJOB*: Just allocate seasons.
* *getdates.sps*	For *wJOBHIST* files: $Y_{t<}Y_{t+1}$ or last spell Max=12 $Y_t=Y_{t+1;}$ Not last spell Max=M_{t+1} $Y_{t-1}<Y_t$ or first spell Min=1 $Y_{t-1}=Y_t$; Not first spell Min=M_{t-1} Min>Max Max=Min M_t=(Min+Max)/2
Maré	• Month initially allocated to June • Retain minimum and maximum feasible values of January to December. • Where the assignment to June causes a date conflict, adjust month as if it were missing (see Table 9)
Oskrochi & Crouchley *(p. 14)*	The month was generated as a random number (uniformly distributed) between 3 and 9, on the assumption that respondents had remembered correctly dates of events at the end and at the start of the year better than the dates of events at the middle of the year.
Paull	• Missing months prior to 1990 were initially allocated to July (P8, fn 34). • Other 'year only' dates were treated the same as missing dates (see below)

Table 9: Dealing with totally missing dates

Researcher	*Spell Start and/or End dates missing**	*End and subsequent start dates missing [E_{t-1}; S_t]*
Halpin	• Missing start dates assigned to equal previous end date (*plus one*) • Missing end dates assigned to equal subsequent start date (*minus one*)	• Assign date to half-way between prior start and subsequent end. • Drop if string of missing is repeated eg: ESES
Maré	• Missing start dates assigned	• Assign date to half-way

	to equal previous end date • Missing end dates assigned to equal subsequent start date • Assigned dates inherit the min and max feasible values as well as the assigned value from the source date.	between prior start and subsequent end. • Where there are repeated consecutive missing [E_{t-1}; S_t] pairs, divide the period between the prior start and subsequent end equally between the intervening spells
Oskrochi & Crouchley	• Missing start dates assigned to equal previous end date • Missing end dates assigned to equal subsequent start date	• unknown
Paull	• Missing start dates assigned to equal previous end date if • spells are of different type; and • spell end dates are within 12 months of each other • Spells with missing start dates merged with previous spell if • spells are of same type; and • spell end dates are within 12 months of each other • Employment spells with missing end-dates merged with next spell if: • next spell started with or before initial spell • Non-Employment spells with missing end-dates merged with next spell if they are of the same type.	• Missing dividing dates between two spells were replaced with the midpoint of the feasible period for the start date. • $S_t = 0.5*(S_{t-1} + E_t)$)

* *Isolated missing start or end dates arise as a problem only in CLIFEJOB and wINDRESP, since a) in wJOBHIST only start dates are collected, and end dates are assigned to be equal; b) in BLIFEMST, a change date is collected, so that end dates and subsequent start dates are assigned to be equal.*

3.1.3 Negative Durations

Table 5 shows that negative durations arise in only a few (22) cases in the base data. Recall, however, that negative durations are shown only for dates that are neither partially nor totally missing. The allocation of seasons or unknown months to dates reveals a greater number of negative durations. Some can be reconciled within the bounds of the known season or known year. Others cannot be reconciled in this way, and the dates must be reallocated in a way that conflicts with the survey responses.

A negative duration can arise directly in the data only in *CLIFEJOB,* where both start and end dates are reported for a spell. Negative durations appear elsewhere only due to inconsistencies in a series of start or end dates across more than one spell.

The approaches taken by different researchers differ. Halpin reverses the start and end-dates, possibly generating overlaps with adjoining spells. Maré creates zero duration spells that will not contribute to accumulated experience measures, and which minimise the number of overlaps generated. Oskrochi and Crouchley expand the spell to fill the available space. This

will not create overlaps, but may not resolve the negative duration (in which case the observation is dropped).

Table 10: Dealing with Negative Durations

Researcher	*Negative Duration*
Halpin	Reverse the start and end dates
Maré	• Reconcile within date ranges if possible • a zero-duration spell is created with start and end-date equal to the midpoint of the overlap (from Min(start) to Max(End) • If not possible: • create a zero duration spell that starts and ends at the midpoint of Max(End) and Min(Start). (If Max(End) is higher than Min(End_{t-1}) use Min(End_{t-1}) instead of Max(End). If Max($Start_{t+1}$) is lower than Min(Start), use Max($Start_{t+1}$) instead of Min(Start))
Oskrochi & Crouchley	• If the start (end) date of the event is inconsistent (eg: the event starts after it finishes), we use the end (start date of the previous (next) event (if it is available) • If duration is still negative, drop the spell.
Paull	• Because negative durations (other than in *CLIFEJOB*) are due to inconsistencies across spells, they were adjusted as for overlapping spells

3.1.4 Overlaps

The final column of Table 5 shows that overlapping spells arise only in the *CLIFEJOB* file. As noted above, overlaps do not arise in *BLIFEMST* or *wJOBHIST* files because start and end dates are not independently collected. *BLIFEMST* collects only consecutive dates when a respondent's situation changed, and *wJOBHIST* collects only consecutive spell start dates (in reverse chronological order).

The rules to deal with overlaps become more important when different data sources are combined. eg: Paull's rules are designed primarily for combining *wJOBHIST* spells with previous *(w-1)INDRESP* spells.

Table 11: Dealing with overlapping spells

Researcher	
Halpin	• Give priority to the end date of a spell over the start date of a subsequent spell, if necessary overwriting the beginning of the subsequent spell

Maré	• Reconcile within date ranges where possible
	  • Set the break mid-way through the overlap = half way between Start and End_{t-1} (With the added condition that the allocated date cannot be earlier than Min(End_{t-1}) or later than Max(Start))
	• Where reconciliation is not possible:
	  • Complete overlap: If one spell lies entirely within a prior or subsequent spell, drop it.
	  • Partial overlap:
	  • Set the break mid-way through the shortest gap between date ranges = half way between Min(End_{t-1}) and Max(Start) (With the added condition that the allocated date cannot be earlier than $Start_{t-1}$ or later than End)

Oskrochi & Crouchley	If the spells had the same start date: • If same employment status, drop the shorter spell • If different status, replace the start date of the longer spell with the end date of the shorter spell.
Paull	• Where possible, reconcile within season • Otherwise, change the start date to equal the end-date of the previous spell. (trust the more recent date) • If the spells had the same start date, or the first spell began after the second, the second spell is dropped

3.1.5 Gaps between spells

Table 5 shows that only *CLIFEJOB* contains gaps between spells. Recall that *CLIFEJOB* does not contain any records for non-employment periods. Gaps arising from this exclusion are not included in the count in Table 5. It refers only to gaps between *CLIFEJOB* employer spells within the same employment status spell (as defined in *BLIFEMST*).

As in the case of overlaps, more gaps become apparent when partially or totally missing dates are allocated. Again, this issue arises only in *CLIFEJOB*, which is the only source to collect both start and end dates.

An alternative way of dealing with gaps between spells would be to create a 'filler spell' with missing status and job information

Halpin creates a spell file from *CLIFEJOB* alone (file *ljobe*). He deals with gaps by coding status by the reason that the last job ended.

Table 12: Dealing with Gaps between spells

Researcher	
Halpin	• Change start-dates to equal the previous end-date. ["Give priority to the end date of a spell over the start date of a subsequent spell, if necessary overwriting the beginning of the subsequent spell"]
Maré	• Reconcile within date ranges if possible • Set the break mid-way through the gap = half way between Start and End_{t-1} (With the added condition that the allocated date cannot be earlier than Min(Start) or later than Max(End_{t-1})) • If reconciliation is not possible, reassign ES dates to be mid-way between Min(Start) and Max(End_{t-1})
Oskrochi & Crouchley	• unknown
Paull	• Change end-dates to equal the following start date (trust the more recent date) where gap is between spells from same or consecutive waves.

3.1.6 Zero or short duration spells

There are a number of spells in the base data that start and end in the same month. While these are quite legitimate, they contribute nothing to measures of accumulated duration, and are generally excluded from the parametric analysis of spell durations.

The minimum spell length varies in the different datasets:

- *CLIFEJOB* - 'lasting one month or more'
- *BLIFEMST* - 'even if it was only for a month or two'
- *wINDRESP* - current main job, any duration
- *wJOBHIST* - 'even if they were just a few days'

Some of the adjustments that have been described above generate additional zero-duration spells. While these might be useful to record the occurrence of a particular type of spell, they can be excluded for the purposes of most analyses.

While it is possible to use 'day of the month' information to generate more precise spell lengths, only one researcher (Bryan) uses this information, and then only to 'round' start and end dates to the nearest month-end.

3.2. Consistent Spell and Job characteristics

3.2.1 Employment Status

The categories for coding of non-employment spells are consistent across the various data sources (apart from the different numbering of *wJBSTAT* and *wJHSTAT* response codes in Wave A). The coding of employment spells varies. We wish to be able to distinguish the types of spells shown in Table 13.

This level of detail is available for all datasets except *wJOBHIST*, for which the employee spells can be categorised as full-time or part-time only where a job is the final one with an employer. For all *wJOBHIST* spells ending with a move to another employer, hours of work information is missing.

Table 13: Full Coding of Employment Status

Employment Spells	*Non-employment Spells*	
• Self-	• Unemployed	• Long term sick or

Employed • Full-time employee • Part-time employee	• Retired from paid working altogether • Maternity leave • Looking after family or home • Full-time Student/At School	disabled • On a government training scheme • National Service/War Service • Something else (please give details)

In *wINDRESP* and *wJOBHIST*, the categorisation of employment spells into full-time and part time requires combining information from more than one variable.

As discussed above, there is also a choice of employment status variable for *wINDRESP*. Most researchers use *wJBSTAT* to identify the current employment status (supplemented with *wJBFT* to separate full-time and part-time). One problem with this approach is that the responses to *wJBFT* are collected for people who have a current job which is an overlapping but not identical group to those who report employment as their main status in *wJBSTAT*. The variables described in Table 14 under the heading 'approach two' ensure a closer link and greater consistency with the job information collected in *wINDRESP*.

Table 14: Employment Status - Sources of Information

Dataset	*Main source variable(s)*
BLIFEMST	BLESHST
CLIFEJOB	CLJSEMP
wJOBHIST	• wJHSTAT to identify spell type • wJHSEMP to identify FT/PT/SE for employment spells that ended with a change of employer • wJHSEMP or wJBSEMP from the (chronologically) next spell, for employment spells followed by another job with the same employer • For Wave A, assume that all spells are full-time
wINDRESP	Approach One • wJBSTAT to identify type of spell; • (possibly use wJBHAS and wJBOFF to identify employment spells, which will over-ride the wJBSTAT coding) • wJBSEMP and wJBFT to identify FT/PT/SE for employment spells Approach Two • wNEMST to identify the type of non-employment spells (except for Wave A, when wJBSTAT is used); • wJBHAS and wJBOFF to identify employment spells; • wJBSEMP and wJBFT for employment spells, to identify FT/ PT/ SE Refinements: • Paull: wJBSTAT (and rely on wNEMST to identify non-employment spells)) • Bardasi: uses wJBHRS to determine FT/PT if wJBFT is missing

3.2.2 Employer spells

A consistent history of spells with different employers can in principle be built up by combining information in *CLIFEJOB*, *wJOBHIST* and *wINDRESP*. The employment spells from *CLIFEJOB* are collected as employer spells. The *wJOBHIST* spell records contain an

indicator for whether the spell is with the same employer as the (chronologically) following spell

There are some issues when combining a *wJOBHIST* spell with an (overlapping) spell from *(w-1)INDRESP*. Ways of deciding whether the overlapping spells are the same spell (and thus with the same employer!) are discussed in more detail below (section 4.3.2). The employment spells in *BLIFEMST* are spells of the same status, and a single *BLIFEMST* employment spell may contain spells with more than one employer. It is also possible that consecutive *BLIFEMST* employment spells may be with the same employer, if one spell is full-time and the other is part-time.

3.2.3 Job spells

It is not possible to recover job spell information from *BLIFEMST* or *CLIFEJOB*, since these collect status and employer spells respectively, both of which are more aggregated than job information. *wJOBHIST* and *wINDRESP* spells are collected as job spells, and can thus be used to derive a work-life history over the sub-period covered by the panel (1 Sept 1990 to interview date).

3.2.4 Job type - Occupational, Industry, permanent v temporary, firm size, unionisation

Industry and occupation

Industries are coded according to the Standard Industrial Classification 1980 (SIC). Occupations are coded according to the Standard Occupation Classification (SOC) scheme. Wherever occupational information is collected, it is also possible to get various occupation-related scales (SEG, RGSC, Goldthorpe, Hope-Goldthorpe, Cambridge Scale and ILO-ISCO 88.)

For *CLIFEJOB* and *wJOBHIST*, the occupation and industry are measured as at the beginning of the spell. For *CLIFEJOB* spells, the industry coding should apply throughout the employer spell (unless the employer changed industry), although the occupation may change if the respondent changes jobs while with the employer. For *wJOBHIST* spells, the industry and occupation codes should apply for the duration of the spell, since a new spell should be recorded if the job changed. *wINDRESP* information on occupation and industry is recorded as at the date of interview, although it should apply to the entire spell

Table 15: Occupation and Industry Coding - Sources of information

	Main source	*Timing of information*
BLIFEMST	not available	
CLIFEJOB	CLJSOC; CLJSIC	at the beginning of the (employer) spell. (Occupation may change within spell)
wJOBHIST	wJHSOC; wJHSIC	Applies to the entire spell - collected as at the beginning of the (job) spell
wINDRESP	wJBSOC; wJBSIC	Applies to the entire spell - collected as at the (right censored) end of the spell

Permanent/ Temporary Jobs

A question on whether a job is permanent, temporary, or casual is explicitly asked only in *wINDRESP* and *CLIFEJOB* questionnaires. As noted above, the coding of the variable in *wINDRESP* changed in wave 9. The coding for the *CLIFEJOB* variable is the same as *wJBTERM* in waves 1-8.x

Table 16: Permanent and Temporary employment - Information Sources

	Main source	*Alternatives*
BLIFEMST	not available	
CLIFEJOB	CLJTERM	Some information may be obtained from cLJYLFT (as outlined Table C3)
WJOBHIST	not available	Some information may be obtained from wJHSTPY (as outlined in Table C3)
WINDRESP	wJBTERM (waves 1-8) wJBTERM1 & wJBTERM2 (from wave 9)	

Both *wJOBHIST* and *CLIFEJOB* also contain information about why respondents left their jobs, although the relevant variables are not consistently coded, as shown in Appendix Table C4. It is also not clear from the questionnaire whether the question refers to the first job with the employer, or the entire (possibly multi-job) spell with the employer. The word 'job' is used to refer to the first job that the respondent started with the employer. The preamble to the question also defines job as meaning employer spell.

Firm size

Firm size information is available for *wINDRESP* and *wJOBHIST* spells, in the variables *wJBSIZE* and *wJHSIZE* respectively. Responses are coded in size bands [1-2; 3-9; 10-24; 25-49; 50-99; 100-199; 200-299; 300-499; 500-999; 1000 or more] with possible 'don't know' responses of 'fewer than 25' and '25 plus'.

Firm sector

Firm sector information is available for *wINDRESP* and *wJOBHIST* spells, in the variables *wJBSECT* and *wJHSECT* respectively. Response codes are [Private firm/ Company; Civil Service/ Central government; Local government/ Town hall; NHS or higher education; Nationalised Industry; Non-profit organisation; Armed forces; Other]. The *wJOBHIST* questionnaire has slightly abbreviated category descriptions.

Other Job characteristics

wINDRESP contains more detailed information on a range of job characteristics, including unionisation, training, employer pension schemes; job satisfaction, time of day when work is done; etc.

3.2.5 Earnings

Information on labour earnings is available only for *wJOBHIST* and *wINDRESP* spells. The information available from *wINDRESP* is considerably more detailed. Table 17 summarises the available information. The *wJOBHIST* information relates to the start of the job spell, or Sept 1 of previous year, whichever is later. The most consistent measures that are available across the two datasets are (usual) gross or net monthly pay.

Table 17: Sources of Earnings information (derived variables)

	Main source
BLIFEMST	not available
CLIFEJOB	not available
wJOBHIST	Monthly gross pay (wJHGPAY)

	Monthly net pay	(wJHNPAY)
wINDRESP	Gross rate of pay per month:last payment	(wPAYG)
	Net rate of pay per month: last payment	(wPAYN)
	Usual gross pay per month: current job	(wPAYGU)
	Usual net pay per month: current job	(wPAYNU)
	Usual monthly gross pay: Sept this year	(wPAYGTY)
	Usual monthly gross pay: Sept year ago	(wPAYGLY)
	Usual monthly net pay: Sept this year	(wPAYNTY)
	Usual monthly net pay: Sept year ago	(wPAYNLY)
	Monthly self-employed profit	(wJSPROF)
	Monthly self-employed gross pay	(wJSPAYG)

4. Combining the Files

This section focuses on the issues that arise when combining work history segments from the different files.

The previous section has described issues in deriving consistent variable coding across the different data sources, and consistent spell sequences within the various files. This section provides more detail on ways of combining work history segments from the different files while creating or maintaining consistent spell sequences. This is the most difficult challenge in deriving consistent WLH files - reconciling what are essentially conflicting recorded histories.

An additional objective in deriving consistent histories is to enable the histories to be linked to the annual *wINDRESP* records, which contain the full set of BHPS individual questionnaire responses, including more detailed job information for those who are employed. This linking is valid only if the activity reported in the *wINDRESP* file is the same as the wave *w* interview date activity as recorded in the WLH file.

4.1. Combining different concepts

4.1.1 Types of spell

Some of the difficulties in deriving consistent work-life histories arise from the differences in the sort of spells about which information is collected. In particular, the long-term retrospective work history contained in *CLIFEJOB* contains information on spells with each employer, and not on each job. The job-type information (industry, occupation, etc.) relates to the start of the employer spell. It will therefore be a less accurate summary of the spell characteristics than is the available summary of *wJOBHIST* spells, which are spells in the same (main) job, during which, by definition, characteristics (other than pay) should not change. The following list summarises the different sort of spells collected.

Table 18: Types of spells

Source	*Type of spell*	*Minimum duration of spell*
BLIFEMST	Spell with same employment status	'even if it was only for a month or two'
CLIFEJOB	Spell with same employer	'lasting one month or more'
wJOBHIST	Spell in same main job	'even if they were just a few days'
wINDRESP	Spell in current main job	any duration

Employment Status spells are the lowest common denominator of these spells - job and employer spells, combined with non-employment spells, can always be aggregated to employment status spells. However, use of status spells alone does not allow any analysis of job characteristics. Similarly, job spells can be used to create employer spells.

4.1.2 Multiple Job-holding

Another difficulty arises from trying to obtain a simple sequential summary of spells that may in fact be non-sequential (ie: spells may in fact be overlapping). Both *wJOBHIST* and *wINDRESP* ask for information about 'main job, although the distinction between job and *main* job is only an issue where a respondent holds multiple jobs.

The only source of information on multiple job-holding in the BHPS data is from the annual interviews. All other sources of information ask for main job or activity only. In *INDRESP*, respondents are asked about their current main job and also about any other jobs that they may have. The actual question is:

> *Do you currently earn any money from (a second job) odd jobs or from work that you might do from time to time (apart from your main job)?*

Because having a second job under this definition does not necessarily mean that two jobs are being undertaken concurrently, the count of people with second jobs is an upper bound on the extent of concurrent multiple job holding. The following table summarises the wave 13 patterns of multiple job holding. Of those who hold a main job, around 7 percent report also having another job

Table 19: Multiple Job Holding - Summary (Wave 13 (M))

Number *(row percent)*	*Second Job*		
	0	*1*	*Total*
Main Job			
FT Employee	6,145	380	6,525
	(94)	*(6)*	*(100)*
PT Employee	1,769	214	1,983
	(89)	*(11)*	*(100)*
FT Self Employed	804	60	864
	(93)	*(7)*	*(100)*
PT Self Employed	190	36	226
	(84)	*(16)*	*(100)*
Total	8,908	690	9,598
	(93)	*(7)*	*(100)*

The definition of a 'main job' (most hours, or else highest pay rate) imposes a hierarchy of spell types - a part-time spell can be interrupted by a concurrently-held full-time job, or by a higher-paying part-time job. Among the differences in spells is that some types will be, on average, longer than others. Spells from *wJOBHIST* will be the shortest, since they may be interrupted by a *(w-1)INDRESP* full-time or high-paid spell.

By relying on job information from *wINDRESP*, an unemployment spell will also be interrupted by any short spell of employment, even if it does not change an individual's self-defined status as unemployed.

Other sorts of concurrent spell are more difficult to deal with. For instance, should a period of maternity leave be treated as part of a job or employer spell, or as a break in that spell? This can have significant effects on average spell duration. A one-year maternity leave spell in an 11-year spell with an employer could reduce the average spell length from 11 to 5.

4.2. Who has complete history data?

Not all respondents will have participated in all interviews. We therefore need to identify respondents for whom we have complete information. Essentially, we need to be confident that there are no gaps in reported work-life histories.

A minimal set of respondents for whom we have full information is the set of original sample member (OSM) respondents observed in every wave. This selection rule would exclude some usable records, which could be added:

- OSM respondents who completed full-time education during the life of the panel, and who have had full interviews every year since then. It does not matter if they did not have a full interview prior to leaving full-time education.
- Extended sample members [ECHP from Wave 7 (G); Scotland & Wales from Wave 9 (I)] who completed full-time education since joining the panel, and who have had full interviews every year since then.
- OSM respondents for whom we do have complete histories, even though they have missed an interview. For instance, if someone is absent for 3 years, but reappears with a current spell of more than 3 years duration, they could, in principle, be included in the

sample of respondents with complete histories. Including them would, however, lead to the inclusion of a non-randomly selected subsample.

An alternative approach would be to include all respondents, and code periods of unknown employment status separately. It would, however, be difficult to interpret results from analyses with this sort of censored data. I do not include them in the files that I have derived.

The following table shows, for each wave, the number of original sample members who have a full interview in the wave, and the number who have had full interviews up to and including the wave. (Membership of the original sample is defined by the variable *MEMORIG* in the file *XWAVEID*, and confusingly, is not synonymous with OSM, which is specified in the variable *wSAMPST* in *wINDRESP*.)

The first point to note is that the number of original sample members increases during the period. This is due to the addition of children of Wave 1 OSM respondents (themselves categorised by *wSAMPST* as OSM respondents), and of adults with whom the original sample member has lived. (categorised as temporary sample members TSM). Parents of the children of original sample members are also classified as TSMs, and are followed even if they do not live with the OSM respondent.

By wave 13, there had been 16,082 people who had given a full interview at some point during the life of the panel. This number is the number of people who contribute at least one spell to the WLH files.

The second panel of Table 20 shows, separately for each wave, the number of people classified as original sample members (*MEMORIG*=1) who give a full interview. We find, for instance, that 8,264 original sample members provided full interviews in wave 13. Not all of these respondents were part of the fully responding wave 1 OSM group. That number is shown in the third panel. Of the 9,912 OSM full-interview respondents in wave 1, 5,481 gave a full interview in wave 13. However, as shown in the fourth panel, only 4,653 of them had given a full interview in each of the 13 waves.

The table also shows the number of respondents contributing records to *BLIFEMST* and *CLIFEJOB*. A more detailed summary of the wave by wave response patterns of these respondents is contained in Halpin (1997, Table 3).

Table 20: Original sample and OSM sample with full interviews

	Wave 1	*Wave 2*	*Wave 3*	*Wave 4*	*Wave 5*	*Wave 6*	*Wave 7*	*Wave 8*	*Wave 9*	*Wave 10*	*Wave 11*	*Wave 12*	*Wave 13*
Cumulative count of 'original sample' respondents													
number	9912	10803	11401	11982	12487	13095	13636	14093	14516	14932	15354	15706	16082
Cross section (includes all 'original sample' respondents)													
number	9912	9459	9032	9060	8827	9137	9118	8940	8820	8701	8590	8383	8264
w/ blifemst	8340	9026	7875	7518	7104	7020	6793	6532	6294	6039	5790	5574	5361
w. clifejob	6701	7074	7074	6571	6203	6105	5905	5688	5494	5278	5057	4868	4675
w/ both	6682	7050	7050	6551	6182	6085	5888	5670	5479	5264	5044	4854	4662
Cross section (includes only OSM respondents with full interview in wave 1)													
number	9912	8568	7847	7577	7183	7132	6900	6647	6396	6143	5914	5694	5481
w/ blifemst	8340	8340	7423	7138	6744	6704	6487	6250	6029	5795	5564	5367	5162
w. clifejob	6701	6701	6701	6283	5952	5865	5675	5477	5293	5089	4878	4699	4517
w/ both	6682	6682	6682	6267	5935	5849	5661	5462	5281	5078	4868	4688	4507
Continuously responding OSM members													
number	9912	8568	7630	7138	6685	6418	6171	5924	5674	5405	5123	4883	4653
w/ blifemst	8340	8340	7423	6947	6490	6259	6028	5789	5547	5283	5006	4772	4548
w. clifejob	6701	6701	6701	6283	5871	5665	5454	5237	5013	4774	4520	4304	4096
w/ both	6682	6682	6682	6267	5855	5651	5442	5225	5003	4765	4512	4296	4088
Continuously responding OSM members (plus children becoming full respondents)													
number	9912	8721	7914	7539	7171	7064	6911	6738	6558	6332	6073	5831	5643

Notes: *Continuous counts refer to respondents who have had full interviews in every wave up to and including the wave specified in the column. Cumulative counts refer to respondents who have had a full interview in any wave up to or including the wave specified in the column. Extended sample members are excluded. Original sample members are identified by memorig=1. The term OSM is used in the table to denote original sample members who were present in the first wave.*

4.3. Steps in deriving consistent histories

I divide the task of deriving consistent histories into three separate stages:

- Standardise the base data (as discussed in Section 3.2)
- Regularise the date sequences (as discussed in Section 3.1)
- Combine the different sources (as discussed in this section)

This section is organised according to the different merges that need to be carried out

- Merging *wINDRESP* and *wJOBHIST* from the same wave
- Merging *wINDRESP* and *wJOBHIST* from different waves
- Merging *BLIFEMST* and *CLIFEJOB*
- Merging *BLIFEMST*/*CLIFEJOB* with w*INDRESP*/*wJOBHIST*

4.3.1 Merging wINDRESP *and* wJOBHIST *from the same wave*

Overlaps & timing of information

Because of the way that the information is collected, the *wINDRESP* and *wJOBHIST* spells from the same interview should be consistent and sequential.

- if the current spell started before 1 September of the previous year, there is no *wJOBHIST* record.
- if there is a *wJOBHIST* record, it is provided in response to the question "which of the descriptions comes closest to what you were doing immediately before then?", so should not overlap with the *wINDRESP* spell.
- consecutive *wJOBHIST* records are provided in response to the same question, so should similarly not overlap.

In practice, there are 27 cases where the most recent *wJOBHIST* spell has a missing end-date, suggesting that the spell may be still current at the time of the interview. In 8 of these cases, the start date of the *wINDRESP* spell falls before the start date of the most recent *wJOBHIST* spell.

In the case of multiple job-holding, information on some spells may be lost or censored, as in the following two hypothetical examples in which a current spell overlaps with prior spells.

a) current part-time job started 18 months ago and the respondent held another job for a few months during this 18 month period.)

   Information on the second job would not be collected, since the wJOBHIST questions are only asked of respondents whose current spell started after 1 Sept of the previous year.

b) current part-time job lasted from May to September; previous full-time job lasted from January to August (ie: overlap from May to August)

   By collecting only start-dates in wJOBHIST, the recorded sequence would be:

   Jan to May: FT job
   June to Sept: PT job (duration is understated)

Resolution

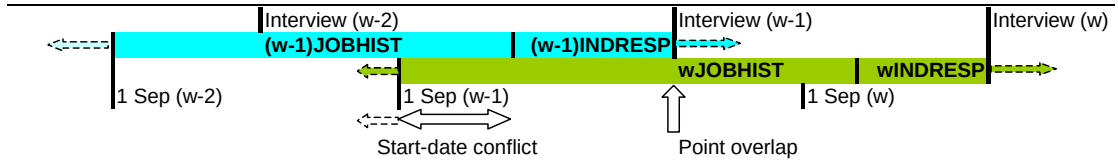
Where there is an overlap, priority is given to the *wINDRESP* reported spell. Answering the *wJOBHIST* question literally would have the same effect as applying this priority treatment.

4.3.2 Merging wINDRESP *and* wJOBHIST *from different waves*

Overlaps & timing of information

Figure 4 illustrates (some of) the overlaps the may arise when combining [*wINDRESP* or *wJOBHIST*] with [*(w-1)INDRESP* or *(w-1)JOBHIST*]. The overlaps shown are for the case where *JOBHIST* records exist for both periods *w* and *w-1*. Where they do not, the relevant *INDRESP* and *JOBHIST* spells would appear as a single *INDRESP* spell.

Figure 4: Overlaps between wINDRESP/wJOBHIST and (w-1)INDRESP/(w-1)JOBHIST



With the exception of respondents interviewed in the previous *(w-1)* August or September, the time span covered by the combined *wJOBHIST*/*wINDRESP* sequence overlaps with the date of the previous interview. Ideally, the spell that was current at the last interview would appear as the oldest spell in *wJOBHIST* (or as the current spell from *wINDRESP*, if the current spell started before the prior year's *(w-1)* September). (shown as 'point overlap' in Figure 4)

In practice, the reports are not always consistent. Spells must agree not only on timing but also on type of activity. A problem arises when the respondent retrospectively reports (as recorded in a *wINDRESP* or *wJOBHIST* record) doing something at the time of the previous interview that is not the same as what was reported when the previous interview took place (in *(w-1)INDRESP*). This may be due to a misreported start date for the *wINDRESP*/*wJOBHIST* spell (where the spell is reported as starting prior to the *(w-1)* interview date when it in fact started after that date). It may also occur because more than one job or activity was current at the time of the *(w-1)* interview, and different activities are given priority by the respondent at different times. As noted in the previous section, the questionnaire structure may cause some spells to be suppressed or censored when there is multiple job-holding.

Where the oldest *wINDRESP*/*wJOBHIST* spell is reported to have started prior to the previous interview, it is possible that it will overlap with earlier *(w-1)JOBHIST* spells (shown as 'start-date conflict' in Figure 4[4]). Ideally, the spell reported as current at the previous interview *(w-1)INDRESP* should be the same spell, and should thus have the same job characteristics. Even if this is the case, an incorrectly recalled start date for the *wINDRESP* spell could cause it to overlap with a *(w-1)JOBHIST* spell.

Figure 4 shows the case where the oldest *wINDRESP*/*wJOBHIST* spell starts before the *(w-1)INDRESP* spell. It is also possible that this ordering of start-dates is reversed. For the case shown, this would mean that the *(w-1)INDRESP* start date fell before the start of the *wJOBHIST* spell. Because, for the case shown, the *(w-1)INDRESP* start date is after 1 Sep *(w-1)*, a second *wJOBHIST* spell would have been recorded, and would occur entirely prior to the *(w-1)INDRESP* spell.

Resolving dates

There are two main approaches that researchers have taken to deriving consistent chronological sequences:

a) Rely on the report that is closest in time to the event (in order to minimise recall error). This has the effect of creating transitions at the date of the previous interview. These are known as 'seam effects'. Halpin (1997, pp. 28 ff.) has a good discussion of the seam effect problems that arise in the BHPS, when taking this approach.

b) Rely on the report that is collected at the same time as a sequence of events (on the basis that the responses will be more accurate because of the context of the response). Oskrochi and Crouchley (2000) rely on retrospective (*wJOBHIST*) reports, and use the interview-date (*(w-1)INDRESP*) reports as 'support' (it is not clear to me exactly what

[4] It is the information from the overlap between 1 Sep *w-1* and the *(w-1)* interview date that is exploited by Paull (2002) to investigate biases in the reporting of labour market dynamics.

this means in practice). It has the advantage of improving the (internal) consistency of reports, but gives greater weight to retrospective than to interview-date reports.

Approach a) implies that *wINDRESP* reports always over-ride *(w+1)JOBHIST* reports that overlap. In many cases, the overlap is valid because the two reports refer to the same spell. The spells can then be merged, as long as they agree on some selected dimensions, such as are shown in Halpin's seam codes (see below).

Approach a) also implies that *wINDRESP* spells that overlap with previous *(w-1)JOBHIST* spells should be truncated because the *(w-1)JOBHIST* report was closer to the actual start date than is the later *wINDRESP* report. In the data, it is clear that start dates of long spells are recalled with error. For jobs lasting several years, the start date of the current job (in *wINDRESP*) may be reported slightly differently from year to year, often differing by exactly one year or one month, or one decade.

Approach b) does not create as many seam effects, but relies more on more remote retrospective responses, and does not permit easy integration with the detailed employment information that is available from *wINDRESP* and which relates to the job that was current at the time of interview for *wINDRESP*.

One particular problem is in determining whether an overlap does in fact refer to the same spell. Halpin considers agreement at three different levels. For each record that follows a *wINDRESP* record (from either *(w+1)JOBHIST* or *(w+1)INDRESP*)), he assigns a 'seam code' to indicate the likelihood that the spell is the same spell, and could therefore be merged. The estimated likelihood is based on whether the two spells have the same characteristics along the following dimensions:

- Detailed Employment status [FT Employee; PT Employee; Self Employed; specified type of non-employment spell]
- Broad employment status [Employed; Not employed]
- Industry [at 2-digit level]
- Occupation [at 2-digit level]

The industry and occupation matching is particularly challenging, since these variables have relatively low reliability. Bushnell (1998) (drawing on Martin et al (1995)) reports reliability of 74% for the CASOC system of occupational coding, which is used for the BHPS. The same occupational description was given to two independent coders. The same CASOC 5-digit code was chosen in only 74% of cases. For consecutive BHPS interviews, the same job may be described in different ways, which would lead to even more false apparent changes in occupation. The BHPS documentation notes that:

> *Attention should be paid to the fact that there is a degree of change found in the coding of occupations and industry between waves for respondents who have not apparently changed jobs. These cases could be the result of respondent recall error or misunderstanding, or of coding or keying errors. A blind re-coding exercise indicated a low level of absolute error, however, and if this is randomly distributed throughout the sample, this should not be problematic for analysis. Researchers will have to make their own decisions in these cases, as to whether they consider a job has changed, perhaps by looking at changes in other variables such as pay and hours.*

Given the imprecision of industry and occupation coding, it may be worth looking at using other job information to gauge whether two spells are the same. Many of the spell-pairs that are split by seam effects agree on start date and status, although not on occupation and industry. It seems reasonable to infer that it is the same job, even if occupation or industry codes do not match. Other potential match variables include firm size and sector, and wage levels.

As shown in the following table, which is based on Halpin's *newpan* file, 56 percent of the seam transitions agreed on detailed employment status, and on 2-digit industry and 2-digit occupation. A further 31 percent agreed on detailed employment status.

Table 21: Seam Effects

seam status [variable seamcode]	*Code*	*Agreement on*	*Freq.*	*Percent*	*Cum.*
tight extensive match	1	employment status; and 2-digit industry; and 2-digit occupation	42818	56.11	56.11
tight empstat match	2	employment status	23947	31.38	87.49
loose empstat match	3	employed v non-employed	2553	3.35	90.84
empstat mismatch	4		6537	8.57	99.41
apparent gap	10		454	0.59	100
Total			76309	100	

Source: newpan file. Seams that match on employment status, industry code, and occupation code have been merged, and are not included in this table.

There are alternative methods of resolving dates that have not been used to date by researchers. Oskrochi & Crouchley note the possibility of randomising dates as a means of resolving seam effect problems. Where there is a conflict between a *(w-1)INDRESP* spell and a *wINDRESP* or *wJOBHIST* spell (ie: they are of different type), it is possible to assume that this is due to misreporting of the start-date of the wave *w* spell. In this case, the true start date must lie between the *(w-1)* interview date and the end date of the overlapping wave *w* spell. The true date could be imputed by randomly choosing a date in this feasible interval. Presumably it is more likely that the true start date is somewhere close to the *(w-1)* interview date than that it is close to the *w* interview date although the actual distribution from which a random date is selected is not clear. This approach would be inappropriate where the conflict arose as a result of multiple activities rather than as a result of an incorrectly recalled start date.

A further variant that has not been employed to date is the application of the 'main job' hierarchy to resolving conflicts. The questionnaire asks for 'main job' to be reported. Thus, where a reported part-time job and a reported full-time job overlap, it would be consistent to give priority to the full-time job. Similarly, for two jobs with the same hours, the most highly paid should take precedence.

Resolving spell characteristics

Where there is not an exact match of job characteristics (eg: industry and occupation) between *wINDRESP* and *(w+1)JOBHIST*, there is a choice of whether to use the industry and occupation as measured at the beginning of the spell, or as measured at the end of the spell. Recall that *wJOBHIST* industry and occupation information relates to the start of the retrospective spell. There is clearly a trade-off between consistently measuring characteristics at the beginning of all spells, and using the interview-date report available from *wINDRESP* in order to link it with the richer set of job characteristic information available in *wINDRESP*. *wINDRESP* reports of occupation may differ from the *wJOBHIST* reports not only because of a change in job but also because of inaccurate recall in the retrospective *wJOBHIST* report.

Summary of Approaches

The following table summarises the way that different researchers deal with the overlap between a *wJOBHIST* spell and a *(w-1)INDRESP* spell. These same rules apply to the overlap between *wINDRESP* and *(w-1)INDRESP* spells, which are discussed in the next subsection.

Table 22: Merging *wINDRESP/wJOBHIST* with *(w-1)INDRESP/(w-1)JOBHIST*

Researcher	*Treatment*
Halpin	• Split overlapping spells at the date of the (w-1)INDRESP interview and provide a seam code that allows users to decide on whether to merge them.

Maré	• Delete wJOBHIST spells that end before the (w-1)INDRESP end date
	• Provide a range of 'continuity indicators' that show whether consecutive spells have the same (or similar) status, occupation, industry, start date, employer.
	• Provide a Stata program that applies user-defined rules for merging seam-code spells. Users specify what constitutes the 'same characteristics' and the program then applies the following rules:
	  • For seam spells that overlap but don't have the same characteristics, split them at the date of the (w-1)INDRESP interview
	  • Merge wINDRESP/wJOBHIST and (w-1)INDRESP spells that overlap and have the same characteristics
	  • Use (w-1)INDRESP spell characteristics and start date, since they were collected earlier.
	  • Optionally, merge spells with the same characteristics whether occurring at a seam or not. (This can be used to create a dataset of status spells rather than job spells.)

Oskrochi & Crouchley	• If the start dates of the wINDRESP/wJOBHIST and (w-1)INDRESP spells are the same, merge them if the statuses are the same
	• Merge wINDRESP/wJOBHIST and (w-1)INDRESP spells that overlap and are of the same status if:
	  • the start of the wINDRESP/wJOBHIST spell is earlier than the start of the (w-1)INDRESP spell; and
	  • the difference in start dates is 3 months or fewer.

Paull	• Inconsistencies between (w-1)INDRESP and wJOBHIST and across wINDRESPs are dealt with in the same way as other inconsistencies across spells. (wINDRESP spells have missing end-dates)
	  • Unemployment and out of the labour force spells with missing end-dates were merged with subsequent spells of the same type.
	  • Employment spells with missing end dates were merged with subsequent spells of employment if the subsequent spell started at or before the initial spell.

4.3.3 Merging BLIFEMST *and* CLIFEJOB

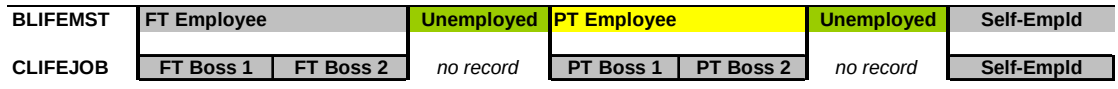
No attempt has been made to enforce consistency with the lifetime employment status history collected at wave two, and contained in record type BLIFEMST. However this history was available to the respondent, and the variable CLJESFV on record type CINDRESP indicates whether the respondent believed this to be correct. In this case, the variable CLJESFN will indicate the number of the employment status spell which should correspond to this record (=BLESHNO on record type BLIFEMST). Respondents may have held more than one job in a single employment status spell, and hence there may be more than one CLIFEJOB record corresponding to the period covered by a single BLIFEMST record. In the employment status history, full-time and part-time spells were distinguished. It is therefore possible in a limited number of cases that a single CLIFEJOB record may correspond to the period of more than one BLIFEMST record.

Source: BHPS documentation (Taylor et al (2005))

Overlaps & timing of information

This is by far the most complex step in the process of combining work history information. The following diagram illustrates how, in principle, the *BLIFEMST* and *CLIFEJOB* records should link together. *BLIFEMST* records sequential and non-overlapping spells of different employment statuses. For the employee or self-employed spells, *CLIFEJOB* provides more detailed information on the sub-spells with each different employer. Employment spells in *CLIFEJOB* can (should!) appear in *BLIFEMST*. In fact, there can be multiple matches.

Figure 5: Merging *BLIFEMST* and *CLIFEJOB* (in theory)



The reality is considerably more complicated, in about every way imaginable!:

- *CLIFEJOB* subspells do not always totally account for the underlying *BLIFEMST* spell (ie: there are gaps)
- *CLIFEJOB* subspells overlap with each other
- *CLIFEJOB* subspells extend into non-employment *BLIFEMST* spells
- *CLIFEJOB* subspells exist that are not matched to *BLIFEMST* spells with which they overlap
- *CLIFEJOB* subspells are missing for some *BLIFEMST* spells

Some but not all of the inconsistencies can be accounted for by the fact that some respondents failed to have a full interview in one or other of the years, or because the two datasets cover different periods. *BLIFEMST* contains records for spells up to and including the spell that was current at wave 2 (B). *CLIFEJOB* in theory (but not in practice) does not contain any spells that start after 1 Sept 1990, since this period is covered by information collected in wave 1(A) and 2(B) *wINDRESP* and *wJOBHIST* files.

When answering the *CLIFEJOB* questions, respondents were provided with a calendar of their responses to the *BLIFEMST* questions. The responses are recorded as variable CLJESFV, which show that only 49 percent of valid responses confirmed that the calendars were correct, with a further 39 percent saying that they were partially correct. The challenge of reconciling *BLIFEMST* and *CLIFEJOB* spells is in many cases one of reconciling histories that respondents themselves identified as inconsistent.

In principle, the *CLIFEJOB* spells should span a shorter period than the *BLIFEMST* spells. The final *BLIFEMST* spell should be the spell that appears in *BINDRESP*. The final *CLIFEJOB* spell should be the spell that precedes the oldest *AINDRESP* or *AJOBHIST* spell, (or which is with the same employer as was current in *CINDRESP* if that started earlier). In

practice, there are 1,682 *CLIFEJOB* spells that start after 1 Sept 1990, which is the latest start-date for *AJOBHIST* spells.

Resolving spell characteristics

In principle, *CLIFEJOB* subspells should have the same employment status characteristics as the underlying *BLIFEMST* spell. They should be full-time employee, part-time employee, or self-employed spells. In practice, the match is imperfect. Halpin (1997, s. 4.4.1) compares the sequence of monthly statuses derived from *BLIFEMST* and from *CLIFEJOB*, and reports that the level of agreement is high – with agreement on broad status (employed, unemployed, not employed) for over 95% of the observations. The disagreements are, however, numerous and varied enough to constitute a significant challenge for integration.

Spell characteristics for employment spells are taken from *CLIFEJOB* spells, since these include more detailed information (eg: industry and occupation) about the spells, and also allow for the characteristics to change between subspells in the same underlying *BLIFEMST* spell.

Note that Halpin initially combines the panel (w*INDRESP* & *wJOBHIST*) files with each of the retrospective files (*BLIFEMST* & *CLIFEJOB*) separately, and only subsequently merge these two files (See Table E1). When creating a longitudinal file from panel and *CLIFEJOB* information only, the non-employment spells appear only as gaps in the long term retrospective history. Halpin codes these gaps on the basis of responses to why the (employment) spell ended. Each period of non-employment is thus created as a single continuous spell of only one non-employment type.

Summary of Approaches

Table 23: Merging *BLIFEMST* with *CLIFEJOB*

Researcher	*Treatment*
Halpin	• Ignore CLIFEJOB information that relates to periods that BLIFEMST reports as non-employment. • Proportionally lengthen all 'within-BLIFEMST-spell' subspells so that they fill the period covered by the underlying BLIFEMST spell.
Maré	• For CLIFEJOB subspells that have been matched to a BLIFEJOB spell, truncate them at the BLIFEJOB start and/ or end dates. • For CLIFEJOB subspells without a BLIFEMST match, keep only those that fall entirely within a BLIFEMST spell • Fill gaps in BLIFEMST spells by creating CLIFEJOB subspells with the same characteristics as the underlying BLIFEMST spell. (This also accommodates CLIFEJOB employment spells within BLIFEMST unemployment spells)
Oskrochi & Crouchley	• Because they analyse only employment status spells, they need use only BLIFEMST. Industry and occupation information is merged back from wINDRESP. • It appears that they do, however, use information from CLIFEJOB as the main source of information (where it exists, which according to the questionnaire wording, but not in practice, is only for jobs current at wave 3 that started prior to 1Sep90) for the period between waves 2 and 3, and with BLIFEMST as the second source of information for the period covered by wave 1 (in preference to using AINDRESP and AJOBHIST)
Paull	• Rely primarily on BLIFEMST. • To obtain employer spells, identify the dates of employer-to-employer moves from CLIFEJOB and create an employer-to-employer spell division in the employment spells from BLIFEMST if these moves fell within an employment spell in BLIFEMST (fn 32, p8).

Note that for the creation of a WLH file of employment status history only, this arduous step can be avoided.

4.3.4 Merging BLIFEMST/CLIFEJOB *with w*INDRESP*/wJOBHIST*

The main issue in combining the retrospective (*BLIFEMST*/*CLIFEJOB*) spells and the annual (*wJOBHIST*/*wINDRESP*) spells is deciding the date at which to start using the more detailed annual data. (Note, however, that Oskrochi and Crouchley use the retrospective information in preference to the annual information whenever it is available.)

According to the wording of the questionnaire, combined retrospective information is available only for the period covered by *CLIFEJOB*, which is shorter than that covered by *BLIFEMST*. The retrospective spells should thus finish with the spell prior to that covered by *AJOBHIST*, although the final retrospective spell may include one or more spells from *AJOBHIST*, *AINDRESP*, etc if they were with the same employer as the final *CLIFEJOB* spell.

Overlaps & timing of information

As noted above, the two retrospective sources cover slightly different periods. The last *CLIFEJOB* spell should be one that starts before 1 Sept 1990, which is the period after which

AJOBHIST information is available. The *BLIFEMST* spells continue later - until the spell is reached that was current at wave 2 (which is included).

Halpin (1997, s. 4.4.2) examines the overlap between *BLIFEMST* and *AJOBHIST* to gauge consistency. He reports rates of agreement over a 12 month overlapping period (9/90 to 8/91) and for 3 different status groupings. Agreement is high, ranging from 87.5% to 93.3%, depending on length of recall and coarseness of grouping.[5]

Resolving Dates and spell characteristics

By relying on the most proximate report of an event, most researchers use annual (*wINDRESP*/ *wJOBHIST*) spell information where it is available, and use the less detailed employer and status information from retrospective sources for periods prior to that covered by *AINDRESP* and *AJOBHIST*. Oskrochi and Crouchley are the exceptions, in relying primarily on retrospective sources.

Seam Effects

The merging of retrospective (*BLIFEMST*/ *CLIFEJOB*) spells with panel (*wINDRESP*/ *wJOBHIST*) spells creates seam effects in the same way that merging *wINDRESP* with subsequent *wINDRESP* or *wJOBHIST* does. Because wave A and wave B responses are given precedence over retrospective reports, any inconsistency between the two sets of information results in a truncated retrospective spell. In some cases, the final retrospective spell will be the same spell as the first panel spell, although as with the seam effects discussed earlier, it is not always straightforward to tell whether spells are 'the same'.

[5] Longer recall and more detailed status groupings generate lower match rates.

5. The Maré files: *wlh.dta*

This section provides a summary of the programs used and derived datasets created in implementing the adjustments described in earlier sections.

5.1. Details of construction

My programming has followed the same stages outlined in sections 3 and 4:

- Standardise the base data
- Regularise the date sequences
- Combine the different sources

Figure 6 shows the structure of programs and derived datasets, grouped into these three stages. Note that for combining *BLIFEMST* and *CLIFEJOB* records, the second and third of these steps is performed by the same program, since the two sources cover a significantly overlapping period, providing different information. Dealing with the dates is therefore best done by using information from both datasets.

Programming was carried out using Stata. Programs thus have a '.do' extension and datasets have a '.dta' extension.

The shaded datasets are the main derived datasets:

- *tempall.dta* that contains all records from the base data, in a consistently coded form. This is particularly useful for recovering the base records from which the reconciled histories are derived.
- *wlh.dta* that is a file containing consistent work-life histories.

Utility programs needed for creation of wlh.dta

The suite of programs that I use also contains a number of 'utility programs' that are called by the programs shown in Figure 6. These programs are:

progs.do: contains repeatedly used code to perform the following functions:

- Recalculating minimum and maximum feasible dates, which needs to be done whenever a date is imputed.
- Adjusting inconsistent dates within feasible ranges
- Classifying and counting inconsistent dates

labels.do: contains variable coding labels (formats) for variables in *tempall* and *wlh*.

clean_bcsort.do: sorts *BLIFEMST* and *CLIFEJOB* spells into approximate chronological order (called by *clean_clifejob.do* and saved as a separate file only because Stata cannot hold a program of more than 32Kb.

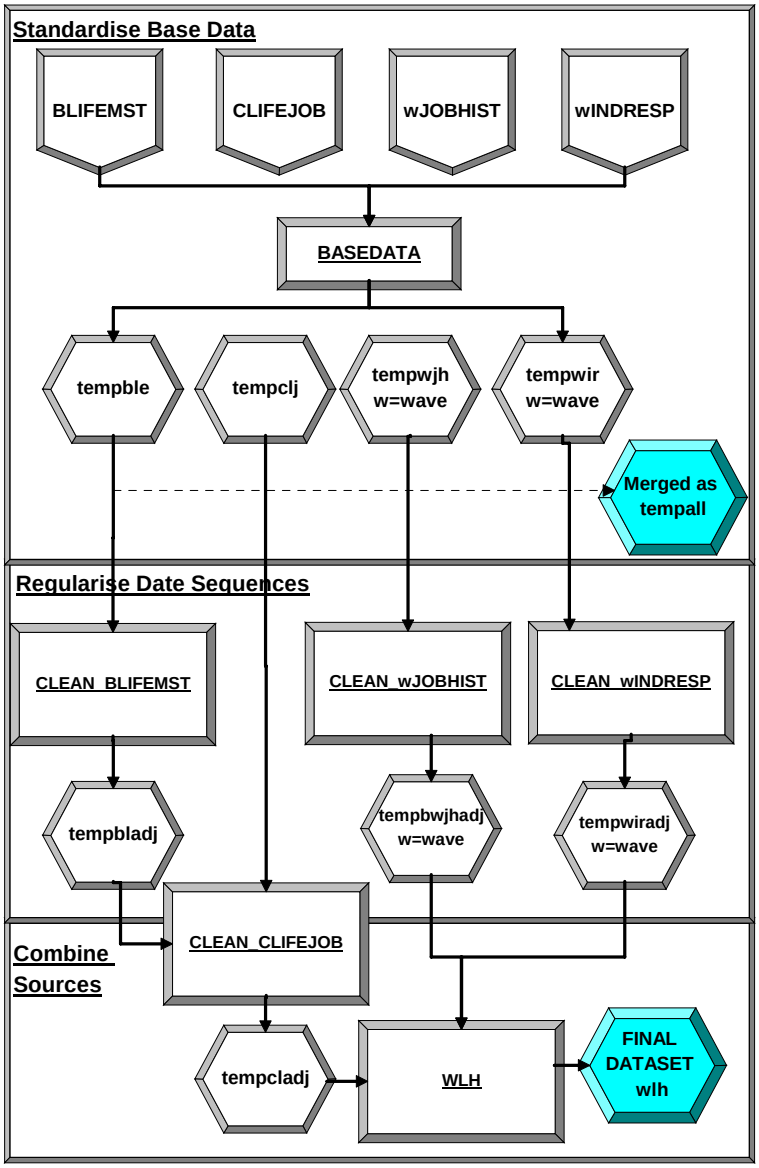
setpath.do: defines Stata macros referring to file locations

dashboard.do: A single program that invokes all of the other programs, other than setpath.do (for ease of execution)

Other utility programs

deseam.do: A program to merge consecutive spells that match on specified criteria, either just at seams, or across all spells. A further explanation and instruction is included below.

Figure 6: Maré files – Structure of programs



Notes: Rectangular boxes denote programs. Pentagonal boxes denote base data. Hexagonal boxes denote derived datasets.

Installation

Installing and executing the programs is relatively straightforward:

1. Unzip the file wlh.zip into an empty folder
2. Edit the file *setpath.do* to record a) the name of the folder containing the file; b) the location of the BHPS data; c) the number of waves of data; and d) (optionally) the location of Halpin's work history files.[6]
3. Run the file *setpath.do* from Stata, and then run *dashboard.do* (which invokes all of the other programs.)[7]

[6] To obtain versions of the Halpin files in a form similar to *tempall.dta*, follow the instructions towards the end of *basedata.do* (the code is included after the *exit* statement).

[7] The entire suite of programs takes about 5-10 minutes to run on a 2.0 GHz Pentium 4 computer.

The *wlh.dta* dataset (and *tempall.dta*) will be in the *\data* subfolder of the folder used in step 1. The following sections provide guidance on the use of the resulting dataset, and some further documentation of the files and programs.

5.1.1 Standardise the Base Data

The first step is to get all of the work-life-history data into a common format, with consistent coding schedules. Section 3.2 summarised the issues involved in obtaining consistent measures of variables of interest from the various datasets. The program file *basedata.do* contains the code to implement the standardisation. It creates separate output datasets for each data source, which is then also pooled as *tempall.dta*. This pooled dataset contains information on all work history spells from all the data sources. The spell data are included in a standardised form but there has been no attempt to reconcile the various reports. The pooled dataset is particularly useful in identifying whether inconsistencies that become apparent during processing were present in the base data, or have arisen in the course of processing. (I referred back to this dataset often!)

The following is a list of the variables included in *tempall.dta*.

Table 24: Standardised Dataset Variables – tempall.dta

Variable Name	*Description and coding*
Pid	Cross-wave personal identifier
Srcename	Records the file from which the spell is drawn. Coding as in Halpin: wINDRESP = 10*[wave number] wJOBHIST = 10*[wave number]+1 BLIFEMST = 22 CLIFEJOB = 32 [formatted to display as filename]
Source	Srcename with a sequential spell number appended to the right of the decimal point: wJOBHIST: spell=1,2, . . . 9 (1=most recent spell) BLIFEMST: spell=01, 02, . . . 38 (1=oldest spell) CLIFEJOB: spell=bbcc where: bb=corresponding BLIFEMST spell number (numbered above 50 for unmatched spells) cc=sequential within PID (=CLJSEQ)
Sort	A variant of source, which orders the spells in chronological order [with source ordered as: BLIFEMST; CLIFEJOB; AJOBHIST; AINDRESP; BJOBHIST; BINDRESP, etc.]
Smonth	Spell start month – coded as wJHBGM
Syear	Spell start year – coded as wJHBGY4
Start	Start date – expressed as number of months since January 1900 Set to missing if month or year is not reported exactly.
Startmiss	Indicator of type of missing date: 0=month and year known; 1=year alone is known; 2=season & year known; 3=year unknown – some month information; 4=missing month and year.
Emonth	Spell end month – coded as wJHENDM
Eyear	Spell end year – coded as wJHENDY4
End	Start date – expressed as number of months since January 1900 Set to missing if month or year is not reported exactly.
Endmiss	Coded as for startmiss
Status	Employment Status – See Table C1 for uniform coding. Other values: 98=*wJOBHIST* employment spells with (*wJHSTAT*=1) and no further information from *wJHSEMP*; 99=*wJOBHIST* records for other employment spells (*wJHSTAT*=2) with missing *wJHSEMP*, or

	wINDRESP employment records with missing *wJBSEMP* or *wJBFT*.
Term	Term of employment (Permanent/ temporary/ casual) See Table C3 for uniform coding (missing for BLIFEMST)
Sic	industry (coded as in wJBSIC)
Soc	occupation (coded as in wJBSOC)
Sect	Sector of employer (coded as in wJBSECT)
Size	Firm size (coded as in wJBSIZE)
Whyleft	Reason for leaving. See Table C4 for uniform coding (missing for wINDRESP and BLIFEMST)
Paygu	Usual Gross Monthly Pay (wJOBHIST – as at the later of spell start and 1Sept (w-1); wINDRESP - at interview date)
Paynu	Usual Net Monthly Pay (wJOBHIST – as at the later of spell start and 1Sept (w-1); wINDRESP - at interview date)

5.1.2 Regularise the Date Sequences

The second step is to ensure that date sequences for the records *within* each data file are consistent. This is done for *BLIFEMST*, the *wJOBHIST* files, and the *wINDRESP* files by the programs *clean_blifemst*, *clean_wjobhist* and *clean_windresp* respectively. What is done in these files is summarised in sections 3.1 and 3.2.

The regularising of the *CLIFEJOB* spells is done at the same time as they are combined with the (already regularised) *BLIFEMST* records, due to the overlapping nature of the data from these two sources. Because the combining of *BLIFEMST* and *CLIFEJOB* spells is by far the most complex stage in the whole exercise of creating consistent histories, I describe the procedure in some detail in Appendix D below.

5.1.3 Combine the different sources

There are two issues that arise as a result of combining all of the different data sources in a single file. Conceptually different spell types are combined, and seam transitions are created.

Conceptually different spells

Spells of employment exist in the data in the form of employer spells (from *BLIFEMST*), as well as job spells (from *wJOBHIST* and *wINDRESP*). There are also some spells that are available only as employment status spells (from BLIFEMST, where CLIFEJOB spells are not available). Spells of non-employment are available from all data sources, with a similar categorisation (as detailed in Appendix Table C1).

Seam Effects

Seam effects lead to artificially shortened spells. A more detailed exposition of the nature and extent of the problem, and examples of practical solutions are contained in section 5.2.3.

5.2. The final reconciled dataset – wlh.dta

Table 25 shows the variables that are included in the final dataset – *wlh.dta*. Many are the same as were defined for *tempall.dta*. Additional variables are those indicating coverage (whether a complete history exists for an individual in a given wave), imputation, and spell continuity (whether the spell is with the same employer as, has similar characteristics to the spell that follows, or ends at an interview date).

Table 25: Reconciled Dataset Variables – *wlh.dta*

Variable Name	*Description and coding*
Variables from tempall.dta	
Pid	Unique individual ID
Srcename	Records the file from which the spell is drawn. Coding as in Halpin: *wINDRESP* = 10*[wave number] *wJOBHIST* = 10*[wave number]+1 *BLIFEMST* = 22

	CLIFEJOB = 32 [formatted to display as filename]
Source	Srcename with a sequential spell number appended to the right of the decimal point: *wJOBHIST*: spell=1,2, . . . 9 (1=most recent spell) *BLIFEMST*: spell=01, 02, . . . 38 (1=oldest spell) *CLIFEJOB*: spell=*bbcc* where: *bb*=corresponding *BLIFEMST* spell number (numbered above 50 for unmatched spells) *cc*=sequential within *PID* (=*CLJSEQ*)
Hist	Indicates whether the spell is from: 0) A *wINDRESP* file 1) A *wJOBHIST* file 2) A retrospective file (*BLIFEMST/CLIFEJOB)*
Sort	A variant of source, which orders the spells in chronological order [with source ordered as: *BLIFEMST*; *CLIFEJOB*; *AJOBHIST*; *AINDRESP*; *BJOBHIST*; *BINDRESP*, etc.]
Start	Start date – expressed as number of months since January 1900 Set to missing if month or year is not reported exactly.
End	Start date – expressed as number of months since January 1900 Set to missing if month or year is not reported exactly.
[Other date variables]	Start date: *smonth* and *syear* contain the start month and year The reported start and end dates for the spell (as they appear in the base data), are included as *sraw* and *eraw*.
Status	Employment Status – See Table C1 for uniform coding
eun	Broad employment status: Employment; Unemployment; Not in Labour Force
Term	Term of employment (Permanent/ temporary/ casual) See Table C3 for uniform coding (missing for *BLIFEMST*)
SIC	industry (coded as in *wJBSIC*)
SOC	occupation (coded as in *wJBSOC*)
Sic2d	2-digit industry code
Soc2d	2-digit occupation code
Sect	Sector of employer (coded as in *wJBSECT*)
Size	Firm size (coded as in *wJBSIZE*)
Whyleft	Reason for leaving. See Table C4 for uniform coding (missing for *wINDRESP* and *BLIFEMST*)
Paygu	Usual Gross Monthly Pay (*wJOBHIST* – as at the later of spell start and 1Sept (w-1); *wINDRESP* - at interview date)
Paynu	Usual Net Monthly Pay (*wJOBHIST* – as at the later of spell start and 1Sept (w-1); *wINDRESP* - at interview date)
Coverage [(w)ok]	[aok, bok, cok, etc.] PID-level variable: Indicate the wave up to which full interviews are available

Table 25 (cont): Reconciled Dataset Variables – *wlh.dta*

Additional variables	
Imputation/Edit Flags	SFLAG: identifies where (and why) start date has been altered to achieve consistency. EFLAG: identifies where (and why) end date has been altered to achieve consistency. DFLAG: identifies individuals for whom one or more recorded spells have been dropped to achieve consistency. MFLAG: identifies spells that have been merged with other spells (Coding described below)
Samemp	Indicates whether a *wJOBHIST* employment spell is with the same employer as the next spell
Empstart	The date that the respondent started with the current employer. (not used to distinguish spell types)
Nextsame	Defined for *wINDRESP* records that might be the same spell as the following spell. Indicates how good the match is (based on status, 2-digit industry, 2-digit occupation) (Similar to Halpin's seamcode)
Seam	Identifies *wINDRESP* spells whose end-dates occur at a seam (interview date)
Seam	Identifies *wINDRESP* spells whose end-dates occur at a seam (interview date)
E_seam	Identifies final retrospective spells whose end-dates occur at the earliest *aINDRESP* or *aJOBHIST* spell.
Spell continuity indicators	*samestat* equals one if a spell is followed by a spell with the same employment status (as defined in Table C1) *sameeun* equals one if a spell is followed by a spell with the same broad employment status (employed, unemployed, not in labour force) *samesoc* equals *i* if a spell is followed by a spell with the same *i*-digit occupation. (i=1, 2, 3) Eg: samesoc>2 captures spells that match at 2-digit or higher *samesic* equals *i* if a spell is followed by a spell with the same *i*-digit industry. (i=1, 2, 3, 4) Eg: samesic>2 captures spells that match at 2-digit or higher *samejob* equals one if *samesoc*=3 and *samesic*=4 *same2d* equals one if a spell is followed by a spell with the same 2-digit industry *and* 2-digit occupation. *Samestart* equals one if a spell is followed by a spell that started before the seam date (with start date as reported in the raw data), and has the same start date as the seam spell. (The start dates are treated as matching if they differ by exactly 12 months or by up to three months, to accommodate the impact of common recall errors.) *Nextsame* (as shown in Table 26) is an alternative (hierarchical) indicator of seam match, and corresponds the other variables as follows: *sameeun* =1 (& samestat = *samestat* = 1 (& *same2d*=0) *samestat* = 1 and *same2d* = 1 (& samejob~=1) *samejob* = 1

Note: The dataset created from waves 1 to 13 contains 148,747 spell observations on 9,017 individuals

5.2.1 Coverage

As indicated earlier, one of the intended uses of the final *wlh* file is that it could be merged with wave-specific (*wINDRESP*) information. It is therefore useful to know whether the history of an individual is complete, in the sense of being completely observed, up to the date of each interview. If a respondent is not fully interviewed in any one year, their completely observed history is interrupted, and cannot be reliably merged with subsequent *wINDRESP* files. The variables *aok*, *bok*, *cok*, etc, are indicator variables that take the value of one if the history is completely observed up to wave a, b, c, etc. respectively. For some individuals, the indicators are initially zero, and get 'turned on' at a wave later than wave 1(*A*). These are individuals who are children initially, but whose transition to a fully interviewed adult is observed, together with subsequent full interviews (See section 4.2 for a discussion).

Note also that respondents giving full interviews in Wave A or B have fully observed histories only if they also fully responded to Waves B and C, since these are the waves in which the retrospective data were collected.

Depending on the type of analysis undertaken, (eg: using number of prior unemployment spells as a correlate of spell duration) it may be necessary to restrict the sample to respondents who have complete histories (using the [*w*]*ok* variables).

5.2.2 Imputation

There are four imputation flags, indicating whether observations have been merged, dropped, or have had start or end dates imputed.

mflag: Equals one if the observation contains information from more than one recorded spell. This flag equals zero unless seam transitions have been merged, as described in the next section.

dflag: Equals one for all observations on an individual if a spell has been deleted from the history for that individual, because it overlapped totally with another spell, or because it was a CLIFEJOB spell that was dropped (See section 5.1.2). Overall, 18% of respondents have had one or more spells dropped from their histories.

sflag: Indicates whether start date has been imputed. Coded as follows:

0	not imputed
11	date imputed by moving a 'winter' from January to December
12	date has been altered so as to truncate the spell length
13	date has been altered so as to lengthen the spell
21	totally missing date has been imputed
22	date with missing month has been imputed
23	date with month reported as season has been imputed.

eflag: Indicates whether end date has been imputed. Coded as for sflag.

5.2.3 Seam Effects

The assumptions applied when combining the different regularised data sources are summarised in section 4. The only issue that is dealt with further in the current section is the treatment of seam effects. Seam effects were described in section 4.3.2, with reference to Halpin's *newpan* file.

To recap, there are two sources of seam transitions in the *wlh.dta* file. First, because we have retained all *wINDRESP* spells, with the interview date as their end-date, they almost all truncate the following (*w+1)JOBHIST* or *(w+1)INDRESP* spell. The only exceptions arise with final interviews. There is thus a seam at almost every interview date. Second, the first *aJOBHIST* or *aINDRESP* spell takes precedence over the final retrospective spell, and many retrospective spells therefore have truncated end-dates.

As shown in Table 25, there are several 'spell continuity' indicators that help in identifying spells that might be considered for merging. Table 26 shows the relationship between the status and job characteristics of each spell with those of the spell following, using the *nextsame* indicator variable. We would expect that, in many cases, *wINDRESP* spells would refer to the same spell as the following (*wJOBHIST* or *wINDRESP*) spell. We therefore need to decide when the spells are the same. The top-left box in Table 26 indicates that there are 44,092 *wINDRESP* spells that appear to be identical to the spell that follows. They are identical in that they have the same status, industry, and occupation codes. (For non-employment spells, industry and occupation are missing, and thus always match exactly.)

It is notable that 50 percent of *wINDRESP* spells match the spell that follows on the three dimensions. This is much higher than the 9% or 14% in the other datasets. Even though they match on status, industry, and occupation, *wJOBHIST* and many retrospective spells are distinct. They are reported as separate spells during the same interview, and may arise where a respondent gets a promotion within the same firm without changing occupations, or where a job change does not lead to any change in status, occupation and industry.

Some of the pairs that fail to match on industry and occupation still match on 2-digit industry and occupation (3% of spells). This could reflect inconsistent reporting or coding of the same job. An even larger proportion of spells (18%) match on status but not on 2-digit industry and occupation. Some of these may be the same job, coded differently.

Table 26: Identifying Spell-continuity

	Source			
Next spell has . . .	*wINDRESP*	*wJOBHIST*	*Retrospective*	*Total*
same status, SIC, and SOC	44,092	3,374	3,227	50,693
Nextsame=4	50%	14%	9%	34%
same status, 2-digit SIC & 2-digit SOC	2,586	369	520	3,475
Nextsame=3	3%	2%	1%	2%
same status	21,560	5,826	13,652	41,038
Nextsame=2	25%	24%	37%	28%
same broad status [Emp/Un/NILF]	5,655	2,399	3,893	11,947
Nextsame=1	6%	10%	11%	8%
Different characteristics	13,674	12,617	15,303	41,594
Nextsame=0	16%	51%	42%	28%
Total	71598	20506	37005	129109

Notes: *Top-left cell indicates the wINDRESP spells that are most likely to be merged with subsequent spells because they appear to be the same spell. The variable nextsame is created by the program seamcode, stored in the file deseam.do*

5.3. Some Practical Examples of the use of *wlh.dta*

The final dataset is in a form that can readily be merged with *wINDRESP* files. For use as a source of accumulated experience of various types, the dataset is in a 'ready-to use' format. For analysis of spell durations, some additional manipulation is needed. In either case, care needs to be taken to select only individuals for whom valid histories are available.

5.3.1 *Merging with* wINDRESP

wlh.dta intentionally includes all *wINDRESP* spells, to enable researchers easily to merge other variables from *wINDRESP* onto the file. This section shows one way of performing such a merge.

Selecting a sample

When merging with *wINDRESP*, we need to limit the sample to those respondents who have complete histories up to and including wave *w*. This can be done by:

- select from *wlh.dta* only observations for which variable [*w*]*ok*=1.

- merge with *wINDRESP* and restrict the sample to cases that contain a matched *wlh.dta* observation. (in Stata, _merge==3).

Note that the sample in *wlh.dta* is already restricted, so that histories are included only up to the point where there is a gap in the history. There are, however, cases where [*w]ok*=0 and [*w+1*]*ok*=1 (children becoming fully interviewed adults).

Sample Stata code to merge wlh_2 onto *d*indresp:

```
use dindresp, clear
keep pid ddoiy4 ddoim /* and other vars (eg: dage dsex djbhrs dpaygu) */
sort pid
merge pid using wlh
keep if _merge==3 /* drop if no wlh record (merge==2) or if no indresp
                     record (_merge==1) */
keep if dok==1 /* drop if history to wave d is incomplete */
ge doi=((ddoiy4-1900)*12)+ddoim   /* for wave a, need to generate
      adoiy4=1991 */
keep if start<doi
sort pid sort

by pid: replace end=doi if _n==_N
```

Obtaining accumulated experience measures up until the *wINDRESP* interview date

To calculate accumulated experience, it is a straightforward matter of summing up the durations of all spells of interest for each individual (*PID*). A little more programming is required if you want to extract experience from only the previous x years.

```
* to obtain total unemployment experience
egen totun=sum((end-start)*(status==3)), by(pid)

* to obtain a count of months unemployed in the previous 12 months
gen recent_un=max((status==3)*(end-max(doi-12,start)),0)
egen tot_recent_un=sum(recent_un), by(pid)
```

5.3.2 Merging spells at seam transitions

This section documents the prevalence of seam transitions in *wlh.dta*, and shows the number that remain after applying various plausible rules for merging. In the interests of making such analyses easier for other users, I illustrate the use of the *deseam* command to join together seam spells that 'match' in the sense of meeting user-defined matching criteria. The *deseam* command is a specially-written program that I will make available with the dataset. I also show how the command can be used to join non-seam spells that users may wish to treat as a single spell.

The decision about which seam spells to join involves a trade-off between making false joins and leaving false transitions. Failing to join spells that are in fact the same will lead to artificially shortened durations. Joining spells that actually differ will lead to misclassification of spells and artificially lengthened spells.

The first row of Table 27 shows the prevalence and nature of seams in the *wlh.dta* data file for the first 13 waves of the BHPS. The file contains 148,738 spells, 78,550 (53%) of which are spells that end on an interview date, and a further 7.077 (5%) of which are final retrospective spells that have been truncated at the start date of the earliest spell reported in wave 1 (from *aINDRESP* or *aJOBHIST*). The great majority of seam transitions are thus from the *indresp*

file. In fact, the only *indresp* spells that are *not* classified as seam transitions are those from the final wave, for which the spells are incomplete.

Table 27: Impact of different treatments of seam effects

Treatment of Seams	*Seams from Indresp*	*Seams from retrospective spells*	*Non-seam spells*	*Total number of spells*
No adjustment	78,550 (53%) VI Start: 77,036	7,077 (5%) VI End: 2,321	63,120 (42%) VI Start: 3,802 VI End: 2,802	148,747 (100%)
Merge if samejob==1 & samestart==1	47,162 (40%) VI Start: 45,669	6,254 (5%) VI End: 2,102	63,120 (54%) VI Start: 3,802 VI End: 2,802	116,536 (100%)
Merge if samejob==1	34,458 (33%) VI Start: 33,620	5,674 (6%) VI End: 1,933	63,120 (61%) VI Start: 3,802 VI End: 2,802	103,252 (100%)
Merge if same2d==1 & samestart==1	45,329 (40%) VI Start: 43,836	6,220 (5%) VI End: 2,070	63,120 (55%) VI Start: 3,802 VI End: 2,802	114,669 (100%)
Merge if same2d==1	31,872 (32%) VI Start: 31,059	5,586 (6%) VI End: 1,861	63,120 (63%) VI Start: 3,802 VI End: 2,802	100,577 (100%)
Merge if samestat==1 & samestart==1	30,314 (31%) VI Start: 28,839	5,213 (5%) VI End:1,809	63,120 (64%) VI Start: 3,802 VI End: 2,802	98,647 (100%)
Merge if samestat==1	10,312 (14%) VI Start: 9,743	2,856 (4%) VI End: 949	63 120 (83%) VI Start: 3,802 VI End: 2,802	76,288 (100%)
Merge if sameeun==1 & samestart==1	26,709 (28%) VI Start: 25,238	4,952 (5%) VI End: 1,709	63,120 (67%) VI Start: 3,802 VI End: 2,802	94,781 (100%)
Merge if sameeun==1	4,657 (7%) VI Start: 4,269	1,897 (3%) VI End: 595	63,120 (91%) VI Start: 3,802 VI End: 2,802	69,674 (100%)

Note: VI refers to 'violent imputations', whereby a start or end date is forced to be inconsistent with a reported date. It is identified by sflag or eflag taking values 12 or 13.

The two types of seam-spells are flagged by the variables *seam* (=1 for *wINDRESP* seam spells), and *e_seam* (=1 for final retrospective spells).

Not all seam transitions are inconsistent with reported start dates for the subsequent spells. The subsequent spell could report the interview date as the start date, or the start date could be reported as a range (eg: season), or could be missing. Table 27 shows the number of seam spells that lead to a contradiction of the stated start date of the following spell. These are labelled *VI* for 'violent imputation'. They are identified by the value of the variable *sflag* (for spells following *wINDRESP* seams) or *eflag* (for final retrospective spells). A value of 12 or 13 is counted as a violent imputation. Inconsistencies in the reporting of work-life histories

result in 6,604 violent imputations in non-seam spells – a rate much lower than for the seam spells. Almost all of the *wINDRESP* seams result in violent imputations.

The second row of Table 27 shows the impact of applying a fairly strict criterion to identifying which seam spells should be merged – requiring that seam spells agree on employment status, industry, occupation, and start date (with some allowance for reporting errors in start dates, as described in Table 25). The number of *wINDRESP* seams reduces by about a third, to 47,162. The original number of *wINDRESP* seams is roughly halved, to 34,458 if we require status, industry, and occupation match but do not require start dates to match.

The remainder of the table shows the impact of applying less stringent conditions for merging. Requiring industry and occupation to agree only at the 2-digit level allows for some known measurement error in these variables, but achieves a relatively small number of additional merges. We also report the impact of merging job spells that agree only on either detailed (samestat) or broad (eun) employment status. The final row shows that the number of *wINDRESP* seams is greatly reduced by merging seam spells where sameeun==1. However, this is achieved only by allowing full-time employment, part-time employment, and self-employment spells to be merged together.

The use of the *deseam* command

The application of different seam merging rules, as summarised in Table 27, can be easily effected with the use of the *deseam* command, which has been written specifically to be applied to the *wlh.dta* data file.

The syntax of the *deseam* command is:

deseam *"join condition"* [, scope of join]

"join condition" is a logical expression. When the expression is true, the seam spells are joined. The join condition expression can be easily constructed using the continuity indicators shown in Table 25.

Scope of join options are: seam, indresp, retrospective, non, and all

seam [default] joins seam spells from both *wINDRESP* and retrospective spells

indresp joins seam spells from *wINDRESP* only

retrospective joins seam spells from *BLIFEMST and CLIFEJOB* only

not joins all consecutive non-seam spells that meet the join condition.

all joins all consecutive spells, whether they are seam spells or not.

EXAMPLES

Each of the 8 'merge' rows of Table 27 was obtained by issuing the following commands:

```
use wlh, clear
run deseam

* and then one of the following:
deseam (samejob==1 & samestart==1)
deseam (samejob==1)
deseam (same2d==1 & samestart==1)
deseam (same2d==1)
deseam (samestat==1 & samestart==1)
deseam (samestat==1)
deseam (sameeun==1 & samestart==1)
deseam (sameeun==1)
```

following any one these with the command 'seamanal' provides the analysis of seam patterns. Note that, in each case, the scope of the merge takes on the default value of 'seam', and so merges only seam spells.

In each case, issuing the *deseam* command does the merging of spells and leaves a modified version of the *wlh.dta* data in memory, to be further analysed on saved.

The Impact of different rules on duration distributions

Figure 7 shows the impact of different seam merging rules on the duration distribution of different types of spells. Panel (a) shows the distribution of job spells. In the base *wlh.dta* file, 18 percent of job spells have a duration of exactly 12 months, with a similar proportion at either 11 or 13 months. Merging seam spells that agree on status, industry and occupation (samejob==1) reduces the peak somewhat, to 12 percent. Relaxing the merge requirements to permit merging of seam spells that agree on detailed employment status reduces the peak to around 5 percent.

The remaining three panels of Figure 7 show the duration distribution of broad employment status spells. The non-seam *wlh.dta* spells were first merged to form 'eun' spells (employment, unemployment, not in labour force) using the command *deseam sameeun, not*. In this modified file, 28 percent of employment spells, 8 percent of unemployment spells, and 36 percent of not in labour force spells were exactly 12 months in length, almost entirely as a result of seam effects. (The only spells longer than 12 months were from retrospective files.)

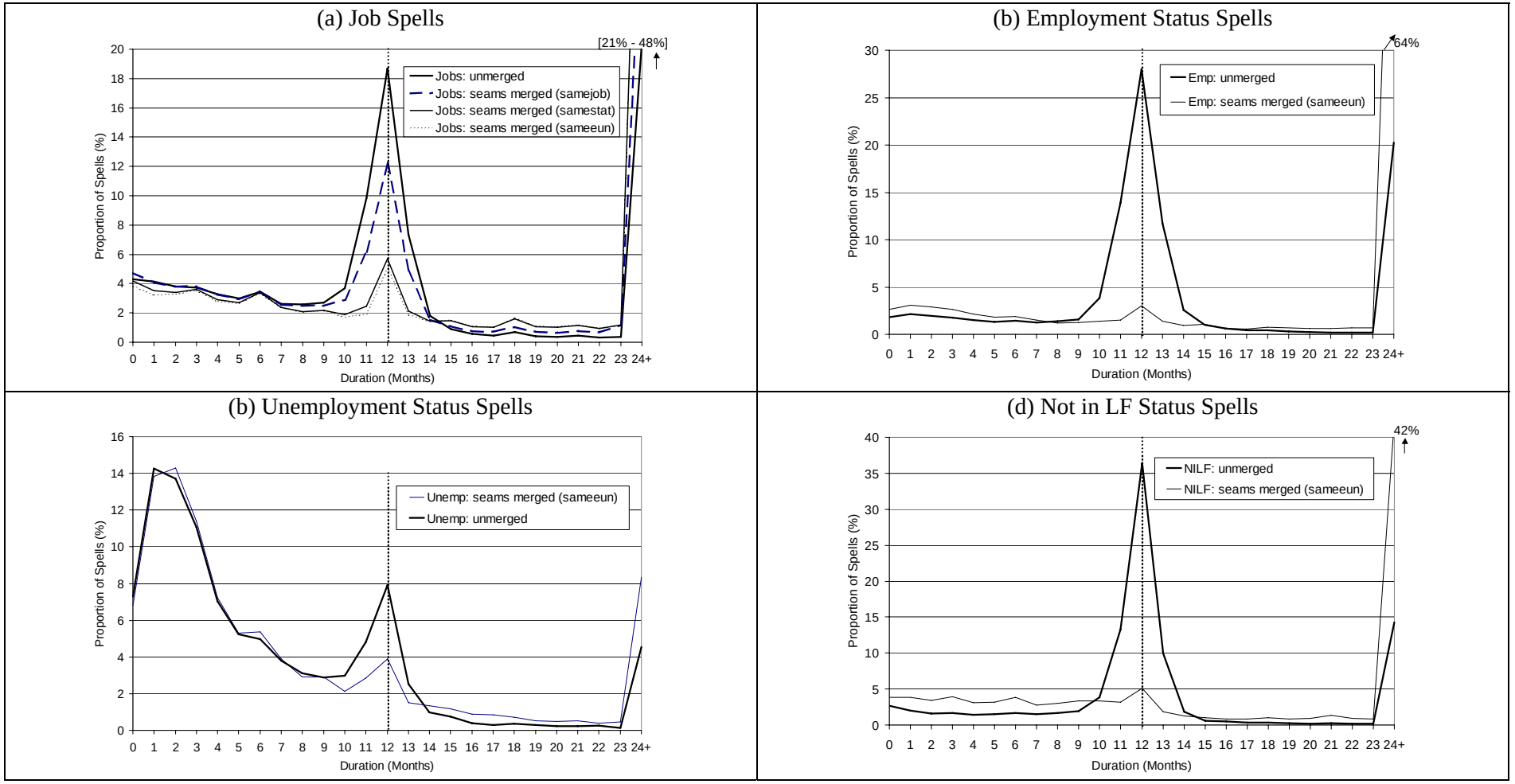
Merging the seam spells if theyagreed on broad employment status significantly reduced the height of the 12-month peak, although there still appears to be an inflated proportion of spells with 12-month duration. The impact of seams is less pronounced for unemployment spells because spells of less than 12 months are relatively common. In contrast, not in labour force spells are greatly affected, especially spells of retirement, many of which last for many years, and thus contribute several 12-month segments to the unadjusted spell count.

A casual inspection of the remaining seam effects reveal many cases of apparently long job or status spells punctuated by short spells of other activity for one wave, or else cases where repeated spells of self-employment or employment with the same industry and occupation coding are interspersed with occasional short unemployment or part-time spells. I have not undertaken a systematic analysis to check whether these impressions are representative of the data as a whole.

Overall, it appears that seam transitions are still problematic for the analysis of the duration of job spells. The challenges are somewhat less severe for the duration of employment status spells, and are likely to be less problematic for the analysis of accumulated experience.

Further analysis of the prevalence and impacts of seam effects in the analysis of BHPS data can be found in Halpin (1997), Paull (2002), and Jäckle and Lynn (2006).

Figure 7: Impact of Seam merging rules on spell duration



5.3.3 Creating files of different spell types

As illustrated in the previous section, the *deseam* command can also be used to create spell files with different types of spells, by converting the spells in *wlh.dta* into different types:

- Deriving consistent job spells requires that retrospective spells be dropped (*drop if hist==2*)
- Deriving consistent employer spells requires that *wJOBHIST* spells with the same employer be merged (*deseam samemp==1, all*)
- Deriving consistent employment status spells requires that all spells with the same status be merged (*deseam samestat==1, all*)

Deriving consistent (broad) employment status (ie: EUN) spells requires that all spells with the same broad status be merged (*deseam sameeun==1, all*)

Job spells (Detailed Employment Status)

Retrospective (*BLIFEMST/ CLIFEJOB*) spells need to be dropped since they are more aggregated than jobs, being employment status, or employer spells. Seam-merging rules need to be applied. Here is sample code to create a file of job spells:

```
use wlh, clear
drop if hist==2 /* keeps only wINDRESP and wJOBHIST spells */
deseam samejob /*     merges seam transitions that apply to the same job */
               /*     Users can choose alternative merge rules   */
seamanal       /* to summarise the impact of the merging – as per Table 27 */
```

Employer spells

WJOBHIST spells that are with the same employer as for the (chronologically) subsequent spell must be merged. Some seam-merging rules need to be applied. Here is sample code to create a file of job spells:

```
use wlh, clear
deseam samejob /*     merges seam transitions that apply to the same job */
               /*     Users can choose alternative merge rules   */
deseam samemp, all /* merges same-employer jobs.  The ,all option merges */
                   /* same employer spells even if they are not */
                   /* at the seam */
seamanal       /* to summarise the impact of the merging – as per Table 27 */
```

Broad employment status (EUN) spells

Consecutive spells that have a common employment status need to be merged, whether they occur at seams or elsewhere.

```
use wlh, clear
deseam sameeun, all  /* merges seam and non-seam transitions that */
                     /*have the same broad employment status */
seamanal       /* to summarise the impact of the merging – as per Table 27 */
```

5.4. Comparison with Halpin's files

As noted in the Section 1, I began this exercise by using Halpin's work-history files. I have ended up creating a new file that is somewhat different from Halpin's files. It is worth briefly noting some of the main differences between the two files.

5.4.1 Coverage

- Halpin's files contain all available history information.
- My file contains only completely observed histories (ie: history segments following a wave of non-response are excluded).
- Only one of Halpin's four files (LJEMP) contains information from all sources (See Table 28).
- My file contains information from all sources.

Table 28: Coverage and Spell definition

Dataset	*BLIFEMST*	*CLIFEJOB*	*wJOBHIST*	*wINDRESP*
Halpin				
• *NEWPAN*			Job	Job
• *LEMPE*	Status			
• *XLEMPE*	Status		Status	Status
• *LJOBE*		Employer (fpsun)		
• *XLJOBE*		Employer (fpsun)	Employer (fpsun)	Employer (fpsun)
• *LJEMPE*	Status	Status	Status	Status
Maré				
• *WLH*	Employer (Stat(2))		Job (Stat(2))	Job (Stat(2))

Note: *Status denotes a breakdown of employment status coded as in BLIFEMST*
FPSUN denotes a breakdown into Full-time, Part-time, Self-Employed, Unemployment, Other Not Employed. The last two of these are derived from the reported reasons for leaving a job spell.
Stat(2) denotes a uniform breakdown as shown in Table C1.

5.4.2 Spell definition

- Each of Halpin's files has a consistent spell definition
- My file contains a mix of spell types. Job characteristics are measured with more noise for employer spells, since they are measured only once for what may be several different jobs As an indication of the extent of within-employer moves, around a quarter to a third of *wJOBHIST* employment spells are with the same employer as the spell that follows.
- My file contains job characteristics wherever available.

5.4.3 Content differences

- Halpin's newpan file includes derived measures of social class, occupational prestige, and socioeconomic status, whereas mine currently does not.
- Halpin's newpan file contains date information on two definitions of spell: 'splits', which are equivalent to 'job' in the Maré file; and 'episode' which is a spell in which broad employment status (employed, unemployed, other) is unchanged.
- Halpin contains flags to indicate left-censored spells (due to missing start dates)
- My file includes sample selection variables, indicating incompletely observed histories.
- My file includes imputation flags, indicating where and how dates have been altered.

6. Final comments

I see three main contributions of this paper. First, it documents in one place relevant features of the BHPS data files that are used as the source of work-life histories. Second, it documents in a comparative way the different approaches that I and others have taken to generating consistent work histories. Third, it provides documentation for the specific dataset of consistent histories that I have created.

There is much more that could be done!

The major remaining gap in what I have done is that I have not examined what difference the choices that I have made make for particular analyses or estimates, beyond the simple presentation of duration distributions. A more thorough examination was undertaken and reported by Halpin (1997, 2001) and Paull (2002), in addition to their summaries of match or alteration rates. Halpin (1997) gauges the impact of seam codes by estimating a spell hazard model, for all spell types combined. Interview dates not only increase the probability of a transition, but also affect other parameters in the model. Paull (2002) performs an impressively broad range of analyses to gauge the impact of different approaches to deriving consistent histories on various estimates of labour market dynamics.[8] She analyses these impacts in terms of biases that she demonstrates are present in the data (recall bias, and spurious transition bias), and are treated differently under different integration approaches. She examines, for a selection of age/ gender groups, the impacts of different integration approaches on:

- The estimated division of time between different activities;
- The estimated probability of different types of employment status transitions;
- Estimated median survival times for different spell types;
- Estimated accumulated experience;
- Estimated employer tenure; and
- Estimated wage returns to tenure.

The reader is referred to her paper for a very good discussion of the issues.

My final dataset, *wlh.dta*, has not been able to resolve all of the inconsistencies resulting from reporting errors or from trying to reduce people's rich life experiences to a sequence of consecutive non-overlapping spells. The evidence of remaining seam effects, even after applying various deseaming algorithms, show that there are still challenges to be met by researchers wishing to use the data. In particular, the remaining seams will lead to shorter spells and more transitions than occur in actual life histories. As Paull has shown, this is likely to cause (*inter alia*) biases in the analysis of durations and transitions, and possibly overestimates of the return to tenure. The impact is likely to be less damaging for estimates of accumulated experience, apart from the fact that recall bias leads to an understatement of 'atypical' types of experience in retrospective files. (eg: young people tend to forget or relabel employment spells, older men tend to forget or relabel non-employment spells, and women tend to lengthen spells out of the labour force.)

I believe, however, that *wlh.dta* has ironed out, in as light-handed a way as possible, many of the inconsistencies in the base BHPS work-life history data, and represents a great improvement on analysing the original data directly.

It would be useful to know more about the sensitivity of findings based on work history files to different treatments of inconsistencies in dates and spell characteristics. Hopefully this paper (and the availability of the programs that I have used) makes it easier for users of the file to test the robustness of their results to different assumptions and treatments. I look

[8] Paull does not examine the sensitivity of findings to seam code effects that arise due to measurement error in the characteristics of job spells (primarily industry and occupation). Her focus was on employment status and employer spells.

forward to hearing feedback and comments from fellow researchers' about their experiences in using *wlh.dta*.

I hope also that my efforts will make it easier for other researchers to understand and use the BHPS work-life history files, which are a valuable and, in my opinion, underused resource.

7. References

a) Works referred to in the text

Bushnell, Diane. 1998. "Computer-Assisted Occupation Coding," Presentation given to Association for Survey Computing Compstat Satellite Meeting on 'New Methods for Survey Research' 21 – 22 August 1998, Chilworth Manor, Southampton. Available online at http://www.assurcom.demon.co.uk/Events/C98.

Crouchley, R and G. Oskrochi. 2000. "Testing for Origin Dependence in Socio-Economic Duration Data," University of Lancaster Centre for Applied Statistics Working Paper 2000/01. Available online at http://www.cas.lancs.ac.uk/wpapers.html .

ESRC Research Centre on Micro-social Change. 2005. "Data File: British Household Panel Survey; Waves 1-13, 1991-2004," June 2005, SN: 5151, Distributed by UK Data Archive, Colchester, Essex.

Halpin, Brendan. 1997. "Unified BHPS Work-Life Histories: Combining Multiple Sources into a User-Friendly Format," No 13, ISER, University of Essex, Colchester.

Halpin, Brendan. 2000. "BHPS Work-Life History Files, Version 2," mimeo, ISER, University of Essex, Colchester. Available online at UKDA (documentation for study 3954).

Jäckle, Annette and Peter Lynn. 2006. "Dependent Interviewing and Seam Effects in Work History Data," ISER Working Paper, ISER, University of Essex, Colchester.

Martin, J; Bushnell; P Campanelli and R Thomas. 1995. "A Comparison of Interviewer and Office Coding of Occupations," Proceedings of the Survey Research Methods Section, American Statistical Association. Available online at http://www.amstat.org/sections/srms/Proceedings .

Oskrochi, G. and R Crouchley. 2000. "Using SPSS and Gauss to Unify the BHPS Work History Data," University of Lancaster Centre for Applied Statistics Working Paper 2000/03. Available online at http://www.cas.lancs.ac.uk/wpapers.html.

Paull, Gillian. 2002. "Biases in the Reporting of Labour Market Dynamics," IFS Working Paper W02/10, Institute for Fiscal Studies, London. Available online at http://www.ifs.org.uk/publications.php?publication_id=2012.

Taylor, Marcia F.; John Brice; Nick Buck and Elaine Prentice-Lane. 2005. British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices, Colchester: University of Essex [http://www.iser.essex.ac.uk/ulsc/bhps/doc/]. Available online at http://www.iser.essex.ac.uk/ulsc/bhps/doc/.

b) Other works using the BHPS work history files

The following list is a list of papers that I found which have used the BHPS Work-Life History files. I am sure that it is not comprehensive.

Alexandros N. Zangelidis (????), " An Empirical Study of the Contribution of Employer Tenure on an Individual's Earnings Profile", Department of Economics, University of Warwick, Coventry, CV4 7AL, UK.

Arulampalam, W (2001) "Is Unemployment Really Scarring? Effects of unemployment experiences on wages" *Economic Journal* 111, pp F585-F606

Bardasi, E Jenkins, SP (2002). Income in Later Life: Work history matters. mimeo - *Joseph Rowntree Foundation*

Booth A. & Taylor M (1996) "The changing picture of male unemployment in Britain" *Institute for Labour Research Discussion Paper* 96-03, http://www.essex.ac.uk/ilr/discussion/discuss.htm.

Booth, AL Francesconi, M Garcia-Serrano, C (1999). "Job tenure and job mobility in Britain"*Industrial & Labor Relations Review*, 53, 43-70.
Booth, AL & Francesconi, M. (1999) Job Mobility in 1990s Britain: Does Gender Matter?
Contini, B Pacelli, L. & Villosio, C. (1999) "Short Employment Spells in Italy, Germany and Great Britain: Testing the 'Port-of-Entry' Hypothesis" *LSE CEP Discussion paper 426*
Elias, Peter (1997). Who Forgot They Were Unemployed? . ISER Working Paper, 97-19,
Gershuny J. (2002) "A New Measure of Social Position: Social Mobility and Human Capital in Britain" Working Paper 2002-2 [uses 4 years before each survey]
Gershuny, Jonathan & John Brice (1994) 'Looking backwards: family and work 1900 to 1992' Nick Buck, John Gershuny, David Rose & John Scott (eds) *Changing households: the British household panel survey 1990-1992* Essex: ESRC Research Centre on Micro-Social Change.
Halpin, B. (1999) "Is Class Changing? A work-life history perspective on the salariat" Institute for Social and Economic Research University of Essex September 27, 1999
Kan, Man Yee and Gershuny Jonathan. 2006. "Human Capital and Social Position in Britain: Creating a Measure of Wage-Earning Potential From BHPS Data," *ISER Working Paper 2006-03*, ISER, University of Essex, Colchester.
McGinnity, Frances. 2002. "The Labour-Force Participation of the Wives of Unemployed Men: Comparing Britain and West Germany Using Longitudinal Data," *European Sociological Review*, 18, pp. 473-88.
Maite Martinez-Granado , "Self-employment and Labour Market Transitions: a Multiple State Model" Universidad Carlos III de Madrid [*BLIFEMST* ONLY]
Malo, Miguel A. and Muñoz-Bullón. 2006. "Employment Status Mobility From a Life-Cycle Perspective: A Sequence Analysis of Work Histories in the BHPS," *Demographic Research*, 9:article 7. Available online at http://www.demographic-research.org/volumes/vol9/7/9-7.pdf.
Malo, M. (????) "Involuntary job separations and labour mobility from a life-cycle perspective" Universidad de Salamanca (Spain)
Martin, J., Bushnell, D., Campanelli, P. & Thomas, R. (1995) "A Comparison of Interviewer and Office Coding of Occupations" Proceedings of the Survey Research Methods Section, ASA. http://www.amstat.org/sections/srms/Proceedings
McGinnity, F & Hillmert, S. "Social Classes at Risk? Class-specific unemployment in British and German life courses since the 1970s" Max-Planck-Institute for Human Development 9/2002 http://www.mpib-berlin.mpg.de/de/forschung/bag/projekte/PolisBer.pdf
Pollock, G. (1997), "Individualization and the transition from youth to adulthood" http://www.alli.fi/nyri/young/1997-1/artikkelPollock1-97.htm
Scherer, Stefani (2001). Early Career Patterns a Comparison between Great Britain and West Germany. European Sociological Review, 17(2), 119-144.
Stewart, Mark B. 2006. "The Relationship Between the Financial Position of Pensioners and Their Working-Life Earnings Levels," *mimeo*, University of Warwick. Available online at http://www.warwick.ac.uk/res2003/papers/Stewart1.pdf.
Taylor, M. P. (1999) "Survival of the Fittest? An analysis of self-employment duration in Britain", *The Economic Journal*, 109(March), pp. 140-55.
Thomas, Claudia; Michaela Benzeval and Stephen A. Stansfeld. 2005. "Employment Transitions and Mental Health: an Analysis From the British Household Panel Survey," *Journal of Epidemiology and Community Health*, 59, pp. 243-9. Available online at http://jech.bmjjournals.com/cgi/content/full/59/3/243.
Upward, R (1999) "Constructing data on unemployment spells from the PSID and the BHPS" [Appendix to : Haynes, M. Upward, R., and Wright, P. *Smooth and Sticky Adjustment: a Comparative Analysis of the US and UK*. GEP Research Paper 99/07 Leverhulme Centre for research on globalisation and economic policy.]

Appendix AInformal Survey of users of the BHPS Work-Life History data

A1 Summary of Responses

The number of responses was very small (10), so it is difficult to draw too much from the results of the informal survey. This low number does not imply that the work history files are little used. Responses were not received even from all of the 'main' users identified in this paper, or from all of those whose names appear in the references.

- The main reported uses of the work history files were in the analysis of spell durations or transitions, or in deriving measures of accumulated experience.
- The existing work history files were suitable or mostly suitable for most people's needs (7/10). While there were some specific concerns raised, only 2 respondents reported that the files were unsuitable for their needs.
- Half of the users reported creating their own versions of the work history files, although it is not clear whether this meant just creating derived variables or linking the work history files with other files, or whether they returned to the base datasets to create an alternative file of work histories.
- Half of the users used Stata as their main programming tool.

There were a lot of positive comments about the BHPS and about the work-history files in particular. This is despite the fact that there was a question asking explicitly about weaknesses but none asking explicitly about strengths.

A2 QUESTIONS (with summarised responses in boxes):

10 Usable Responses (plus 6 non-users, 3 of whom are planning to use the files in the future)
One (very positive) user may have used only the main BHPS files and not the work-history data, but it is not clear.

1. What type of information did you need from the files (eg: spell durations (transition analysis); accumulated experience measures; detecting occurrences of work-life events, etc.)?

• spell durations/ transitions*	5
• accumulated experience	5
• first spell information	2
• full work histories	1
• migration and mobility	1

* *includes job tenure/ unemployment duration/ employer tenure*

2. Were the BHPS Work history files suitable for your needs? Y / N

• Yes	5
• Yes with small exceptions	2
• Not entirely	1
• No	2

a) What (if any) major weaknesses did you have to deal with?

summarised below with other comments

ii) did you use the standard work history files for your analysis? Y / N
iii) did you create your own versions of work or life history files? Y / N

		Used own dataset	
		Yes	*No*
Used std dataset	*Yes*	3	4
	No	2	1 (prob NY)

What programs did you use (SPSS/ Gauss/ Stata/ SAS/ other)

Stata	5	SAS	2	SPSS	2	TDA	1

Are you willing to share your programs? Y / N

Yes	4
Possibly (suggested sharing files)	1
Yes, if funded to document them	1
No	2

3. Other Comments (continue on separate sheet if necessary):

Strengths

- Excellent and well-constructed dataset. We have had little difficulty in using the file and have two papers accepted for publication on residential mobility and migration.
- It is very helpful to have [standard files] as they can a) save a lot of work, and b) (more importantly) [permit comparability]
- The strength of the work-history files is in combining the two retrospective parts.
- I really appreciated the fact that Halpin made the effort combining these datasets although it did not perfectly match with what I was wishing to have
- We are really lucky with the BHPS.
- the vector of activity start dates with parallel activity codes, industries, etc is the most useful data structure as this can be used for both duration analyses and measures of accumulated experience . . . a data file of this format should be retained.
- these data are extremely important and should be used more by researchers.
- This survey (and any consequent conclusions) is obviously a very useful step forward to finding some sort of general guidelines!

Experiences

- We just used the files (and asked Brendan for help during the process). We did not check whether the codings in the original files were OK. We assumed that they were.
- I started combining the two retrospective parts by myself but finally used the second version of the Halpin files.
- Mean accumulated experience [from a user-derived file] turned out to be pretty close to those derived from the Halpin files.

Weaknesses

- The durations were too short - it appeared that spells were divided by a change in job characteristic (occ/ ind) that introduced spurious transitions either through arbitrary changes in definitions/ descriptions in the data or genuine random measurement error;
- I discussed the problems with several other users . . . and, at least for [looking at work experience and job tenure] it was quickly realised that we could not use the archive data.
- employer tenure info inconsistent with LFS info
- gaps and overlaps in histories
- missing duration values
- The data was useful for experience measures but not for duration analysis or employer tenure.
- I wouldn't be entirely sure I'd use [the two retrospective files] when drawing on the information from the panel and focusing just on transitions rather than sequences.

Unavoidable problems

- One reason why users may not [use the files] may be the complex data structure that needs some advanced SPSS or other statistical package knowledge to access.
- the BHPS collected employment/ family status and job history in two different waves (and of course they don't match perfectly - but it is solved in the combined files);
- [the BHPS] uses different definitions of job spells in the panel part and the retrospective part - which will remain as a problem.
- [unspecified] problems with pre-panel data meant they were unusable

Linking

- I would have liked to have additional information about single job spells coming from the panel part but this is in principle no problem as you can link them back.
- I had to match the work-life histories to the individual characteristics contained in the main dataset

Other comments

- Having a pre-prepared file, it is important to have transparent documentation of the underlying steps and that it is possible to link back to the original datasets and to potentially add information from there.

A3 Cover letter

BHPS Work-Life History files

I am contacting you as a potential user of the BHPS Work-History files that are deposited at the UK Data archive (SN3954; Halpin (1997, 2000)). If you have not used these files, please disregard this message. My apologies for the interruption

If you have used these files, I am seeking information from you about your use of the files, their suitability for your purposes, and ways of further facilitating the use of BHPS work-life history data.

My name is Dave Mare and I am currently a visiting research fellow at the Institute for Social and Economic Research (ISER) at the University of Essex. I too am a user of the work-history files. The aim of the standard work-history files was to facilitate use. However, I ended up returning to the base data and creating new work history files because of difficulties I faced when using the standard files. Others here at ISER and elsewhere (Oskrochi & Crouchley (2000); Paul (2002)) have done likewise.

I doubt whether any single file can meet all users' needs. It may, however, be possible to further facilitate use by other researchers through some sharing of programs or documentation, or through amending the form of the 'standard' dataset. My colleagues here at ISER are interested in hearing about the feedback that I get from other users, to see if there is a need to change the form or documentation of the work history files.

I would appreciate it if you would answer the questions that I have attached as a Word file, and return your responses to me:

Email: iserg011@essex.ac.uk
Mail: Dave Mare, ISER, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom
or Fax: +44 (0) 1206 873151

I have tried to keep the questions short. Please feel free to give additional comments or suggestions.

Thanks

Dave Mare

Appendix BSelected Extracts from BHPS Documentation

B1 Record Type BINDRESP

B1.1 (Selected) w*INDRESP* Questions

D17. **SHOWCARD D2**
Please look at this card and tell me which best describes your current situation? ***wJBSTAT***

Self employed	01	Looking after family or home	06
In paid employment		Full-time student/ at school	07 **GO TO D19**
(full or part-time)	02 **ASK D18**	Long term sick or disabled	08
Unemployed	03	On a government training scheme	09 **ASK D18**
Retired from paid work altogether	04	Something else **(PLEASE GIVE DETAILS)**	10
On maternity leave	05		

E1 . Can I just check, did you do any paid work last week -that is in the seven days ending last Sunday - either as an employee or self employed? ***wJBHAS***

Yes	1 **GO TO E4**
No	2 **ASK E2**

E2. Even though you weren't working did you have a job that you were away from last week? ***wJBOFF***

Yes	1 **ASK E3**
No	2 **GO TO E83 (page 47)**
Waiting to take up job	3 **GO TO E95 (page 50)**

E3 . What was the main reason you were away from work last week? ***wJBOFFY***

Maternity leave	01
Other leave/holiday	02
Sick/injured	03
Attending training course	04
Laid off/on short time	05
On strike	06
Other personal/family reasons**(GIVE DETAILS)**	07
Other reasons **(GIVE DETAILS)**	08

Prior to Wave 9(I)

E4. Is your current job ***(READ OUT) wJBTERM***

A permanent job	1
A seasonal, temporary or casual job	2
Or a job done under contract or for a fixed period of time	3

From Wave 9(I)

*E4. Leaving aside your own personal intentions and circumstances, is your job . . .****READ OUT wJBTERM1 .***

A permanent job	1 **GO TO E5**
or Is there some way that it is not permanent?	2 **ASK E4a**

*E4a. In what way is the job not permanent, is it . . .****READ OUT***
CODE ONE ONLY
wJBTERM2 .

Seasonal work	1
Done under contract for a fixed period or for a fixed task	2
Agency temping	3
Casual type of work	4

or Is there some other way that it is not permanent? **(SPECIFY)**

E5 . What was your (main) job last week? Please tell me the exact job title and describe fully the sort of work you do.
IF MORE THAN ONE JOB: MAIN = JOB WITH MOST HOURS.
IF EQUAL HOURS: MAIN JOB = HIGHEST PAID

ENTER JOB TITLE: _____

DESCRIBE FULLY WORK DONE: (IF RELEVANT `WHAT ARE THE MATERIALS MADE OF?') ***wJBSOC***

E7. Are you an employee or self-employed? ***wJBSEMP***
Employee 1 **ASK E8**
Self-employed 2 **GO TO E73 (page 53)**

J6. **SHOWCARD J1**
Please look at this card and tell me which best describes your current situation? ***wNEMST***
NON EMPLOYED ONLY

Unemployed	03	Full-time student/ at school	07
Retired from paid work altogether	04	Long term sick or disabled	08
On maternity leave	05	On a government training scheme	09
Looking after family or home	06	Something else **(PLEASE GIVE DETAILS)**	10

J7. On what date did your present spell of being **(CODE AT J6)** begin?
Day Month Year
WRITE IN:
wCJSBGD . wCJSBGM . wCJSBGY4 .
IF DON'T KNOW DAY OR MONTH ENTER `98' AND CODE YEAR
IF DON'T KNOW YEAR ENTER `9998'

B1.2 Variable List (Selected Fields only)

PID	Cross-wave person identifier
wHID	Household identification number
wPNO	Person number
wDOID	Date of interview: day
wDOIM	Date of interview: month
wDOIY	Date of interview: year
wDOBM	Month of birth
wDOBY	Year of birth
wSEX	Sex
wJBSTAT	Current economic activity
wEDENDM	Month left education institution
wEDENDY	Year left education institution
wSCEND	School leaving age
wFEEND	Further education leaving age
wJBHAS	Did paid work last week
wJBOFF	No work last week but has job
wJBOFFY	Reason off work last week
wJBTERM*	Current job: permanent or temporary
wJBSOC	Occupation (SOC): current main job
wJBSIC	Industry (SIC) of employer: current job
wJBSEMP	Employee or self-employed: current job
wJBSIZE	No. employed at workplace: current job
wJBHRS	No. of hours normally worked per week
wJBOT	No. of overtime hours in normal week
wJBOTPD	No. of hours worked as paid overtime
wJBBGD	Day started current job
wJBBGM	Month started current job
wJBBGY	Year started current job
wJBBGLY	Started job after 1.9.91
wJBSECT	Employing organisation: current job
wJBONUS	Pay includes bonuses or profit share
wJBRISE	Pay includes annual increments
wTUJBPL	Union or staff association at workplace
wTUIN1	Member of workplace union
wTUIN2	Member of any trade union or similar
wPAYS	Starting pay, job start after 1.9.91
wPAYSW	Starting pay period (weeks): current job
wPAYSG	Starting pay gross or net: current job
wPAYLY	Usual pay at 1.9.91
wPAYLYW	Usual pay at 1.9.91: pay period (weeks)
wPAYLYG	Usual pay at 1.9.91:gross/net deductions
wPAYGYR	Total gross earnings in last 12 months
wPAYGYA	Accuracy of Total gross earnings
wJSBOSS	S/emp: hires employees
wJSSIZE	S/emp: number of employees
wJSHRS	S/emp: hours normally worked per week
wJSHRLK	S/emp: preference over hours worked
wJSTIME	S/emp: times of day usually work
wJSTYPE	S/emp: nature of employment
wJSPL	S/emp: work location
wJSBGD	S/emp: day started present job
wJSBGM	S/emp: month started present job
wJSBGY	S/emp: year started present job
wJ2HAS	Has a second paid job
wJ2SOC	Occupation (SOC), second job
wJ2SEMP	Employee or self employed, second job
wJ2HRS	No. of hours worked per month, 2nd job
wJ2PAY	Gross earnings from 2nd jobs last month
wNEMST	Current labour force status
wCJSBGD	Day current labour force status began
wCJSBGM	Month current labour force status began
wCJSBGY	Year current labour force status began
wCJSBLY	IC:current l.f. spell began after 1.9.91
wIVFIO	Individual interview outcome
wJBFT	Employed full time
wPAYG	Gross rate of pay per month:last payment
wPAYN	Net rate of pay per month: last payment
wPAYGU	Usual gross pay per month: current job
wPAYNU	Usual net pay per month: current job
wPAYGTY	Usual monthly gross pay: Sept this year
wPAYGLY	Usual monthly gross pay: Sept year ago
wPAYNTY	Usual monthly net pay: Sept this year
wPAYNLY	Usual monthly net pay: Sept year ago
wJSPROF	Monthly self employed profit
wJSPAYG	Monthly self employed gross pay
wCJSTEN	Length (days) of current labour market
wCJSWK9	Weeks in current spell: year to 1.9.92
wJLID	Identifier of latest job
wJBSTATL	Employment status on Sept 1, year ago
wJBSTATT	Employment status on Sept 1, this year
wLRWGHT	Longitudinal respondent weight
wLEWGHT	Longitudinal enumerated individual weight
wXRWGHT	X-sectional respondent weight
wXEWGHT	X-sectional enumerated individual weight
wDOIY4	Date of interview: 4 digit year
wJBBGY4	Year started current job: 4 digit
wCJSBGY4	Year labour force status began: 4 digit

B2 Record Type wJOBHIST

B2.1 wJOBHIST Questions

READ OUT
I'd like to ask you a few questions now about what you might have been doing since September 1st last year in the way of paid work, unemployment, or things like time spent retired or looking after your family. We can use this calendar to help sort it out.
HAND CALENDAR
As we need to get as complete a picture as possible I'd like you to tell me about any spells you may have had in or out of paid employment, even if they were just a few days when you were waiting to take up another job. I'd also like you to tell me about any changes that might have happened while you were working like getting promoted or starting a different job with the same employer.
I'll start by asking what you were doing immediately before {your current job which you told me earlier started on (DATE J8)/your current spell of (CODE J9) which started on (DATE J10)}

J12 . **SHOWCARD 26**
Can you look at this card please and tell me which of the descriptions comes closest to what you were doing immediately before then?
ENTER CODE FROM SHOWCARD 26 ON GRID AT J12

SHOWCARD 26

01 Doing a different job for the same employer 02 Working for a different employer In paid employment (not self employed) Working for myself (self employed) 03 Unemployed/looking for work 04 Retired from paid work altogether 05 On maternity leave	06 Looking after family or home 07 In full-time education/student 08 Long term sick or disabled 09 On a government training scheme 10 Something else **(PLEASE GIVE DETAILS)**

J13 . And on what date did you start doing that?
IF DON'T KNOW DAY OR MONTH, ENTER `98'

J14 . **IF `PAID EMPLOYMENT' (01 or 02): RECORD BRIEF JOB TITLE**
IF `SOMETHING ELSE' (10): RECORD BRIEF DETAILS

- Continue with J12 - J14 until date reported at J13 is September 1st 199LY or before

	J12	J13			J14
SPELL NO BEFORE CURRENT STATUS	STATUS CODE (FROM CARD 26)	DATE SPELL BEGAN (DON'T KNOW = 98)			IF `EMPLOYMENT' (01 or 02): ENTER BRIEF JOB TITLE IF `SOMETHING ELSE' (10) WRITE IN DETAILS
		DAY	MONTH	YEAR	
1	☐☐	☐☐	☐☐	☐☐	
2					

For Job Spells:

	J16	J 17
Spell Number	**WRITE IN** Transfer details for relevant spell no. of status code and	Could you give me some details of the job which you started in **(DATE at J1**6). Please tell me the exact job title and describe fully the

	date began (Month/year only)	sort of work you did. **NB. IF MORE THAN ONE JOB, MAIN = MOST HOURS IF EQUAL HOURS THEN HIGHEST PAID**
wJSPNO	STATUS CODE (01/02) ***wJHSTAT*** MONTH YEAR ***wJHBGM*** ***wJHBGY***	ENTER JOB TITLE: ***wJHSOC*** DESCRIBE FULLY WORK DONE: (AND MATERIALS USED, IF RELEVANT)

J18	J19	J20	J21	J22
CHECK STATUS CODE (J16)	Were you a full-time employee a part-time employee or self-employed?	Did you have any employees?	**SHOWCARD 27** Which of the types of organisation on this card did you work for?	Did you have any managerial duties, or did you supervise any other employees?
STATUS CODE (01/02) **0** ***wJHSTAT*** 0**1 GO TO J22** 0**2. ASK J19**	***wJHSEMP*** F/T Employee .....1 **GO TO J21** P/T Employee ..... 2 **GO TO J21** Self-Employed ..... 3 **ASK J20**	***wJHBOSS*** Yes ..... 1 No ..... 2 **NOW GO TO J25**	***wJHSECT*** Private firm 01 Civil Service 02 Local Government 03 NHS 04 Nationalised Ind 05 Non-Profit Org. 06 Armed Forces 07 Other **(SPECIFY)** 08	***wJHMNGR*** Manager 1 Foreman/ superv 2 **NOT** manager or supervisor 3

	J23	J24	J25	J26	J27
Spell No. **WRITE IN**	**CHECK STATUS CODE (J16)**	Was this job at the same workplace as the job you told me about before?	What did the firm/organisation you worked for actually make or do? **DESCRIBE FULLY**	How many people (were employed/ did you employ) at the place where you worked?	Please think back to September 1st last year or to when you started that job, if that is more recent. At that time, how much were you usually paid?
wJSPNO	STATUS CODE (01/02) ***wJHSTAT*** 01 **...ASK J24** 02 **GO TO J25**	***wJHPLDF*** Same.....1 **GO TO J27** Different .....2 **ASK J25**	***wJHSIC***	***wJHSIZE*** 1-2 01 3-9 02 10-24 03 25-49 04 50-99 05 100-199 06 200-499 07 500-999 08 1000 or more 09 DK, fewer than 25 10 DK, 25 plus 11	**ENTER TO NEAREST £** ***wJHPAYL*** **ASK J28** Don't know 8 Refused 9 **GO TO J30**

J28	J29	J30	J31
How long a period did that cover?	And was that before or after any deductions for tax, national insurance, union dues and so on or were there no deductions at all usually made from your pay?	**SHOWCARD 28** Would you look at this card please and tell me which of the statements on the card best describes why you stopped doing that job?	**FOR SPELL ONE ONLY IF J30 IS 02 ASK:** What was the main thing about your present job that attracted you to it? **OTHERS RETURN TO J16 IF THERE ARE MORE JOB SPELLS TO COMPLETE**
wJHPYLW Week.....1 Fortnight.....2 Four Weeks.....3 Calendar Month .....4 Year.....5 Other.....6 **(WRITE IN)** OFFICE CODE	***wJHPYLG*** Before deduction 1 After 2 No deductions 3 Don't know 8	***wJHSTPY*** Promoted 01 Left for better job 02 Made Redundant 03 Dismissed/sacked 04 Temporary job ended 05 Took retirement 06 Health reasons 07 Left to have baby 08 Look after family 09 Look after other person 10 Other reason 11	**WRITE IN:** ***wJBLKY*** OFFICE CODE

B2.2 wJOBHIST Variables

BHID	Household identification number
BPNO	Person number
BJHSTAT	Labour force status code
BJHBGD	Day labour force spell began
BJHBGM	Month labour force spell began
BJHBGY	Year labour force spell began
BJSPNO	Spell number of labour force status
BJHSOC	Occupation (SOC): previous job

BJHSEMP	Employee or self-employed: previous job
BJHBOSS	Had employees: previous job
BJHSECT	Employing organisation: previous job
BJHMNGR	Managerial duties: previous job
BJHPLDF	Change of workplace: previous job
BJHSIC	Industry (SIC): previous job
BJHSIZE	No. employed at workplace: previous job
BJHPAYL	Pay at 1.9.91: previous job
BJHPYLW	Pay period (weeks) 1.9.91: previous job
BJHPYLG	Pay gross or net at 1.9.91: previous job
BJHSTPY	Reason for stopping previous job
BJBLKY	Attraction of present job
BJHENDD	Day labour force spell ended
BJHENDM	Month labour force spell ended
BJHENDY	Year labour force spell ended
BJHA9LY	Whether job started after 1.9.91
BJHSEG	Socio economic group: previous job
BJHGOLD	Goldthorpe Social Class: previous job
BJHRGSC	RG Social Class: previous job
BJHISCO	International SOC : previous job
BJHCSSM	Cambridge Scale males : previous job
BJHCSSF	Cambridge Scale females : previous job
BJHHGS	Hope - Goldthorpe Scale : previous job
BJHSPW	Weeks in year to 1.9.92: prev job spell
BJHGPAY	Monthly Gross Pay: previous job
BJHNPAY	Monthly Net Pay: previous job
BJHGPAYI	Imputation flag - BJHGPAY (derived)
BJHNPAYI	Imputation flag - BJHNPAY (derived)
BJHBGY4	Year labour force spell began: 4 digit
BJHENDY4	Year labour force spell ended: 4 digit

B3 Record Type BLIFEMST

B3.1 *BLIFEMST* Questions

L50 . ***SHOWCARD 18***
Which description on this card comes closest to what you first did after leaving full-time education?

SHOWCARD 18 [recorded as variable BLESHST]

01	Self-employed	07	Looking after family or home
02	Full-time paid employment	08	Full-time Student/At School
03	Part-time paid employment	09	Long term sick or disabled
04	Unemployed	10	On a government training scheme
05	Retired from paid working altogether	11	National Service/War Service
06	Maternity leave	12	Something else (please give details)

L51 . When was the next time your situation changed, I mean a change to one listed on the card, even if only for a month or two.
WRITE IN MONTH AND YEAR. IF MONTH NOT KNOWN, CODE SEASON:

IF STATUS STILL APPLIES CIRCLE `Not Ended' AND GO TO E1 (page 41)

L52 . Which of the descriptions on the card best describes what you did next, even if it was only for month or so.
ENTER UNDER `Status Code' ON GRID

REPEAT L51 AND L52 UNTIL THE STATUS THAT IS APPLICABLE TODAY IS REACHED. THEN CIRCLE `Not Ended' AND GO TO E1 (page 41).

B3.2 *BLIFEMST* Variables

BHID	Household Identification number
BPNO	Person number
BLESHNO	Lifetime employment history event number
BLESHST	Lifetime employment history status
BLESHEM	Month lifetime emp. hist. status changed
BLESHEY	Year lifetime emp. history status changed
BLESHNE	Lifetime emp. history status not ended
BLESHSM	Month lifetime emp. hist. status started
BLESHSY	Year lifetime emp. history status started
BLESLEN	Length of emp. history spell (months)
BLESHEY4	Year emp. history status changed: 4 digit
BLESHSY4	Year emp. history status started: 4 digit

B4 Record Type CLIFEJOB

B4.1 *CLIFEJOB* Questions

I'd like to ask you a few questions about any jobs you have done, since leaving full-time education. I will ask you when you changed jobs, by that I mean when you started working for a different employer. I am interested in all the jobs you have had lasting one month or more. Can I just check, have you ever had a paid job at all other than your current job? By that I mean either a full-time or part-time job that lasted at least one month.

L5	L6	L7	L8
ESF SpellNumber	In your first/next job, were you self-employed, a full-time employee or a part-time employee?	On what date did you begin (your first paid job/working with your next employer)? **IF DATE BEGAN ON OR SINCE September 1st 1990 GO TO V1 (page 71)**	Could you give me some details of the exact job you started in **(DATE AT L7)**. Please tell me the exact job title and describe fully the sort of work you did. And what did the firm or organisation you worked for actually make or do? **IF MORE THAN ONE JOB, MAIN = MOST HOURS. IF EQUAL HOURS THEN HIGHEST PAID**
Spell Number ***CLJESFN***	***CLJSEMP*** Self Employed 1 F/T Employee 2 P/T Employee 3	Month Year ***CLJBGM*** ***CLJBGY*** **IF MONTH NOT KNOWN ENTER SEASON CODE** Winter (13) Spring (14) Summer (15) Autumn (16)	Job Title/Work done: ***CLJSOC*** Nature of business: ***CLJSIC***

L9	L10	L11	L12	L13
Did you supervise other employees? **INCLUDE SELF-EMPLOYED**	Was this job a permanent job, seasonal job, temporary or casual job or a job done under contract for a fixed period of time?	What was the date you left working with that employer? **SHOWCARD 29**	Which of the reasons on the card best describes why you stopped doing that job?	**ASK IF NEEDED** Have you had any more paid jobs since then? **PROBE FOR WHETHER PRESENT EMPLOYER IF L12 = 01 or 02 CODE 1 or 2, AS APPROPRIATE**
CLJMNGR Yes ..... 1 No..... 2 Don't know 8	***CLJTERMR*** a permanent job 1 a seasonal, temp or casual job 2 contract job for fixed period of time 3	Month Year ***CLJLFTM*** ***CLJLFTY*** **IF MONTH NOT KNOWN ENTER SEASON CODE** Winter (13) Spring (14) Summer (15) Autumn (16)	***CLJYLFT*** Left for better job (promoted) 01 Left for different job 02 Made redund/ comp bankrupt 03 Dismissed/sacked 04 Temporary job ended 05 Took retirement 06 Health reasons 07 Left to have baby 08 Look after family/other 09 Other reason **(SPECIFY)** 10	***CLJOTHJ*** Yes, not present employer ..... 1 Yes, with present employer ..... 2 No..... 3

B4.2 *CLIFEJOB* Variables

CHID	Household Identification number
CPNO	Person number
CLJESFN	Employment status spell number
CLJSEMP	FT, PT or s/employed-lifetime
CLJBGM	Month job spell began-lifetime
CLJBGY	Year job spell began-lifetime
CLJSOC	Occupation (SOC)-lifetime
CLJSIC	Industry (SIC) of employer-lifetime
CLJMNGR	Supervised other employees - lifetime
CLJTERM	Perm. or temp. contract - lifetime
CLJLFTM	Month left employer - lifetime
CLJLFTY	Year left employer - lifetime
CLJYLFT	Reason for leaving job - lifetime
CLJOTHJ	Had further paid jobs - lifetime
CLJSEQ	Job spell number
CLJSEG	Socio economic group - lifetime
CLJGOLD	Goldthorpe social class - lifetime
CLJRGSC	RG social class - lifetime
CLJISCO	Internatonal SOC: previous job
CLJCSSM	Cambridge Scale males - lifetime
CLJCSSF	Cambridge Scales females - lifetime
CLJHGS	Hope-Goldthorpe scale - lifetime
CLJLEN	Length (months) of job history spell
CLJENST	End status of job history spell
CLJBGY4	Year job spell began-lifetime: 4 digit
CLJLFTY4	Year left employer - lifetime: 4 digit

Appendix CComparative variable coding Schedules

Table C1 : Employment Status - Coding Variations

DATA SET	*wINDRESP*					*wJOBHIST*			*BLIFEMST CLIFEJOB*		*Maré uniform code*
Variable name	*wjbstat* Wave One	*wjbstat* From Wave 2	*wjbsemp*	*wjbft*	*wnemst*	*wjhstat*	*wjhsemp* Wave One	*wjhsemp* from Wave 2	*bleshst*	*cljsemp*	*status*
FT Self employed	1	1	2	1	n/a	2	2	3	1	1	0
PT Self employed	1	1	2	2	n/a	2	2	3	1	1	0
FT – same employer	2	2	1	1	n/a	1	1	1	2	2	1
FT – Different employer	2	2	1	1	n/a	2	1	1	2	2	1
PT – same employer	2	2	1	2	n/a	1	1	2	3	3	2
PT – Different employer	2	2	1	2	n/a	2	1	2	3	3	2
Unemployed	3	3	n/a	n/a	3	3	n/a	n/a	4	n/a	3
Retired	4	4	n/a	n/a	4	4	n/a	n/a	5	n/a	4
On maternity leave	8	5	n/a	n/a	5	5	n/a	n/a	6	n/a	5
Family care	5	6	n/a	n/a	6	6	n/a	n/a	7	n/a	6
Full time student	6	7	n/a	n/a	7	7	n/a	n/a	8	n/a	7
Long term sick/disabled	7	8	n/a	n/a	8	8	n/a	n/a	9	n/a	8
Govt training scheme	9	9	n/a	n/a	9	9	n/a	n/a	10	n/a	9
National / War Service	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	11	n/a	11
Something else	10	10	n/a	n/a	10	10	n/a	n/a	12	n/a	10
Missing or Wild	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9	-9
Inapplicable	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8	-8
Proxy respondent	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7	-7
Refused	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2	-2
Not answered/ Don't know/ Can't remember	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Table C2 : Employment Status – Coding from Halpin files

DATASET	*NEWPAN*			*XLJOBE*	*XLEMPE*	*LJEMPE*
Variable Name	*empstat*	*seftpt*	*djse*	*stnemp*	*stempl*	*stemp*
FT Self employed	1	3	2	0	1	1
PT Self employed	1	3	2	0	1	1
FT – same employer	2	1	1	1	2	2
FT – Different employer	2	1	2	1	2	2
PT – same employer	2	2	1	2	3	3
PT – Different employer	2	2	2	2	3	3
FT/PT unknown – same employer	2	5	1		2.5	2.5
FT/PT unknown – different employer	2	4	2		2.5	2.5
Same/ different employer unknown			3			
Unemployed	3	43	n/a	3	4	4
Retired	4	4	n/a	4	5	5
On maternity leave	**8**	**5**	n/a	4	6	6
Family care	**5**	**6**	n/a	4	**7**	**7**
Full time student	**6**	**7**	n/a	4	**8**	**8**
Long term sick/disabled	**7**	**8**	n/a	4	9	9
Govt training scheme	9	9	n/a	4	10	10
National / War Service	n/a	n/a	n/a	4	11	11
Something else	10	10	n/a	4	12	12
?????????				5		
Missing or Wild	-9	-9	n/a	-9	-9	-9
Inapplicable	-8	-8	-8	-8	-8	-8
Proxy respondent	-7	n/a	n/a	-7	-7	-7
Refused	n/a	n/a	n/a	-2	-2	-2
Not answered	n/a	-1	n/a	-1	n/a	n/a
Don't know	-1	n/a	n/a	n/a	-1	n/a
Can't remember	n/a	n/a	n/a	n/a	n/a	-1

Table C3 : Permanent or Temporary employment - Coding Changes

*Variable: wJBTERM**	*Waves 1-8*	*From Wave 9*	*From Wave 9*	*Maré Uniform Coding*
	wJBTERM	*wJBTERM1*	*wJBTERM2*	
Permanent	1	1	n/a	1
Seasonal	2 "seasonal, temporary or casual"	2	1	2
Done under contract for a fixed period or for a fixed task	3	2	2	3
Agency temping	No specific code	2	3	4
Casual type of work	2 "seasonal, temporary or casual"	2	4	2
Other non-permanent? (Specify)	No specific code	2	5	4

Table C4 : Differences in coding of reasons for leaving a job

wJOBHIST (Variable wJHSTPY)	*CLIFEJOB (Variable CLJYLFT)*	*Maré uniform Coding*
1 Promoted		1
2 Left for better job	1 Better job	1
	2 Different job	1
3 Made redundant	3 Redundancy	3
4 Dismissed or sacked	4 Dismissed	4
5 Temporary job ended	5 Temp job ended	5
6 Took retirement	6 Retired	6
7 Stopped health reas	7 Health reasons	7
8 Left to have baby	8 To have baby	8
9 Children/home care	9 Family care	9
10 Care of other person		11
11 Other reason	10 Other	11
	11 War/Nat Service	11
12 Moved away (from Wave 5)	12 Moved from area	11
13 Start College of Uni (from Wave 5)	13 F/T education	11
-9 Missing or wild	-9 Missing or wild	-9
-8 Inapplicable	-8 Inapplicable	-8
	-2 Refused	-2
	-1 Don't know	-1

Appendix DDetailed description of my reconciliation of *BLIFEMST* and *CLIFEJOB* spells

Table D2 summarises the steps that I took to achieve a consistent WLH segment from the *BLIFEMST* and *CLIFEJOB* records. It shows the extent of date issues that needed to be addressed, and how I addressed them, indicating where I chose to drop observations, or assign dates that contradicted the responses in one or other of the sources.

The columns of Table D2 indicate the number and type of inconsistencies that exist in the dataset at each stage of reconciliation. Note that missing dates do not contribute to the counts of conflicts, gaps, or overlaps. The measures shown are:

- End and Start Conflicts: These provide a count of start dates and end dates that are 'inconsistent'. Inconsistent start dates are in spells that start after they end, or start after the start or end of a subsequent spell. Inconsistent end dates are in spells that end before they start, or end before a prior spell starts or ends. By this measure, a single wrong date can generate multiple instances of both start and end conflicts. The measure is thus an indicative summary that something is wrong.
- Negdur conflicts: This is a count of spells where the start date is later than the end date
- Overlap in Bspell: This is a count of overlaps between spells in the same underlying *BLIFEMST* spell.
- Overlap in consecutive spell: This is a count of overlaps between spells in consecutive *BLIFEMST* spells.
- Overlap in other spells: This is a count of overlaps between spells in different non-consecutive *BLIFEMST* spells.
- Gap in Bspell: This is a count of gaps between spells in the same underlying *BLIFEMST* spell.
- Gap in consecutive spell: This is a count of gaps between spells in consecutive *BLIFEMST* spells. (Gaps between non-consecutive spells are legitimate as a result of non-employment *BLIFEMST* spells.)
- Add obs: Indicates the number of observations that have been added to the dataset.
- Drop obs: Indicates the number of observations that have been dropped from the dataset.
- Move Start and End : This is a count of start or end dates that have been allocated in a way that contradicts the originally reported date.

The following sections labelled a) to r) summarise the adjustments made to the *BLIFEMST* and *CLIFEJOB* records.

a) Read in base data

CLIFEJOB contains 32,773 observations on 7,074 individuals.

There are 116 cases where start dates and end dates are 'inconsistent'. Inconsistent start dates are in spells that start after they end, or start after the start or end of a subsequent spell. Inconsistent end dates are in spells that end before they start, or end before a prior spell starts or ends. By this measure, a single wrong date can generate multiple instances of both start and end conflicts. The measure is thus an indicative summary that something is wrong. Partially and totally missing dates are at this stage treated as valid, and do not contribute to the count of inconsistent spells.

There are 113 cases of overlapping spells. Thirteen occur between *CLIFEJOB* spells that fall within the same *BLIFEMST* (employment status) spell. Twenty-four occur between spells from consecutive *BLIFEMST* spells and a further 76 occur between *CLIFEJOB* spells from non-consecutive *BLIFEMST* spells.

There is a large number of gaps between *CLIFEJOB* spells. Gaps that occur between non-consecutive *BLIFEMST* spells are not shown here, since they will legitimately arise where a non-employment *BLIFEMST* spell occurs. The 1582 intra-*BLIFEMST*-spell gaps and the 930

gaps between *CLIFEJOB* spells from consecutive *BLIFEMST* spells thus represent gaps in what should be continuous accounts of activity.

b) Assign months

The first step is to assign partially missing months. This is done 'within range', and the minimum, guessed, and maximum feasible values are remembered. The counts of inconsistencies shown in Table D2 are inconsistencies that could not be resolved by adjusting dates within their feasible ranges, although this process of nudging is not carried out until later (see below).

Table D1 : Allocation of partially missing months

Month	*Minimum feasible*	*Guess*	*Maximum feasible*
Winter	November of prior year	January	March
Spring	February	April	June
Summer	May	July	September
Autumn	August	October	December
Missing	January	June	December

Note: The same allocation rule is used for other data sets.

c) Renumber intra-spell C-only spells

Some spells in *CLIFEJOB* have not been linked with a corresponding *BLIFEMST* spell *CLJESFN*=0). In some instances, the timing of these '*CLIFEJOB*-only' spells would put them between two matched *CLIFEJOB* spells from the same underlying *BLIFEMST* spell, the '*CLIFEJOB*-only' spell is renumbered so that it is matched to that *BLIFEMST* spell. 127 such changes are made. The mix of inconsistencies changes as the spells are reordered.

d) Merge BLIFEMST & sort

The full extent of the challenges becomes apparent when the *BLIFEMST* data are merged. They are merged by person ID and *BLIFEMST* spell number, which is also included on the *CLIFEJOB* file as *CLJESFN*. The combined dataset contains 55,688 observations on 9,050 individuals. 34 observations are dropped – 29 because they were '*CLIFEJOB*-only' spells with all date information missing, and 5 observations were dropped from the history for the respondent with PID=11111828 due to irresolvable inconsistencies.

The sorting of the spells in the merged dataset is a crucial step in deciding whether the implied history is a consistent one. In principle, spells should be easily ordered by *BLIFEMST* spell number, and within that, by *CLIFEJOB* spell number (*CLJSEQ*). Problems arise for the '*CLIFEJOB*-only' spells. Their order among the *CLIFEJOB* spells can be seen from the spell number but it is not always clear whether they fall before or after particular *BLIFEMST* spells. In many cases, the sequence number of '*CLIFEJOB*-only' spells suggests that they occur after other spells, although their dates clearly show that they occurred much earlier. Furthermore, it is not possible to rely on either start date or end-date alone, since either may be missing

In some instances, '*CLIFEJOB*-only' spells have missing start and end dates. In such cases, date information from sequentially prior or subsequent spells (as identified by *CLJSEQ*) is used to bound the dates, and the '*CLIFEJOB*-only' spells are moved as a block.

Note that the level of agreement between the two data sources appears to be slightly lower than in Halpin's (1997, s. 4.4.1) monthly comparison of *BLIFEMST* and *CLIFEJOB*. Here there are over 4,200 conflicts in 55,688 observations. Some of these are, however, due to duplicates and overlaps that would have been resolved in the process of creating Halpin's two monthly series. When focusing on resolving inconsistencies, it is easy to lose sight of the fact that there is agreement for over 80% of the spells!

e) Drop final duplicates

In 734 cases, the final *BLIFEMST* spell commenced at the same time as the final '*CLIFEJOB*-only' spell (the feasible ranges overlap). The *BLIFEMST* spells are dropped.

f) Accept 1st & last C-only spells

'*CLIFEJOB*-only' spells that fall totally before any merged spells, or totally after any merged spells are retained, and numbered so that they appear in the appropriate position.

g) Remove duplicates

In 282 cases, *BLIFEMST* spells exactly coincide with '*CLIFEJOB*-only' spells (the feasible ranges overlap for both their start dates and their end dates). The *BLIFEMST* spells are dropped.

h) Renumber spells & pad

In many cases, one or more '*CLIFEJOB*-only' spells fall entirely within the period covered by an unmatched *BLIFEMST* spell. The '*CLIFEJOB*-only' information is used, except where it does not fully account for the period spanned by the underlying *BLIFEMST* spell. I fill gaps in *BLIFEMST* spells by creating new subspells with the same characteristics as the underlying *BLIFEMST* spell. (This also accommodates *CLIFEJOB* employment spells within *BLIFEMST* unemployment spells). The net change in the number of observations is 343 as a result of adding 596 subspells and deleting 253 *BLIFEMST* spells that are entirely accounted for.

i) Use BDATES

Although *CLIFEJOB* spells are generally linked to particular *BLIFEMST* spells, they frequently extend beyond the period covered by the *BLIFEMST* spell. The *BLIFEMST* dates are relied on because they were collected closer to the event, and because this removes a great deal of inconsistency between employment and non-employment spells. In all, 3,022 start dates and 2,503 end dates are reallocated so that the *CLIFEJOB* spells lie entirely within the relevant *BLIFEMST* spell.

The change in date is minimised. Start dates that are moved forward in time are moved to the earliest feasible date for the *BLIFEMST* start. End dates that are moved back in time are moved to the latest end-date for the *BLIFEMST* end.

This step dramatically reduces the number of inconsistencies

j) Drop C-only spells

Remaining '*CLIFEJOB*-only' are present only because they have not been matched to any *BLIFEMST* spell (Generally because they overlap more than one *BLIFEMST* spell). 960 such spells are deleted.

This step also dramatically reduces the number of inconsistencies.

k) Adjust Wrong Winters

As discussed in the text earlier, winters are assigned to January but may in fact relate to either end of the calendar year. I identify cases where winter dates generate negative durations, and where reallocating the date to a date in the October-December period would not cause any other inconsistencies. These adjustments lead to 111 start dates and 67 end dates being reallocated.

l) Drop total overlaps

There are 146 spells that lie entirely within the previous or subsequent spell. The shorter spells are dropped. (An alternative would be to allow short full-time jobs to split a part-time spell in two, which would respect the definition of 'main job')

m) 'nudge' to resolve negdur & overlaps

Start and End-dates are adjusted within range to remove negative durations and overlaps. This nudging is done repeatedly, to ensure that resolving one of these problems does not give rise to more of the other. All except 2 negative durations are resolved.

n) Move dates to resolve negdur

The remaining 2 negative durations are resolved by moving the end date of the spell to equal the start date.

o) Move dates to resolve overlaps

The remaining 58 overlaps are removed by reallocating 58 start dates 58 end dates. They are allocated to the midpoint of the overlap, unless that would lead to inconsistencies with earlier or later spells.

p) Drop one obs

There remains one inconsistent history that cannot be resolved [PID=13487582]. At one point, the history shows five concurrent spells! The 14 spells in that individual's history are retained, but all dates are set to 9999.

q) Move dates to resolve gaps

Finally, 1,416 gaps in histories are closed up by setting start and previous end dates to equal the midpoint of the gap.

r) FINAL

The final dataset has 53,875 observations on all 9,050 individuals.

Table D2 : Regularising date sequences in LT retrospective files (*BLIFEMST&CLIFEJOB*)

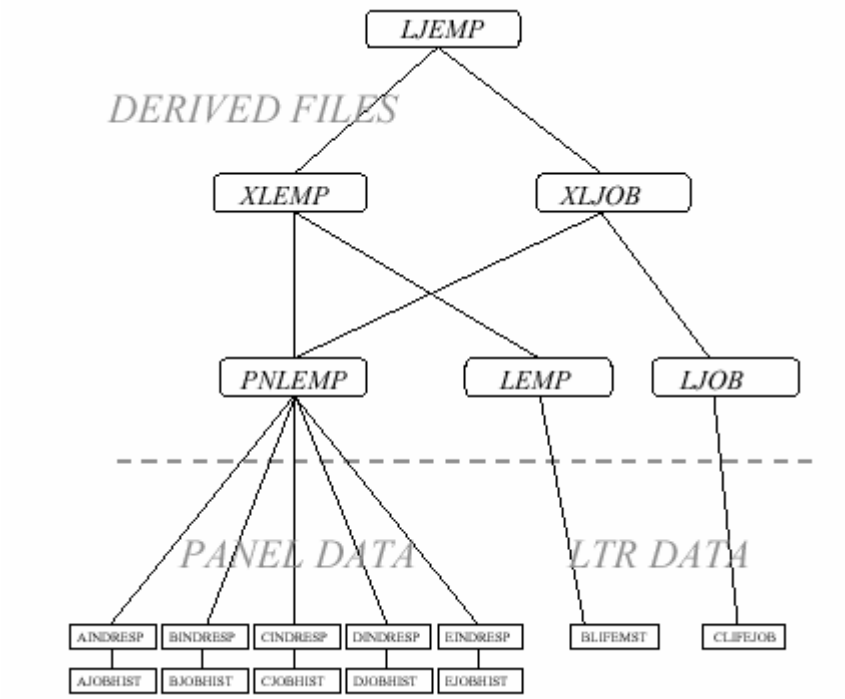
				conflict			*overlap*		*Gap*		*Add*	*Drop*		*Move*
ACTION	*Obs/ Indivs*	*end*	*start*	*neg dur*	*in Bspell*	*consec Bspell*	*other*	*in Bspell*	*consec Bsp*		*Obs*	*Obs*	*Start*	*End*
Read in base data: CLIFEJOB	32773/ 7074	116	116	0	13	24	76	1582	930					
Assign months		309	308	144	123	67	183	2093	1377					
Renumber intra-spell C-only spells		309	308	144	128	66	179	2168	1371					
Merge BLIFEMST & sort	55688/ 9050	4228	4241	144	148	1756	2785	2187	2156			34		
Drop final duplicates		4227	4240	144	148	1785	2751	2187	2175			734		
Accept 1st&last C-only spells		4227	4240	144	145	1731	2808	2176	1978					
Remove duplicates		3968	3981	144	144	1737	2524	2176	1997			282		
Renumber spells & pad		2955	2971	144	174	1777	1415	2214	1810		596	253		
Use BDATES		1580	1574	91	165	55	1563	2129	135				3022	2503
Drop C-only spells		199	197	90	165	55	124	2129	154			960		
Adjust Wrong Winters		121	119	32	122	46	119	2171	163				111	67
Drop total overlaps		62	62	13	101	40	90	2169	177			146		
'nudge' to resolve negdur & overlaps		62	62	2	29	1	28	1874	37					
Move dates to resolve negdur		60	60	0	29	1	28	1873	37					2
Move dates to resolve overlaps		2	2	0	0	0	0	1873	37				58	58
Adjust one obs		0	0	0	0	0	0	1873	37				14	14
Move dates to resolve gaps		0	0	0	0	0	0	0	0				1416	1416
FINAL	53875/ 9050													

Notes: See text for an explanation of column headings and description of actions

Appendix EThe Halpin files

This section illustrates problematic features of the Halpin files. In particular, it gives examples from the Halpin files of overlaps between spells, gaps between spells, and spells with negative durations.

Table E1 : Halpin Files - Structure of Programs



Source: Halpin (1997)

In general, the difficulties in the Halpin files arise due to:

- imposing the requirement that the end date for one spell strictly precedes the start date for the next spell [Start=end[_n-1]+1]. The base data frequently have zero duration spells [Start=End] and the algorithm used to impose the inter-spell gap creates problems.
- reconciling only with adjacent spells and not with the full sequence of spells. There are many complicated histories in the data and failing to take into account the full sequence of spells results in amendments that incompletely fix inconsistencies in the base data, or even worse, in some cases, create other inconsistencies.

To illustrate the problems, I present tables showing:

- the raw data (obtained by pooling all *wINDRESP*, *wJOBHIST*, *BLIFEMST* and *CLIFEJOB* observations);
- the relevant records from the Halpin files *NEWPAN*, *XLJOBE*, *XLEMPE*, and *LJEMPE*:
- the relevant records from my own reconciled work life history file.

The coding shown in the *Source* column is consistent with that used by Halpin. Records from *wINDRESP* have *Source*=10*w. Thus, Wave 5(E) records show *Source*=50. The *Source* code for *wJOBHIST* records equals one plus the corresponding w*INDRESP Source* code. Thus, EJOBHIST records are coded 51. Sequential spell numbers are included after the decimal point, in reverse order. *BLIFEMST* and *CLIFEJOB* records have *Source*=22 and *Source*=33 respectively. Spell numbers are appended to the right of the decimal point.

E2 Inconsistencies

E2.1 Overlaps

The file *NEWPAN* has 653 instances of overlapping spells. In all cases, these are of length one. (ie: the end date of a spell equals the start date of the subsequent spell.) This is inconsistent with the general rule used in the dataset that Start=End[_n-1]+1.

Table E2 shows the worst case, in which there are 3 overlaps in a single history. In all three cases, the problem is associated with a missing spell start date. The algorithm that was used to create the Halpin files does not deal successfully with this circumstance. The spell end dates are sufficiently far apart that imposing an inter-spell gap of one period would not cause any problems.

The only differences between the Halpin sequence and the sequence from the Maré file is that Halpin adjusts end-dates to strictly precede the subsequent start date. This generates the three one-period overlaps in *NEWPAN*, two of which are carried over to *XLJOBE* and *XLEMPE*. It also generates a negative duration for the record with *Source*=61.1.

E2.2 Negative Durations

NEWPAN has 2,010 instances of negative spell durations. The majority of these (1965/2010) are of duration –1, like the one shown in Table E2.

Table E3 shows two other histories that contain spells with negative durations. The first history is a worst-case, with 7 negative durations within the single history. All have duration of –1. The second history shows a spell with negative duration of 5 periods. Table E4, which shows gaps between spells, contains two further examples of negative durations of more than one period.

The one-period negative durations arise where the raw data has two consecutive spells with the same end-date. The attempt to enforce a one period gap between spells leads to the negative durations. The longer (5-period) overlap is a result of allocating a missing end-date to precede the subsequent start date, without checking that the imputed end-date for the prior spell falls after that spells start date.

The start and end dates in the Maré file differ by one period from the Halpin dates in many instances. Note also that the Maré file excludes the initial spells when the respondent was a student, since the histories are meant to capture the time since the respondent first left full-time education.

E2.3 Gaps

There are 2,825 gaps between spells in the *NEWPAN* dataset. Almost all of them can be accounted for as legitimate or intended gaps. Gaps may legitimately occur in the data if:

- the respondent was not interviewed in a wave, in which case the gap represents a break in responses;
- the spell is left censored. Halpin (2000, p. 3) notes that when a spell start date is missing, the variable *lcens* is set to 1, and the start date is set to equal the spell end date.

Left-censoring accounts for 2,371 of the 2,825 gaps between spells. Failure to be interviewed accounts for a further 452 cases, leaving only 2 unexplained gaps. These two cases are shown in Table E4. In both cases the second of the two consecutive spells has a negative duration, possibly due to two script interpretation or data input errors (99 instead of 79 in the first case; 95 instead of 45 in the second).

The Maré file does not contain any spell information for the first case shown because the history is incomplete. The respondent did not respond in waves 2 and 3, and therefore has missing retrospective spell information.

Table E2 : Halpin files – Overlaps between spells

		Raw			*newpan*			*xljobe*			*xlempe*			*ljempe*			*Maré*		
PID	*Source*	*Start*	*End*	*Status*	*Start*	*End*	*Status*	*Start*	*End*	*Status*									
11117427	51.1	Apr-95	??-95	FT	Apr-95	May-95	FT	Apr-95	May-95	FT	Apr-95	May-95	FT	Apr-95	May-95	FT	Apr-95	Jun-95	FT
11117427	50	??-95	Sep-95	UN	Jun-95	**Sep-95**	UN	Jun-95	Sep-95	UN	Jun-95	Sep-95	UN	Jun-95	Sep-95	UN	Jun-95	Sep-95	UN
11117427	61.4	Sep-95	Dec-95	FT	**Sep-95**	Nov-95	FT	Sep-95	Nov-95	FT	Sep-95	Nov-95	FT	Oct-95	Nov-95	FT	Sep-95	Dec-95	FT
11117427	61.3	Dec-95	Jan-96	UN	Dec-95	Dec-95	UN	Dec-95	Dec-95	UN	Dec-95	Dec-95	UN	Dec-95	Dec-95	UN	Dec-95	Jan-96	UN
11117427	61.2	Jan-96	Apr-96	FT	Jan-96	Mar-96	FT	Jan-96	Mar-96	FT	Jan-96	Mar-96	FT	Jan-96	Mar-96	FT	Jan-96	Apr-96	FT
11117427	61.1	Apr-96	Apr-96	UN	Apr-96	Mar-96	UN	Apr-96	Mar-96	UN	Apr-96	Mar-96	UN				Apr-96	Apr-96	UN
11117427	60	Apr-96	Sep-96	FT	Apr-96	**Sep-96**	UN	Apr-96	Sep-96	UN	Apr-96	Sep-96	UN	Apr-96	Sep-96	UN	Apr-96	Sep-96	FT
11117427	70	??-??	Sep-97	FT	**Sep-96**	**Sep-97**	FT	Sep-96	Nov-98	FT	Sep-96	Nov-98	FT	Oct-96	Nov-98	FT	Sep-96	Sep-97	FT
11117427	80	??-??	Oct-98	FT	**Sep-97**	Oct-98	FT										Sep-97	Oct-98	FT
11117427	91.5	Jun-98	Sep-98	UN															
11117427	91.4	Sep-98	Dec-98	FT	Nov-98	Nov-98	FT										Oct-98	Dec-98	FT

Table E3 : Halpin Files – Negative Spell Durations

		Raw			*newpan*			*xljobe*			*xlempe*			*ljempe*			*Maré*		
PID	*Source*	*Start*	*End*	*Status*	*Start*	*End*	*Status*	*Start*	*End*	*Status*	*Start*	*End*	*Status*	*Start*	*End*	*Status*	*Start*	*End*	*Status*
14021692	50	??-84	Oct-95	Stud	Jun-84	Oct-95	Stud	Jun-84	Mar-96	Other	Jun-84	Mar-96	Stud	Jun-84	Mar-96	Stud			
14021692	61.6	Sep-84	Apr-96	Stud	Nov-95	Mar-96	Stud												
14021692	61.5	Apr-96	Apr-96	FT	**Apr-96**	**Mar-96**	FT	**Apr-96**	**Mar-96**	FT	**Apr-96**	**Mar-96**	FT				Apr-96	Apr-96	FT
14021692	61.4	Apr-96	Apr-96	FT	**Apr-96**	**Mar-96**	FT	**Apr-96**	**Mar-96**	FT							Apr-96	Apr-96	FT
14021692	61.3	Apr-96	Jul-96	UN	Apr-96	Jun-96	UN	Apr-96	Jun-96	UN	Apr-96	Jun-96	UN	Apr-96	Jun-96	UN	Apr-96	Jul-96	UN
14021692	61.2	Jul-96	Aug-96	FT	Jul-96	Jul-96	FT	Jul-96	Jul-96	FT	Jul-96	Jul-96	FT	Jul-96	Jul-96	FT	Jul-96	Aug-96	FT
14021692	61.1	Aug-96	Sep-96	UN	Aug-96	Aug-96	UN	Aug-96	Aug-96	UN	Aug-96	Aug-96	UN	Aug-96	Aug-96	UN	Aug-96	Sep-96	UN
14021692	60	Sep-96	Sep-96	FT	Sep-96	Sep-96	FT	Sep-96	Sep-96	FT	Sep-96	Sep-96	FT	Sep-96	Sep-96	FT	Sep-96	Sep-96	FT
14021692	71.8	Aug-96	Sep-96	FT													Sep-96	Sep-96	FT
14021692	71.7	Sep-96	Oct-96	FT	**Oct-96**	**Sep-96**	FT										Sep-96	Oct-96	FT
14021692	71.6	Oct-96	Jan-97	UN	Oct-96	Dec-96	UN	Oct-96	Dec-96	UN	Oct-96	Dec-96	UN	Oct-96	Dec-96	UN	Oct-96	Jan-97	UN
14021692	71.5	Jan-97	Aug-97	FT	Jan-97	Jul-97	FT	Jan-97	Jul-97	FT	Jan-97	Jul-97	FT	Jan-97	Jul-97	FT	Jan-97	Aug-97	FT
14021692	71.4	Aug-97	Aug-97	UN	**Aug-97**	**Jul-97**	UN	**Aug-97**	**Jul-97**	UN	**Aug-97**	**Jul-97**	FT				Aug-97	Aug-97	UN

14021692	71.3	Aug-97	Aug-97	FT	**Aug-97**	**Jul-97**	FT	**Aug-97**	**Jul-97**	FT	**Aug-97**	**Jul-97**	UN				Aug-97	Aug-97	FT
14021692	71.2	Aug-97	Sep-97	UN	Aug-97	Aug-97	UN	Aug-97	Aug-97	UN	Aug-97	Aug-97	UN	Aug-97	Aug-97	UN	Aug-97	Sep-97	UN
14021692	71.1	Sep-97	Sep-97	FT	**Sep-97**	**Aug-97**	FT	**Sep-97**	**Aug-97**	FT	**Sep-97**	**Dec-97**	FT	Sep-97	Dec-97	FT	Sep-97	Sep-97	FT
14021692	70	Sep-97	Oct-97	FT	Sep-97	Oct-97	FT	Sep-97	Dec-97	FT							Sep-97	Oct-97	FT
14021692	81.3	Aug-97	Sep-97	UN															
14021692	81.2	Sep-97	Jan-98	FT	Nov-97	Dec-97	FT										Oct-97	Jan-98	FT
14021692	81.1	Jan-98	Mar-98	UN	Jan-98	Feb-98	UN	Jan-98	Feb-98	UN	Jan-98	Feb-98	UN	Jan-98	Feb-98	UN	Jan-98	Mar-98	UN
14021692	80	Mar-98	Sep-98	FT	Mar-98	Sep-98	FT	Mar-98	Oct-98	FT	Mar-98	Oct-98	FT	Mar-98	Oct-98	FT	Mar-98	Sep-98	FT
14021692	91.4	Mar-98	Nov-98	FT	Oct-98	Oct-98	FT										Sep-98	Nov-98	FT
14021692	91.3	Nov-98	Nov-98	UN	**Nov-98**	**Oct-98**	UN	**Nov-98**	**Oct-98**	UN	**Nov-98**	**Oct-98**	UN	Nov-98	Jan-99	FT	Nov-98	Nov-98	UN
14021692	91.2	Nov-98	Feb-99	FT	Nov-98	Jan-99	FT	Nov-98	Jan-99	FT	Nov-98	Nov-99	FT	Feb-99	Apr-99	FT	Nov-98	Feb-99	FT
14021692	91.1	Feb-99	May-99	FT	Feb-99	Apr-99	FT	Feb-99	Apr-99	FT				May-99	Nov-99	FT	Feb-99	May-99	FT
14021692	90	May-99	Nov-99	FT	May-99	Nov-99	FT	May-99	Nov-99	FT							May-99	Nov-99	FT
10373977	11.3	Aug-90	Oct-90	UN	Aug-90	Sep-90	UN	Aug-90	Sep-90	UN	Aug-90	Sep-90	UN	Aug-90	Sep-90	UN	Aug-90	Oct-90	UN
10373977	11.2	Oct-90	Dec-91	Emp	Oct-90	Nov-91	Emp	Oct-90	Nov-91	FT	Oct-90	Nov-91	Emp	Oct-90	Nov-91	Emp	Oct-90	Nov-91	FT
10373977	11.1	Dec-91	??-??	UN	**Dec-91**	**Jul-91**	UN												
10373977	10	Aug-91	Nov-91	SE	Aug-91	Nov-91	PT	Aug-91	Nov-91	PT	Aug-91	Nov-91	PT				Nov-91	Nov-91	SE

Table E4 : Halpin files – Gaps between spells

		Raw			*newpan*			*xljobe*			*xlempe*			*ljempe*			*Maré*		
PID	*Source*	*Start*	*End*	*Status*	*Start*	*End*	*Status*	*Start*	*End*	*Status*									
10755799	10	Jun-78	Sep-91	PT	Jun-78	Sep-91	PT	Jun-78	Sep-96	PT	Jun-78	Oct-99	PT	Jun-78	Oct-99	PT			
10755799	40	??-79	Oct-94	PT	Oct-91	Oct-94	PT												
10755799	50	??-78	Nov-95	PT	Nov-94	Nov-95	PT												
10755799	60	??-79	Sep-96	PT	Dec-95	**Sep-96**	PT												
10755799	70	??-99	Sep-97	PT	**Jun-99**	Sep-97	PT	Jun-99	Oct-99	PT									
13703323	22.01	Dec-45	??-??	FT															
13703323	10	Dec-45	Oct-91	FT	Dec-45	Oct-91	FT	Dec-45	Oct-93	FT	Dec-45	Sep-96	FT	Dec-45	Sep-96	FT	Dec-45	Oct-91	FT
13703323	20	Dec-45	Sep-92	FT	Nov-91	Sep-92	FT										Oct-91	Sep-92	FT
13703323	30	Dec-45	Oct-93	FT	Oct-92	**Oct-93**	FT										Sep-92	Apr-94	FT

13703323	40	Dec-95	Sep-94	FT	**Dec-95**	Sep-94	FT	Dec-95	Sep-96	FT						Apr-94	Sep-94	FT
13703323	50	Jan-45	Oct-95	FT	Oct-94	Oct-95	FT									Sep-94	Oct-95	FT
13703323	60	Dec-45	Sep-96	FT	Nov-95	Sep-96	FT									Oct-95	Sep-96	FT
13703323	71.1	Dec-45	Oct-96	FT	Oct-96	Sep-96	FT									Sep-96	Oct-96	FT
13703323	70	Oct-96	Oct-97	Ret	Oct-96	Oct-97	Ret	Oct-96	Sep-99	Other	Oct-96	Sep-99	Ret	Oct-96	Sep-99	Ret	Oct-96	Oct-97	Ret
13703323	80	Oct-96	Sep-98	Ret	Nov-97	Sep-98	Ret										Oct-97	Sep-98	Ret
13703323	90	Oct-96	Sep-99	Ret	Oct-98	Sep-99	Ret										Sep-98	Sep-99	Ret
13703323	100	Oct-96	Sep-00	Ret													Sep-99	Sep-00	Ret