The British Household Panel Survey: Introduction to a longitudinal data resource

Working Paper 2 of 'Longitudinal Data Analysis for Social Science Researchers', ESRC Researcher Development Initiative training programme

http://www.longitudinal.stir.ac.uk/

Dr Paul S. Lambert

Applied Social Science
University of Stirling
Paul.lambert@stirling.ac.uk

Last updated 18.9.06

Acknowledgements:
This paper was prepared under funding from the ESRC Research Development Initiative training programme 'Longitudinal Data Analysis for Social Science Researchers', grant reference PTA-035-25-0042. A previous version of this paper was prepared under funding from the ESRC Research Methods Programme Phase II project 'Longitudinal Data Analysis in the Social Sciences', grant reference RES-333-25-0007.
The BHPS data used in this report was provided bu the UK Data Archive, University of Essex, study numbers 5151 (June 2006 edition), 3954, 3909, 4901 and 5038. The BHPS data was collected under ESRC funding through the Institute for Social and Economic Research at University of Essex. Neither the data collectors nor distributors bear any responsibility for errors in subsequent usage.

Preface

These notes are written by a BHPS user and are intended to familiarise social scientists with the survey resource, to introduce the main practical issues as they are typically experienced by researchers new to the survey, and to provide some brief examples of analyses using the BHPS over a range of its data resources. These notes use a relatively informal style and readers should note that they have not been 'officially' endorsed by the primary data producers or suppliers of the BHPS, the ISER/ULSC and UK Data Archive at the University of Essex. **The web-pages of the project for which this text is produced** contain links to several relevant internet resources; access to copies of this text and related report files; access to a number of syntax example files which illustrate analyses of the type referred to in this text; and access to other related training materials in longitudinal data analysis. See: **http://www.longitudinal.stir.ac.uk/** .

1. Introduction: Structure and Origins of the BHPS

The British Household Panel Survey (BHPS) is a major government funded survey in the UK. It is a nationally representative panel survey which has been supported continuously since 1991. Originally, a stratified random sample of households was drawn at the start of the survey, then all residents of those households were traced and re-interviewed each year, to generate annual panel records which have been collected ever since. As the BHPS developed since 1991, its original sample has been supplemented with a number of 'boost' groups, including major additional subsamples from Wales and Scotland (1999 onwards), and Northern Ireland (2001 onwards – the BHPS is, strictly speaking, no longer 'British'). The BHPS represents an enormous investment in the social science research infrastructure of the UK, and a powerful resource for social science analysts. There are, however, numerous complexities to understanding and working with BHPS data. This text is written to help social science users to understand and overcome those obstacles.

The BHPS has been designed, coordinated and released annually since 1991 by staff at the University of Essex. The UK Longitudinal Studies Centre (ULSC) manages the production and promotion of the survey, whilst the UK Data Archive at Essex University manages its dissemination. (The ULSC forms part of the Institute for Social and Economic Research at Essex; formerly, the ISER directly managed the BHPS; earlier still, the same group were known as the ESRC Research Centre on Micro-Social Change). All datasets associated with the BHPS are processed by the ULSC and released to the University of Essex's Data Archive, where they are made freely available to the academic community. The main datasets are re-released every year with the addition of the latest panel's records (in some cases, errors made on previous records are also corrected in the latest release, so users should try to use the most recently released records at all times).

Internet resources play a crucial role in the dissemination of the BHPS:

The ULSC: *http://iserwww.essex.ac.uk/ulsc/*

The UK Data Archive: *http://www.data-archive.ac.uk/*

Each annual survey panel is referred to as a BHPS 'wave'. Interviews with respondent individuals begin each year at the start of September, and the bulk of fieldwork is usually completed over September / October of the relevant year. However, smaller numbers of interviews are conducted much later in the year – in some analyses of BHPS data, the date of interview can be an important contextual variate.

In September 2006, interviewing for the 16th wave of the survey began. After data collection, there is then a time lag between the conduct of interviews and the processing and release of the BHPS datasets. At present (September 2006), the latest BHPS data release covers files from wave 1 (first interviews September 1991) to wave 14 (first interviews September 2004), whilst wave 15 BHPS data is likely to become available around April 2007. Throughout the data collection and processing phases of its production, the BHPS is widely known as the 'Living in Britain' study (or 'Living in Wales', etc).

Funding for the continuation of the BHPS has hitherto been secured through a sequence of long-term research council grants. The existing format of the survey is currently funded to continue data collection at least until 2009. Moreover, in 2006, the conclusions drawn from the ESRC's strategic review of longitudinal survey data resources (see the 'Longview' report available from www.longviewuk.com) led to an increased commitment by the ESRC to the funding of micro-social panel surveys in the UK. At time of writing, the exact implications of this commitment are being negotiated. It is widely expected, though not yet confirmed, that the funding will lead to the current BHPS procedures being continued indefinitely, whilst being supplemented by a substantial additional new annual micro-social panel survey in the UK, which will be designed for maximum comparability with the existing BHPS procedures.

The BHPS data is released to users through a number of related 'studies'. These are distinct collections of data files themed around different aspects of the survey data. These studies comprise, first. a series of twelve collections of data files released through the UK Data Archive. The most influential of these studies are described shortly below. By far the most important study is the 'core' or 'main' BHPS panel data collection, the collection of files which includes all BHPS panel survey interviews from all previous years, and which is updated and re-released each year. In fact, the contents of all other BHPS studies are in some way derived specialist data extracted from part of these core data files. In addition to the UK Data Archive files, BHPS data is also available through a number of international studies, in the form of

certain cross-national survey datasets arising from the BHPS's harmonisation with other national panel surveys (described briefly below).

Extensive documentation describes the content and data collection procedures of the BHPS and the various studies through which its data is released. Most of the documentation is available online as well as being released as component materials to the relevant BHPS studies. With regard to the main data collection (the year-on-year panel studies) the documentation is supplied online, and in paper or 'pdf' file formats when the data is ordered. This documentation is updated every year in line with the data release updates. The latest version of documentation for the main datasets, which covers waves 1 to 14, can be cited as Taylor et al (2006a and 2006b). The online version of this documentation is particularly helpful due to its ease of navigation and rapid linkage to details on the BHPS content and studies. It is organised around two volumes. Volume A includes a thorough introduction to the BHPS and its sampling methods, discussion of access to and the representativeness of the data, the coding frames of the more complex variables included in the survey, and practical guidelines to working with the data. Volume B covers a lengthy listing of all BHPS variables, cross linked in the webpages to enable browsing and searching of variables by subject areas or data sources; in most cases, it is also possible to access the univariate distribution of each particular variable at every BHPS wave. Lastly, it is also worth pointing out that the BHPS is one of the surveys covered by the UK 'Question Bank' project, whose internet resources include copies of the BHPS questionnaires which may be searched by keywords and topics.

The BHPS documentation online: *http://iserwww.essex.ac.uk/ulsc/bhps/doc/*

The Question Bank: *http://qb.soc.surrey.ac.uk*

By far the most analyses of the BHPS datasets have been conducted by staff at or associated with the ISER/UCLS, and a large list of relevant publications is assembled in their working, occasional and technical paper series'. An index to these publications is available on the ISER/UCLS website, including links to abstracts, summaries and often full versions of the papers. Aside from being practical examples of the analysis of BHPS data, most of these publications also describe the BHPS in detail, whilst some of the earlier papers are explicitly methodological discussions of the BHPS. Nevertheless, sociologists often find such publications hard going, since many of them, produced by economists, employ quite advanced quantitative methods with limited background discussions. For social scientists, much more approachable introductory reading on the BHPS can be found in two edited collections which 'showcase' sociological analyses of the data as well as providing accessible introductions to the survey. These are the collections of Buck et al (1994), and Berthoud and Gershuny (2000)[1]. Another recent ISER produced edited book, that of Rose (2000b), takes a more methodological perspective in examining the use of the

[1] Ermisch and Wright (2005) provide a similar edited collection showcasing the use of the Scottish BHPS subsample, whilst there are future plans for similar illustrations of the Welsh and Northern Irish data.

BHPS, as well as comparable panel surveys from other countries, and also includes several segments of an introductory nature.

Listings of ISER publications : *http://iserwww.essex.ac.uk/pubs/*

The ISER/UCLS staff are also the primary point of contact for BHPS users. The senior organisers of the BHPS are Prof's Nick Buck and Stephen Jenkins, and the current survey manager at Essex is Dr Heather Laurie. The researcher charged with most user liaison is John Brice, *bricj@essex.ac.uk* . In addition, a BHPS user group exists holding annual meetings for users, as does a (currently seldom used) JISCmail listserv. Biennial BHPS research conferences has also been taking place in Colchester since July 2001. Details of all of these resources can be found through the ISER/UCLS websites.

When users apply for access to the BHPS datasets, the first impression is usually of a daunting number of studies, data files and associated variables. As mentioned above, the first and most important BHPS study is the **primary or 'core' collection of BHPS panel data records**, updated and re-released each year with each new wave of data. This study, which is usually supplied as a single zip archive, contains within it a large number of different data and documentation files (143 different data files in the 2006 release for waves 1-14). The numerous data files (also known as 'records'), reflect the supply of BHPS data according to a structure of different data files for different survey waves, plus, within waves, the supply of different types of survey data in different files. The files are organised firstly in terms of the waves to which their data refers, with filenames being preceded by sequential letters, starting with 'a' at wave 1, to indicate the year they apply to. Users should get used to this convention, since it also characterises all BHPS panel data variables, which are prefixed a/b/c/d/e/f/g/h/i/j/k/l/m/n from waves 1 to 14 respectively. For instance, the variable indicating a respondent's current job status is named ajbstat in the appropriate wave 1 data file; bjbstat in the wave 2 file, and so on. Following the conventions of the BHPS documentation, generic file and variable names are indicated '{w}fileorvarname', as in {w}jbstat.

Table 1 lists all the files contained in the current wave 14 release of the core BHPS dataset, and gives a brief description of their contents; a more lengthy guide can be found in the BHPS online documentation:

http://iserwww.essex.ac.uk/ulsc/bhps/doc/volb/allrecs.php .

In the first instance, Table 1 illustrates that most files are associated with particular waves, excepting three, prefixed with an x, which are designed specifically to aid the linking of records between waves (the 'x' standing for the cross in 'cross-wave'!). Complications then arise because there is more than one data file per wave, and because there are some data files which are not found in every wave. First, the multiple files reflect the separation of collected data by the subject of analysis to

which they apply, as well as a pragmatic separation of some of the more complex subsets of data. The BHPS survey design, see section 2, works by interviewing all members of a selected household, so some of the subsequent data files, which include the letters 'ind', apply to individual level data for each respondent (household member), whilst others, indicated 'hh', contain variables which apply at the household level (for instance, the amenities of the dwelling). It may or may not reassure users to know, however, that the vast majority of BHPS analyses make use of just 2 files for any given wave : individual level data from the '{w}indresp' files, supplemented with a few household level details obtained from the '{w}hhresp' files.

Outwith the explicitly individual or household level files, other wave specific files are used to link data between individuals within households ({w}egoalt), to separate out a complex set of variables measuring income components ({w}income), and to store event history data which provides continuity between interview points in information on employment circumstances ({w}jobhist, see the next paragraph). Additionally, since the wave 4 data collection, children aged 11-15 have been interviewed with a separate survey design, generating the highly promising, but seemingly underused, '{w}youth' datasets.

In addition to the core BHPS datasets, there are also at present several further studies containing BHPS data which are developed and released separately (though their records can be linked with cases from the core files). One example which is also frequently updated in line with the addition of new panels, is the collection of **"Derived current and annual net household income variables"**, at present available for waves 1 to 12 as UK Data Archive study number 3909 (latest release October 2004). This dataset, described in its technical report, Bardasi et al (2001), extends the available information on household income by computing a number of 'net' derivations, for instance utilising regional information on local tax rates.

The BHPS's primary longitudinal data collection takes the form of annual panel records, allowing for the comparison between individuals' positions at discrete points in time. Another method of longitudinal analysis, of course, involves the analysis of continuous time records. Such is the popularity and analytical scope of such 'event history' or 'survival' data analysis techniques in the assessment of demographic and employment transitions, that the BHPS survey instruments also include the collection of retrospective life history files designed to 'fill in the gaps' on such events between , and before, survey interview points. All such relevant data is released within files from the core BHPS dataset, including the intra-wave job history records '{w}jobhist' and the pre-1991 life history records on employment and demographic experiences recorded in additional files collected at waves 2, 3, 11 and 12[2]. However, navigating all of these sources is an extremely complex task, in recognition of which the ISER have also produced a separate BHPS dataset, latest release Essex Data Archive study number 3954, the **"BHPS combined work-life history data"**. This file collates information from the multiple records to produce much more accessible life history

[2] Restrospective questions for original BHPS respondents were asked in waves 2 and 3, whilst the retrospective questions of waves 11 and 12 were given to new BHPS respondents who were not present in the earlier years of the survey.

data files (a 'case' is usually an employment or demographic event with additional variables indicating the relevant individual and the circumstances, starting and ending times of that event). Again the dataset is re-released periodically to incorporate the latest additions of BHPS data. The production, and indeed use, of this dataset, is a complex process described fully by the researcher involved, in Halpin (2002, 1998). It represents, nevertheless, a huge step forward from working with the 'raw' BHPS life history files of the core dataset[3]. In particular the combined file improves the retrospective reliability of the records by cross-checking and correcting for any mismatches between related reports collected at different points over time (cf Solga 2001, Elias 1997 for more general discussions of the methodological issues in working with such retrospective data resources). One drawback with this dataset however is that full life history records are only complete for BHPS respondents who were interviewed at waves 2 and 3, or else who have grown into the sample (see section 2). Thus for example, new adult entrants to the BHPS, such as those incorporated by recent extension samples (again, see section 2), do not at present have full life history information available for them within this dataset. Additionally, Crouchley and Oskrocki (2001, 2000a) have argued that several endogeneous sample selection effects seriously compromise the representativeness of the BHPS's life history records. At time of writing (September 2006) it is not clear whether future updates to this derived dataset will definitely take place; the present data ends at time points around 2001, and there is clear scope for incorporating more recent information into this valuable resource.

The BHPS is the only major panel study of its nature in Britain, but several other countries have closely comparable panel studies, and indeed several research projects focus on making cross-national comparisons in panel survey research. The best known of these is the European Community Household Panel (ECHP) dataset, which is an amalgamation of data from 14 annual panel studies in Europe. The ECHP's first data collection began in 1994, but the project was actually ended with a final sweep of data in 2001 (a BHPS subsample constitutes the British contribution to the ECHP from 1997-2001; initially, the datasets were separated). The ECHP is produced by EUROSTAT and can be accessed by application to that organisation[4]. The European Panel Analysis Group, whose webpages are linked from the ISER/UCLS site, is a research network of users who primarily utilise the ECHP database. Elsewhere, the CEPS institute in Luxembourg coordinates two projects which utilise the BHPS in cross-nationally comparable work. The PACO (Panel Comparability) project generates a database of a similar nature to the ECHP, covering a wider range of countries and years though with a smaller range of data; it can be accessed directly through contact with the CEPS institute. The CHER (Consortium of Household Panels for European Socio-economic Research) project, also hosted by CEPS, also involves harmonising a selective range of data panel studies across different countries; the CHER data promises to be more contemporary than PACO. In place of the discontinued ECHP, a new international panel project – EU-SILC – is currently in its early stages of data collection and processing (2003 onwards). Additionally the

[3] Crouchley and Oskrochi (2000b) describe alternative procedures which collate life history information from the source BHPS core files, though it is the datasets released by ISER which have been more widely used.

[4] Although a preview of the ECHP's contents can be found via its questionnaires' storage at the UK 'Question Bank', at http://qb.soc.surrey.ac.uk/surveys/echp/echpintro.htm

CNEF project (Cross-National Equivalence File) represents a particularly powerful and long-run coordination of international panel survey datasets, collating data from the BHPS for Britain, the PSID for the US, the GSOEP of Germany, and the SLID panel study of Canada. The CNEF attains particularly high standards of data harmonisation and documentation, with a focus on income and health measures, and a longer time span than many other harmonised projects (commencing 1980 for the US data and 1984 for the German data). Indeed, the pioneering US Panel Study of Income Dynamics (PSID), first wave 1968, is worth highlighting simply because of its high usage in American research, and its seminal influence on the design and analysis of subsequent panel studies, including the BHPS.

European Community Household Panel (via EUROSTAT) :
http://www.forum.europa.eu.int/Public/irc/dsis/echpanel/home

CEPS (information on PACO and CHER) : *http://www.ceps.lu/*

EU-SILC:
http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1913,47567825,1913_58814988&_dad=portal&_schema=PORTAL

Cross-National Equivalence File:
http://www.human.cornell.edu/che/PAM/Research/Centers-Programs/German-Panel/Cross-National-Equivalent-File_CNEF.cfm

Panel Study of Income Dynamics : *http://www.isr.umich.edu/src/psid/*

The BHPS (and its substnatial funding) has been justified with reference to the many arguments for the superiority of longitudinal data resources for social science analysis (e.g. Rose 2000b, Scott 1993 with regard to the BHPS). The basic contention is that processes of change lie at the heart of interesting social science investigation, and that they cannot be studied without reliable longitudinal data. The BHPS as a representative sample of the British population was designed to provide just such a resource for British research, and to open avenues for more thorough analyses of processes across the range of social studies. At this period, it is an open question whether such research has grown appropriately in response to the data availability. Certainly there has been a rapidly increasing range of publications using BHPS data, many of them demonstrating examples of longitudinal research investigations which have been able to adjudicate more confidently on previously irresolvable issues (e.g. Jarvis & Jenkins 1995). Equally, a number of papers have begun to introduce and then examine innovate social science questions which could not have been asked without longitudinal household panel data (eg Gershuny 2002a, Brynin 2000).

Nontheless there has also been some concern over a lack of wider usage of the BHPS in the UK's social science community. For one thing, the large majority of current BHPS analyses can be associated with researchers at or connected with the ISER/UCLS. For another, a surprising number of analysis using BHPS data have actually not made extensive use of its longitudinal panel elements (instead using the

survey for its cross-sectional, repeated cross-sectional, or retrospective resources). This is regarded as disappointing, not least because it is the panel aspect of the data which is the most expensive to collect, but also because panel data offers by far the most reliable longitudinal evidence. There are several possible explanations for reluctance to adopt the BHPS as a data resource. From recourse to a particularly thorough longitudinal data resource, namely this writer's anecdotal experience of the UK social science research community, by far the greatest deterrent is the perceived complexity of both dealing with the BHPS data, and communicating results from its analysis.

Other plausible explanations for lower usage rates do reflect genuine, albeit moot, shortcomings of the data. For instance, despite the large scale of the BHPS, it is often noted that its sample size remains relatively small for many specialist purposes, and the coverage of its questions relatively limited - whereas populist fields of social science research are often interested in more specialised debates or subpopulations. Another reason for a lack of analyses using panel data trends in the presently available data is that in the early years of the survey, the number of BHPS waves has not yet spread over lifetime spells which are long enough to be of interest to certain longitudinal research questions. Similarly, the BHPS has a slightly greater time lag between the point of data collection and the availability of the data as distributed files – typically a couple of years, which can deter those eager for highly contemporary resources. Finally, we can note that there are a number of genuine methodological concerns over the use of BHPS data. Many of these relate to generic problems of household panel surveys, discussed in section 3, though at least one group of researchers have consistently argued that the sampling basis of the BHPS is fundamentally flawed (eg Davies & Crouchley 1989, Crouchley & Oskrochi 2000a, Crouchley & Oskrochi 2001).

2. Basics of working with the survey

In the next section we discuss some key issues which most new BHPS users are likely to confront. Other information covering similar such operational issues can be obtained from the BHPS documentation, or through contacting relevant staff or groups at the ULSC. In addition, for several years staff involved with the BHPS have run introductory BHPS training workshops, both internally and at the annual 'Essex Summer School in Social Science Data Analysis' (where other courses on techniques relevant to longitudinal data analysis are also offered).

ULSC training courses: *http://iserwww.essex.ac.uk/ulsc/bhps/courses/*
The Essex Summer Schools: *http://www.essex.ac.uk/methods/*

2.1) The BHPS sampling design, extension samples and weighting methods

The BHPS can generally be regarded as a random sample which is nationally representative after weighting by population level variables supplied with the survey. However all users should have a basic understanding of the components of the survey sample, particularly because many analyses proceed (defensibly) without the use of sampling weights. Volume A of the BHPS documentation (Taylor et al 2006a) covers these issues in greater detail.

The BHPS began as a stratified random sample of the population of British households, drawn and first contacted at wave 1 in 1991. The 'stratified' design involves a device intended to reduce survey costs – a random selection of relatively small geographical locations was first drawn (namely, around 250 'primary sampling units'), then only households within those units were subject to random selection. This is a common strategy in survey research, although there was some debate over its adoption in the BHPS (eg Coxon 1991). Around 5,500 households, covering some 14,000 respondents were contacted by the initial random sample. It is worth noting that these contacts represented an approximate 65% successful contact rate from the intially selected sampling frame, a proportion fairly typical in survey research, but which carries obvious implications for the potential representativeness of the responding BHPS sample. Taylor (1994) reviews issues of BHPS initial wave non-response.

The basic principle of the survey was then to declare all initial members of the households selected at wave 1 to be 'original sample members' (OSM's), and to recontact them annually throughout the duration of the survey, regardless of whether they should move between locations or households. Any new born descendants of OSM's would also become OSM's themselves, a following rule which in principle should offset the loss from the sample of any deaths of OSM's. Additionally, because information on all household sharers of all respondent members is desired, in later waves the BHPS also interviews any household sharers of OSM's who were not themselves selected in the original sample. These cases are referred to as 'temporary sample members' (TSM's), who themselves remain in the BHPS sample for as long as they share a household with an OSM[5]. It is important to remember that all members of the originally selected households were treated as survey members, including children, although the main BHPS interviews are conducted with over-16 year olds only. (This factor explains the difference between the number of individuals listed in the overall BHPS sample files at each wave, and the number in the main interview sample files). Thus, when child OSM's pass their 16th birthday, they 'grow into' the full BHPS interview pool (known as 'rising 16's'). Table 2 shows the numbers of individuals interviewed in some way during each BHPS wave, giving the sample groups from which individuals were drawn (the numbers on the left panels of Table 2 include children and those enumerated but not given a full interview). For the core sample, often known as the 'Essex sample', Table 2 also shows the division between

[5] One complication is that a TSM may be reclassified to a 'Permanent Sample Member' if they have a child with an OSM.

the number of cases who are classified as Original Sample Members, and those being interviewed as Temporary Sample Members[6]. Lastly Table 2 also givens the total number of contacts each year, and the smaller total number of contacts which resulted in a full adult interview. Analytically, BHPS research is designed to be conducted on the combined sample of OSM's plus TSM's. However it is worth remembering that the TSM's are not necessarily the same people between waves.

A great deal of emphasis is placed at the BHPS data collection stages on the need to maximise recontact rates at each wave, and by and large the BHPS has been very successful in this regard (illustrated in table 2 by the stable numbers of OSM's in the first column, although the precise statistics on attrition rates require further details). The 'wave-on-wave attrition' of cases was highest in the transition from waves 1 to 2 (around 11% of OSM's were not recontacted), but since that stage, recontact rates have stayed very high, routinely well over 95% between waves. Nevertheless a small number of cases do drop out at each wave, some of them being successfully recontacted again at later waves, but the bulk of them being lost permanently. Taylor (1994) reported specifically on attrition in the early waves of the BHPS, suggesting it to be empirically less of a concern that in some other longitudinal designs. (Indeed, in practice, a more problematic issue for researchers is that the BHPS's recontact successes do tend to mask often moderate levels of within-contact item non-response, namely circumstances where an individual was contacted but did not answer the relevant variable or section of the dataset. Item non-response is a generic problem to all survey research and is not discussed further here).

An important complexity was added to the BHPS samples at Waves 7, 9, and 11, namely the introduction of 'extension' (or 'boost') samples, incorporating cases used specifically for the ECHP (wave 7 onwards), then new contacts from Scotland and Wales (wave 9 onwards), and from Northern Ireland (wave 11 onwards). The figures involved are shown in Table 2. The former boost was undertaken in large part for administrative reasons to cooperate with the ECHP design, but it is important to note that the new cases introduced via the ECHP were not representative at the national level, instead over-representing low-income households. The latter regional boosts had a firm substantive intention, namely to increase the number of survey contacts from Scotland, Wales and Northern Ireland in order to ensure sufficient cases within each country for more detailed analyses both within and between the UK's countries. Indicator variables named {w}memorig and {w}hhorig on the main BHPS files show which parts of the combined BHPS files particular individuals or households (respectively) originate from.

The various additions and drop-outs from the BHPS original sample, and in particular the introduction of the later extension samples, means that the basic BHPS dataset is never clearly a random sample, and is markedly biased when the extension samples are incorporated. Nevertheless it is neither uncommon, nor unreasonable, for certain analyses to proceed on the various BHPS datasets with no weighting of cases attempted, since it is a general expectation in quantitative research that multivariate

[6] Technically (see the preceeding footnote), the numbers of "OSM's" reported in the table are actually the combination of OSM's and any PSM's.

analyses of relative effects are robust to smaller sampling imperfections. However, analysts who wish to present their results as representative of national averages should apply a series of wave-specific weighting factors designed by the BHPS producers - this is particularly important if simpler univariate or bivariate statistics are being presented. Primarily, the available weights divide between cross-sectional weights, which align core demographic features of a particular wave's respondents with the current national averages, and longitudinal weights, which undertake the same process only for those BHPS respondents who have been re-interviewed for every wave up to the one in question. Since waves 7 and 9, those weights take on greater complexity in order to deal with the incorporation of the extension samples, but the net result remains essentially the same. Further details on all weighting variables can again be found in the documentation, Taylor et al (2006a). Training files associated with the 'Longitudinal Data Analysis' project include software illustrations of applying different BHPS weights (see the web site).

2.2) Access media, inter-file identifiers and working practices

The main facility for accessing BHPS studies involves applying for access from the UK Data Archive, then obtaining electronic files from that organisation. The data can be supplied by remote transfer or CD, and in a choice of data formats – in plain text ASCII, or files specifically suitable for SPSS or STATA. The large majority of BHPS users obtain their data in this way from the Essex Data Archive, storing the transferred files on a private computer and analysing from there. A problem in this method, however, is the large size of the various BHPS datasets : the 2006 release of the 'core' datasets requires around 750MB of storage space, whilst the latest derived net income files occupy a further 30MB, and the life history files 100MB. Bearing in mind that most users will later produce their own derived working files from the 'raw' versions, and moreover that, every year, more files are released, there is a clear recipe for filespace congestion.

The number of files associated with the BHPS datasets (eg Table 1), the number of waves, and the number of topics covered, proves a fairly accurate indicator of the relative complexity of data management required in typical BHPS analyses! In this regard, two factors prove most important. The first is to know that a number of 'key linking variable' exist in all BHPS records, which identify people and / or households with unique numeric values. The name of the individual level identifier used across files is 'pid', a single number which is unique to each individual, the first digit of which indicates the wave number at which the case first joined the BHPS. Additionally, wave specific variables {w}hid serve as household level identifiers, unique values for each household in a given wave. Typically, multiple 'pid' cases are found within each household. To improve linkage between individuals within a household, a third key linking wave specific variable, {w}pno, is assigned to each individual, indicating their 'person number' within the household (any given {w}hid to {w}pno combination is necessarily unique to an individual. Thus, BHPS users link records between waves and related data files by merging computer files on the basis of these key linking variables. Several examples of SPSS and STATA command

language operations which illustrate such matching, with explanatory notes, can be accessed from the website of the 'Longitudinal Data Analysis' project.

The second issue for successful data management with the BHPS is to appreciate the importance of good programming practice. With the high degree of file manipulation routinely required for work with the BHPS, it is essential that users retain a record of the transformations and file mergers they make. The most efficient way of achieving this with modern analyses packages is to record the command file syntax used, perhaps adding annotations to explain relevant techniques. Adequate recording of command file syntax is a habit many social science researchers do not have, which in this writer's belief goes a long way to explaining the difficulties many have experienced in working with the BHPS and other complex data resources. Assorted examples of SPSS and STATA command syntax, for instance, are found as downloadable files available from the website relating to this document.

2.3) Key issues in cross-wave data matching

The basic principle of the panel data files is that equivalent variables are re-recorded, each year, with the same variable suffix name and definitions. Thus for instance the occupational title held by a working respondent at the point of interview in wave 1 is named ajbsoc, the title at the point of interview in wave 2 bjbsoc, and so on. By matching together data files which include individual identifiers and a selection of variables over waves, we can quickly generate files which show sequences of variable values over time (see figure 1 an example of a 'wide' format panel dataset).

Figure 1: Illustration of a cross-wave linked individual level file, 5 respondents

pid	*asex*	*aage*	*ajbsoc*	*bjbsoc*	*cjbsoc*
10101	1	40	220	220	222
10102	2	38	-1	-1	-1
10103	1	26	872	950	-9
10104	1	54	140	140	140
10105	2	49	320	320	320

However there are some exceptions to this model. Many variables are not asked at every wave of the panel study, but included intermittently, primarily in order to allow more space elsewhere in the survey for other variables (information on which waves a variable does appear in is found in volume B of the documentation). Additionally, a number of variables can be related to each other between waves without being strictly identical. This occurs, for example, when the form of the question asked varied slightly between waves. Such occurrences are usually highlighted by a variation in the variable name suffix, but in a few examples there are cases when an apparently equivalent variable between waves is actually recorded in separate ways (a widely known example concerns the difference in categories used for the employment status indicator variables ajbstat and bjbstat). Additionally, there are of course some

circumstances where variables are nominally equivalent between waves, but might be considered substantively different. Examples include income measures in pounds, where users should remember to consider the effects of annual inflation and the date of interview; and more subjective records, for instance political views, where users might consider that to hold a certain stance means different things at different points in time. This issue also applies to the timing of the BHPS survey interviews, which are seldom exactly annual but tend to occur within a three month window each year – it is possible that seasonal factors affect some individuals' variable responses. Nevertheless, there is no clear prescription for dealing with such issues, other than suggesting users take care to remember the context of their data records.

A final point worth mentioning in this area is that researchers are regularly caught out by the BHPS's handling of variables which are 'fixed' in time. To save time during data recording, respondents are only asked the relevant variable if they have not responded to the same question previously – a typical pattern emerges in the example of the ethnic group indicator variable {w}race, where most cases have a positive response at wave 1, then only a few (mainly new entrants) have positive responses in later waves. In order to maximise the information on such variables, it is necessary to create a file which pools the variables between waves, then calculate an aggregate variable which borrows from whichever wave specific variable had a positive response. An example exercise accessible via the website related to this document contains an illustration of this process for the case of the parental occupational information measures. Increasingly, the BHPS data release includes pre-prepared versions of this harmonisation of fixed-in-time variables on the 'xwavedat' data file (this file was not available in earlier years of the survey).

Many panel data analyses involve matching BHPS records between waves in an alternative format (known as 'long' format) - illustrated by figure 2.

Figure 2: Illustration of a 3-wave panel file, 2 respondents

pid	*Wave*	*sex*	*age*	*jbsoc*	*pay*
10101	1	1	40	220	950
10101	2	1	41	220	990
10101	3	1	42	222	1120
10102	1	2	38	-1	-1
10102	2	2	39	-1	-1
10102	3	2	39	-1	-1

In this case, it is necessary to add records together from the relevant wave specific files whilst distributing out fixed information from single variables to the multiple records. Then, wave varying variables can be calculated by computing the new variable as a function of the relevant contributing wave specific variables. The SPSS and STATA syntax examples on the 'Longitudinal Data Analysis' website include illustrations of several such examples of long format panel file data constructions. Note that figure 2 illustrates a hypothetical circumstance where the second respondent had just passed her September birthday before being interviewed for waves 1 and 2, but was interviewed before her birthday at wave 3.

3) Pro's and cons of the BHPS

The next section runs through a selective series of attractive, and unattractive, features in working with the BHPS resources.

3.1) Positive assets of the BHPS

- It's the panel, stupid

Social scientists in the UK have never previously has access to social science panel data records collected and distributed on such a scale. The possibilities for substantive and methodological advances are considerable (c.f. Berthoud and Gershuny 2000; Rose 2000b; Gershuny 2002b; Gershuny 2004).

Moreover, the BHPS represents a large-scale panel data collection into which a tremendous effort has been made in ensuring data quality at every stage of the data processing. The standards of survey data collection and documentation associated with the collection of the BHPS and its supply to secondary analysts are widely regarded as amongst the highest (cf Dale 2006).

- Household sharer information

Many sociological approaches lay great emphasis on the roles of household contacts in structuring individuals' life experiences, yet surprisingly many empirical analyses work at the level of isolated individual actors. The excessive detail available on BHPS respondents' household sharers - including longitudinal information – makes for a very rich resource in the analysis of household interrelations[7]. Additionally, the BHPS's following rules mean that information on the continuing development of previous household sharers is also often available. For instance it is often possible to link the records of young adults' who have left the parental home, with data on their no-longer co-resident parents, if the family group initially shared a household earlier in the BHPS. On the other hand, the BHPS has less detail on wider social networks such as family connections which have never coincided with household sharing, and whilst the BHPS does include some information on the characteristics of the 'closest friends' of respondents, it is of a relatively low quality[8].

- The wave specific 'youth' records

[7] The analysis of every possible intra-household relationship can become extremely complex. Lambert (2001) discusses possibilities in the BHPS data for analysing different forms of intra-household ties ('person-group' clusters), and provides (very long) command file macros which can be used to operationalise representations of such relationships.

[8] Unfortunately a design error in dealing with data on the characteristics of friends falsely conflates information on 'best friends' and 'any named friends'.

An often neglected feature of the datasets are the youth surveys conducted with all BHPS respondents aged between 11 and 15 from wave 4 onwards. These questionnaires cover a variety of topics including attitudes and expectations of school and work, relationships with family and friends, and health related behaviours. Indeed, as the panel interviews continue to be collected, an increasing number of adult respondents filling out the full BHPS response will also have filled out youth interview questionnaires in the past, allowing expectations and outcomes to be compared.

- Occupational information and work history files

The BHPS records the occupational titles of current and previous jobs to a high level of detail, as well as collecting information on a host of variables related to employment situation. The job history files extend this detail to periods between the panel interviews, and indeed for the entire post-schooling period in the case of the life history files. The BHPS also collects comparable details on the occupations of a respondent's parents, as well as by definition the occupations of household sharers, and some limited detail on the occupations of friends. Occupational unit analysts will recognise that, in combination with the survey's breadth of coverage of other topics, this constitutes a valuable resource.

- High quality of income information

Extensive efforts have been made in the production of the BHPS panel records to maximise the information available on income measures, including extensive data collection across a mass of sources of income, and comprehensive work on imputing total income values based on combinations of information on constituent properties. The result is a series of closely comparable, widely understood variables with high coverage over the sample.

- Subpopulation analyses

Because the BHPS is designed as a moderate scale nationally representative sample, a number of subpopulations of interest to sociologists can be identified in numbers, although equally the rarity of other subpopulation groups means some analyses remain largely out of bounds. Recently, much attention has focussed upon the extension samples which allow for comparisons between Scotland, Wales and England, but equally any grouping which is reasonably common in the national population can be identified in the BHPS and then a dataset with extensive cross-sectional and longitudinal records considered. As an example, the full panel histories of a subsample of over 200 teachers are covered by the BHPS. Perhaps even more effectively, the crosstabulation of BHPS variables could allow us to identify and analyse the longitudinal histories of particular subpopulations who display a certain combination of circumstances. Sociologically interesting examples might be, say, women who have at some stage given up on jobs in nursing, or adults who once returned to education as a mature student.

3.2) Negative features of the BHPS

- Complexity of the data records

As we have alluded to above, work with the BHPS datasets can become quite complex. The large number of, regularly re-released, data and documentation files, quirks of the data with regard to some particular variables, and the complexity of integrating retrospective life history records with panel data resources, all combine to produce considerable work for old and new analysts alike. Arguably, this complexity has reached a stage where many potential users are deterred, rightly or wrongly, from using the BHPS.

- Dropout and item non-response

Although as mentioned above attrition rates between BHPS waves are relatively low, there are still some cases of dropout at each wave, whilst there is also moderate item non-response in the surveys amongst those who are recorded as contacted respondents. The effects of these missing data can be particularly disconcerting for panel data models, where the number of respondent cases who have supplied valid responses across a particular range of variables of interest declines rapidly when cross-wave records are matched. (This is especially pronounced if the model requires a 'balanced' panel design, ie one where all records have exactly the same number of valid response points). Indeed, because of such effects, panel model sample sizes in practice often reach the point where sparse representation of cases becomes a problem. Increasingly in modern analyses, imputation of missing data is used to lessen this problematic.

- Interviewer effects and panel conditioning

Like most large scale UK social surveys, the BHPS recruits interviewers from major market research firms. It can be observed apparently unambigiously that those people who work as interviewers are not a mixed cross-section of the population, but have very distinctive characteristics and personalities. One observer described them as: *"polite and presentable, apparently middle class, middle-aged women ..[suffused with].. Daily Mail/Telegraph can do conservatism"*[9]. This uniformity may be expected to lead to some degree of interviewer bias; although the impact of intra-interviewer effects on BHPS records have been investigated (eg O'Muircheartaigh &

[9] Author's personal observation, quoted from a report on attendance at an interviewer debriefing session and BHPS interview (arranged through the ESRC "survey link" scheme), available from http://www.staff.stir.ac.uk/paul.lambert/downloads.html at 1.9.06.

Campanelli 1998), it is much harder to allow for the aggregate effects of 'typical' interviewer types. Moreover, another feature distinct to panel datasets may also have an unknown impact upon the BHPS sample. 'Panel conditioning' refers to the circumstances where both interviewers and respondents become increasingly familiar with the survey contents between waves and may begin to structure their interactions in reaction to this (van der Zouwen & van Tilburg 2001, Holt 1989). In practice, the bulk of BHPS interviews involve people who have had the same conversations with the same interviewers once a year for ten years, and to a casual observer, evidence of panel conditioning is clear[10].

- Regional sampling base

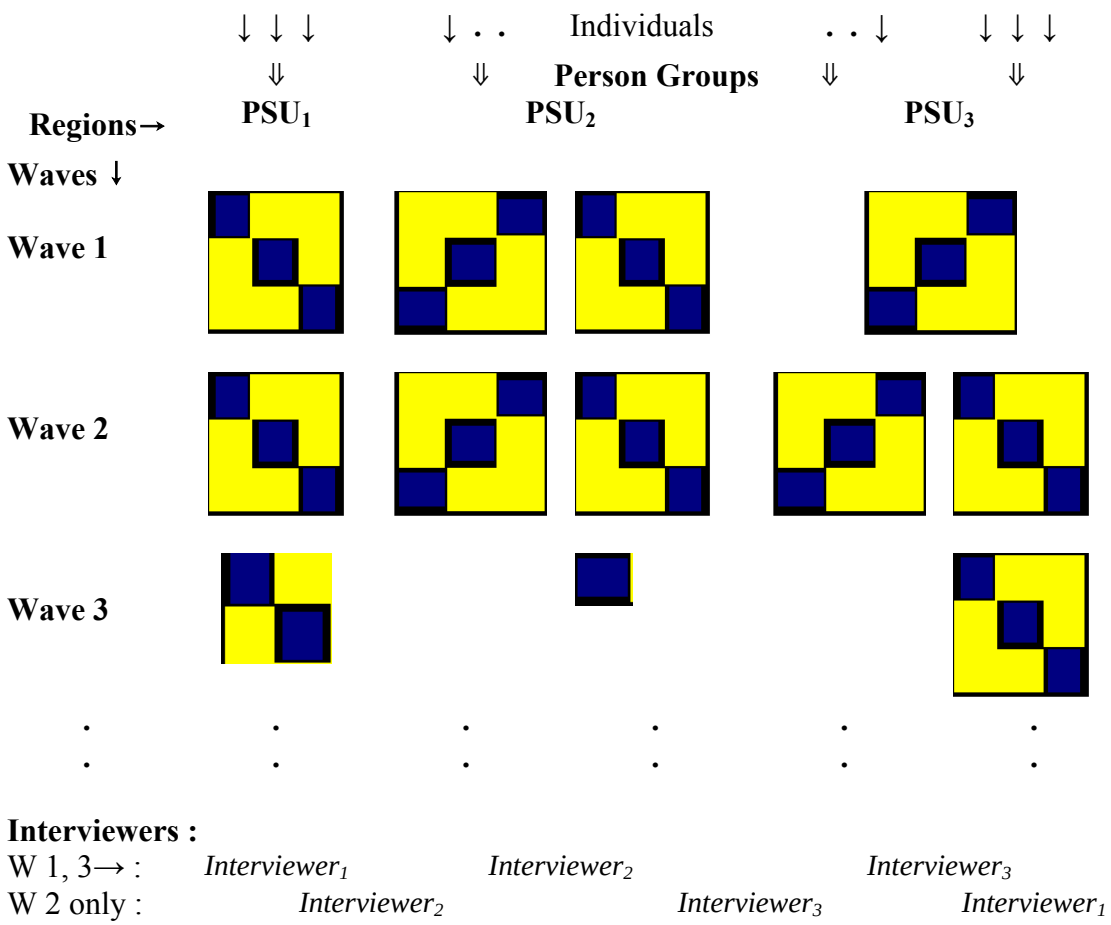
Also as mentioned above, the BHPS's design has received criticism from some quarters for the nature of its regional stratification (eg Davies & Crouchley 1989). One point is that the convenience regional clustering means that membership of the sample in any area is correlated with (endogenous to) any other variables which have some regional influence, such as most measurements of economic positions in the UK. Such a relationship can jeopardise the validity of a wide class of statistical models which might otherwise be attempted for the relevant processes. A second issue is that after the selection of a relatively low number of random geographical sampling units, regionally based analyses using the BHPS becomes problematic due to skewed sub-regional representations. Purchase on the latter problem can be seen from an oft-quoted anecdote with regard to the original Scottish members of the BHPS. Because of an uneven allocation of BHPS sampling districts within Scotland, for many years there were more BHPS households contacted in Dundee than in Glasgow.

- Complex clustering patterns

As a household panel dataset, BHPS records are clustered such that multiple panel records are nested within individuals, and multiple individuals are nested within households. The stratified random design means that a household's cases are also nested within interviewer and geographical unit ('PSU') clusters. The set of relations are illustrated in figure 3, taken from Lambert (2001), a downloadable conference paper which discusses these issues in greater detail.

[10] see preceding footnote.

Figure 3 : Illustration of clustering in the BHPS



The basic problem with this degree of clustering is that individual records are likely to be both similar to and dependent upon the properties of their cluster sharers, and these properties may complicate and / or bias analyses which try to deal with individual level processes. (Though, of course, the 'clustering' of multiple responses within an individual record is considered desirable and specific longitudinal accounts of this type of clustering are well developed). Lambert (2001) suggests that the most important of the non-panel clustering influences are those associated with the household, or 'person group' (although more than one definition of a 'person group' may be considered). Some methods of modelling these complex relationships may make a contribution to better individual level model specification – for instance hierarchical random effects models are widely used to account for unobserved hierarchical similarity between cluster sharers, eg O'Muircheartaigh and Campanelli (1998). However the extent of potential clustering interrelationships in the BHPS are so considerable that it is unlikely that mainstream analytical techniques will deal comprehensively with them for some time yet.

4) Modes of analysis with BHPS data

In the next section a selection of methods are discussed, supplemented in some cases with brief results from a selected contemporary application.

4.1) Cross-sectional and repeated cross-sectional analyses

Despite the extensive nature of its multiple longitudinal records, one of the most common uses of BHPS data remains the conduct of cross-sectional analyses. Users should beware that the BHPS was not designed as a cross-sectional survey and the patterns of attrition, wave-on-wave recontact rules, and the incorporation in later waves of extension subpopulations all add their complications. Nevertheless, for each wave record, individual and household level cross-sectional weighting variables are supplied which the datasets, serving to align the survey samples with a nationally representative distribution of core demographic features. Indeed, because of the breadth of the BHPS survey remit, it is often the case that a particular wave's sample can offer a researcher cross-sectional resources which are not available elsewhere. Areas in which the BHPS stands out in such regard include extensive personal and household income information; the ability to link responses between household sharers; the existence of the youth surveys; and the more general ability to link factual and attitudinal variables between a wide variety of topics.

Additionally, a longitudinal element to BHPS cross-sectional analyses (sic) can often be productively adopted. By this, we mean simply that the depth of the BHPS's longitudinal records mean that it is very easy to supplement what is basically a cross-sectional analysis, with a little background information derived from longitudinal sources. Of course such techniques are long established and it seems unneccesary to refer to them as longitudinal. Common examples include the use of variables indicating previous work experience as predictors in human capital models, or the use of information on whether a respondent has ever been, say, unemployed, or has had a child[11]. We mention this here, though, because the wide scope of the BHPS includes access to many such possible variables – when a researcher is familiar with the dataset, it becomes very easy to quickly investigate cross-sectional patterns with supplementary information from the panel files.

[11] Additionally, some more advanced examples of such uses of historical information are at the forefront of progressive social science methodologies and require a depth of information only available from surveys on scales close to the BHPS. In one example, Gershuny (2002, 2000) advocates the construction of aggregate measures of the likely social position of an individual as assessed through retrospective evidence on the likely progression of life circumstances. In another, DiPrete (2002) has suggested that a significant variable, missing from previous econometric analyses of social differentiation, may be a representation of 'cumulated advantage'. This concept reflects (perceived) inertia in individual's 'breaks' in life, which might be operationalised (through a latent variable) via panel data summaries of prior circumstances.

All of these points are then doubly true of the use of the BHPS resources for repeated cross-sectional analyses, namely when conducting analyses within a particular wave but then comparing similar analyses between different waves. The flip side, however, of the survey's non-cross-sectional design is particularly important in this respect: non-random changes in the sample composition by later waves must always be appreciated by researchers using the data in this way.

We take as a contemporary illustration a claim from a recent MORI survey, that Briton's are increasingly describing themselves as 'working class', in contradiction to the class or social position categories that traditional social science methods would assign them to (see appendix 1 to this document, a summary of a report given by the BBC, 21.8.02). The MORI report argues that there has been a big jump in this trend in the last 5 years, which one academic, Richard Scase, attributes to recent cultural trends : "*It's fashionable to be anti-establishment and speak with an accent and we're influenced by working class role models like Jonathan Ross and Ben Elton*" (as quoted on the BBC site). On the other hand, to read through the many comments on the subject posted through the BBC site, it is clear that an interpretation of parental background is a fundamental element of many social stratification self-descriptions.

It is possible to examine some of these claims very quickly from the nationally representative BHPS samples. Respondents were asked questions about subjective social class identity at three waves, A (1991), E (1995) and J (2000). The first panel of table 3.1 summarises descriptive statistics for the proportion of the adult population at each wave who describe themselves as working class after weighting with the cross-sectional sampling weights. The aggregate statistics do not seem to show the same swing as suggested by the MORI poll[12], though there is a small upturn in the percentage who describe themselves as working class in the 2000 wave. Significantly however, there is no trend in the aggregate pattern of association between self description and occupational position as assessed by the CAMSIS score of male's occupations (CAMSIS scoring measures can be used to indicate the relative social stratification location typically associated with an occupation, see Prandy & Lambert 2002). In all waves, an approximately equivalent, and relatively small, Eta association statistic between higher CAMSIS score (greater advantage) and the tendency to self-description as not-working class, is observed.

The second panel of table 3.1 shows results from some simple logistic regression models predicting the odds of describing oneself as working class, using a number of variables which attempt to operationalise some of the concepts mentioned in the MORI report. There is some fluctuation in the relative importance of the different predictors between the waves, but there are at least a few trends in the most recent model which might agree with the claims of the report. Whilst the influence of parental background remains constant and strong, the influence of current occupational position declines between waves 6 and 10, whilst the significance of the interaction term between age and the CAMSIS measure of current occupational advantage on the last wave only, suggests that it is older rather than younger people

[12] They are also of a different magnitude to the MORI 2002 report, but since we do not know the nature of the MORI questionnaire it is unwise to make firm conclusion on this comparison.

who have a more linear relationship between employment advantage and a lower chance of regarding themselves as working class. Additionally, the significant associations between the variable indicating a belief that it is unfair if money helps buy an individual better heath service treatment (which is included as a possible indicator of 'fashionable values'), but the lack of significance of labour party support on subjective class identity, might both correspond with the alleged shift from overt economic to 'cultural' value determinants.

4.2) Panel data analysis

The BHPS's major selling point is the ability through the panel data to link records on the same individuals between waves, and thus to ask questions about the processes and transitions experienced by individuals and their families (see Rose 2000a for a review). A convenient distinction can be drawn between two methods of analysing the BHPS panel resources. The first involves analysing transitions for cases by collating several variables with response properties at multiple points in time into an individual level variable by case matrix. This uses the 'wide' data format described above in Figure 1: a single case in a data file contains information on the respondent's state in selected variables at two or more points in time, then properties of cases are analysed in terms of patterns of differences. This style of analysis has probably involved the larger number of BHPS research investigations. A clear attraction is the clarity with which transition state categories (for instance, 'those moving out of unemployment between waves A and D'), can be defined, communicated and analysed against other information. Some fields in which such analyses are well established include the assessment of characteristics of individuals against the nature of any recent income transitions (eg Jarvis & Jenkins 1995), employment transitions (eg Taylor 2000), or movements in geographical locations (eg Buck 2000).

Whilst methods for the analysis of transitions can be presented as essentially descriptive (though it is certainly the case that very complex models of transition relations can be developed, eg Jenkins 1999), an alternatively more analytical method of utilising the BHPS panel records is to build up predictive models for individual or household outcomes in terms of multiple pieces of information on the circumstances of cases at each recorded time point. This uses the 'long' data format described by Figure 2. An advantage of this approach is that the analysis of the combined role of multiple variables in predicting the outcomes of interest is likely to lead to better assessments of the relative role of each factor over time. However, a disadvantage is the relative complexity of the statistical methods needed to adequately specify a model which involves multiple information contributions from the same cases. A widely used analytical technique, the variance components panel model (Laird & Ware 1982), allows the specification of models where multiple cases from one individual are 'nested' within the cluster, using a random error term to account for similarities between the cases involved. However, there are many other features to the potential explanation in panel data models which are not dealt with by the variance components model, most notably the desirability of incorporate a reflection of dynamic state dependency (it has been argued that an unavoidable component of any

social explanation is simply information on the circumstances at an earlier point in time, Davies 1987). Literature on the adequate specification of such predictive models with panel data is available (eg Engel & Reinecke 1996, Hsiao 1986), but the increasingly complex methods have been less widely adopted in most social science disciplines, with the clear exception of economics.

Returning to the example above concerning subjective social class categorisation, it is clear that many of the hypotheses discussed in the reporting of the MORI findings involve assumptions about processes of change, namely that individuals have changed in their self defined class locations. It is suggested that changes may be due to, for instance, current or changing occupational positions, or the fickle adoption of 'fashionable' values. Tables 4.2 and 4.3 summarise brief results from investigation of these issues using the two panel analysis techniques mentioned.

First, Table 3.2 shows that there is a very high level of instability in subjective social class descriptions, as a large proportion of the sample are involved in definition transitions (although not shown, their occupational and income transitions over the same period are nothing like as varied). The second panel of Table 3.2 then examines any patterns of association found between explanatory variables and an indicator of transitions in self-definition, namely whether or not an individual defines themselves as working class in wave 10 when they did not do so in either wave 1 or 6. We find that measures of father's advantaged occupational position, and of own advantage in occupational position at any given wave, are all associated with a lesser chance of making a transition to self-definition as working class, a finding which is not consistent with the expectations from the MORI report (no relation would be expected). Surprisingly however, only a very small and non-significant association between a decline in the level of advantage in own occupational position, and changing self-definition, is seen, and this is more in line with the expectation that self-definition is becoming divorced from occupational location. A similar set of patterns is seen with the definition of occupations in terms of manual or non-manual status (as operationalised by reclassifying Registrar General's social class categories). The finding that both transitions into manual, *and* those into non-manual jobs, are associated with a shift to self-definition as working class, is consistent with a hypothesis of the declining influence of current occupational position. Lastly, the association between a putatively 'fashionable' value system at wave J (the belief that it is wrong for wealth to influence medical provision), and transition into self-definition as working class, is also in line with the claims of the MORI report.

Next, table 3.3 reports results from regression models predicting subject identity as working class on the pooled unbalanced panel dataset. Unlike the repeated cross-sectional models of table 3.1, these regressions attempt to distinguish the effects of age and period on attitudes. They suggest over the 3 waves concerned, that after controlling for the effects of age and, in the variance components model, intra-unit clustering, there is a distinguishable trend over time towards in increase in the chances of a 'working class' self-description. There is no interaction effect detected between time period and age (or occupations, not shown), but there is an interaction effect whereby older respondents give more weight to their occupational advantage in not

describing themselves as 'working class', than do younger respondents – again in line with some of the claims of the MORI report. These models are cursory examples however, and a more thoroughly evaluated panel design, possibly with a more accurate indicator of time period or an allowance for lagged dependence effects, may yield different results.

4.3) Life history analyses

'Survival' or 'event history' models lay a slightly different emphasis on the analysis of social change, namely the exploration of which individual factors are associated with a greater risk of an event occurring sooner or later. In this way researchers have analysed the factors associated with, for instance, transitions between labour market participation status's (McCulloch & Dex 2001), and between alternative demographic household situations (Chan & Halpin 2000). As one simple example, the schema below shows factors highlighted by McCulloch and Dex (2001) as being associated with women's employment status transitions.

Increases risk of leaving work for housekeeping
-Older at age of marriage; has a child; has a pre-school child; married more recently
Decreases risk of leaving work for housekeeping
-Already been working for > 5 years; Husband higher education level
Increases risk of leaving housekeeping for part-time work
-Recently moved into housekeeping; child is of school age; married more recently
Decreases risk of leaving housekeeping for part-time work
-Younger age at marriage; higher number of children; Husband's poorer background;
(Source: Selected models from McCulloch & Dex 2001)

The analysis of event history records from the BHPS, typically using the separately supplied life history dataset, is achieved by accessing and constructing 'event' databases. These can be structured in two ways. In an event oriented dataset, each record indicates a distinctive life event, with associated information on the starting and ending situation, the duration of the event, and any other characteristics associated with the event or the individual in question. In a discrete time framework, each case indicates an equivalent period of time (typically a month), which is then marked with information about the nature of any event which occurred within that month and the individual to which it refers. Indeed, the BHPS files supplied in the Life History datasets include examples of both formats. However, approximately equivalent forms of survival or event history models can be estimated in either circumstance.

It is also worth noting that survival models are not the only legitimate uses of the information contained on the BHPS event history files. The BHPS's records may also be used to summarise or overview aggregate structures to the career or demographic histories of individuals (cf Munoz-Bullon & Malo 2003; Pollock et al 2002; Lambert & Prandy 2002, Gershuny 2000, Halpin & Chan 1998). They can also serve a purpose in simply locating specific information on respondent's backgrounds (or those of their

household sharers), such as with records of aggregated labour force experience. In this respect, one possible role for the life history files in analysing our information on changes in subjective social class definition, would be to include information on the duration in the current employment position as a predictor of subjective evaluation. It might be expected for example that more stable employment trajectories would correspond with more 'conventional' subjective positions.

5) Summary

It is perhaps fitting that the 'analytical' section of this writing is cursory – to become so preoccupied with understanding and describing the BHPS's complexities that an intended substantive investigation gets abbreviated, is a pretty typical researcher experience of the survey! It is indeed a widely expressed concern in certain circles that too many of the BHPS's current users are 'technical, problem solving types', whilst those researchers who might do a better job substantively with the BHPS's hugely promising resources, largely remain distant, apparently scared by an image of the complexity of the survey. Yet, much of that fear of the BHPS is itself misplaced – the truth is that many of the 'great complexities' of the BHPS and options for its analysis can be communicated and understood reasonably quickly, and it is with such a purpose in mind that media such as these notes are produced.

Tables referred to in the text

Table 1: Data files in the 2006 core BHPS dataset,
Essex data archive study number 5151, waves 1 to 14 (1991-2004) inclusive

File names (extensions include *.por *.exp, *.dta etc, depending on format files are supplied in):

aegoalt	begoalt	cegoalt	degoalt	eegoalt	fegoalt	gegoalt
ahhresp	bhhresp	chhresp	dhhresp	ehhresp	fhhresp	ghhresp
ahhsamp	bhhsamp	chhsamp	dhhsamp	ehhsamp	fhhsamp	ghhsamp
aincome	bincome	cincome	dincome	eincome	fincome	gincome
aindall	bindall	cindall	dindall	eindall	findall	gindall
aindresp	bindresp	cindresp	dindresp	eindresp	findresp	gindresp
ajobhist	bjobhist	cjobhist	djobhist	ejobhist	fjobhist	gjobhist
	bindsamp	cindsamp	dindsamp	eindsamp	findsamp	gindsamp
	bchildad	clifejob	dyouth	eyouth	fyouth	gyouth
	bchildnt					
	bcohabit					
	blifemst					
	bmarriag					
hegoalt	iegoalt	jegoalt	kegoalt	legoalt	megoalt	negoalt
hhhresp	ihhresp	jhhresp	khhresp	lhhresp	mhhresp	nhhresp
hhhsamp	ihhsamp	jhhsamp	khhsamp	lhhsamp	mhhsamp	nhhsamp
hincome	iincome	jincome	kincome	lincome	mincome	nincome
hindall	iindall	jindall	kindall	lindall	mindall	nindall
hindresp	iindresp	jindresp	kindresp	lindresp	mindresp	nindresp
hjobhist	ijobhist	jjobhist	kjobhist	ljobhist	mjobhist	njobhist
hindsamp	iindsamp	jindsamp	kindsamp	lindsamp	mindsamp	nindsamp
hyouth	iyouth	jyouth	kyouth	lyouth	myouth	nyouth
			kmarriag	lmarriag	mchild	
			kcohabit	lcohabit		
			kchildad	lchildad		
			kchildnt	lchildnt		
			klifemst	llifemst		
				lchild		
xwaveid	xwlsten	xwavedat				

File suffix references:
egoalt: indexes for intra-household links
hhresp: basic characteristics of household
hhsamp: respsone levels by household
income: measures of income and sources
youth: Interviews with 11-16 year olds
childad, childnt, cohabit, marriag: life history information on key demographic events, pre-dating first panel entry
child: specialist data on the children of a respondent
indall: basic characteristics of all household members (includes children, non-interviewed)
indresp: full dataset of all adults' responses
jobhist: details of employment history through period since last interview or last 12 months
indsamp: information on inter-wave links
clifejob: pre-1991 life history information on employment circumstances

Cross wave files:
xwaveid: Information on the cross-wave response patterns of each BHPS individual
xwlsten: Information on the latest known sample status of each BHPS individual
xwavedat: Harmonised data on certain fixed-in-time variables (e.g., ethnicity; parental occupation)

Table 2: Sample membership of the BHPS, individuals in Waves 1-14

	Sampling design membership*							
	'Essex' sample		Extension samples					
	OSM	TSM	ECHP boost	Scotland boost	Wales boost	N. Irel boost	Total sample	Total adults interviewed
Wave:								
A: 1991	13,840						13,840	10,264
B: 1992	12,567	584					13,151	9,845
C: 1993	12,219	885					13,104	9,600
D: 1994	11,821	1030					12,851	9,481
E: 1995	11,425	1124					12,549	9,249
F: 1996	11,412	1308					12,720	9,438
G: 1997	11,251	1301	2490				15,042	11,193
H: 1998	11,161	1300	2374				14,835	10,906
I: 1999	10,996	1339	2258	3395	3577		21,565	15,623
J: 2000	10,773	1481	2193	3582	3573		21,602	15,603
K: 2001	10,624	1610	2125	3516	3523	5188	26,586	18,867
L: 2002	10,470	1664		3327	3385	4589	23,435	16,597
M: 2003	10,173	1701		3177	3313	4210	22,574	16,238
N: 2004	10.063	1740		3099	3285	3940	22,127	15,791
All years†	16,971	5380	2852	4502	4627	5655	39987	

* OSM: Original sample member (includes Permanent Sample Members); TSM: Temporary Sample Member (Essex sample only). After the first wave of collection, the boost (extension) and ECHP sample figures shown include their own TSM's when relevant.
† Total number of individuals ever contacted by the BHPS at any wave between waves 1-14.

Table 3.1: The BHPS as a repeated cross-sectional resource – Subjective social class in 1991, 1996 and 2000

	Wave A 1991	Wave F 1996	Wave J 2000
Descriptive statistics			
Percent 'Working Class':			
All adults	44.5 (9149)	43.7 (8362)	47.1 (7842)
Men only	45.4 (4432)	45.4 (3957)	48.6 (3548)
Eta-correlation between higher occupational level, CAMSIS scores, and category 'not WC':			
Male employees	+0.30 (2987)	+0.35 (2585)	+0.33 (2384)
Values weighted by cross-sectional BHPS weighting factors (weighted n)			
Logistic regressions predicting probability of describing self as 'Working Class' (Adult male employees)			
Continous:			
Own higher CAMSIS	{-0.02}	-0.05**	-0.02
Father's higher CAMSIS	-0.03**	-0.03**	-0.03**
Age in years	{0.00}	{-0.02}	{0.01}
(Age*own CAMSIS) / 100	{-0.04}	{-0.01}	-0.05*
(Age*Support labour)	{0.01}	{0.01}	0.01
Categorical dummy variables:			
Supports labour party	{0.37}	{0.12}	{-0.12}
Agree that it is "unfair that wealth buys medical priority"	{0.09}	0.26*	0.16*
Nagelkerke R-2 approximation	0.22		
N	2,472	1,796	2,971

* /** / {} : Unstandardised parameter estimates significant at 95 / 99 / lt 90 percent significance criteria

Table 3.2: Associations with transitions between subjective social class categories, BHPS waves A (1991), F (1996) and J (2000)
(uses approx 6,000 cases who are interviewed at each wave)

Aggregate Transitions between subjective position:
all adults, N: not working class; W: working class;

	F-N	F-A	I-N	I-A
A-N	2350	675	2085	753
A-W	659	1767	585	1735
F-N			2179	621
F-A			509	1886

Associations with transition to 'working class', defined as W at wave I and N at either A or F (no such transition for 4311 adults, transition to working class for 1061)

	Eta-correlation		Phi association
CAMSIS advantage in..†		**Manual job A**	+0.07**
Fathers job	-0.12**	**Manual job F**	+0.10**
Wave A job	-0.07**	**Manual job J**	+0.10**
Wave F job	-0.10**	**Move to manual**	+0.06**
Wave J job	-0.10**	**Move to non-manual**	+0.03
J job – (A or F)	{0.02}	**Vote labour J**	{-0.00}
		Unfair rich get medic.	+0.02*

* / ** / {} : Association significant at 95 / 99 / lt 90 percent significance criteria
† : Working males only

Table 3.3: Panel data models for subjective social class identity, BHPS waves 1, 6 and 10, male workers

	No VC (SPSS)	Variance components panel model (MlwiN) 1st Order MQL			PQL-2
Wave of record as years	0.04**		0.04**	0.04**	0.06**
Own job CAMSIS	-0.02**	-0.04**	-0.04**	-0.04**	-0.05**
Father's job CAMSIS	-0.03**	-0.03**	-0.03**	-0.03**	-0.05**
Age in years	{0.00}	-0.01**	-0.01**	-0.02**	-0.02**
Votes labour	{0.18}	0.51**	0.49**	0.48**	0.65**
Unfair that money buys indvs health service	0.16**	0.14**	0.16**	0.16**	0.23**
Age*Wave				{0.00}	{0.00}
Age*Own CAMSIS	-0.04			-0.04*	-0.06*
Age*Labour voting	0.01			0.01	0.01
Nagelkerke R-2	0.224				
Person level variance†		0.843	0.859	0.861	3.588

Explanatory variables were 'centred' for variance components model but not for SPSS estimate
* / ** / {} Unstandardised parameter estimates significant at 95 / 99 / lt 90 percent significance criteria
† Person level ('level 2') variance estimate, repeated measures (level 1) estimate=1. Relative size of estimates can be related to Intra-cluster correlations, see eg Goldstein (1995)

Appendix 1: Summary of report from BBC webpage, http://news.bbc.co.uk , as at 21.8.02

What makes you feel working class?

Most British adults like to call themselves working class even though half of them fit the middle class definition, Mori research has shown.
The popular sentiment shows a shift in thinking, as in 1994 only 51% felt they were working class compared with 68% today. And despite more people thinking they belong to it, the social group has actually shrunk in size.

Working class men live off their skills and wits instead of their education and women live near to their mothers and share childcare with them, according to former Labour MP Joe Ashton's definition.

It's fashionable to be anti-establishment and speak with an accent and we're influenced by working class role models like Jonathan Ross and Ben Elton, says sociologist Professor Richard Scase.
What makes you feel working class? Why do we want to be working class? What makes it fashionable? Are class perceptions influenced by a Labour government?

I was bought up in working class Neasden by parents who thought education was important. Ever since I've realised that a belief in education is the single biggest separator between the two classes. Academic children are definitely discriminated against in working class society. I've lived in other cultures, notably Germany, where this doesn't apply. There everybody thinks education is a good thing. There is far more social mobility as a result.
Andy Edmonds, England
Ben Elton and Jonathan Ross working class? Just like Jamie Oliver is a cockney? The real working class are the only ones who don't feel it is cool to be so as we cannot afford to live in a decent manner.
Seamus, UK

Just like Jamie Oliver is a cockney?

Seamus, UK

Working class is only fashionable when you can afford it. Otherwise, for those who genuinely meet the criteria, it often means reduced life opportunities through lack of money and lack of education: Not desirable and certainly not fashionable. Those with money and education can be fashionably working class. But of course, they're only pretending.
Chris Parker, Germany (UK really)
Surely the definition applies to how you were brought up, not what you are now?
Kieron, Germany
I consider myself to be working class because my job involves producing something concrete rather than sitting behind a desk moving paper around. I live in a rented flat not a semi with a mortgage my grandchildren will have problems paying. I wear jeans and a t-shirt all the time, and I don't worry about whether I should say "toilet" or "serviette".
Jim Allen, Scotland

I don't worry about whether I should say 'toilet' or 'serviette'

Jim Allen, Scotland

The British obsession with class in a society where it is meaningless is one of the big sources of social division today. Time our nation grew up, I think.
Simon Richardson, UK
Carpenters, electricians, plumbers are all probably considered to be working class in the definition above. However, a carpenter learns his trade as an apprentice - it's all education in my book. What's the difference between learning your skills from a tradesman or teacher/lecturer? I think the working class has all but ceased to exist.
David, UK

I am public school and university educated. Unfortunately, I regularly have to hide my background by masking my accent. Whether I am taking a cab, shopping or ordering a drink in a pub, I have found a degree of discrimination if I use my real accent and do not splatter my conversation with "mate" or "cheers". It has also been a problem in interviews, as well, with more people being put-off by the Queen's English than discriminate in favour of it.

I have to hide my background by masking my accent

Tim, UK

Tim, UK

I do not feel working class, I do not want to be working class, and if it is fashionable to be working class, I am too much of an individual to follow the herd. If class perceptions are fashioned by a Labour government, maybe it is because so many of our 'beloved' rulers come from such privileged backgrounds that they feel embarrassed.

Marie Cameron, UK

I've always thought of myself as working class and am proud of it. My dad was a coal miner and my mum a housewife. We lived in a council house throughout my childhood and my parents are still there. I am now a graduate and a houseowner, but I still think of myself as working class, perhaps because of the values instilled in me by my parents. I see no need to aspire to middle class-dom. I am proud to be a 'working class achiever'.

I am a graduate and houseowner but think of myself as working class

Lesley, Scotland

Lesley, Scotland

Class is one of those things that baffles me - how do you know what class you fit into? I know I don't have a clue.

Gillian, Northern Ireland

Why would anyone want to be working class? Surely everyone with any sort of pride or self worth aspires to become middle or upper class? Whoever could be proud of themselves in a dead end job, living on a council estate and claiming off the state? People who claim to be proud of being working class seem to me to be the ones who have accepted that they are failures.

Why would anyone want to be working class?

Ed, UK

Ed, UK

My dad is a bus driver, my mum used to work in a factory, I am working class and proud. I support the idea of unions, the ideals of fair wages and employee rights. The idea that being middle class is going out of fashion is quite scary. Where were all these so-called working class people when Thatcher spent the 80s destroying the country?

vish, UK

This is just victim mentality shining through. The cliche of "fighting against all odds" to succeed is still admired. Even those who have enough money to live comfortably would rather portray themselves as a working class hero done good.

Matt, Amsterdam, The Netherlands

I come from an aristocratic background - my family used to own Tredegar House in Wales but lost all their money. I now consider myself to be working class - I married a man from Yorkshire, trained to be a nurse and live in a small terraced house. I could give myself airs and graces and pretend to be upper class but at the end of the day I'm living a working class life so that must be what I am.

I come from an aristocratic background

Louisa Morgan, UK

Louisa Morgan, UK

Is it really necessary to fit each individual into some class category? I work for a living, am not laden with cash, do not have any pretences about my social standing, but due to my reasonable income, probably would be considered working class. It's about time we started moving away from these social stigma's of what we're supposed to be and face the fact that we're all British, all equal, and proud of it!

Mitch, UK

Class isn't easy to define in this country anymore. Nowadays the divide is less class based and more a clear divide between rich and poor people. With greater numbers getting educated to a higher standard the working class element doesn't really apply any more. I know lots of people who say there working class but have money for holidays, cars etc. Working class is being short of money, plain and simple.

Working class is being short of money, plain and simple

Lu, England

Lu, England

References

Bardasi E, Jenkins SP, Rigg JA. 2001. *Documentation for Derived Current and Annual Net Household Income Variables, BHPS Waves 1-9*. Colchester: Institute for Social and Economic Research, University of Essex. 17 pp.

Berthoud R, Gershuny J, eds. 2000. *Seven Years in the Lives of British Families*. London: Policy Press (in association with The Institute for Social and Economic Research, University of Essex)

Brynin M. 2000. Political values: a family matter? In *Seven Years in the Lives of British Families*, ed. R Berthoud, J Gershuny, pp. 193-214. Bristol: Policy Press and Institute for Social and Economic Research

Buck N. 2000. Housing, location and residential mobility. In *Seven Years in the Lives of British Families*, ed. R Berthoud, J Gershuny, pp. 133-60. Bristol: The Policy Press and the Institute for Social and Economic Research

Buck N, Gershuny J, Rose D, Scott J, eds. 1994. *Changing Households : The BHPS 1990-1992*. Colchester: ESRC Research Centre on Micro Social Change

Chan TW, Halpin B. 2000. Who marries whom in Great Britain? Colchester: Working paper 2000-12 of the Institute for Social and Economic Research, University of Essex

Coxon APM, ed. 1991. *Sample Design Issues in a Panel Survey : The Case of the British Household Panel Study*. Colchester: Working Paper 3 of the Institute for Social and Economic Research, University of Essex

Crouchley R, Oskrochi G. 2000a. *Testing for Sample Selection Bias in the Labour Market Behaviour of Respondents from the British Household Panel Survey*, Working paper 2000-02 of the Centre for Applied Statistics, Lancaster University, and http://www.cas.lancs.ac.uk/wpapers.html

Crouchley R, Oskrochi G. 2000b. *Using SPSS and Gauss to unify the BHPS work history data*, Working paper 2003-03 of the Centre for Applied Statistics, Lancaster University, and http://www.cas.lancs.ac.uk/wpapers.html

Crouchley R, Oskrochi G. 2001. *A random effects treatment of dropout in the multi-spell multi-state labour market panel data of the BHPS*, Working Paper 2001-02 of the Centre for Applied Statistics, Lancaster Univerity, and http://www.cas.lancs.ac.uk/wpapers.html

Dale A. 2006. Quality issues with survey research. *International Journal of Social Research Methodology.* 9: 143-158.

Davies RB. 1987. The limitations of cross-sectional analysis. In *Longitudinal Data Analysis*, ed. R Crouchley. Aldershot: Averbury

Davies RB, Crouchley R. 1989. Sampling for a Panel Study : The Consequences for Statistical Modelling. In *Sample Design Issues in a Panel Survey : The Case of the British Household Panel Study*, ed. APM Coxon. Colchester: University of Essex

DiPrete TA. 2002. *The Process of Cumulative Advantage as a Mechanism for Trends in Inequality: An Exportation of Possible Linkages*. Presented at ISA 15th World Congress of Sociology, Research Committee 28 On Stratification and Mobility, Brisbane, Australia

Elias P. 1997. *Who Forgot They Were Unemployed?* Colchester: Working Paper 97-19 of the Institute for Social and Economic Research, University of Essex

Engel U, Reinecke J. 1996. *Analysis of Change : Advanced Techniques in Panel Data Analysis*. New York: Walter de Gruyter

Ermisch J and Wright, RE. 2005. *Changing Scotland – Evidence from the British Household Panel Survey.* Bristol: Policy Press.

Gershuny J. 2000. Social Position from Narrative Data. In *Renewing Class Analysis*, ed. R Crompton, F Devine, M Savage, J Scott. Oxford: Blackwell Publishers

Gershuny J. 2002a. *A New Measure of Social Position: Social Mobility and Human Capital in Britain*. Colchester: University of Essex, Working paper 2002-02 of the Institute for Social and Economic Research

Gershuny J. 2002b. Editorial: Co-ordinating longitudinal survey data in the UK: towards a national strategy. *Journal of the Royal Statistical Society, Series A.* 165: 3-7.

Gershuny J. 2004. Domestic equipment does not increase domestic work: A reply to Bittman, Rice and Wajcman. *British Journal of Sociology.* 55: 425-431.

Goldstein H. 1995. *Multilevel Statistical Models*. London: Edward Arnold

Halpin B. 1998. Unified BHPS work-life histories : Combining multiple sources into a user-friendly format. *Bulletin de Methodologie Sociologique* 60: 34-79

Halpin B. 2002. *British Household Panel Survey Combined Work-Life History Data, 1990-1999 [computer file]. 3rd ed*, Economic and Social Research Council Research Centre on Micro-Social Change, University of Essex, Institute for Social and Economic Research; distributed by The Data Archive, University of Essex, Colchester

Halpin B, Chan TW. 1998. Class Careers as Sequences : An optimal matching analysis of work-life histories. *European Sociological Review* 14: 111-30

Holt T. 1989. Panel Conditioning : Discussion. In *Panel Surveys*, ed. D Kasprzyk, GJ Duncan, G Kalton, R Singh. New York: Wiley

Hsiao C. 1986. *Analysis of Panel Data*. Cambridge: Cambridge University Press

Jarvis S, Jenkins SP. 1995. *Do the poor stay poor? : New evidence about income dynamics from the Brisish Household Panel Survey*, Occasional Paper 95-2 of the Institute for Social and Economic Research, University of Essex, Colchester

Jenkins SP. 1999. *Modelling Household Income Dynamics*, Working paper 99-01 of the Institute for Social and Economic Research, Univeristy of Essex, Colchester

Laird NM, Ware JH. 1982. Random-effects models for longitudinal data. *Biometrics* 38: 963-74

Lambert PS. 2001. *Individuals in household panel surveys: dealing with person-group clustering in individual level statistical models using BHPS data*. Presented at British Household Panel Survey Research Conference, Colchester, UK, and www.iser.essex.ac.uk/activities/bhps-2001/docs/pdf/papers/lambert.pdf

Lambert PS, Prandy K. 2002. *Intra-generational Mobility and Social Distance: Work History Analysis and Occupational Structure*. Presented at ESRC Seminar Series: Occupational Information in Social and Economic Research: Occupational Careers, 10th May, London School of Economics

McCulloch A, Dex S. 2001. Married Women's Employment Patterns in Britain. In *Careers of Couples in Contemporary Societies*, ed. H-P Blossfeld, S Drobnic, pp. 177-200. Oxford: Oxford University Press

Munoz-Bullon F, Malo MA. 2003. Employment status mobility from a life-cycle perspective: A sequence analysis of work histories in the BHPS. *Demographic Research Online* 9: 119-162.

O'Muircheartaigh C, Campanelli P. 1998. The relative impact of interviewer effects and sample design effects on survey precision. *Journal of the Royal Statistical Society Series A : Statistics in Society* 161: 63-77

Pollock G, Antcliff V, Ralphs R. 2002. Work orders: Analysing employment histories using sequence data. *International Journal of Social Research Methodology.* 5: 91-105.

Prandy K, Lambert PS. 2006. *CAMSIS project webpages, http://www.camsis.stir.ac.uk* . University of Stirling.

Rose D. 2000a. Household Panel Studies. In *Researching Social and Economic Change*, ed. D Rose, pp. 3-35. London: Routledge

Rose D, ed. 2000b. *Researching Social and Economic Change : The uses of household panel studies*. London: Routledge

Scott J. 1993. *Using Household Panels to Study Micro-Social Change*, Working Paper 25 of the Institute for Social and Economic Research, University of Essex, Colchester

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

Swift A. 2000. Class analysis from a normative perspective. *British Journal of Sociology* 51: 663-79

Taylor A. 1994. Appendix : sample characteristics, attrition and weighting. In *Changing Households*, ed. N Buck, J Gershuny, D Rose, J Scott. Colchester: ESRC Research Centre on Micro-Social Change, University of Essex

Taylor M. 2000. Work, non work, jobs and job mobility. In *Seven years on the lives of British families*, ed. R Berthoud, J Gershuny, pp. 73-106. Bristol: The Policy Press and the Institute for Social and Economic Research

Taylor MF, Brice J, Buck N, Prentice-Lane E, eds. 2006a. *British Household Panel Survey User Manual Volume A: Introduction, technical report and appendices*. Colchester: University of Essex

Taylor MF, Brice J, Buck N, Prentice-Lane E, eds. 2006b. *British Household Panel Survey User Manual Volume B: The Codebook*. Colchester: University of Essex

van der Zouwen J, van Tilburg T. 2001. Reactivity in Panel Studies and Its Consequences for Testing Causal Hypotheses. *Sociological Methods & Research* 30: 35-56