



Quality Profile: British Household Panel Survey

Version 2.0:
Waves 1 to 13: 1991-2003

Editor:
Peter Lynn

Contributors:
Nicholas Buck, Jonathan Burton, Heather Laurie, Peter Lynn, S C Noah Uhrig

February 2006

Institute for Social and Economic Research
University of Essex
Wivenhoe Park
Colchester
Essex CO4 3SQ

Institute for Social and Economic Research

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

- ESRC Research Centre on Micro-social Change. Established in 1989 to identify, explain, model and forecast social change in Britain at the individual and household level, the Centre specialises in research using longitudinal data.
- ESRC UK Longitudinal Studies Centre. This national resource centre was established in October 1999 to promote the use of longitudinal data and to develop a strategy for the future of large-scale longitudinal surveys. It is responsible for the British Household Panel Survey (BHPS) and for the ESRC's interest in the National Child Development Study and the 1970 British Cohort Study
- European Centre for Analysis in the Social Sciences. ECASS is an interdisciplinary research centre which hosts major research programmes and helps researchers from the EU gain access to longitudinal data and cross-national data sets from all over Europe.

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

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

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

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

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

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

Table of Contents

1.	Statement of Core Research Purposes	9
1.1	Original design.....	9
1.2	British Youth Panel.....	10
1.3	ECHP low-income and Northern Ireland sub-sample.....	10
1.4	Scotland and Wales.....	11
1.5	Northern Ireland.....	11
2.	Overview of the Survey Design.....	12
3.	Sample Design – Main Sample.....	14
3.1	Target population.....	14
3.2	Sample selection process	14
3.2.1	Sample selection at wave 1	14
3.2.1.1	Selection of PSUs.....	14
3.2.1.2	Selection of addresses	15
3.2.1.3	Selection of households/persons	16
3.2.2	Sample additions and losses at subsequent waves.....	16
3.2.3	Quality control of sample selection process	17
3.3	Sample size	17
3.4	Selection probabilities.....	19
3.4.1	OSMs	19
3.4.2	OSM(D)s.....	20
3.4.3	PSMs and TSMs	20
4.	Sample Design – Additional Samples.....	22
4.1	European Community Household Panel sub-sample.....	22
4.1.1	Sample selection process	22
4.1.2	Sample size	23
4.2	Scotland and Wales extension samples.....	23
4.2.1	Sample selection process	23
4.2.1.1	Selection of PSUs.....	24
4.2.1.2	Selection of Addresses	25
4.2.1.3	Selection of Households / Persons	25
4.2.2	Sample size	25
4.3	The British Youth Panel.....	25
4.3.1	Sample selection process	25
4.3.2	Sample size	25
4.4	Northern Ireland sample	26
4.4.1	Sample selection process	26
4.4.1.1	Selection of Households / Persons	26
4.4.2	Sample size	26
5.	Content of Data Collection Instruments.....	27
5.1	Summary of instruments	27
5.2	Instrument design: principles and procedures.....	28
5.2.1	Pre-testing and piloting	30
5.3	Content of instruments	31
6.	Data Collection	32
6.1	Data collection mode	32

6.2	Fieldwork - general	32
6.2.1	Interviewing	32
6.2.1.1	Interviewers	32
6.2.1.2	Interpenetrating sample experiment	33
6.2.1.3	Interviewer Briefing	34
6.2.2	Field work quality control.....	34
6.2.3	Field work dates	35
6.2.3.1	The ECHP sub-sample	36
6.2.3.2	Scottish and Welsh booster samples	36
6.2.3.3	Northern Ireland sample.....	37
6.2.3.4	Pre-testing and piloting	37
6.3	Fieldwork procedures.....	37
6.3.1	Advance mailing	38
6.3.2	Respondent incentives	38
6.3.3	Sample management	38
6.3.3.1	Active sample.....	39
6.3.3.2	Inactive sample.....	39
6.3.4	Panel maintenance	39
6.3.5	Tracking	40
6.3.6	Refusal conversion.....	40
6.4	Field outcomes and unit response rates	41
6.5	Item non-response	42
6.5.1	Overall item non-response	42
6.5.2	Item non-response for specific variables	43
6.5.3	Wealth, assets and debt component - waves 5 and 10	43
7.	Data Preparation.....	45
7.1	Coding.....	45
7.2	Editing procedures	46
7.3	Data capture	48
7.4	Quality control procedures during data processing	48
8.	Statistical Adjustment Procedures	50
8.1	Weighting procedures	50
8.1.1	Wave 1 weights.....	50
8.1.2	Longitudinal weights	50
8.1.3	Cross-Sectional weights.....	52
8.2	Imputation procedures	53
8.2.1	Hot-deck imputation	54
8.2.2	Regression imputation	54
9.	Documentation and Data Accessibility.....	56
9.1	Data Organisation	56
9.2	Documentation.....	56
9.2.1	BHPS Waves 1-13, 1991-2003	57
9.2.2	BHPS Combined Work-Life History Data, 1990-2003	58
9.2.3	BHPS Derived Current and Annual Net Household Income Variables Waves 1-12, 1991-2002	59
9.2.4	Improving Survey Measurement of Income and Employment (ISMIE) Data.....	59
9.3	Formats, Cost and Access conditions of the BHPS data sets	60
9.4	The Data Documentation Initiative.....	60
10.	Data Usability and Interpretability.....	62

10.1	Data structure	62
10.2	Identifiers	62
10.3	Variable names.....	63
10.4	Missing values	63
10.5	Response data.....	64
10.6	Design, field and processing data	64
10.7	Weights	65
10.8	Data usage	65
11.	Coverage Error.....	66
11.1	Under-coverage of sampling frame/method	66
11.1.1	Cross-sectional populations	66
11.1.2	Longitudinal populations	67
12.	Sampling Error.....	69
12.1	Sampling Variance.....	69
12.2	Sampling bias	71
12.2.1	Bias arising from inclusion of TSMs	72
12.2.2	Bias arising from inclusion of OSM(D)s	73
13.	Non-Response Error.....	74
13.1	Wave 1 non-response	74
13.2	Attrition.....	74
14.	Measurement Error	77
14.1	Unemployment.....	77
14.2	Gross flows	78
14.3	Income sources.....	80
14.4	Dependent interviewing	81
15.	Confidentiality and Ethics.....	82
15.1	Ethical considerations	82
15.2	Informed consent	82
15.3	Collecting contact details	83
15.4	Confidentiality	83
15.5	Feedback to respondents	84
15.6	Refusal conversion.....	84
16.	External Comparisons	85
16.1	Socio-demographic comparisons	85
16.2	Other specific comparisons.....	86
17.	References.....	88
Table 1: Stratification of Postal Sectors.....		92
Table 2: Distribution of selected addresses per PSU		92
Table 3: Adult population in institutions, by institution type and sex		93
Table 4: Numbers of achieved individual interviews at each wave.....		93
Table 5: Sample status and longitudinal response status of interview respondents at each wave		94
Table 6: Number of sequential waves response starting at Wave 1 by age at Wave 1.....		95
Table 7: Number of sequential waves response starting at Wave 5 by age at Wave 5.....		95
Table 8: Number of consecutive pairs of waves response by OSM or PSM, by age in 1991 ..		95
Table 9: Number of observed wave-on-wave employment transitions by OSM or PSM up to wave 13		95

Table 10: Number of observed wave-on-wave demographic transitions by OSM or PSM up to wave 13	96
Table 11: Distribution of within-address selection probabilities of households.....	96
Table 12: Distribution of design weights.....	97
Table 13: Numbers of achieved individual interviews: ECHP sub-sample.....	97
Table 14: Stratification of postal sectors (Scotland).....	98
Table 15: Stratification of Postal Sectors (Wales).....	98
Table 16: Numbers of achieved individual interviews: Scottish booster sample	98
Table 17: Numbers of achieved individual interviews: Welsh booster sample.....	98
Table 18: Numbers of achieved individual interviews: British Youth Panel	99
Table 19: Numbers of personal, proxy and telephone interviews	99
Table 20: Stability of BHPS interviewing panel.....	100
Table 21: Distribution of individual interviews by month.....	100
Table 22: Dates of wave 1 pre-tests and pilots	101
Table 23: Percentage of respondents giving contact details	101
Table 24: Outcome of field work conversion attempts.....	102
Table 25: Wave 1 field outcomes	103
Table 26: Wave 2 Individual outcomes	103
Table 27: Wave 3 Individual outcomes	104
Table 28: Wave 4 Individual outcomes	104
Table 29: Wave 5 Individual outcomes	105
Table 30: Wave 6 Individual outcomes	105
Table 31: Wave 7 Individual outcomes	106
Table 32: Wave 8 Individual outcomes	106
Table 33: Wave 9 Individual outcomes	107
Table 34: Wave 10 Individual outcomes	107
Table 35: Wave 11 Individual outcomes	108
Table 36: Wave 12 Individual outcomes	108
Table 37: Wave 13 Individual outcomes	109
Table 38: Wave 1 Field outcomes for Scotland and Wales extension samples (BHPS wave 9)	109
Table 39: Wave 2 Individual outcomes for Scotland extension sample (BHPS wave 10)....	110
Table 40: Wave 3 Individual outcomes for Scotland extension sample (BHPS wave 11)....	110
Table 41: Wave 4 Individual outcomes for Scotland extension sample (BHPS wave 12)....	111
Table 42: Wave 5 Individual outcomes for Scotland extension sample (BHPS wave 13)....	111
Table 43: Wave 2 Individual outcomes for Wales extension sample (BHPS wave 10).....	112
Table 44: Wave 3 Individual outcomes for Wales extension sample (BHPS wave 11).....	112
Table 45: Wave 4 Individual outcomes for Wales extension sample (BHPS wave 12).....	113
Table 46: Wave 5 Individual outcomes for Wales extension sample (BHPS wave 13).....	113
Table 47: Wave 1 field outcomes for Northern Ireland sample (BHPS wave 11)	114
Table 48: Wave 2 Individual outcomes for Northern Ireland sample (BHPS wave 12)	114
Table 49: Wave 3 Individual outcomes for Northern Ireland sample (BHPS wave 13)	115
Table 50: Mean levels of item non-response - individual questionnaire	115
Table 51: Mean levels of item non-response - household questionnaire.....	115
Table 52: Item non-response levels for selected variables	117
Table 53: Item non-response; wealth, assets and debts questions; core BHPS sample.....	118
Table 54: Item non-response; wealth, assets and debts questions; Scottish boost sample	119
Table 55: Item non-response; wealth, assets and debts questions; Welsh boost sample.....	120
Table 56: Verbatim items coded at each wave	121
Table 57: Validation re-coding results for selected items	123

Table 58: Some distribution statistics for weights	123
Table 59: Analysis of BHPS Usage from Data Archive Records.....	124
Table 60: Published papers, books and reports based on BHPS.....	124
Table 61: BHPS Registered Users by discipline.....	124
Table 62: Complex Standard Errors, DEFTs and 95% Confidence Intervals for a Range of BHPS Variables	125
Table 63: Actual and ideal weights for OSM(Ds)	126
Table 64: Wave 1 household response status by region (row percentages)	126
Table 65: Wave 1 household response status by housing type (row percentages)	126
Table 66: Wave 1 response to individual interview; responding households.....	127
Table 67: Summary outcomes for wave 1 respondents	127
Table 68: Summary outcomes for persons enumerated in wave 1 households	127
Table 69: Impacts of attrition I	128
Table 70: Impacts of attrition II	129
Table 71: Impacts of attrition III.....	130
Table 72: Wave 1 – 2 comparison of SOC major groups	131
Table 73: Wave 1 – 2 comparison of SIC major divisions	131
Table 74: False negative rates in the report of income sources: ISMIE study	131
Table 75: Sample distribution by age and sex, Census and BHPS	132
Table 76: Employment status, Census and BHPS, persons aged under 60	132
Table 77: Socio-economic group by sex, all current employees and self-employed.....	133
Table 78: Ethnic group, Census and BHPS, aged 16 and over.....	133
Table 79: Marital status, BHPS and GHS, aged 16 and over	134
Table 80: Household characteristics, GHS and BHPS, 1991 and 1994/5	135
Figure 1: BHPS tracking procedures	136
Figure 2: Measurement error regarding unemployment – from Elias (1997).....	137
Annex A: ISER Code of Ethics.....	138

1. Statement of Core Research Purposes

1.1 Original design

The central purpose of the British Household Panel Survey (BHPS) is to understand the dynamics of change experienced by the population of Great Britain and its evolution over the lifetime of the survey, from its start in 1991. In contrast to most cohort studies, its sample covers the whole population, and not simply narrowly defined age ranges. While BHPS aims to provide cross-sectional population estimates for the lifetime of the study, its central aim is to facilitate longitudinal research. For related reasons it is primarily produced in order to make micro-data sets available to a wide range of secondary analysts across a range of social science disciplines, and for policy research, rather than to generate reports specified at the time the study was started.

Thus, the aim of the BHPS is to provide high quality, timely data on the short-term processes of change at the individual and household level for the domains with which it is concerned. These domains are: labour markets, income, savings and wealth, household and family organisation, housing, consumption, health, social and political values, education and training. Through the collection of this short-term data, they also allow construction of longer sequences of high quality biographical information across a range of domains. This general statement can be broken into a range of research uses:

- The analysis of the incidence of states and events such as poverty or unemployment over time, which provides a very different understanding of their distribution in society from that provided by cross-sectional data.
- The measurement of the rates of transition between states, and the factors associated with these transitions. This analysis may be based either on repeated annual measures or on the construction of complete histories based on monthly calendars for the periods between waves. This short-term retrospective data is much more reliable than that collected from longer period life histories.
- The design in which all household members are interviewed and followed permits the analysis of associations between the life course of different household members, and how their individual decisions may impact on each other. It also makes the household panel study particularly suited to the analysis of the dynamics of household formation and dissolution, and associated events and outcomes.
- The analysis of the association between change in the different domains (eg health and the labour market), in order both to understand causal ordering, and to understand the wider social impacts key events and processes.
- The analysis of associations between measures in a modelling context which take account of unobserved heterogeneity through the use of repeated measures fixed and random effects models.
- The accumulation of life history data, both within the panel itself, and in the retrospective life histories collected in the early waves makes the panel particularly suited to the analysis of the long-term accumulation of resources (personal and financial), and their impact on later outcomes.

By continuing into a second decade and beyond it will start to provide data on period differences in some of the short-term transition processes (for example, whether jobs and families are becoming still more unstable - something difficult to explore with retrospective data).

Certain new advantages start to emerge as the panel increases in length. For example it becomes possible to analyse some of the longer spells and sequences with full data on antecedents and the evolution of events during the spell. It becomes possible to analyse the impacts of earlier life stages, including for example the impacts of childhood poverty and family disruption on later life circumstances, with information from multiple measurement points, allowing inference about impacts and instability. It also becomes possible to analyse other lifetime acquisition processes, especially of wealth accumulation, allowing prediction of economic circumstances in retirement. It is also becoming an important resource for the analysis of mortality.

Analyses of within household relationships and influences are also informed by better historical information on all household members. It also becomes possible to undertake analyses of inter-generational influences.

An extensive network of household panel studies, throughout Europe, North America and Australasia is already in existence. One key goal is to ensure that BHPS can contribute to international comparative research using such studies.

1.2 British Youth Panel

The BHPS was supplemented in wave four to include children in sample households and this has been maintained in subsequent waves. The age band is 11 to 15 inclusive. In the early years the BYP was funded by the Health Education Authority. The BYP aimed to exploit three important research advantages:

- (1) Because of the transitional nature of adolescence, youth panels are a scarce resource. The BYP is an on-going panel with an increasing pool of transitions which can be studied as new 11-year olds are added and as the cohorts move upwards one year. Every year the number of wave-on-wave and longer transitions increases.
- (2) Equally, as respondents move into the adult survey analysis of their responses in the BYP can be linked to their responses in the BHPS.
- (3) As with the BHPS, the full range of household information is available to enable analysis of the impacts of both home context and of specific relationships, whether with parents, siblings, or other household members.

1.3 ECHP low-income and Northern Ireland sub-sample

From Wave Seven the BHPS began providing data for the United Kingdom European Community Household Panel (ECHP). As part of this, it incorporated a sub-sample of the original UKECHP, including all households still responding in Northern Ireland, and a 'low-income' sample of the Great Britain panel. The low-income sample was selected on the basis of characteristics associated with low income in the ECHP. The sub-samples aimed to extend

coverage to the whole UK, and to increase the sample of lower income households available for analysis. Such households have proved to be a priority group for many BHPS research agendas.

1.4 Scotland and Wales

A major development at Wave 9 was the recruitment of two additional samples to the BHPS in Scotland and Wales. There were two main aims of the extensions. First, to increase the relatively small Scottish and Welsh sample sizes (around 400-500 households in each country in the initial BHPS sample) in order to permit independent analysis of the two countries. Second, to facilitate analysis of the two countries compared to England in order to assess the impacts of the substantial public policy changes that may be expected to follow from devolution. The additional samples contained over 2,000 extra households in each of Scotland and Wales.

1.5 Northern Ireland

At Wave 11, the survey was extended to Northern Ireland with the introduction of a sample of around 2,900 households (5,200 persons). This sample is jointly funded by the ESRC and government departments in Northern Ireland and permits both separate analysis of Northern Ireland and comparative analysis of Northern Ireland with the rest of the UK. From 2001 onwards the survey has therefore been a truly UK-wide survey.

2. Overview of the Survey Design

The BHPS was started as a result of a proposal to the ESRC to establish an Interdisciplinary Research Centre (IRC) at the University of Essex. This bid arose from a call for such IRC's across the whole range of disciplines. The IRC, originally called the Research Centre on Micro-Social Change, had a remit to provide a data resource of value across a range of social science disciplines.

The proposal for the BHPS was very consciously seeking to emulate the success of household panel studies in other countries in Europe and North America. In particular the original inspiration was the Panel Study on Income Dynamics carried out by the Institute for Social Research at the University of Michigan since 1968. However different household panel studies offered somewhat different models. For example, the PSID was based on one interview per family, while the German SOEP had separate interviews for each household member. At various points in this document we draw attention to different design decisions.

The Research Centre on Micro-Social Change was established in April 1989, with key staff coming into post in the summer and autumn of that year. These staff developed the design taking advice from a range of advisory groups including five groups advising on substantive areas of the questionnaire, and one group advising on panel design issues. These groups included a range of academics, and representatives of government department research divisions, representatives of private sector and voluntary organisations, and representatives of panel studies in other countries.

In developing the survey design a number of issues had to be resolved. They are summarised here, and in some cases further discussion is contained in later sections.

The sample design for wave one was based on a clustered, stratified sample of addresses throughout Great Britain south of the Caledonian Canal. There was early discussion of whether an unclustered design could be used, but this was rejected on grounds of costs and difficulties of survey organisation. The initial sample was intended to have around 5000 households, approximately the same as the sample size with which other national panels started. (Issues of sample design are discussed in more detail in Sections 3.2 and 3.4)

Members of original sample households are followed as they move between different households, and if aged 16 and over are eligible for interview. People with whom they form new households are temporary sample members while they live with original sample members, and again will be eligible for interview if aged 16 and over. These following rules, allow the panel to track the processes of family and household change experienced by the population as a whole. (Issues of longitudinal population representativeness are discussed in Section 5.)

In common with most of the other national household panels, waves were set at annual intervals. This is considered the minimum frequency for collecting reasonably reliable data on incomes and labour market activities between waves. It was also decided that fieldwork should be concentrated in a relatively short period within the year, rather than being spread evenly through the year, as it is sometimes the case with cross-sectional surveys.

In common with the GSOEP, but not the PSID, BHPS seeks to interview all adult members of each household. This is in part because of a rather broader focus than that of PSID, and in particular a focus on areas where data collected by proxy from household member would be unreliable or invalid (eg socio-economic values).

The questionnaire was also designed so that most information was collected at the individual level, rather than the household level. There is for example no attempt to collect global summary estimates of household incomes. The questionnaire was somewhat broader in scope, and also somewhat longer than the other national panels, with an interview for each individual of slightly less than 40 minutes. The first three waves of the panel collected a substantial range of information to establish the 'initial conditions' of respondents, including lifetime demographic and employment histories. (Questionnaire design is discussed more fully in Section 5).

3. Sample Design – Main Sample

3.1 Target population

The BHPS is primarily designed to be representative of all persons who are resident in Britain at multiple time points corresponding to the waves of data collection. Technically, this means that there are a large number of potential target populations, corresponding to all possible combinations of waves, including individual waves. Section 3.3 below discusses the extent to which the BHPS adequately represents each of these populations, and related issues.

The BHPS can also be considered to be a study of households and a number of studies based on BHPS data have assumed, either explicitly or implicitly, that it is representative of a population of households. The conceptualisation of longitudinal populations of households and the assessment of coverage of such populations are complex issues beyond the scope of this document.

3.2 Sample selection process

Sections 3.2.1 and 3.2.2 of this document describe the rules defining eligibility for sample membership and the procedures used to select the sample. It should be noted that eligibility for sample membership and eligibility for interview at any given wave of data collection are two separate concepts. There are a number of situations in which an eligible sample member is not eligible for interview and in which a non-sample member is eligible for interview. Eligibility for interview will be described in Section 6.3.3.

3.2.1 Sample selection at wave 1

The initial sample for Wave One of the BHPS in 1991 consisted of an equal-probability clustered sample of 8,167 addresses drawn from the Postcode Address File (Lynn and Lievesley 1991). There were three stages to the selection process – selection of Primary Sampling Units (PSUs), addresses, and households. Here we provide an outline of these three stages. The sampling procedures are described in greater detail in part IV of Taylor *et al* (2001).

3.2.1.1 Selection of PSUs

A list was generated of all postcode sectors on the small users Postcode Address File (PAF) for Great Britain south of the Caledonian Canal (ie excluding Northern Ireland and the North of Scotland). Sectors with fewer than 500 delivery points (addresses) were grouped with an adjacent sector and thereafter treated as a single sector. Sectors were then stratified by region, SEG profile, proportion of pensionable age, proportion of employed persons working in agriculture (in non-metropolitan areas) and proportion of persons living in single-person non-pensioner households (in metropolitan areas). In detail, the stratification proceeded as follows:

- (a) Sectors were sorted into 18 regions (see Table 1).
- (b) Within each region, sectors were ordered by the proportion of heads of households in socio-economic groups 1 to 5 and 13 (that is, the proportion of heads of households in

professional or managerial positions). Within each region, sectors were then split into major strata of approximately equal size (in terms of delivery points). The number of major strata varies by region from two to three, depending on region size (see Table 1), resulting in a total of 41 major strata.

- (c) Within each major stratum, sectors were then re-ordered by the proportion of the population of pensionable age (ie females over 60 and males over 65). The order of sorting alternated between ascending and descending within each successive major strata (ie a serpentine listing), in order to ensure that potential periodicity problems in the frame could be minimised. Each major stratum was then split into two minor strata each of approximately equal size.
- (d) Finally, within each of the 82 minor strata, sectors were re-ordered, using serpentine listing as follows:
 - (i) Sectors in non-metropolitan regions were ranked by the proportion of the employed sector population working in agriculture (denoted AGEMP in Table 1) and
 - (ii) Sectors in metropolitan regions were ranked by the proportion of the sector population that was both under pensionable age and living in single person households (denoted SPH in Table 1).

Two separate factors were used, since they discriminated overall population characteristics more effectively in the two types of area.

A size measure was generated for each sector. In England and Wales this was simply the number of delivery points in the sector. In Scotland the measure was the sum of the MOI for all delivery points in the sector. A systematic sample of 250 sectors was then selected, with probability proportional to this size measure, using a random start and a fixed interval. As well as summarising the definition of strata, Table 1 shows the number of postal sectors selected per region, as an indicator of the relative population size of the strata

3.2.1.2 Selection of addresses

An average of approximately 33 delivery points were sample per selected sector. The number varied slightly (minimum 21, maximum 36) in order to keep the overall selection probabilities of addresses approximately equal. The reason for this variation was that the number of delivery points per sector had changed, for some sectors, between the time when sectors had been selected (some months in advance of field work, in order to allow for field work planning and allocation of PSUs to interviewers) and the time when the delivery points were selected (as close as possible to the field work period). The number of addresses selected per sector was:

$n_i = \frac{N_{i2}}{N_{i1}}$, where N_{i1} was the number of delivery points in sector i at the time when sectors were selected and N_{i2} is the corresponding number at the time when delivery points were selected. The distribution of n_i is shown in Table 2 below. It can be seen that the variation is small.

Delivery points were selected systematically from throughout the whole sector, using an appropriate random start and fixed interval.

3.2.1.3 Selection of households/persons

Any sample address that was found in the field to be non-residential or an institution was excluded from the survey. An institution was defined as

an address at which four or more unrelated people sleep; while they may or may not eat communally, the establishment must be run or managed by a person or persons employed for this purpose by the owner.

At residential addresses, selection of households was carried out by interviewers. At all sample addresses with up to three households present (97.3% of sample addresses), all households were included in the sample. If there were more than three households at an address, a random selection procedure was used to select three households for inclusion in the sample. This was implemented using a Kish Grid procedure.

Ideally, the procedures would have been modified slightly in Scotland to reflect the variation in selection probabilities of addresses. In practice this proved impossible due to fieldwork and organisational constraints. The design weights allow for this by correcting for the realised selection probabilities of households (see Section 8).

The standard OPCS definition of a household was applied:

one person living alone or a group of people who either share living accommodation OR share one meal a day and who have the address as their only or main residence.

Six months continuous residence during the year was a minimum requirement, thus excluding students who might have been at a parental home during vacation. Students sampled at their term-time address were included if this was non-institutional (ie not a hall of residence).

All individuals enumerated in respondent sample households (regardless of age) became part of the longitudinal sample. All these sample members are known as Original Sample Members (OSMs).

3.2.2 Sample additions and losses at subsequent waves

All OSMs identified at wave 1 (regardless of whether they were successfully interviewed at wave 1) remain sample members at all subsequent waves until they die. This includes members of households that were non-contacted at wave 1. (Such cases were issued at wave 2, but in practice a large proportion were non-respondents at wave 2 and no further attempts to contact them were made at subsequent waves.)

It should be noted that OSMs who move into institutions or who move north of the Caledonian Canal in Scotland remain in the sample and attempts are made to interview them where possible. OSMs who move out of England, Scotland or Wales also remain in the sample, though no attempts are made to interview them until or unless they return to England, Scotland or Wales.

Additionally, two categories of new permanent sample members join at all waves subsequent to wave 1. The first category consists of babies born to (or adopted by) an OSM subsequent to wave 1. We shall refer to these as OSM(D), viz. original sample members by virtue of descent. (Note that babies born to an OSM prior to wave 1 will themselves be OSMs if they are in the same household as the OSM at wave 1.) The second category of new sample members consists of parents of an OSM(D), who have joined the household of the OSM parent. We shall refer to these as PSM, viz. permanent (but not original) sample members.

It should be noted that the BHPS also uses the concept of Temporary Sample Members (TSMs). A TSM is someone who is not themselves an OSM but is living in the same household as an OSM (at any wave). TSMs are eligible to be interviewed so long as they are in the same household as an OSM (because their experiences are important to understanding the experiences of the OSM), so they are therefore considered sample members for survey administration purposes (see Section 6.3.3). However, they are not themselves members of the BHPS sample so they are not considered in this section of the quality profile. (Though, of course, some individuals will, over time, change status from TSM to PSM.)

3.2.3 Quality control of sample selection process

At wave 1, a number of quality control checks were carried out on the selection of addresses. These included checks designed to ensure that the stratification of postal sectors had been implemented correctly, checks that the sampling interval for sectors had been calculated and applied correctly; and checks that the distribution of selected sectors across strata was plausible vis à vis the distribution of PAF delivery points across strata. However, the details of these checks have not been documented.

Quality control checks on the wave 1 selection of households and persons within addresses formed part of the postal and telephone recall interviews carried out as part of the general fieldwork quality control process – described in Section 6.2.2.

Similarly, quality control checks of the identification of new sample members at subsequent waves (OSM(D), PSM and TSM) were carried out as part of the fieldwork quality control process at each wave: the field call-backs checked the household composition information and in-office checks ensured that individuals had been assigned the correct sample status.

Although some (possibly most) deaths of sample members will be identified in the course of field work, some will not. Ad-hoc checks with official death registrations have been carried out in order to identify sample members who have died.

3.3 Sample size

At wave 1, 8,167 addresses were selected. The wave 1 field work identified 13,840 persons at these addresses. These constitute the OSMs. Of these, 10,751 were aged 16 or over and therefore eligible to be interviewed. The total number of interviews achieved at wave 1 (including proxies) was 10,264. These interviews encompassed 5,505 households. (A further 33 responding households are documented on the wave 1 data file as containing OSMs – 27 where

there is no individual response data and 6 where the household composition form was completed – see Section 5.1 – but no interviews.) The number of interviews achieved at each wave is summarised in Table 4. It should be noted that response rates can not be derived from the numbers presented here. Response rates are discussed separately in Section 6.4. The distribution of interviews by mode (personal, proxy, telephone) is discussed in Section 6.1.

A further breakdown, by patterns of unit response over waves, is presented in Table 5. This indicates the sample sizes available for key longitudinal analysis bases.

However, the adequacy of the sample size of a panel study to address any particular research question cannot necessarily be assessed simply from total sample size, since it depends both on initial sample size, and particular patterns of attrition. Moreover, analysts will want to structure their analyses in many different ways. This section gives some indicative numbers for the first thirteen waves of the BHPS original sample. Table 6 shows the number of first wave respondents of various ages who respond at every wave up to various thresholds. Thus there were 1,091 respondents aged between 25 and 34 in 1991 who responded at each wave up to wave 13. The thirteen wave sample is smallest relative to the starting sample for those aged 60 and over, since this group also experiences higher losses through mortality.

Analysts may want to follow sequences starting from waves after the first wave. In this situation, the number of cases available may be higher than implied by Table 6, since people with an incomplete response pattern in previous waves, or who have reached the minimum eligible aged for interview can also be included. Table 7 shows the number of fifth wave OSM respondents of various ages who respond at every wave up to various thresholds.

In many cases analysts, rather than examining sequences of waves, may want to pool responses from adjacent waves in order to examine transitions. Table 8 shows the total number of pairs of consecutive waves where the respondent gave a full interview at each wave, by age in 1991. These numbers are clearly much larger than the number of individual respondents, and statistical tests must adjust for the non-independence of observations.

The data of the type described in Table 8 is used for analysis of events and transitions, their antecedents and consequences. Table 9 and Table 10 show the total number of events of types that are often explored using panel data. Table 9 focuses on transitions between employment statuses. Thus there are over 60,000 pairs of waves where the respondent was employed at both waves, with over 11,000 respondents with at least one such pair. There were 1,400 moves into unemployment, but the fact that almost 1,300 respondents experienced at least one such move indicates that there will be relatively few respondents with multiple moves of this sort.

Table 10 focuses on demographic events. Some of these are also associated with attrition, but it may be possible to identify that an event has occurred. For example interviewing parents of adult children who were living in the same household should identify cases where the child has left home, even if they have not been successfully interviewed. All instances of events are shown in the first column, and instances where the subject is a respondent at the next wave are shown in the second column. In principle, the first column shows cases that can be used for the analysis of antecedents of events, the second, cases for the analysis of consequences. In the case

of partnership formation and parenthood, the difference reflects individual non-response within responding households. As in Table 9, the final column shows the number of respondents experiencing the event at least once. In the case of births, this is defined more narrowly, to show the number of first births, ie events of becoming a parent.

3.4 Selection probabilities

3.4.1 OSMs

The wave 1 selection procedures described in section 3.2.1 above resulted in all residential addresses in Britain having equal selection probabilities, apart from Scottish addresses with a multiple output indicator of greater than 1. At sampled addresses, all households were included unless there were more than three, in which case three were sampled. This led to slightly smaller overall selection probabilities for households at multi-household addresses. Once households had been selected, all persons within those households were included in the BHPS sample.

Selection probabilities for individuals are therefore proportional to $MOI_i \times \frac{H_i^S}{H_i^T}$, where

$MOI_i = 1$ for addresses in England and Wales;

= PAF multiple output indicator for addresses in Scotland

and H_i^S households at address i were selected for inclusion in the sample out of a total of H_i^T households at the address.

The distribution of $\frac{H_i^S}{H_i^T}$ amongst wave 1 responding households (defined as all households where at least the household composition information was complete – see Section 6.4 for a full break-down of field outcomes) is presented in Table 11. The 5,286 households where this quantity has the value of 1.00 include 5,162 at single-household addresses, 64 at 2-household addresses and 60 at 3-household addresses. The full distribution of design weights (reciprocal of relative selection probabilities) across all wave 1 responding individuals is presented in Table 12. It can be seen that the variation in weights is not large, though there are a few cases with relatively large weights and a few with small weights. The five cases with weights of 4.0 come from a sample address at which 12 households were found (and 3 selected for inclusion). In this case the PAF MOI took the value 1. At the other extreme, there were nine individuals at addresses with an MOI of 9, where the number of households found was no more than 3 (5 addresses, all in Scotland). The predicted average design effect due to variable sampling fraction is 1.036¹.

¹ Assuming equal variances within weighting classes, ie $DEFF_{VSF} \cong \frac{n \sum_{h=1}^{24} n_h (w_h^2)}{\left(\sum_{h=1}^{24} n_h w_h \right)^2}$, where w_h is the design

weight for weighting class h and n_h is the number of sample cases in weighting class h , the 24 weighting classes being as shown in Table 12.

3.4.2 OSM(D)s

All children resident in Britain at the time of wave 1 had an equal chance of being selected as an OSM as described above. Children born subsequent to the time of wave 1 fieldwork can only enter the sample as OSM(D)s. The selection probabilities for these children depend upon whether their parents had a chance of selection as an OSM at wave 1 (ie whether they were OSE – original sample eligible) and whether both parents were in the same household at that time. Specifically, P_j , the probability of selection of person (child) j , for all persons born subsequent to wave 1 can be written:

$$P_j = \frac{n \sum_{i=1}^N I_{ij}}{N},$$

where $I_{ij} = 1$ if the inclusion of household i at wave 1 would lead to the inclusion of person j (0 otherwise), there being n households selected out of a total of N at wave 1. We therefore see that:

$$\sum_{i=1}^N I_{ij} = 2$$

if both parents of person j were OSE and living in 2 separate households at the time of wave 1;
 $= 1$ if both parents were OSE and living in the same household;
 $= 1$ if only one parent was OSE
 $= 0$ if neither parent was OSE

Of these four categories, the last will obviously not be represented in the BHPS sample. Unfortunately, for sampled OSM(D)s it is only possible to identify those falling into the second category; categories 1 and 3 cannot be distinguished as the non-OSM parent is not asked whether they were OSE.

3.4.3 PSMs and TSMs

PSMs and TSMs are primarily of interest for the contextual information that they provide regarding OSMs. For that purpose, selection probabilities are not relevant. However, the standard BHPS weighting schemes (see Section 8) are designed to allow PSMs and TSMs to be included in their own right in cross-sectional analysis. In this context, selection probabilities are important.

Each (OSE) person who becomes a parent subsequent to wave 1 of a child whose other parent was not in the same household at the time of wave 1 has an overall selection probability twice as great as that of other (OSE) people. Put simply, they had one chance to be selected as an OSM (by selection of their wave 1 household) and one additional chance to be selected as a PSM (by selection of the wave 1 household of their (future) partner). All non-OSE persons who become a parent of a child whose other parent was OSE have an equal selection probability.

At wave m , say, each person resident in Britain has a probability of being included as a TSM. This probability is proportional to the number of households at the time of wave 1 that contained at least one member of that person's household at the time of wave m . These probabilities are

unknown, as it is not feasible to ask details of autumn 1991 household membership of all persons entering the sample subsequently. However, the weight-share method that is used to provide weights for cross-sectional analysis (see section 8.1.3) deals adequately with the variation in selection probabilities and provides unbiased cross-sectional estimates.

4. Sample Design – Additional Samples

4.1 European Community Household Panel sub-sample

As described in Section 2, from wave 7 the BHPS began providing data for the United Kingdom European Community Household Panel (ECHP). As part of this, it incorporated a sub-sample of the original UKECHP. This sub-sample had two components:

- All sample households in Northern Ireland;
- All “low-income” sample households in Great Britain.

From the point of view of the BHPS this constitutes a new sample whose first wave is wave 7. However, their sample membership status depends in part on their membership status within the ECHP. Thus, members of the original 1994 ECHP sample are defined for our purposes as OSMs, while joiners to ECHP households after the first wave of ECHP, including joiners at waves 7 to 10 of BHPS are defined as TSMs or PSMs according to standard BHPS rules. There are also a small number of ECHP original sample members who rejoin selected households after Wave 7. These are also classified as OSMs.

4.1.1 Sample selection process

The Northern Ireland sample consisted of all households in Northern Ireland in which all adult members responded at ECHP wave 3 (1996).

The low-income sample was selected on the basis of characteristics associated with low income in the ECHP. As the ECHP Wave 3 income data were not available to carry out the selection, proxy indicators (correlates) had to be used. The sample consisted of households in which all adult members responded at ECHP wave 3 (1996) and which fell into one or more of the following categories:

- Household reference person unemployed at interview or within the last year;
- Household reference person in receipt of lone parent benefit;
- Household reference person in receipt of means tested benefit;
- Household in rented accommodation.

Household level criteria were used because individual level criteria would have created complications with the probabilities of selection of households, as the household level probabilities would have been affected by the numbers of relevant individuals they contained.

Data collection for ECHP wave 4 was carried out in 1997 by Social Survey Division of the Office for National Statistics (ONS) and Social and Community Planning Research (SCPR, now known as the National Centre for Social Research), using the BHPS wave 7 questionnaire. A question was included asking respondent households for permission for their data and details to be passed to the University of Essex. For both components of the sub-sample, only respondent households who agreed to have their data passed to the University of Essex were incorporated into the BHPS.

4.1.2 Sample size

The procedures outlined in Section 4.1.1 above resulted in a total of 1,076 households in Britain meeting the eligibility criteria and being issued to the field for BHPS wave 7. Including new split-off households identified during field work, 1,086 households were eligible for interview. Of these, 930 (85%) were interviewed (household interview and at least one individual interview) and agreed to be incorporated into the BHPS. Additionally, there were 169 households in Northern Ireland interviewed at wave 7. For these households, permission was not sought for data to be passed to Essex, as future waves of field work would continue to be carried out by Central Survey Unit of the Northern Ireland Statistics and Research Agency. Thus, a total of 1,099 households were added to BHPS at wave 7.

Amongst these 1,099 households, a total of 1,727 individual interviews were achieved. The number of interviews achieved at each wave, broken down by sample type, is summarised in Table 13.

4.2 Scotland and Wales extension samples

4.2.1 Sample selection process

The sample selection procedure for the Scottish and Welsh boosters required comparability with the BHPS. The starting point for the design was therefore the original 1991 BHPS design. The key elements of the original BHPS design also implemented for the booster samples are below:

- the target sample was the resident population living in private households
- the frame was the 1999 Postcode Address Small Users File
- the primary sampling units were postcode sectors selected with probability proportional to size (ie number of addresses).
- stratification was implicit, by selection from a list ordered by sub-region and socio-economic characteristics
- addresses were selected at a second stage, on the basis of a random start and a fixed selection interval calculated as nearest integer to the total number of addresses in the PSU, divided by the number of addresses to be selected.
- the survey agency was expected to make all reasonable efforts to contact all issued addresses, and there was no substitution.

The design for Scotland departed from the original BHPS design in one key respect. For Scotland, the target population was the whole of Scotland at the date of the fieldwork ie with no exclusion of areas north and west of the Caledonian Canal as per the original BHPS design.

In each of Scotland and Wales, 75 PSUs were selected with between 30 and 33 addresses selected in each PSU. In selecting the PSUs the selection of an original BHPS PSU was allowed but the selected addresses were checked prior to fieldwork to ensure no existing BHPS addresses had been sampled by chance. This checking found no addresses that required exclusion.

Oversampling of specific areas or populations was not carried out even though some consideration was given to this for the Highlands and Islands of Scotland in particular. As the core purpose of the booster sample was to represent the whole population of Scotland as well as possible, it was decided that oversampling was not desirable. Moreover it was felt that oversampling to represent the main candidate areas, the Highlands and Islands sufficiently for separate analysis would be so costly as to prejudice the entire project.

4.2.1.1 Selection of PSUs

A list was generated of all postcode sectors in Scotland/Wales. Sectors with fewer than 500 delivery points (addresses) were grouped with an adjacent sector and thereafter treated as a single sector. Sectors were then stratified as follows.

Scotland:

Sectors were sorted into 6 regions defined by council areas (see Table 14).

- (a) Within each region, sectors were ordered by the proportion of heads of households in socio-economic groups 1 to 5 and 13 (that is, the proportion of heads of households in professional or managerial positions). Within each region, sectors were then split into major strata of approximately equal size (in terms of delivery points). The number of major strata varies by region from two to three, depending on region size (see Table 14), resulting in 15 major strata.
- (b) Within 8 of the major strata (3 regions), sectors were then re-ordered by the proportion of the population of pensionable age (ie females over 60 and males over 65). The order of sorting was serpentine, in order to ensure that potential periodicity problems in the frame could be minimised. Each of these 8 major strata was then split into two minor strata each of approximately equal size.
- (c) Finally, within each of the 23 minor strata, sectors were re-ordered, using serpentine listing as follows:
 - (i) Sectors in 20 of the strata were ranked by the proportion of the employed sector population working in agriculture (denoted AGEMP in Table 14 Table 14) and
 - (ii) Sectors in the other 3 strata were ranked by the proportion of the sector population that was living in single person households (denoted SPH in Table 14).

Wales:

- (a) Sectors were sorted into 5 regions defined by electoral constituencies (see Table 15).
- (b) Within each region, sectors were ordered by the proportion of heads of households in socio-economic groups 1 to 5 and 13 and split into 2 or 3 major strata (as for Scotland) resulting in 12 major strata.
- (c) Within 6 of the major strata (2 regions), sectors were then re-ordered by the proportion of the population of pensionable age (ie females over 60 and males over 65) and split into two minor strata. Within the other 6 major strata (3 regions), sectors were re-ordered by the proportion of the employed sector population working in agriculture and split into two minor strata. Thus, 24 minor strata were created.

- (d) Finally, within each of the 24 minor strata, sectors were re-ordered, using serpentine listing as follows:
- (i) Sectors in 6 of the strata were ranked by the proportion of the employed sector population working in agriculture (denoted AGEMP in Table 15) and
 - (ii) Sectors in the other 18 strata were ranked by the proportion of the adult population who speak Welsh (denoted WELSH in Table 15).

A size measure was generated for each sector. In Wales this was the number of delivery points in the sector. In Scotland the measure was the sum of the MOI for all delivery points in the sector. Systematic samples of 75 sectors were then selected in each country, with probability proportional to this size measure, using a random start and a fixed interval.

Table 14 and Table 15 show the number of postal sectors selected per region, as an indicator of the relative population size of the strata.

4.2.1.2 Selection of Addresses

From each selected sector, 33 addresses were sampled systematically, using an appropriate random start and fixed interval.

4.2.1.3 Selection of Households / Persons

This proceeded in exactly the same way as for BHPS wave 1, as described in Section 3.2.1.3.

4.2.2 Sample size

Just under 2,500 interviews were completed in each of Scotland and Wales at wave 9, and slightly larger numbers at wave 10 due to the addition of temporary sample members and the re-issue of refusals and non-contacts from wave 9. The numbers are summarised in Table 16 and Table 17.

4.3 The British Youth Panel

4.3.1 Sample selection process

The Youth Panel began in 1994. All children aged 11 - 15 years on or before December 1st of the survey year living in a household with at least one Permanent Sample Member (PSM) are eligible for interview. The child does not have to be a PSM to be eligible.

Children living in former ECHP households have been interviewed since wave 8. Children living in households in the Scottish and Welsh booster samples have been interviewed since wave 10 (their wave 2).

4.3.2 Sample size

The number of youth interviews achieved in core BHPS households has been between 700 and 800 each wave from wave 4 to wave 10. The addition of the ECHP and Scottish and Welsh

booster samples in latter waves increased the total number of youth interviews to 1,414 in wave 10. The numbers of youth interviews, broken down by wave and sample type, are shown in Table 18.

4.4 Northern Ireland sample

4.4.1 Sample selection process

A systematic random sample of addresses in Northern Ireland was selected by the Central Survey Unit of the Northern Ireland Statistics and Research Agency, who also carried out the survey field work in Northern Ireland. The sampling frame was the Valuation and Lands Agency's (VLA) list of domestic properties in Northern Ireland. Addresses were selected using a random start and fixed interval. There is no geographical clustering of the sample addresses. The VLA database is supplemented with Postcode Address File information in order to provide complete post code and geographical information to assist interviewers in the location of sample addresses.

4.4.1.1 Selection of Households / Persons

This proceeded in exactly the same way as for BHPS wave 1, as described in Section 3.2.1.3.

4.4.2 Sample size

At the sample addresses, 2,885 eligible households were identified, of whom 1,979 (69%) co-operated with the survey. Interviews were successfully completed with just under 3,500 individuals at wave 1 (BHPS wave 11) and with just under 3,000 individuals at each of waves 2 and 3 (BHPS waves 12 and 13). The numbers are summarised in Table 47, Table 48 and Table 49.

5. Content of Data Collection Instruments

5.1 Summary of instruments

The main data collection instruments used on each wave of BHPS are the following:

Household coversheet. This contains an interviewer call record, observations on the type of accommodation and the final household outcomes.

Household composition form. This is usually administered at the interviewer's first contact with an adult member of the household. It involves a listing of all household members together with brief summary data regarding sex, date of birth, marital status, employment status and relationship to the household reference person (HRP). Additional checks are made on presence in the household of natural parents or spouse or partners, in order to unambiguously establish all relationships (for instance, secondary or 'hidden' couples).

Household questionnaire. This is administered to the household reference person and takes 10 minutes to complete on average. It contains questions about the accommodation and tenure and some household level measures of consumption.

Individual questionnaire. This takes around 40 minutes to complete on average and is administered to each member of the household aged 16 or over. The individual questionnaire covers the following topics:

- neighbourhood
- individual demographics
- residential mobility
- health and caring
- current employment and earnings
- employment changes over the past year
- lifetime childbirth, marital and relationship history (Wave 2 only)
- employment status history (Wave 2 only)
- values and opinions
- household finances and organisation

Self-completion questionnaire. This too is administered to each person aged 16 or over. It contains questions that are subjective or particularly vulnerable to the influence of other people's presence during completion, and potentially sensitive questions requiring additional privacy. It includes a reduced version of the General Health Questionnaire (GHQ) plus attitudinal items and questions on social support. It takes about five minutes to complete.

Proxy schedule. This is used to collect information about household members absent throughout the field period or too old or infirm to complete the interview themselves. It is administered to another member of the household, with preference shown for the spouse or adult child. The questionnaire is a much shortened version of the individual questionnaire, collecting some demographic, health, and employment details, as well as a summary income measure.

Telephone questionnaire. This is based upon the proxy schedule. It is administered by an experienced interviewer employed by ISER, when all efforts to achieve a face-to-face interview have failed.

From waves 1 to 8 all the above instruments were paper documents. From wave 9 onwards, the household and individual questionnaires were administered by CAPI, while the other instruments remained paper documents. From wave 13, the proxy questionnaire has also been administered by CAPI. In Northern Ireland, all questionnaires apart from the self-completion questionnaire have been CAPI since the survey's inception in 2001. Copies of all the instruments are available in down-loadable PDF form from http://www.iser.essex.ac.uk/bhps/doc/pdf_versions/index.html.

5.2 Instrument design: principles and procedures

The interview for each individual lasted, on average, 45 minutes, with an additional short household level questionnaire for one individual in the household. The design includes three main components:

- Core questions repeated at each wave
- Rotating core questions repeated on a two or three year cycle
- Variable component questions

Approximately three quarters of the questionnaire was made up of core and rotating core questions. The rotating core items covered topics where there was no expectation of rapid change and there is therefore no need to ask questions on them every year. This also allowed a means to deal with competing demands for limited space within the questionnaire. The remainder at each wave included the variable component. The variable component was designed for:

- Questions which needed to be asked less frequently than core or rotating core items;
- New questions engendered by changing policy and research issues;
- Questions to elicit retrospective data on panel members' life history before the first interview.

The design of the BHPS questionnaire was influenced by a number of distinctive features and opportunities of longitudinal research design. Firstly, the design allows for analysis at the level of the individual or the household. In comparison, research on successive cross-sectional surveys focuses on change at the population level. While the BHPS questionnaire includes many similar questions to those used in cross-sectional surveys, the questions included had a clear focus on characteristics, behaviour or values that are either expected to be subject to change, or are significant factors affecting the likelihood of change.

The second and more distinctive feature of longitudinal research is that it allows an analysis of individuals over time. The research instrument should allow the construction of continuous measures, of, for example, income, employment histories and labour market participation, household structure and residential mobility over the life-cycle. This is collected much more reliably than in long term retrospective history surveys so many questions in the panel survey

were concerned with events in the twelve months between interviews, rather than with the current situation at the time of interview. A further use of panel data collection is to compare expectations about change in the subsequent year with change that actually takes place. Important areas in this context are occupational change and residential mobility.

The following paragraphs fill out these points further by indicating how BHPS research priorities and these research design considerations affected the focus in the six broad topic areas represented in the questionnaire: household organisation, the labour market, housing and residential mobility, income and wealth, health and socio-economic values.

The focus on the areas outlined below was established after extensive consultation within the British academic and policy research community. Prior to designing the questionnaire, an advisory group for each substantive area was established. These groups met throughout 1988 and were responsible for reviewing the data requirements for each area, prioritising the measures critical for longitudinal analysis, and recommending the appropriate balance between the different substantive areas within the questionnaire. This consultation process culminated in the Panel Design Conference held at the University of Essex in July, 1989. The purpose of the conference was to review and discuss the conclusions and recommendations of all of the advisory groups and make final recommendations for the content and design of the BHPS questionnaires. Following the conference, the design work on the questionnaires began, taking forward the recommendations received. The advisory groups were replaced by an Advisory Committee of some 40 academic and policy researchers who met bi-annually to oversee the content and conduct of the panel for the first three waves. A smaller Scientific Advisory Committee provided advice in later waves.

The **Household Organisation** component included detailed demographic information, as well as information on the intra-household division of labour and on household economic organisation, in particular the control of money and access to consumption goods within the household. From a longitudinal perspective, this information permits new research on patterns and processes of household formation and dissolution, while the information on internal organisation can be related to other changes affecting the household, for example how changes in the work patterns or income of individual members influence internal organisation.

The **Labour Market** component was shaped by two main agenda. Firstly, it was concerned to allow research focused on patterns of individual mobility, either by comparing jobs and labour market position at successive waves, or by analysing the detailed work histories collected over the course of the panel. Such career patterns can clearly be related to other components of panel data, such as education, income, health or household organisation. To facilitate this type of analysis the panel collected relatively detailed information on job characteristics, covering such issues as promotion within jobs, job security, training and fringe benefits. Secondly, it was concerned with collecting data on how individual labour market participation decisions relate to the household context and thus, for example, to support economic research on household labour supply models and sociological research on married women's labour market participation. It was particularly concerned with collecting data on transitions.

The **Income and Wealth** component provided data to be used in research for many of the other components. More distinctively it was concerned with income and wealth dynamics at the household level, related to other individual and household changes. In particular it was designed to allow the development of research on transitions in and out of poverty, on life cycle models of income, on the relationships between income change and family composition change, and on lifetime patterns of wealth accumulation and savings. These priorities implied a focus in the questionnaire on income levels over the year, and also, though not until wave 5, on data on wealth and savings.

The **Housing** component collected basic information on housing costs and conditions and other features of housing consumption, as well as data on perceptions of residential neighbourhoods and aspirations and intentions to move. This had a clear value for both housing research and in relation to other research domains. More distinctively longitudinal concerns were housing and tenure mobility and the accumulation of wealth through housing. The BHPS provides a valuable research resource for the analysis of migration and migration decisions.

The **Health** component was concerned with relating aspects of individuals' health and health related behaviour over time to other aspects of their situation and behaviour. This involved data that allowed both analysis of the impact of health and illness on other aspects of behaviour and of social explanations of health and illness. A particular focus was on measures of psychological well-being, in order to explore how changes relate to life events. Data was also be collected on patterns of health service usage over time.

The focus on **Socio-Economic Values** within the BHPS was largely related to the five domains outlined above. The aim was to provide data which allowed an exploration of how value change is affected by changing individual and household situations; how values themselves can be seen to influence behaviour; and also how stable they appear to be over time. A more specific focus of the component was on political values and on social participation.

The breadth of coverage of the BHPS means that on individual subjects the depth of coverage cannot be as great as it is in these other surveys. Moreover, because the sample size is significantly smaller than the LFS and GHS, analysis of small population groups or of the smaller regions is limited to some extent. It was, however, designed to provide some information on the characteristics of the area of residence so that analysis of the impact of change in local contexts can be undertaken.

5.2.1 Pre-testing and piloting

A Pilot Panel, running alongside the main BHPS sample was established for testing and methodological work prior to wave 1. This Pilot Panel comprised 500 households and was extensively used during the first three years of the panel. The Pilot Panel enabled the testing not only of questionnaire items but also provided a vehicle to test the longitudinal components of data collection and the procedures for running the panel. Since Wave 3, the Pilot Panel has been maintained but not used for full-scale Pilots. Sub-groups are selected for pre-testing purposes each year.

Prior to Wave 1 a series of pre-tests and pilots were carried out to test each of the questionnaire components and to streamline fieldwork procedures. Interviewers attended a debriefing session following each of these to report on any problems encountered in administering the survey.

At each subsequent wave, pre-testing and/or piloting of new questionnaire sections or items has been carried out in advance of the main fieldwork period. At waves 2 and 3 a series of pre-tests and a full pilot (500 households) were carried out to test the collection of lifetime marital, fertility, employment status and job histories. Pre-testing techniques have included the taping of interviews for subsequent analysis and the use of behaviour coding (Fowler, 1995) where there was concern about the potential sensitivity of items. For example, behaviour coding was carried out at wave 5 when testing the wealth, assets and debts variable component.

The Youth Panel was introduced at wave 4. The development of the youth questionnaire included a series of focus groups and pre-testing within schools for the relevant age group. The methodology used for collecting the youth questionnaire was extensively tested. The youth questionnaire is administered via a tape recording that the child listened to on a Walkman, completing the answers in a self-completion booklet. The questions and response categories are read out on the tape and pre-testing included testing which type of voice eg male/ female/ accent/ age was liked best by children. It was found that the voice of a woman aged 30 - 40 with no distinctive accent was most acceptable to children. The answer booklet did not include the question wording but just the response categories. In this way the confidentiality of children's responses within the household was guaranteed. The methodology also overcame potential literacy problems for some children. For a full description of the development of the youth questionnaire see Scott *et al* (1995) and Scott (1995).

Wave 9 was the transition year to CAPI and the development of the CAPI script required an extended, iterative pre-testing approach. Sections were pre-tested by a small team of interviewers as they were completed, with interviewer comments on both errors and the usability of the CAPI script feeding into the next iteration of design and testing. This iterative, small scale testing led to a final pilot of 100 households from the Pilot Panel using the full CAPI instrument. Interviewers completed a report sheet on every interview and attended a face-to-face debriefing session. See Banks and Laurie (2000) for a full description of the transition and design of the CAPI questionnaires.

5.3 Content of instruments

The broad topics covered by each of the instruments are summarised in Section 5.1 above. Complete documentation of all questions asked and all data items available in the data sets is available at <http://www.iser.essex.ac.uk/ulsc/bhps/> .

6. Data Collection

6.1 Data collection mode

The main mode of data collection at every wave has been face-to-face in-home interviewing. At waves 1 to 8, all interviews were conducted by PAPI. Since wave 9, household and individual interviews have been conducted by CAPI using In2itive² in England, Scotland and Wales. In Northern Ireland, the CAPI software used is Blaise. Since wave 3, some interviews have been conducted by telephone in cases where it has proved impossible to carry out a face-to-face interview. Additionally, at all waves some interviews have been carried out with a proxy respondent when it has not been possible to interview the target individual. Proxy interviews were only permitted when the target respondent is absent throughout the field period or too old or infirm to complete the interview themselves. The proxy respondent was usually another household member, with preference shown for the spouse or adult child. In a small number of cases the proxy respondent was a relative or carer living outside the household.

Table 19 presents a summary of the number of interviews carried out in person, by proxy and by telephone. Overall, 96.3% of interviews have been carried out in person with the target respondent, 2.9% in person with a proxy respondent and 0.9% by telephone.

6.2 Fieldwork - general

6.2.1 Interviewing

6.2.1.1 Interviewers

The fieldwork for waves 1 to 13 of BHPS in England, Scotland and Wales was carried out on behalf of ISER by NOP Research, London. NOP interviewers have carried out all of the interviews, with the exception of the telephone interviews, which have been conducted by ISER. NOP Research executives are members of the Market Research Society and therefore subscribe to the MRS code of conduct. The company subscribes to the Interviewer Quality Control Scheme (ICQS). The first three waves of fieldwork in Northern Ireland, which took place in parallel with waves 11-13 of BHPS, were carried out by the Central Survey Unit of the Northern Ireland Statistics and Research Agency.

At wave 1, NOP employed 243 interviewers to cover the 250 sample areas, generally one per area. Because of the demanding nature of the BHPS special attempts were made to use interviewers of above average levels of experience and ability. In subsequent waves, the majority of respondents still lived in the same 250 areas, but because of household and individual moves, the sample has become slightly, but progressively more, dispersed over time. Sample members living outside the 250 original areas are allocated to the most appropriate one of those areas (generally, the nearest) for fieldwork management purposes. Generally, it continues to be the case that one interviewer works in each of the 250 areas in the sample. Apart from an interpenetrating sample experiment at wave 2, the same interviewers are employed in the same areas

² A CAI software product owned and supported by SPSS MR.

each year wherever possible. In total, for example, 237 interviewers worked on wave 2, all but 35 of whom had also worked on wave 1. In subsequent years, interviewers leaving the survey had their sample points allocated to experienced BHPS interviewers if possible as long as this did not result in an unacceptably heavy workload or excessive travel.

Table 20 demonstrates the stability of the interviewer panel over the thirteen years. It can be seen, for example, that one-fifth of persons who were interviewed at all thirteen waves were interviewed by the same interviewer on every occasion. (This figure would have been higher had it not been for the interpenetrating design at wave 2 – see Section 6.2.1.2) The number of interviewers working on BHPS for the first time at each wave is small and one third of the wave 1 interviewers were still working on the survey at wave 13.

6.2.1.2 Interpenetrating sample experiment

At Wave 2, an interpenetrating design experiment was implemented to test for interviewer effects separately from area effects. A constrained form of randomisation in which addresses were allocated to interviewers at random within geographic ‘pools’ was used. Each pool consisted of two or three PSUs within reasonable travelling distance of each other. All PSUs whose centroid was a minimum of 10 kilometres away from the centroid of at least one other PSU were eligible for inclusion in the design. One hundred and fifty-three of the 250 PSUs in the BHPS sample were eligible. Mutually exclusive and exhaustive combinations of these 153 eligible PSUs were then formed. This resulted in 70 pools of two or three PSUs each. A systematic sample of 35 pools was then selected for inclusion in the inter-penetrating sample design. Within a given pool, interviewers were randomly assigned to the interviewers working in those PSUs. As one interviewer had typically been assigned to each PSU, this meant that two or three interviewers worked each pool.

O’Muircheartaigh, Campanelli and Smith (1999) and Campanelli and O’Muircheartaigh (1999) examined systematic variations in interviewer influence on survey non-response using the IPS data. Controlling for area effects, significant variations in individual interviewer refusal and non-contact rates were found to be persistent over time and considered somewhat dependent on the individual skills of an interviewer. They found “absolutely no effect of interviewer continuity” from the wave 2 data even though they accept that interviewer continuity has been shown to be significant from a multi-wave perspective and that appreciable benefits in providing the same interviewer wave upon wave might exist on a practical basis (see Laurie, Smith and Scott, 1999). Campanelli and O’Muircheartaigh (2002) extended their earlier work by examining response up to wave 4 and again concluded that there is no evidence of interviewer continuity affecting response rates.

The results from the IPS suggest that interviewers who have a low refusal rate are also likely to have a low incidence of losing track of their respondents. Interviewer effects are particularly appreciable when considered in terms of geographic location indicating the need for enhanced interviewer training in difficult areas such as inner London and other urban areas.

Two commonly cited sources of error, differences between individual interviewers and sample variance have also been examined using the IPS data (see O’Muircheartaigh and Campanelli

1998). This work suggests that within clustered samples errors incurred by deviating from a random sample and other non-sampling errors, such as coder variance and question wording effects, are confounded by assigning a single interviewer to a primary sampling unit (PSU). Using data from the IPS experiment all substantive variables in the BHPS were examined. All types of questions were considered including attitudinal, factual, interviewer checks and quasi-factual. Some significant values for interviewer effects were found but vary dependent upon the element used and the particular question asked. The effects of a clustered sample design lead to homogeneity within the clustered sample, so much so that it shows a “strikingly similar” degree of effect to that ascribed to interviewer effects.

6.2.1.3 Interviewer Briefing

At Wave One, all interviewers and field supervisors were briefed at one of 14 two-day briefing conferences, presented jointly by ISER and NOP at various suitable locations round the country. A special training video was prepared by ISER for use in these briefings. In subsequent waves, interviewers and area managers were again briefed at one of 14 briefing sessions, held in different locations. All interviewers with previous BHPS experience attended one-day briefings, while interviewers new to the survey attended a two-day briefing.

Interviewers received the interviewer instructions and questionnaires prior to the briefing. Home-study exercises were also used for complex substantive sections of the questionnaire. The content of the briefings covered all aspects of the fieldwork process including making contact with respondents, minimising refusals and non-contacts, key definitions such as household membership and eligibility for interview, tracking procedures for movers, progress monitoring and return of work. Any new substantive sections of the questionnaire were explained in detail and complete dummy interviews carried out during the briefing. Since wave 9, the briefings have included the use of the CAPI questionnaire script and management of work using the lap-top computer. ISER staff attended all briefings.

6.2.2 Field work quality control

NOP were required to use interviewers with experience of carrying out random sample work and social surveys. They were selected on the basis of numbers of years of experience and a quality rating of above average. NOP were required to implement the following quality control procedures during fieldwork.

- Accompaniment of all wave 1 interviewers by their supervisor within the first few days of starting
- Post wave 1, all new interviewers to the survey accompanied in their first week; all other interviewers subject to NOP’s standard accompaniment procedures
- Twice-weekly progress monitoring carried out by NOP area managers and weekly progress report sent to ISER.
- Minimum of 6 calls over a minimum period of three weeks until a final household outcome is achieved. At least four of these calls to be made in the evening or weekend. Interviewers encouraged to make “passing” calls and keep trying non-contacts to the end of the fieldwork period.

- Initial contact with household to be made in person. Appointments by telephone allowed only where the respondent has requested this or the interviewer has failed to make contact after three personal calls at the address.
- Interviewers, as far as possible, to ensure that they make personal contact with each individual to determine whether they will be willing to take part at that year of the survey. Proxy refusals should not be accepted.
- Where contact was not made or there was a refusal, full details of the reason for the non-contact/refusal was recorded by the interviewer and the relevant information passed back to the Institute.
- Targets for maximum non-contact rate, partial household coverage and proxy data collection set for interviewers and fieldwork supervisors
- Establishment of an office within NOP to deal with the fieldwork full-time during the survey period
- Postal recall of 10% of all completed interviews (brief self-completion, 604 of 940 were received at W1)
- Telephone recall on 5% of completed interviews by field supervisors. Contact with 633 of 721 attempted households at Wave 1.
- Wave 2 onwards an additional call-back for all new interviewers
- ISER call back at least one household per interviewer, check 5 items of data. Of 246 calls, 72% responded with positive comments on the survey at Wave 1.

The recall questionnaire checked the following details:

- Name of respondent
- Whether the interview took place with named interviewer
- Respondent's place of birth
- Father's occupation when aged 14
- Age of respondent
- Number of people in the household
- Number of people aged 16 and over in household
- General comments about the interviewer and the interview

6.2.3 Field work dates

The fieldwork for the BHPS starts on September 1st every year. At Wave 1, the fieldwork period ran until December 1st 1991. In subsequent years, the main fieldwork period between Sept 1st and Dec 1st has been supplemented by an extended fieldwork period lasting into the early part of the following calendar year. The extended tail is used to carry out tracing of respondents who have moved and a forwarding address is not found by the interviewer, continued attempts for remaining non-contacts, and to carry out the refusal conversion programme.

While the tail-end fieldwork period extends until May of the following calendar year at each wave, the majority of interviews were completed within the main fieldwork period from September to December each year. The only wave where more than 10% of interviews were

carried out between January and May was wave 9 (Table 21). This was the transition year to CAPI where technical difficulties resulted in some fieldwork delays.

6.2.3.1 The ECHP sub-sample

The ECHP sub-sample was included within the BHPS at waves 7 to 10. Wave 7 fieldwork was carried out by ONS-SSD and SCPR, but using the same fieldwork period as for the main BHPS. From wave 8, the ECHP sub-sample was treated in the same way as the main BHPS sample in all respects and subject to the same survey and fieldwork procedures.

6.2.3.2 Scottish and Welsh booster samples

The Scottish and Welsh booster samples were introduced at wave 9. At wave 9 (wave 1 for the boosters) the fieldwork period was the same as the BHPS, starting on September 1st and running to December 1st. The fieldwork for wave 2 of the booster samples was delayed due to lack of available funding to return to the sample in September 2000 together with the wave 10 BHPS fieldwork. Funding became available in December 2000 making February 2001 the first possible fieldwork date to return to the sample.

The decision to carry out wave 2 in February 2001 gave a very short lead-time for both ISER staff and NOP to make preparations. However, the decision to go into the field in February 2001 did allow the sample to be re-contacted without a two year gap and to provide comparable data with the BHPS wave 10, albeit collected a few months after the main BHPS fieldwork.

The very short lead-time for wave 2 meant that no alterations to the content of the questionnaire were possible and the content was identical to that of the main BHPS wave 10 questionnaire. Although there had been plans to carry the lifetime marital, fertility and employment status histories (carried at wave 2 of the BHPS) at wave 2 of the booster samples, there was insufficient time to re-programme the CAPI script. Consequently, the booster sample respondents were asked the same wealth, assets and debts questions being asked of the main BHPS sample. There were concerns that this might have some effect on response as these questions can be sensitive for some respondents. The wealth, assets and debts questions had not been carried before wave 5 on the main BHPS because of worries about potential effects on response. While there was no apparent effect on response at wave 2 for the booster samples, some may be seen at wave 3.

Interviewing in February can be difficult depending on weather conditions and in 2001 there was a severe foot-and-mouth outbreak across the UK, restricting movement in certain areas. While the majority of the sample points were unaffected by this, some rural points in Scotland and Wales did have problems contacting a small number of households. If face-to-face contact was not possible, telephone interviews were carried out. In addition to foot-and-mouth, the UK General Election took place during the latter part of the fieldwork period, increasing interviewer's loads where pre-election surveys were being carried out.

The issued households at wave 2 fell into two categories, those where at least one individual interview was achieved at wave 1 and those which were either non-contact or 'soft' refusals at wave 1. The decision to re-issue some of the non-responding households from wave 1 was based

on our experience of doing this at wave 2 of the main BHPS and the standard practice we follow every year on the BHPS. There is evidence that many refusals are 'situational' in that there is a specific circumstance eg a house move, start of a new job, a recent bereavement, birth of a child or some other life event which means the respondent is unable to take part at that time. Returning at a different point in time, they may be happy to take part. The interviewer comments giving reasons for the wave 1 non-interview were assessed for each case before making the decision to re-issue them to field. The exercise proved to be very successful with 312 individual interviews in these households being achieved in Scotland and 229 in Wales.

6.2.3.3 Northern Ireland sample

Field work for the first wave in Northern Ireland took place almost concurrently with the main BHPS wave 11 fieldwork, starting just one month later in October 2001. 40% of the Northern Ireland interviews were completed by end-December 2001 (compared with 98% of wave 11 interviews with the original BHPS sample). 80% of Northern Ireland interviews were completed by end-February 2002 and the final interviews took place in April.

Fieldwork for the second wave in Northern Ireland again began about four weeks after the commencement of the main BHPS (wave 12) fieldwork. 59% of the Northern Ireland interviews were completed by end-December 2002 (compared with 97% of wave 12 interviews with the original BHPS sample). Fieldwork was completed in Northern Ireland in the first week of May 2003.

In 2003, the start of field work in Northern Ireland was brought in line with that for the main BHPS, in early September. Consequently, 82% of Northern Ireland third wave fieldwork was complete by end-December 2003 (cf. 97% for original BHPS sample) and 98% by end-February with final interviews in April 2004.

6.2.3.4 Pre-testing and piloting

The dates of the seven main stages of pre-testing and piloting carried out at wave 1 are documented in Table 22. At each subsequent wave, pre-testing and/or piloting of new questionnaire sections or items has been carried out in advance of the main fieldwork period. This included, at waves 2 and 3, a series of pre-tests and a full pilot (500 households) to test the collection of lifetime marital, fertility, employment status and job histories.

6.3 Fieldwork procedures

A number of strategies for maintaining the panel and unit response rates were implemented. These included advance letters before the start of each fieldwork period, respondent incentives, panel maintenance and a keeping in touch mailing exercise between interview points, sample management procedures, tracking and tracing procedures, and a refusal conversion programme.

6.3.1 Advance mailing

At wave 1 letters addressed to “The Occupier” were sent to all sample addresses in advance of fieldwork. The letter included a purpose leaflet outlining the aims of the survey and each participating household was given a more a more detailed brochure by the interviewer when contact was made. From wave 2, letters addressed personally to all expected eligible sample members were mailed in advance of calling on the household by interviewers. From wave 3, the text of the advance letter has varied depending on the response outcome at the previous wave ie whether individual interview at last wave, proxy, non-contact, refusal or telephone interview. All sample members turning 16 eligible for an individual interview for the first time also received a specific letter. The text of the letters was provided by ISER.

6.3.2 Respondent incentives

All respondents providing a full individual interview received an incentive in the form of a gift voucher for a major UK chain store. From waves 1 - 5 the value of this voucher was five UK pounds. The voucher was sent to respondents following their interview together with a thank-you letter. From wave 6, the value of the voucher was increased to seven UK pounds per interview. Since wave 6, the voucher was sent in advance of the interview so was not conditional on having taken part. All those respondents completing an individual interview at the previous year of the survey and 16 year olds eligible for interview for the first time had the voucher included in their advance letter. Interviewers were issued with additional vouchers for new entrants to the household and anyone interviewed who was not interviewed at the previous round.

Younger respondents aged 11 - 15 answering the Youth Questionnaire (since 1994), received a four pound gift voucher. This was given to them by the interviewer at the point of interview.

6.3.3 Sample management

The longitudinal sample potentially eligible for interview at any given wave includes all those original and permanent sample members who are not deceased. However, only a sub-set of this potentially eligible sample will be eligible for interview or found at any given wave. In managing the fieldwork process following wave 2, respondents who were known to be out-of-scope, whole households untraced at the previous wave, those in prison and those who had adamantly refused to take part at the previous wave were not issued to interviewers. Issuing these respondents would entail wasted effort trying to interview respondents who were not accessible. These respondents remained potentially eligible for inclusion in the survey and were interviewed if further information about their whereabouts became available, they moved back to Great Britain, or they decided to take part in the survey despite a previous adamant refusal. Management of the longitudinal sample following wave 2 was facilitated by categorising all known sample members according to their current known status or status at the previous wave. Prior to issuing the sample at each wave, sample members were categorised as belonging to either the active or inactive sample as follows.

6.3.3.1 Active sample

- (i) the issued sample - all expected sample members at a given wave. This category includes all members of interviewed households from the previous wave as well as members of non-interviewed households being attempted again. Full details fed-forward and issued to field.
- (ii) the inaccessible sample - includes untraced movers from earlier waves, sample members who are living in an institution where they cannot be interviewed (prison), sample members who are too ill or elderly to be interviewed and no proxy possible, and sample members who are out-of-scope (outside GB). Details fed-forward, not issued to field but are available to interviewers if required.
- (iii) the retiring sample - includes sample members who refused to take part any longer in the panel at the previous wave and are to be withdrawn from the active sample at the following wave. Details fed-forward, not issued to field but are available to interviewers if required.

6.3.3.2 Inactive sample

- (1) the retired sample - includes the deceased, adamant refusals from earlier waves, temporary sample members no longer living with an OSM so ineligible for inclusion and untraced movers who have been withdrawn following a special tracing exercise at wave 6. The details of these sample members are not fed-forward but were archived and not available during fieldwork for interviewers even though they could be restored to the sample if required.

It should be noted that at each wave some members of the active sample turned out to be ineligible for interview (eg because they have moved into an institution or out of Britain), while some persons not in the active sample turned out to be eligible for interview (as OSM(D)s or TSMs).

This method of managing the longitudinal sample allowed details of respondents to be made available to interviewers as required while providing a systematic means of removing sample members from the active sample if required. Periodic checking of the inactive sample was carried out to establish their current status eg checking death registers, checking out-of-scope members via contact names.

6.3.4 Panel maintenance

A custom designed Panel Maintenance Database (PMDB) was used to maintain accurate address records and other between wave information for panel members. The PMDB was maintained as a database of names and addresses of sample members held separately from the survey database containing the interview data for reasons of confidentiality and to comply with the UK Data Protection Act. The PMDB was updated in the year between interview points if notification of a change of address was received. Sample members were issued to the most recently known address that may or may not be the address of interview at the previous wave. Maintaining contact with respondents was facilitated by:

- providing a named contact person, freephone number and answerphone for respondents
- recording details of contacts with respondents between interview points
- passing any relevant information about respondents to the interviewer before each round of interviewing eg news of a family bereavement/illness
- an annual pre-fieldwork mailing of a short Respondent Report of research findings and activities with a confirmation of address card for Freepost return
- providing respondents with a change of address card for Freepost return
- sending a £5 gift voucher incentive to any person returning a change of address card between interview points
- updating address details between interview points
- maintenance of an historical record of all addresses ever occupied for each sample member
- tracing of respondents both during and between fieldwork periods

6.3.5 Tracking

Approximately 10 per cent of the BHPS sample (1,000 individuals) move in a given year. In up to one half of these cases some notification of the change of address was received via the change of address card, the confirmation mailing or by telephone. For the remainder, the tracking process began at the point when the interviewer made their first call at the issued address, found the respondent had moved and was unable to find a new address for them. At each wave of the survey interviewers returned details of between 200 and 250 individual or whole household untraced movers to the office for further tracking. Tracking was carried out mainly through contact names supplied by the respondent in previous years. Every year all respondents were asked for details of a contact name who would know where they are if they happened to move (Table 23). 50 per cent of households for which interviewers found no new address were traced using contact names. From wave 10, respondents were also asked for their own email address and a mobile telephone number. The tracking procedures are summarised in Figure 1.

At wave 6, all untraced movers since wave 2 were re-issued to field for a special tracking exercise. Interviewers were provided with details of the last known address and paid a bonus for finding any of these missing sample members. At wave 6, 96 were successfully found, of whom 54 completed full interviews. Of these 96, 54 remained in the sample at wave 10, of whom 33 completed a full wave 10 interview.

6.3.6 Refusal conversion

Refusal conversion at wave 1 consisted of a letter from the ISER, followed by a visit to the address by the interviewer. Of 685 refusals re-contacted, 62 co-operated. This procedure was revised at wave 2 as the postal/door-step contact was less effective than telephone contact. Wave 1 procedure replaced by initial approach by telephone using an experienced interviewer based at ISER. The converter contacted the household and in the event of a successful conversion the household was reissued, in most cases, to the original interviewer unless the respondent requested a different interviewer. Of 319 households approached over the phone at wave 2, 78 (24%) provided at least one interview. Where there was no phone number ISER approached the household by post. Of the 50 households approached by post 13 produced at least one interview

(26%). The final household conversion rate was 25%. These refusal conversion procedures were followed at waves 3 - 10. From Wave 3, a short telephone interview was introduced during conversion if agreement to a full interview could not be achieved.

From Wave 2 onwards, the BHPS issued all previous wave non-contacts and refusals from the previous wave judged to be 'soft' refusals from reasons recorded by the interviewer.

Table 24 summarises the outcome of refusal conversion attempts from wave 4 onwards. It can be seen that the proportion successfully converted at each wave varies between 42% and 56%, though a majority of these provide only a telephone interview. The final two columns of the table demonstrate that the effects of conversion last beyond a single wave. Of those converted, between 54% and 72% are successfully interviewed again at the following wave. Further details of the BHPS refusal conversion programme and discussion of the outcomes can be found in Burton et al (2006)

6.4 Field outcomes and unit response rates

Table 25 to Table 37 show the cross-sectional fieldwork outcomes for the issued sample at each year of the BHPS for the original (1991) sample. These cross-sectional fieldwork outcomes provide information on the fieldwork operation at each wave of the survey for the issued sample. They do not provide information on the longitudinal response outcomes for original sample members or for the whole sample. See Section 13.2 for a discussion of sample attrition due to non-response.

Following Wave 1, all those living in an interviewed household at Wave 1 were fed forward at Wave 2. In addition, non-contacted addresses at Wave 1 were also issued to field to attempt a contact. At Wave 3, the fed forward sample consisted of all those living in an interviewed household at Wave 2 together with most of the untraced movers. Out of scope respondents were not issued to field and those who had refused to continue with the survey at Wave 2 were placed in the 'retiring' category (see Section 6.3.3.1) and not issued to field.

At each following wave of the BHPS, the sample that is fed forward into fieldwork consists of all those living in a household where at least one interview was achieved at the previous wave, all non-contacted addresses from the previous wave, some whole household refusals from the previous wave and some untraced movers from the previous wave. Respondents who have refused to continue with the survey are placed in the 'retiring' category and are archived out of the main sample in the following year. Untraced movers and out of scope respondents are maintained within the 'inaccessible' category in case they are found and interviewed. At Wave 6, a special effort to trace untraced movers since the start of the survey was carried out. Following that exercise, any untraced movers who remained untraced had their details archived. A similar exercise was carried out at Wave 10. See Sections 6.3.3 and 6.3.5 for further details of the sample management and tracking procedures used on the BHPS.

Equivalent tables of fieldwork outcomes are presented for the Scotland and Wales extension samples (Table 38 to Table 46) and for the Northern Ireland sample (Table 47 to Table 49).

6.5 Item non-response

Levels of item non-response ie missing, refusal and don't know responses to particular questions, provide a measure of data quality. Item non-response may be caused by the respondent's perception that the question is too personal or intrusive, resulting in a refusal to provide a response. Alternatively, respondents may give a 'don't know' response which is more difficult to interpret (Beatty and Hermann, 2002). It may be the case that respondents genuinely do not know the correct answer, especially where an exact amount or specific date is being requested for example. On the other hand, a 'don't know' response may be a polite refusal. The respondent does not want to give the information but rather than refusing outright, says they do not know. Missing responses are often due to interviewer error where the routing has been followed incorrectly and a question that should have been asked is not. With the introduction of CAPI, missing data of this kind should be eliminated even though interviewers could still code a response incorrectly and, as a result, be routed incorrectly by the CAPI programme. The other source of missing data is when the respondent does not understand the question or see that it applies to their particular situation so is therefore unable to respond.

In a panel survey, item non-response, particularly to potentially sensitive questions such as income, might be expected to fall over time. As the respondent grows to trust the survey and build a rapport with the interviewer they might be expected to be more co-operative and more inclined to provide valid responses. Items that require recall over an extended period of time eg the start date of current job, will be subject to recall error, particularly where dates are further from the date of interview. In this case, the recall error would be expected to increase in later waves of the panel for respondents who have not changed their job and for whom the start date is receding into the past.

This section details the overall levels of item non-response on the BHPS waves 1 - 13. The levels of item non-response on some selected variables are then described. Finally, the strategies used to reduce item non-response on the wealth, assets and debts questions carried at waves 5 and 10 are described.

6.5.1 Overall item non-response

The following tables show the mean levels of item non-response across all variables collected in the BHPS for each wave at both the individual (Table 50) and household (Table 51) level. Only substantive questions are included with interviewer checks, derived variables or system variables being excluded. Where a variable contains imputed data, the imputation flag has been used to recode imputed data as item non-response where appropriate. The tables show the mean levels of item non-response for all variables and for variables where more than $n=100$ cases were eligible for a particular question. On a small number of items, the proportion of item non-response was misleading due to the small numbers of the sample eligible for that particular combination of routing.

6.5.2 Item non-response for specific variables

Table 52 gives the mean level of item non-response for some selected variables. The level is consistently low for items such as marital status and health status, suggesting that these are items which respondents do not perceive as sensitive in any way and are items that they can respond to without difficulty. We have also included in the table some of the items with the highest levels of item non-response. These include, notably, measures of income, savings, housing costs and political allegiance.

It is clear that item non-response was higher at wave 1 for many of these variables and in subsequent waves has tended to fall. However, it can also be seen that the levels of item non-response on questions asking for actual amounts eg gross and net earnings for employees, has tended to remain stable since wave 2. The difficulties with recalling the exact day the current spell of employment began can also be seen. For employees, approaching half of respondents do not know the exact day they started their current job, between 5% and 7% cannot recall the month but fewer than 1% fail to remember the year their current job began. It should be noted, however, that the start day is not one of the most important analysis variables. It is included here as an illustration of the type of information that sample members find most difficult to recall. The recall of start dates of the current job was better for employees than the self-employed, something which may be attributed to the less structured forms self-employment can take making recall more difficult for respondents.

6.5.3 Wealth, assets and debt component - waves 5 and 10

The wealth, assets and debts component, first carried at wave 5, was designed with two key issues in mind. First, to avoid upsetting respondents by asking for potentially sensitive information and secondly, to reduce item non-response for those who did not refuse outright to give any details. The PSID in the States has extensive experience of the use of ‘unfolding bands’ to reduce item non-response on such types of questions. Their design principles were followed in constructing the BHPS questionnaire module. This consisted of the following:

- At the first outright refusal, respondents were routed past the follow-up series of questions to minimise offence;
- Respondents who gave a ‘don’t know’ response when asked for an actual amount of savings held, value of investments or debts, were routed through a set of unfolding bands to collect an approximate range for the amount;
- If, during the series of unfolding bands the respondent refused, they were routed past the follow-up questions for that section.

Table 53 shows, separately for the wave 5 and 10 core samples, the percentage of item non-response and valid responses for the ‘Yes/No’ question asking if they had the particular type of savings, assets or debts, the response to initial question asking for the actual amount held or owed, and the percentage of item non-response following the unfolding bands.

For all items, the use of the unfolding bands reduced the overall levels of item non-response. It is likely that the unfolding bands are most effective for those respondents who genuinely cannot

provide an actual amount but can provide a range. It may also be the case that those who provide a covert refusal through giving an initial 'don't know' response to the actual amount question are happier to respond giving a less specific range. The most sensitive items appear to be the actual amounts held in either savings or investments. In contrast, respondents seemed remarkably willing to provide details of debts owed. It is noticeable that for 2 of the 4 questions regarding amounts, the proportion refusing to disclose an actual amount was significantly lower at wave 10 than at wave 5.

Equivalent figures are given for the Scottish and Welsh booster samples in Table 54 and Table 55. The wealth, assets and debt questions were asked at wave 10 for these samples too, though this was only the second wave for these samples. As discussed earlier, this was because the funding for the second wave with these samples was confirmed too late to allow separate questionnaires to be developed. The original plan had been to ask this module of questions at the fifth wave for these samples, ie BHPS wave 13. It is noticeable that the item refusal rates are much higher for the two booster samples than for the core sample. Across the four questions regarding amounts, item refusal rates average 3.1% for the core sample, 6.8% for the Scottish booster sample and 6.0% for the Welsh booster sample.

7. Data Preparation

7.1 Coding

The majority of coding of verbatim responses was carried out by NOP coders post-field. Coded items included occupation and industry, a range of substantive questions such as reasons for moving house, reasons for saving etc. and the coding of 'other' non-standard time periods reported in conjunction with an amount received. Coding frames were supplied by the ISER. Where no coding frame was available, verbatim listings of 200 responses were made and the frame developed by the ISER.

Occupational descriptions were coded to the 1990 OPCS Standard Occupational Classification (SOC). Since wave 11, SOC2000 has also been coded and provided on the data set, in addition to 1990 SOC. Industry was coded to the 1980 Standard Industrial Classification (SIC). At Wave 1, coding of SOC and SIC was carried out manually. From wave 2, the SOC coding was carried out using the Computer Assisted Standard Occupational Classification (CASOC) system developed by Peter Elias (Campanelli and Moon, 1994 – see also Campanelli et al, 1997). Following the adoption of CASOC for occupational coding at wave 2, all of the wave 1 occupations were re-coded using CASOC and it is this version which is included on the BHPS user database. This re-coding exercise was, as far as possible, carried out blind given they were coding from responses on previously coded questionnaires. Coders were instructed to cover the SOC code entered on the paper questionnaire and code using CASOC independent of any previous coding.

A new SIC classification, SIC-92, was introduced in the UK in 1992. The older classification, SIC-80, continued to be used for coding on the BHPS for reasons of longitudinal consistency and comparison. At waves 4, 7 and 11, industry was coded to both SIC-80 and SIC-92 for the main current occupation. At all other waves prior to wave 11, only SIC-80 was used. This is a good example of a particular dilemma faced by panel surveys. Moving to SIC-92 would improve the relevance of the classification and would provide comparability and coherence with other surveys and data sources. However, it would also produce a break-point in the series, over which analysis of change would not be possible at either the micro or macro levels. Nevertheless at wave 12 the decision was taken to move to using SIC-92 and since then, SIC-80 has not been included on the survey data files.

In Northern Ireland, occupation has been coded to SOC2000 and industry to SIC-92 since the survey started, as issues of longitudinal consistency were not relevant for that sample.

The BHPS includes a number of questions that ask for an amount followed by a time period covered by the amount given. In all cases the time period has an 'other' code where interviewers wrote in the response that did not fit the pre-codes. These are coded post field to a standard frame to the number of weeks.

All 'other, specify' items were examined by NOP and back-coded if appropriate. All remaining 'other' answers were listed by NOP. Since the introduction of CAPI, the 'other specify' responses are available in machine readable format.

Since wave 5, the individual questionnaire has carried a final verbatim question on a variety of themes including events of importance to the respondent in the past year, perceptions of generational change and views about their local neighbourhood. These extended verbatim responses required the development of complex coding frames. The development of the frames for these items was led by Dr Jackie Scott (Cambridge) together with ISER staff. Following testing and validation of the frames, the coding was carried out at ISER from the verbatim responses provided by NOP.

A list of all the verbatim items coded on each wave is provided in Table 56.

7.2 Editing procedures

The data were subject to initial editing procedures on receipt at the NOP data processing centre. Following delivery to the ISER, a series of further consistency checks and editing were carried out to produce a clean and consistent data set.

Editing of the paper questionnaires at waves 1 - 8 carried out by NOP included the following elements:

- first-level checks where errors were sufficiently serious to necessitate immediate return to the interviewer. These were critical items such as inconsistencies in date of birth, sex or marital status which could not be resolved from information elsewhere in the questionnaire.
- second-level checks where the interviewer was notified without needing to return the questionnaire eg failure to code an item which could be reconstructed from information elsewhere in the questionnaire.
- checking the consistency of key household and individual identifiers between documents and if necessary, correcting transcription errors.
- checking money amounts and periods. If the period was outside the pre-coded list it was converted into weeks as a coding task.
- checking marginal comments by the interviewer and if appropriate, editing the data accordingly.
- “don’t know” and “refused” written in by the interviewer were converted into standard codes of ‘8’, ‘9’, ‘98’, ‘99’ etc. for punching.
- checking for failure to ring code values associated with “write-in” other responses and editing the questionnaire accordingly.

All items edited by NOP staff were recorded on the questionnaire in red ink to provide a trace of changes made. Following data entry, the only checks carried out by NOP prior to delivery were to ensure that all data for a given household was present. With the introduction of CAPI at wave 9, the elements of visual editing have reduced significantly and were carried out only for the remaining paper documents.

At each wave, the ISER carried out a range of consistency checks and editing procedures to ensure, as far as possible, a clean data-set for users. The primary aim of the data processing carried out at ISER was to ensure the production of a data-set which was internally consistent at

each wave while ensuring the longitudinal integrity of the data over time. Using a suite of checking programs, the data returned from the field were validated against the expected structure of the data given the logic and content of the questionnaire. Any departures from the logic of the questionnaire were flagged for further investigation.

The ISER data processing procedures aim to resolve, as far as possible, any inconsistencies within the data that may cause difficulties during analysis. For example, the interviewer may have made an error in routing which needs correction, a punching error may have been made or there may be inconsistencies between items at different points in the questionnaire which can be resolved. It was not ISER policy to change the information reported by respondents unless there was a clear error that could be corrected. If the respondent reported inconsistent information and there was no clear evidence of which response was correct, the inconsistency would remain for researchers to make their own decisions. All checking involved returning to the paper questionnaires to establish the source of the problem before any editing decision was made. Consistent conventions for distinguishing missing items from inapplicable, don't know or refusal responses were also implemented.

The data processing was split into sequential stages through which each batch had to pass before going onto the next. Data were delivered to ISER in batches so that data processing could start without waiting for the completion of fieldwork. The systems developed for checking the data required the data to pass through a series of checks, with all queries at one point being resolved before moving to the next.

The data processing at ISER comprised the following stages:

- booking-in checks and filing. These checks ensured that all data for a given batch was present and that the paper documents received for that batch were present with correct identifiers.
- primary structural checks. These checked the internal consistency of serial numbers, personal identifiers within the household and other key values such as date of birth and sex.
- secondary structural checks. These checked the substantive consistency of major critical variables and the chronological consistency of data references, for example the sequence of events reported in the annual job history.
- routing consistency checks. These checked for interviewer errors in following the routing through the questionnaire.
- remaining internal consistency checks. These checks validated all remaining items against the logic of the questionnaire.
- plausibility checks. These checks examined outlier amounts and unlikely combinations of responses even though in many cases no information was available to justify any editing of the data.
- editing of data where clear error due to interviewer or punching error was found.
- cross-wave longitudinal consistency checks. These longitudinal checks were primarily to ensure the correct identification of individuals and the accuracy of key demographic data such as sex, date of birth and relationships within the household eg biological parenthood.

The shift to CAPI at wave 9 altered the tasks involved in data processing. Many of the tasks previously carried out post fieldwork were pushed forward and incorporated into the CAPI script used by interviewers. The routing through the questionnaire is controlled during a CAPI interview, reducing the level of interviewer error. Valid codes and ranges are specified in the script and a variety of checks at the point of data collection built in. For example, checking of the entry of dates or amounts that are either very high or very low can be incorporated. While some elements of the data processing post-fieldwork were reduced, the up-front work in writing, testing and checking the CAPI script was significant. The CAPI data were subject to the same checking procedures as in earlier paper waves to ensure that no errors in the CAPI script were present in the data and that the data conformed to the expected structural and substantive elements required by the logic of the questionnaire. Some editing of the data was carried out where clear errors were found even though this process was more limited than in earlier paper waves.

7.3 Data capture

All paper questionnaires were keyed at the NOP data processing centre using a key-to-disk system with 100% verification by a different operator. Only valid codes could be entered. Computer edits ensured that the data was complete and the person number was consistent across individual related documents.

From Wave 9, the introduction of CAPI to collect the household and individual questionnaires reduced the data entry task significantly even though some paper documents remained to be keyed. The CAPI questionnaire data are dialled in by interviewers via modem direct to a central server. Interviewers returned paper documents, including the coversheet, to the NOP data processing unit as in earlier waves. The mix of paper and CAPI instruments required the development of a data management system at NOP to correctly identify and match data from the differing data streams. This system runs a series of checks to ensure that CAPI and PAPI data are correctly matched and that all records required for a given household or individual given the response outcomes, are present (see Banks and Laurie, 2000 for further details).

7.4 Quality control procedures during data processing

Quality control procedures during data processing at NOP included:

- detailed set of written instructions, outlining each task to be performed during editing and coding
- booking-in procedure to verify that all relevant documents had been received for a given household and individuals within the household
- any missing documents were requested from interviewer
- in waves 1 - 8, the first 5 individual questionnaires sent back from each interviewer given a 100% visual edit. If these were satisfactory, subsequent questionnaires from the same interviewer were given a partial visual edit.
- Verbatim items requiring coding were subject to a blind re-coding by a second coder. This constituted a 5% sample or 250 cases (whichever was the greater) for each item requiring coding.
- ISER staff regularly monitored the coding and editing at the NOP data processing centre.

The validation of coded items was carried out by NOP. At waves 1 - 8 when paper questionnaires were used, a random selection of verbatim responses for each question requiring coding was photocopied prior to coding. These cases were then given to a second coder to code 'blind' and the results of the coding on the paper questionnaire and the recode compared. Following the introduction of CAPI at wave 9, an on-line coding system was introduced. A random selection of responses for each verbatim item are selected and again are given to a second coder to code 'blind'. The results of this exercise are given in Table 57 for some key items.

It is noticeable that as the survey has become more established, the level of discrepancy has reduced. For example, at wave 1 there was a 16% discrepancy rate for the reason respondents want to move house but by wave 8 this was only 5%. This is likely to be due to the coders gaining familiarity with the coding frames and tending to code more consistently over the years of the survey. The coding of occupations to SOC has a remarkably high reliability rate, particularly from wave 3 onwards. This is likely to be due to the introduction of CASOC, the computer assisted coding system used by NOP since wave 2. Following entry of the job title, the programme suggests the best code on the basis of word recognition. Coders can then either accept the suggested code as the best code given the verbatim description provided or search using various facilities in the programme for a better match. The results of the coding validation suggest that coders, in the main, tend to accept the code first offered by the programme on the basis of the job title. This enhances reliability but may reduce validity, even though this has not been explored and we have no evidence on this to date.

Differences in coding can be seen to be greater from wave 10 than on earlier waves. There are a number of possible reasons for this. When the survey was PAPI (up until wave 8), validation coding was completely blind, done using photocopies of the pages of the questionnaire containing verbatim responses, with the photocopy made before any coding was done. It was possible for a coder doing the validation to pick up a photocopy from a questionnaire she herself had coded in the first place, and indeed this would have happened on many occasions, given the relatively small number of coders involved. A question being coded by the same person twice is far more likely to be coded the same way as one coded once by one coder and once by another. Under the CAPI system, each coded question includes a coder identifier, and in the validation process coders were only given access to questionnaires they themselves had not coded originally. This would certainly lead to a higher level of variability in the dual-coding.

The second reason is that when coding was done on paper questionnaires, the original coders were able to look elsewhere in the questionnaire for information that might help them resolve difficult cases (but not the second coders, as they had available only the single photocopied sheet, with little or no context information). With CAPI, both codings are more akin to the recoding from previous waves, rather than to the original coding. Some additional items of context information were included at key questions such as the CASOC coding of current occupation, but most questions did not have any. To the extent that CAPI coding made the original coding more difficult, one would expect a general increase in variability levels.

8. Statistical Adjustment Procedures

8.1 Weighting procedures

The BHPS data set contains (longitudinal) weights for each wave of data. The means of calculation of these weights is discussed below. In general, there are separate weights for respondent individuals, for all enumerated individuals and for households. Proxy and telephone respondents have zero respondent weights, but positive enumerated individual weights. There are also cross-sectional weights for use with single wave analyses for each wave.

8.1.1 Wave 1 weights

The weights for use with Wave One data were created using the following sequence of adjustments:

- (1) Weights to adjust for unequal selection probabilities of addresses (design weights);
- (2) Weights to adjust for non-response at the household level;
- (3) Weights to adjust for non-response of individuals within responding households;
- (4) Re-scaling of final weights so that weighted sample sizes equal unweighted (interviewed) sample sizes.

Weights were derived in the above order with all calculations being based on data weighted by the product of all previously derived weights so that the weighting adjustments were made contingent on the already derived weights. The final weights used in analysis are the product of these weights.

For waves after Wave One, two types of weights are derived:

- (a) the longitudinal weights, for those interviewed at all waves up to and including the current wave and for those enumerated in respondent households in all waves up to and including the current wave. These allow the analysis of change between sequences of waves, by adjusting for sample loss between the waves.
- (b) the cross-sectional weights, for those enumerated at each wave and for those giving a full interview. These allow the use of data in cross-sectional analysis by including new entrants and adjusting for within household non-response.

8.1.2 Longitudinal weights

The longitudinal respondent weights for wave j select cases who gave a full interview at all waves up to and including j . At each wave, these cases are re-weighted to take account of previous wave respondents lost at the current wave through refusal or some other form of sample attrition. Thus, the longitudinal weight at any wave will be the product of the sequence of attrition weights accounting for losses between each adjacent pair of waves up to that point, as well as the initial respondent weight at wave one. It should be noted that for this purpose response also includes the deceased, people who have moved into institutions or otherwise gone

out-of-scope. These fail to give an interview not through non-response but due to a terminating event that results in their leaving the population of interest.

Due to varying amount of information available for non-respondents, the longitudinal respondent weights were calculated in two stages. First, all respondents at both waves including those with "terminal events" were weighted to adjust for the attrition of cases whose final status was indeterminate, in that it was not known whether these cases were still eligible for interview or had left the population of interest. These included people who had moved from their previous wave address and were subsequently not traced for interview, as well as refusal households where the interviewer was unable to determine who was still resident and eligible. The second adjustment weighted up the cases interviewed at both waves to take account of those who refused an interview, those for whom only a proxy interview was possible and those who were unable to give an interview at the later wave.

Weighting was carried out using a weighting class method where respondents and non-respondents were classified by a number of variables thought to be informative of non-response or of critical interest in the analysis of BHPS data. The main assumption of this approach is that, within the final cells, the respondents and non-respondents constitute random samples of the population sub-group defined by the cell variables. Since all cases (except new 16 year olds) were respondents who gave a full interview at the previous wave, there were a large number of variables available to define these classes. In order to make this process manageable, an automatic interaction detection programme (SPSS CHAID) was used to aid the splitting of respondents and non-respondents into groups defined by variables associated with non-response. This allowed for the definition of very specific weighting classes and for easy control over the size of the classes and their percentage of non-respondents. The inverse of this non-response ratio defines the weight to be applied to respondent cases within each class. Since some of the most informative variables for non-response had small numbers of missing values these variables were initially imputed using a hot-deck procedure. This method applies to the majority of the weighting factors discussed below.

Variables used in these adjustments included: Whether moved from the previous wave address; individual characteristics such as age, sex, employment status, income total and composition, race, level of organisational membership, educational qualifications, etc. and household characteristics such as region, tenure, number of cars and ownership of consumer durables. The initial attrition weight was defined as the product of the previous wave longitudinal respondent weight (before post-stratification) and the adjustment factors defined on the basis of the two weighting steps described above. After this, a post-stratification adjustment was added so that the Wave One characteristics of the surviving sample corresponded to population marginals for 1991, in terms of age, sex, housing tenure, numbers of cars and household size.

In addition to respondents at both waves, those previous wave children who reach the age of 16 and are interviewed receive longitudinal respondent weights. In order to adjust for this group, children interviewed were given a weight defined as the minimum of the longitudinal respondent weight of their parents, or the minimum longitudinal respondent weight in the household if no parent weight was available. This rule was applied since this group were too small to model adequately for adjustment and since children rising to age sixteen and eligible for interview are

likely to have more in common with other members of their household than with other children in this category. Minimum values were used since this was most likely to reflect the probability of a household response. After this adjustment, the whole group of respondent new 16 year olds was re-weighted back to the number of eligible 16 year olds.

From Wave Three onwards, a small number of respondents out of scope at previous waves return to the sample. These cases are treated for weighting purposes as if they had responded at the previous wave (as they have responded at all waves at which they were eligible to respond). Predictor variables are taken from the most recently available wave.

Weights for individuals enumerated at each wave were derived using the same two-stage method as used for longitudinal respondents with the weighting classes being primarily based on household and head of household characteristics. This weight adjusts the individuals enumerated at each wave (including those experiencing terminating events) for the cases lost through attrition. Longitudinal weights for children and for proxy and telephone respondents as well as within household non-respondents are provided by this weight. New births to the sample are given the mean value of their parents' weights (so that children with two sample parents generally receive a higher weight than those with only one sample parent).

There is no longitudinal household weight since households are not definable as longitudinal entities.

8.1.3 Cross-Sectional weights

For some research purposes, it is desirable to analyse each wave of the BHPS as a cross-section. In order to make this possible, cross-sectional weights have been derived that allow for the inclusion of new entrants who, by definition, do not have a Wave 1 or longitudinal weight. Assumptions are necessary in order to include these individuals since their initial inclusion and response probability are unknown. There are a number of approaches available and we employed a technique called the "fair shares approach" (Ernst 1989, Lavallée & Hunter 1992, Rendtel 1991), which is also used by SLID in Canada. Basically, this approach shares the Wave 1 weights of all enumerated cases, after adjustment for attrition, to all other enumerated members of their later wave household. The sharing of weights was applied so that all members of each household have a weight equal to the sum of the weights of all the Wave 1 enumerated individuals, adjusted for non-response, divided by the number of members of the population at Wave 1 (ie including new entrants who were in that population, but excluding new births since Wave 1).

The first stage of this derivation was to weight the enumerated individuals present at Wave 1 to adjust for attrition up to the latest wave. This used a similar weighting class method to that described above for longitudinal respondents. The weighting classes were defined using head of household, household and individual characteristics (available from the enumeration grid). Once these attrition factors were calculated, weights were defined for original sample members based on the Wave 1 weights (after post-stratification). These were used to define the fair share weights for all eligible enumerated individuals, including new entrants, as described above, giving the final cross-sectional enumerated weight. A cross-sectional respondent weight was also

calculated by adjusting the enumeration weight of all interviewed adults at the wave for those who refused, were proxied or were unable to give an interview. This weight adjusts for within-household non-response for cross-sectional non-respondents. Again a weighting class method was employed using similar variables to that defining the cross-sectional enumerated classes, except that the adjustment depended only on current wave characteristics.

A cross-sectional household weight is available. This is set equal to the cross-sectional enumerated individual weight subject to re-scaling back to the total number of households.

As discussed in the previous section, all weights are trimmed to a maximum value of 2.5, and re-scaled so that the weighted sample is equal to the total number of respondent individuals. A number of distribution statistics for weights at each wave are given in Table 58. As the mean weight is 1.0 in each case, the coefficient of variation of the weights is simply equal to the standard deviation, which is shown in the final column. It can be observed that the coefficient of variation increases over the waves, slightly more so for the cross-sectional weights than for the longitudinal weights.

8.2 Imputation procedures

Missing data on a range of income and housing cost variables have been imputed in all waves of data of the BHPS. This section discusses and evaluates the methods used to carry out these imputations.

Item non-response, where a respondent has given a full interview but where certain items on the questionnaire are missing, is a particular problem in all social surveys (Kalton and Kasprzyk 1986). Imputation is one of a number of possible techniques that can be used to deal with this problem. It is likely to be preferable to the default with standard statistical packages, which is to delete cases with one or more missing values when carrying out modelling procedures. This amounts to a strong assumption that the valid cases are a random sample of all cases, which implies that individuals with item non-response can be adequately represented by cases with complete data. However, this assumption may be problematic, and may could seriously bias results (Hox 1999). For example, refusers on a question asking about their dividend income over the year are likely to be systematically different from those answering this question so that the analysis of complete cases cannot be capturing the true nature of the population. One method of adjustment in such cases is to estimate the true value for missing cases using an imputation technique. Imputation techniques use various models with defined assumptions to obtain a 'best' estimate of the missing values (Martin 1999, GSS 1997).

It is important to stress that the main aim of imputation is to reduce potential bias caused by the elimination of cases with missing data, rather than to increase precision of estimates by increasing the effective sample size. Note that the main problem with imputation as a method of dealing with item non-response is that methods for adjusting estimates of precision such as confidence interval etc. are not easily available so that analysis carried out on data containing imputed values where this fact is not taken into account will tend to give an over-estimate of precision (Rubin 1996).

Each variable subject to imputation has an associated imputation flag variable, in order that imputed values may be excluded from the analysis, or subject to alternative adjustment methods. These flags take three forms. For variables directly associated with a question, the imputation flag takes the missing value code of the original variable (eg Don't Know, Refused, etc.) if imputed, and 0 if not imputed, or -8 if the variable was inapplicable, including cases where the respondent was a proxy. This is to ensure that different missing value categories can still be distinguished. For individual level derived variables, as well as housing related derived variables, the imputation flag takes the value 1 if the variable was imputed, and 0 or -8 otherwise. For household income variables, the flag distinguishes cases where the imputation only concerns some component of a respondent household member's income, from cases where the whole income of one or more household members was imputed.

Two main imputation techniques were used:

8.2.1 Hot-deck imputation

A standard Hot-Deck imputation routine (Ford 1983) is analogous to weighting using weighting classes. This method was applied by firstly dividing the sample into imputation classes found to be predictive of the variable to be imputed. Then, assuming that cases within each class comprise a random sub-sample of the population, a valid value of the variable taken from a non-missing case within a given imputation class was used to impute the value of a missing case in the same class. The validity of this procedure is dependent on how informative the imputation classification for predicting values of missing cases. It ensures also that the imputed value is a possible value for a respondent with the relevant characteristics, and also that some randomness is introduced into the assignment of an imputed value.

This method was used for certain categorical money variables such as Proxy's personal income, banded income from Dividends and Investments, and a number of cases where regression methods appeared inappropriate (eg income from welfare benefits). In order to ensure that the imputation classes were as informative as possible, classes were defined using an automatic interaction detection programme (SPSS CHAID). This procedure allows a high degree of control over the definition and size of these classes and can handle a large number of classification variables.

At various points in both the derivation of weights and for imputation of money amount variables, there was a requirement to impute a small number of missing cases so that certain variables could be used in the definition of weighting classes or within model based imputation procedures. Similar hot-deck procedures were used for this imputation. Since the variables imputed were those with less than two percent of cases missing, these imputation are unlikely to seriously effect the derivation of weights or the more complex regression imputations. These minor imputations are not carried over to the public released data set.

8.2.2 Regression imputation

Money amount variables were imputed using a regression-based imputation technique. First, a regression model was fitted to all valid cases for the variable of interest using predictor variables

which were non-missing (or had themselves been imputed) for both valid and missing cases of the variable to be imputed. Once a well fitting model was obtained, defined in terms of maximal adjusted R-squared, predicted values were defined for valid and missing cases using the model. These were then used to find the closest valid case in terms of the predicted value for each missing case. The missing case was then imputed with the real value of the closest valid case. This form of regression imputation is termed predictive mean matching. The advantage over imputing at the predicted value is both that a possible real value is imputed, and that a random error component is added so that the imputed values are not subject to less variance than reported values.

Having imputed a number of primary variables, a number of other income related variables were computed from these variables, with some additional small scale hot-deck imputation; for example, for the small number of cases where there was a complete refusal to the financial receipts section. Incomes were also imputed for refusers to the whole questionnaire, in order to construct a complete household income.

In a panel study, there are not only variables from the current wave available to use as predictors in an imputation process, but there may also be variables from the same respondent collected at a different wave. It is likely that the best predictor of a missing value for a variable at Wave One, is a value of the same variable at Wave Two. However in using this value in 'cross-wave imputation' it is important to ensure that biases in rates of change in values are not introduced. This is essentially achieved through the methods described above. So in making a cross-wave imputation we are essentially taking a value from a donor who is both similar to the recipient in current characteristics and in the value of the imputed variable at the other wave. The imputed value should therefore imply a rate of change drawn from a randomly selected similar case. This approach will avoid introducing spurious change for panel analysis, which would be likely to arise if only single wave imputation was used.

9. Documentation and Data Accessibility

9.1 Data Organisation

At the time of writing, there are currently three BHPS data sets available at the Data Archive:

- (1) SN 5151 - British Household Panel Survey; Waves 1-13, 1991-2003
<http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=5151>
- (2) SN 3954 - British Household Panel Survey Combined Work-Life History Data, 1990-2003
<http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=3954>
- (3) SN 3909 - British Household Panel Survey Derived Current and Annual Net Household Income Variables Waves 1-12 1991-2002
<http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=3909>

The BHPS is not available in single waves, and the Living in Scotland and Living in Wales sub-samples are not available separately. There is one main file of the BHPS data and when a new wave is released the Data Archive adds it to the existing data set. The data set is then re-catalogued and the old catalogue number is removed.

In addition to the main BHPS data sets, the data from the ISMIE methodological study (referred to elsewhere in this document; see also <http://www.iser.essex.ac.uk/ulsc/methods/research/ismie/index.php>) are also available:

- (4) SN 5157 – Improving Survey Measurement of Income and Employment, 2001-2003
<http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=5157>

9.2 Documentation

Accompanying each data set on the Data Archive web-site are details which give the following information:

- Coverage
- dates of fieldwork
- country
- spatial unit
- observation unit
- Universe sampled
- location of units of observation
- population keywords
- population
- Methodology
- time dimensions
- sampling procedure
- method of data collection

- control operations
- Related studies
- References and publications
- by Principal Investigator
- by Others
- Access conditions
- Availability
- Date of release
- Copyright information
- list of on-line documentation

9.2.1 BHPS Waves 1-13, 1991-2003

For the main survey, BHPS Waves 1-13 (SN5151) the documentation available on-line from the Data Archive mirrors that available from the ISER web-site at <http://iserwww.essex.ac.uk/ulsc/bhps/doc/>. The data were released in April 2005. The on-line documentation has Adobe Acrobat pdf files of Volume A of the BHPS User Manual (“Introduction, Technical Report and Appendices”) and Volume B (“Codebook”) for each wave from one to thirteen. The documentation also includes questionnaires and show cards for each wave consisting of: the coversheet; the household grid the household, individual, self-completion and proxy questionnaire; show cards for all appropriate questions.

There is just one version of Volume A of the User Manual which gets updated when necessary. There is, however, a separate Volume B for each wave of data. So, the full documentation for waves 1-13 consists of one Volume A and thirteen Volume B files.

Volume A of the User Manual, consists of the following sections:

- I. Introduction to the Documentation
- II. Introduction to the British Household Panel Survey
- III. The BHPS Data
- IV. Sampling and Survey Methods
- V. Weighting, Imputation and Sampling Errors
- VI. Data Dissemination
- Appendix 1. Using BHPS Data
- Appendix 2. Notes on Derived Variables
- Appendix 3. Coding Frames
- Appendix 4. Help For Old Friends: Modifications Since Eighth Release
- Appendix 5. Related Publications and Documentation
- Appendix 6. Indexes

Volume B of the User Manual contains variable tables (frequency distributions) for each record type. The information on each variable covers the variable name, a description, the question number and text, the question route, index terms and a list of the waves in which the variable appears. The record types covered are wHHSAMP, wINDSAMP, wINDALL, wHHRESP,

wINDRESP, wJOBHIST, wINCOME, wEGOALT, wYOUTH, wCHILD, XWAVEID, XWLSTEN and XWAVEDAT.

9.2.2 BHPS Combined Work-Life History Data, 1990-2003

The BHPS collects extensive information on respondents' labour market status, (i) at the time of interview at each wave of the panel, (ii) through the period between 1st September a year before and the interview date, and (iii) retrospectively from first leaving full-time education. Because the retrospective information was collected in two sections (one focusing on employment status, the other on occupational information) there are four different types of labour-market history information, located (at Wave 8) in eighteen different files in the BHPS database.

This complexity is a necessary aspect of longitudinal information, but it has inhibited use of the work-life history data. In order to facilitate such use, a set of 'reconciled' files has been created, constituting single continuous records each containing all the information of a particular type in a single location. The first part of the exercise is to take 'current status' information and combine it with the inter-wave history, for each wave, and then to combine the eight waves thus creating a continuous record from September 1990 to the September 2003 (and later). The second stage is to take the life-time employment status history collected at Wave 2, and the life-time occupational history collected at Wave 3, and to combine each of them with analogous information drawn from the combined panel file, thus creating employment and occupational histories that stretch from the labour-market entry to the latest wave. The third stage is to combine these two extended life-time histories into a single record which contains both employment-status information (with good information about non-employed spells) and occupational information (that is, details about the job held during each employed spell).

The on-line documentation for this data set is available on the Data Archive web-site at <http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=3954>. The documentation consists of a text file of notes on the variables, a file which looks at the differences between this data release and the 1997 release and the following technical paper:

Halpin, B. (1997) Unified BHPS work-life histories : combining multiple sources into a user-friendly format, *Technical Papers of the ESRC Research Centre on Micro-Social Change, Technical Paper 13*, Colchester: University of Essex.

There is also a User Manual which is not available for download but is included with the data as an Adobe Acrobat file. The Work-Life Histories data were first made available in November 1997, with a second version in February 2000, which included information for Waves 6 - 8. In February 2002, the lifetime retrospective files were updated to include 1999 information. In October 2004 the files were updated to include information up to 2003. The other data files in the study and the documentation remained unchanged.

9.2.3 BHPS Derived Current and Annual Net Household Income Variables Waves 1-12, 1991-2002

This data set provides derived net income variables for BHPS waves 1-12. It is an unofficial supplement to the set of derived income variables in the official BHPS release (which focus on gross income rather than net income). The purpose of these additional files is to provide an analysis of income and poverty dynamics in Britain, and related topics.

The aim was to produce a longitudinal complement to the cross-section income distribution information provided by the Department of Social Security's Households Below Average Income (HBAI) reports, and to this end the HBAI definition of net income (DSS 1993) has been closely adhered to. Also included are BHPS versions of HBAI family type and economic status variables.

The on-line documentation for this data set is available on the Data Archive web-site at <http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=3909>.

The documentation consists of a User Guide, comprising:

- Introduction
- Sample Selection Criteria
- Variable Lists
- The Definition of Current and Annual Net Household Income
- Deriving Net Incomes Using the BHPS
- References
- Appendix: DSS Monthly Price Indices
- Appendix: Codebook Listings for Files

The data were first released 29 September 1998 and the latest edition available is 30 June 2004 (5th Edition).

9.2.4 Improving Survey Measurement of Income and Employment (ISMIE) Data

The survey data from this methodological study were released by the Data Archive in May 2005, with a 2nd edition in October 2005. The data are available in a series of files, the structure of which mirrors that used for the main BHPS files, except that extra files contain the data from the dependent interviewing versions of questions. Standard BHPS linking variables are included, permitting the ISMIE data to be linked with the previous BHPS responses of the ISMIE sample members. A user guide is available on-line at <http://www.data-archive.ac.uk/findingData/snDescription.asp?sn=5157>, describing in detail the study, the data and file structure. The user guide documents the ways in which the ISMIE questionnaire differed from the BHPS wave 12 questionnaire, listing questions that were dropped from the ISMIE interview and questions that were asked differently (the dependent interviewing experiment). This must therefore be read in conjunction with the BHPS wave 12 questionnaire, which is also available on-line alongside the user guide.

The validation data on income sources, linked from Department for Work and Pensions administrative systems, are not currently available via the Data Archive. It is hoped that a suitably non-disclosive subset of these data can be made available in a future release.

9.3 Formats, Cost and Access conditions of the BHPS data sets

The British Household Panel Survey data is available primarily from the ESRC Data Archive. To acquire data from the archive the user must first register their details. The user is then able to download the data directly over the internet. Alternatively the user may place an order online for the data to be sent to them on CDs or uploaded to an ftp server for the user to access and download. The BHPS data is available in SIR, SPSS and STATA format with documentation files in PDF and HTML format. The Combined Work-Life History Data (SN 3954) is only available in SPSS format.

BHPS data are also available from international data archives who operate an arrangement with the ESRC Data Archive through the Consortium of Household Panels for European Socio-economic Research (CHER). Sub-sets of BHPS variables are supplied to harmonised cross-national data-sets such as the European Community Household Panel (ECHP) (waves 7-11 only), the Cross-National Equivalence File (CNEF), the Panel Comparability project (PACO) and the European Panel Analysis Group (EPAG).

The user must also sign an access agreement before they can acquire the data. There are two access agreements, one is for individuals using the data for their own research, the other is for using the data to teach. The access agreements ensure that the user abides by certain conditions, primarily that the data should be used for non-commercial research and that the confidentiality of the data is preserved. The access agreements may be viewed in full at:
<http://www.data-archive.ac.uk/orderingData/linkAccessAgreement.asp>.

ISER provides BHPS data to the Data Archive free of charge. The data is free for academic users. The Archive may make a charge to non-academic users and for the cost of materials. More information about the Archive's charges are at:
<http://www.data-archive.ac.uk/orderingData/feesExplained.asp>.

9.4 The Data Documentation Initiative

The Data Documentation Initiative (DDI) is an effort to establish an international criterion and methodology for the content, presentation, transport, and preservation of "metadata" about data sets in the social and behavioural sciences. Metadata (data about data) constitute the information that enables the effective, efficient, and accurate use of those data sets. More information about the DDI is available at <http://www.icpsr.umich.edu/DDI/>. The DDI committee has produced a Document Type Definition (DTD) for social science data sets. The DTD uses XML to create codebooks in a uniform, highly-structured format which enables the information to be read easily by computers and internet search engines.

The DTD produces marked-up codebooks with sections on document description, study description, data files description, variable description and other study-related materials. Each

section is divided into sub-sections which have information on a particular aspect of the study. The same type of information is in the same place for every study so that searching across data-sets is more efficient. Whilst an XML version of the BHPS documentation is not available, the information that is needed to complete the DTD fields is, to a large extent, available across the BHPS Documentation web-site.

10. Data Usability and Interpretability

10.1 Data structure

The BHPS is a household survey in which all adult members of each household are eligible to be interviewed. It is also a panel study, so data may be expected from each individual each year. It is also quite likely that individuals may move out of one household, and join with new people. Data are therefore collected at different levels (individual and household), and over time these levels will not fall into neat hierarchies. All these factors mean that the data from the BHPS will have a relatively complex structure.

Moreover, the range of different research uses focusing on different units of analysis means researchers need to be able to restructure the data. The BHPS team took the view that these factors meant that a complete flattening of the data into a rectangular format was unlikely to be either feasible for all data, or particularly efficient where it was practicable.

However the structure has been simplified compared with the questionnaire, which contained significant amount of routing and repeating group structures. These were retained in the database used for data verification and cleaning as described in Section 7, but a User Database was created from this version, with simpler structures, some consistent data presentation features described below, and with a significant number of derived variables created (around 20% of total variables), to minimise the need for users to undertake complex matching, and for example to compute total income values.

The User database is initially created in SIR, but this is transformed directly into a range of other software formats, including SPSS, STATA and SAS. The User Database consists of two types of record (or file or data set using other software terminology). The majority contain the data collected at each wave for different subsets of questions and respondents, and related derived variables. Record types have a consistent naming structure across waves, with a single letter wave prefix. As far as possible a single record type corresponds to a single questionnaire instrument, or a logical group (eg AINDRESP – wave one individual questionnaire and self-completion, BHHRESP, wave two household questionnaire). However, some repeating data structures have been separated out (eg AJOBHIST – wave one annual job history). There are also two record type which contain no substantive data, but whose purpose is to facilitate linkage of data relating to the same individual at separate waves.

10.2 Identifiers

With the BHPS, there are two sorts of primary key variables: first, wave specific key variables which uniquely identify:

- (1) the household which is surveyed at the particular wave; WHID (where W substitutes for the wave identifier letter)
- (2) the individual's number within the household at a given wave, WPNO. Information about relationships between people within households (eg parental, filial, spousal) is given in terms of these WPNO variables.

These wave-specific identifiers are used to link together information from different levels within one wave, but not to connect information across waves. There is no connection between the value of the WHID at one wave and the next, nor, as with some panel studies does the WHID carry any information about household split-off status.

To make connections between information for the same person in successive waves, there is:

- (1) the cross-wave personal identity number PID.

These WHID, WPNO and PID variables are included in most of the files that could have valid "primary" matches.

10.3 Variable names

As with record types, all variable names (except those used in cross-wave operations) begin with a single character wave identifier. The rest of the name is mnemonic which attempts to give some information as to the content of the variable. In general, the second and third characters give some indication of the general subject area of the variable, eg JB for current job characteristics, FI for finance and income variables.

Where question wording remains identical or substantially the same across waves the variable name suffix will remain the same. The documentation draws attention to minor wording differences, and for example to a small number of cases where there are changes in the coding response categories. Where it has been judged that the question wording, or the range of response categories has changed sufficiently that the responses might be treated as equivalent rather than identical, a new variable suffix is used. This is identified in the documentation, and users can also use the cross-wave subject index to identify other potentially comparable variables.

10.4 Missing values

Identical conventions have been used to represent the variety of situations where respondents did not provide data in response to questions, or where a variable could not be computed. These conventions distinguish the following:

- 0 represents 'Not Mentioned' or 'None' (unless it has some other meaning in the coding frame). Thus, where respondents are asked which of a list of items apply to them, those not selected will be coded 0.
- 1 represents a respondent response of 'Don't Know' – these include both pre-codes, and interviewer write-in.
- 2 represents a respondent refusal – these include both pre-codes, and interviewer write-in.
- 3, -4, -5 are reserved for situations arising for particular questions where invalid data are given for other reasons, or data that do not fit into the frame of the main variable – the value label will indicate the particular situation.

- 7 is used on individual respondent records, it indicates that the respondent was interviewed by proxy or by telephone and the relevant question was not asked, or the derived variable could not be computed. It is also used on the household record to indicate that the only household contact was a telephone interview, so that the household schedule was not completed.
- 8 represents data missing because not applicable to that respondent, or because of routing from some previous question.
- 9 represents data missing in error, with no other explanation, or derived variables which could not be computed.

These conventions are used consistently across all waves. There is however one change in the implementation at Wave 9, with the introduction of CAPI. At wave 9 there is a distinction between those variables with explicit ‘don’t know’ and ‘refuse’ responses, which are treated as before (-1 or -2), and others, where the value -1 is used for an interviewer entered ‘not answered’ code (a situation that would result in a code -9 at other waves). For wave 9 interview variables, the value -9 is restricted to data missing because of errors in interview administration, or other processing errors. This change was largely reversed at wave 10 as interviewer entered ‘not answered’ codes were removed from most questions.

10.5 Response data

There is a complete accounting for the response status of all active sample members at each wave. One record type (XWAVEID) contains household and individual response status for each sample member at all waves. This permits identification of both response status, and reason for non-response (distinguishing non-contact, refusal at household or individual level, out of scope and deceased). This information is also carried on individual wave record types, including a household sample record, giving response status of all households issued, and also further data about reasons for non-response. Another record type (XWLSTEN) gives sample status, and wave at which this last changed for all sample members.

10.6 Design, field and processing data

All household sample files contain variables distinguishing both sampling strata and PSUs for the original BHPS sample and the Scottish and Welsh extension samples. These variables are set to missing for the Northern Ireland sample as the sample was a single-stage systematic random sample (see section 4.4).

The files contain interviewer identifiers. These permit situations where an interviewer has changed between waves to be identified. These identifiers also permit analysis of the interpenetrating sample experiment in Wave Two. The files contain data about the number of calls made by the interviewer at each address, and about whether a household was reissued for refusal conversion or tracking, and the outcome.

For all variables where imputation was undertaken (see Section 8.2), the files contain ‘flag’ variables indicating which values were imputed.

10.7 Weights

Section 8 has discussed the weighting strategies used in the BHPS. The public release files contain a range of final weights for use in various types of analysis. These files do not however include variables containing weight components, such as design weights or non-response weights. These are available from the ISER.

In a longitudinal survey with multiple waves there is a potential for generating large numbers of weights for samples based on different combinations of waves. The public release files have focused on the most frequently used sets, ie respondents at all waves up to the latest for each wave, and cross-sectional weights for all respondents at each wave. There are separate weights for full respondents and all enumerated individuals. The cross-sectional weights include TSMs and PSMs, and therefore cannot be used for the analysis of cross-sections of OSMs. For this analysis it is necessary to use longitudinal weights, which will exclude those without a valid response at all waves up to the wave of interest.

With the introduction of new samples, a number of additional weights were introduced. There is a system of cross-sectional weights incorporating the ECHP sub-sample, and both cross-sectional and longitudinal weights incorporating the Scottish and Welsh extension samples.

10.8 Data usage

The BHPS data are released through the UK Data Archive. Table 59, based on UK Data Archive records, shows the evolution of usage since the release of the first data set in 1993. In the nature of the BHPS, regular users may take data from the Archive on more than one occasion, as new waves become available. The table therefore shows both the number of orders and the number of users. The total number of users up until (but not including) the release of the wave 12 data is 1,508. We estimate that this is around eighty to one hundred less than the total number of users, if we also include users within ISER (staff, visiting fellows, visiting students), and other research collaborators who have received data directly. The second column shows that there has been a steady flow of new users, recently around 250 per year. The final column, which analyses all users by the last year they took data, suggests that more than half took data within the last three years. BHPS is the most heavily used single data set that the Data Archive distributes.

Another key measure concerns publications based on the BHPS data. This information is much less complete than information on usage, because of the difficulties of tracking down all publications from a very large number of users. The figures in Table 60 reflect intensive efforts to identify publications, but are likely to be incomplete. The table endeavours to avoid double counting (e.g. of working papers that subsequently become journal articles). It shows a steady increase from the late 1990s, as BHPS became established as research resource in wide use.

A classification of users by discipline and level is available, but only for around 80% of users, and only up until 2001 (just before release of the wave 9 data) as shown in Table 61. It suggests a dominance from economics, but with significant numbers from sociology and a number of other disciplines.

11. Coverage Error

11.1 Under-coverage of sampling frame/method

As noted in Section 3.1, the BHPS has a large number of potential target populations and, therefore, corresponding study populations. In fact, after n waves the number of possible combinations of waves is $\sum_{i=1}^n \frac{n!}{i!(n-i)!} = 2^n - 1$. After thirteen waves, this means that there are 8,191 potential study populations. However, many of these share common characteristics, so we need not discuss each one separately.

An important distinction can be made between cross-sectional target populations (of which there are 13, after 13 waves) and longitudinal populations (of which there are 8,178 after ten waves). Furthermore, some of the longitudinal populations are considerably more important than others.

11.1.1 Cross-sectional populations

The natural starting point for a discussion of BHPS coverage is the relationship between the wave 1 sample and the wave 1 cross-sectional population. If the target population is defined as all persons resident in Britain at the time of wave 1 field work (autumn 1991), then sample under-coverage consists of the exclusion of the following groups:

- (1) *persons resident north of the Caledonian Canal;*
- (2) *persons resident in institutions;*
- (3) *persons resident at a private residential address not listed on PAF;*
- (4) *persons resident neither at a private residential address nor in an institution (eg the homeless, travellers, people with no permanent accommodation).*

In percentage terms, these four groups constitute approximately 0.5%, 2.0%, less than 1.0% and less than 0.1% respectively of the target population. It should be noted that persons in institutions have a rather different age and gender profile to the population as a whole. An indication of the distribution of the institutionalised population, by gender and type of institution, appears in Table 3. Almost by definition, the population in institutions are likely to have rather distinct characteristics in other respects too. Nevertheless, for many purposes, these exclusions might not be very important. They can perhaps be viewed simply as a redefinition of the target population. However, membership of these groups is by no means a permanent characteristic – over time, there is mobility between private households and institutions, across the Caledonian Canal, and so on. Consequently, these exclusions take on greater significance in the context of longitudinal target populations (see Section 11.1.2 below) and in the context of cross-sectional target populations subsequent to wave 1.

At wave j , where $j > 1$, the cross-sectional BHPS sample (OSMs, OSM(D)s, PSMs and TSMs) will exclude:

- (5) *persons who are in one of categories 1-4 above at both wave 1 and wave j ;*

- (6) *persons resident outside of Britain at wave 1 and in one of categories 1-4 above at wave j ;*
- (7) *persons in households where all the household members were resident outside of Britain at the time of wave 1;*
- (8) *persons born since wave 1 whose parents are both in one of categories 1-4 or outside of Britain or in category 7 at both wave 1 and wave j .*

Additionally, children one of whose parents are in one of categories 1 to 4 or 7 will have an inclusion probability half that of others. Thus, while they are not excluded, they will be under-represented in the sample.

It should be noted that the situation regarding persons who were resident outside of Britain at wave 1 is particularly complicated. All such persons who subsequently have a child with a partner who was Original Sample Eligible (OSE, ie resident at a private residential PAF address at wave 1) have an equal chance of becoming a PSM. But this is the only means by which they can become a permanent sample member. They can never become an OSM. They can become a wave j TSM if, at wave j , they are in a household with at least one person who was OSE. But their wave j inclusion probability is proportional to the number of wave 1 households containing at least one member of their wave j household. And their overall probability of inclusion at at least one wave up until wave j is proportional to the sum across waves 2 to j of the total number of wave 1 households containing at least one person who has shared their household at that wave. This should be of concern for analyses that treat TSMs as part of the sample, though it should be noted that the weight-share method (see Section 8.1.3) provides unbiased cross-sectional estimates for single waves.

11.1.2 Longitudinal populations

A longitudinal population is defined as the set of persons resident in Britain at each of a number time points corresponding to waves of BHPS data collection. For example, people resident in both autumn 1991 and autumn 2000 would be the appropriate target population for analysis involving data collected at waves 1 and 10, regarding changes over that time period. Thus, some people who are members of the wave 1 cross-sectional population will not be members of this longitudinal population because they have died between 1991 and 2000. Similarly, some members of the wave 10 cross-sectional population will not be members of this longitudinal population because they were born or moved in to Britain between 1991 and 2000.

With respect to any longitudinal population defined by a combination of BHPS waves including wave 1, the groups excluded are the same four listed above in Section 11.1.1. However, the relative importance of the group and the nature of any coverage bias introduced, could differ between target populations. For example, a substantial proportion of full-time students, particularly those in Higher Education, will have been omitted at wave 1 because they resided in halls of residence, which are classed as institutions. Thus, the wave 1 cross-sectional sample under-represents students, and to some extent therefore young adults generally. However, the wave 10 cross-sectional sample should not under-represent students as few, if any, OSMs who are students at wave 10 are likely to have been living in student residences at wave 1. Thus, the

group under-represented at wave 10 is not students and people aged around 18-22 generally, but rather people aged around 27-31 who had been students nine years earlier.

With respect to any longitudinal population defined by a combination of BHPS waves excluding wave 1, the study sample will exclude the following groups:

- (9) *persons who are in one of categories 1-4 above at wave 1;*
- (10) *persons resident outside of Britain at the time of wave 1;*
- (11) *persons born since wave 1 whose parents were both in one of categories 1-4 or outside of Britain at wave 1.*

It should be noted that the relative size of these groups, though all small, will differ between different target populations. In particular, the size of the last of these groups is likely to increase with passing time, due to the cumulative contribution of immigrants. For example, immigrants (since 1991) are likely to constitute a larger proportion of the wave 12 to 13 longitudinal population (people resident in Britain in both 2002 and 2003) than of the wave 2 to 3 longitudinal population (people resident in Britain in both 1992 and 1993).

12. Sampling Error

12.1 Sampling Variance

Sampling variance is an important component of the precision of survey estimates. Sampling variance occurs due to the fact that a sample is taken rather than a census. As sample selection is governed by a chance mechanism, there are many potential samples that could be selected to a given design. The values of estimates will vary over these potential samples and it is this variation that is measured by sampling variance.

Sampling variance is typically summarised by standard errors associated with survey estimates. The standard error is the square root of the variance. However, when estimating standard errors from survey data, other sources of random error (other than sampling) may become conflated. Commonly, this is true of the random component of interviewer error (as interviewers will to a large extent be confounded with primary sampling units). For a more complete introduction to standard error estimation for complex sample designs, the reader is referred to Butcher and Elliot (n.d.).

The BHPS employed a multi-stage stratified sample design at wave 1 (see Section 3.2.1). Stratification is a technique used on most surveys. It is designed to ensure that specified subgroups are represented in their population proportions. In consequence, the precision of estimates will be increased relative to a simple random sample design. Multi-stage designs are used on most face-to-face interview surveys as they enable the sample to be restricted to a limited number of geographical areas, thereby reducing field costs. The first stage of the BHPS design was to select postcode sectors, small areas that contain an average of 2,500 households (see Section 3.2.1). Designs that lead to a clustered sample can inflate the standard error of an estimate if there is clustering of population characteristics of interest. For example, estimates of tenure type have an inflated standard error (or a decrease in precision) when based on a geographically clustered sample compared with estimates based on an equivalently sized simple random sample since housing tenure is geographically clustered. The effect of clustering on estimates of standard errors is dependent on how homogeneous primary sampling units are with respect to the characteristic of interest. The greater the homogeneity, the more the standard error will be increased.

In general, the effect of a complex sample design (ie a design that departs from simple random sampling) on the standard errors of an estimate $f(\underline{x})$ is given by the design factor, which is defined as:

$$deft(f(\underline{x})) = \frac{s.e._c[f(\underline{x})]}{s.e._{srs}[f(\underline{x})]}$$

where $s.e._c[f(\underline{x})]$ is the standard error of $f(\underline{x})$ with the complex design under consideration and $s.e._{srs}[f(\underline{x})]$ is the standard error of $f(\underline{x})$ that would be obtained from a simple random sample of the same size. Note that $f(\underline{x})$ could be, for example, a proportion, a mean, a ratio, a regression

coefficient, or any other sample-based statistic. To simplify the notation, with no loss of generality, we will refer to estimation of a proportion, p .

The BHPS sample design departs from srs in three ways (see Section 3.2): it is stratified, it is clustered (multi-stage) and it involves variable sampling fractions. Each of these departures potentially invokes a design factor and the overall design factor is the product of these three components:

$$deft = deft_{clus} \times deft_{strat} \times deft_{VSF}$$

As described above, $deft_{clus}$ will typically be greater than 1.0, while $deft_{strat}$ will typically be less than 1.0. The only departure from equal sampling fractions on BHPS (for longitudinal samples – see Section 3.4.3 for discussion of sampling fractions in the context of cross-sectional samples) is that due to addresses which contain four or more households (see Section 3.4), so $deft_{VSF}$ will be very close to 1.0 and is therefore unlikely to have a significant impact on standard errors.

With clustered designs, $deft_{clus}$ can be the dominant component of $deft$. We can express $deft_{clus}$ for any given estimate in terms of the homogeneity of primary sampling units with respect to that statistic:

$$deft_{clus} = \sqrt{1 + (\bar{b} - 1)roh}$$

where roh , the rate of homogeneity, approximates the intra-cluster correlation (ρ), a measure of the relative homogeneity of primary sampling units, while \bar{b} denotes the average number of elements per primary sampling unit, ie sampled persons per postcode sector. The more homogeneous the primary sampling units, the larger the complex sampling error.

Confidence intervals for given point estimates can be obtained assuming that a normal approximation is valid, which is generally true for sample sizes above 30. A 95% confidence interval is then bounded by the points defined below where p is the proportion being estimated:

$$(p - 1.96 \times (deft \times s.e._{srs}(p)), p + 1.96 \times (deft \times s.e._{srs}(p)))$$

For BHPS wave 1, the package SUDAAN was used to calculate estimates of standard errors and design factors. For a range of variables in the BHPS, the complex standard errors, DEFTs and 95% Confidence Intervals were calculated in order to give a general overview of the effect of the sample design on the precision of survey estimates. These are presented in Table 62. Other surveys that use a similar design (households clustered within postcode sectors) and measure similar or related variables can also provide a guide to the likely magnitude of BHPS design effects. Examples include the General Household Survey (Breeze, 1990), the Labour Force Survey, the Expenditure and Food Survey and the British Social Attitudes Survey.

The majority of DEFTs in Table 62 lie between 1.0 and 1.3, indicating that there is a relatively small effect of the sample design on *srs* based tests. For example, the DEFT for the proportion of respondents who had an in-patient stay since 1.9.90 is 1.05, so the complex standard error is only 5% larger than the *srs* standard error. However, for variables that are in some way clustered within postcode sectors and households, the effect on standard errors can be large. For example, the DEFT for local authority tenure is 1.57, reflecting the high degree of clustering of housing tenure. It should be noted that, even with attitudinal variables such as agreement to the statement "All Health Care Should Be Free", the effect of clustering can be marked. In this case, the DEFT is 1.63, indicating that the complex standard error is almost two thirds larger than the *srs* standard error. This is rather larger than the design effects for attitudinal variables on other surveys such as the British Social Attitudes Survey (BSAS) series. The larger design effects on BHPS are almost certainly due to the extra level of clustering within households. BSAS selects just one person per household for interview. (This extra level of clustering is of course also one of the strengths of the BHPS as it allows the study of intra-household relationships. Within-household correlation in attitudes and behaviours is not simply a statistical nuisance, it is also of intrinsic interest to researchers.)

As some design effects can be considerable, care must be taken in the analysis of variables that may have a strong association with area and/or household. The use of *srs* techniques for the analysis of such variables can lead to misleading results due to the inappropriate estimation of precision. For a fuller discussion of the analysis of complex surveys and the problems associated with using simple random sample assumptions as the basis for modelling and testing procedures, the reader is referred to the introductory text by Lee *et al* (1989) and the more comprehensive volume by Skinner *et al* (1989)

The estimates of design effects presented here relate to cross-sectional wave 1 estimates. However, it should be noted that these design effects also provide good approximations to the design effects associated with longitudinal estimates involving wave 1 (eg change in status between wave 1 and wave *j*). Design effects for longitudinal estimates should differ in only two ways. First, due to sample attrition, the sample size, and hence \bar{b} , will be smaller. This will tend to reduce the design effects. Second, due to mobility the geographical locations of the members of each primary sampling unit will become more dispersed. This may have a tendency to reduce ρ_{oh} , which will again reduce design effects. It is therefore likely that cross-sectional design effect estimates, such as those in Table 62, will tend to provide upper bounds for longitudinal design effects. It must be recognised, however, that some longitudinal estimates are inherently different in nature to cross-sectional estimates and it is therefore difficult to know which cross-sectional estimates to use as an approximation. For this reason, we intend to calculate and publish standard errors for a range of longitudinal estimates in the near future.

12.2 Sampling bias

It is important to draw a distinction between coverage bias and sampling bias. Coverage bias is the result of some population units having a zero probability of selection. This is discussed in Section 11 above. Sampling bias is the result of failure to correct for unequal selection probabilities amongst population units with non-zero selection probabilities. In practice, this tends to arise when the selection probabilities are not controlled or not recorded.

In the case of the BHPS, there are two potential sources of sampling bias. The first concerns the inclusion of TSMs (see Sections 3.2.2 and 3.4.3; the second concerns the inclusion of OSM(D)s (see Section 3.4.2).

12.2.1 Bias arising from inclusion of TSMs

Each population member has a probability of inclusion as a TSM at wave j that is proportional to the number of households in the population at wave 1 that contained one or more persons (Original Sample Eligibles - OSEs) with whom they share a household at wave j (and did not share at wave 1). The overall probability of any given person being included as a TSM on at least one occasion between waves 2 and j will be proportional to the total number of households in the population at wave 1 that contained one or more persons with whom they have shared a household on at least one survey occasion from 2 to j . It will be noted that inclusion probabilities for the same individual can be different for different cross-sectional and longitudinal populations. In any case, it is not feasible to obtain the information that would be necessary to establish these probabilities.

Instead, the “fair shares” weighting method (Ernst 1989, Lavallée & Hunter 1992, Rendtel 1991) is used to produce approximate (cross-sectional) design weights for TSMs. Essentially, this involves summing the design weights for each OSM in a household and then dividing this total by the number of OSEs in the household. The resulting weight is applied to all current household members (OSMs, OSM(D)s and TSMs). Note that most (but not all) TSMs will be OSE, but no OSM(D)s will be OSE, so in general the presence of TSMs does not inflate the representation of the household in the sample of persons, whereas the presence of OSM(D)s does. In practice, non-response must also be taken into account, so the weight calculated for OSMs incorporates both the design and non-response components.

This method relies on two assumptions that, if incorrect, could lead to bias. The first is that important characteristics of TSMs are shared by the OSMs with whom they share a household. In the sense that both these groups are people who share a household with persons with whom they did not share a household in 1991, this is true. The second assumption is that non-response predictors are shared between persons within households that contain persons other than OSMs. When the initial weights are calculated for OSMs, they may be different for different OSMs within a household (due to the non-response component being different – design weights cannot differ within a household), but this distinction is lost when the fair shares procedure is applied. Technically, any bias introduced by this is not really a sampling bias. Rather, it is a component of non-response bias that would have been removed by the original OSM weights.

TSMs are not included in longitudinal weighting procedures (their weights are zero), so this issue only affects cross-sectional analysis using the cross-sectional weights. Any resultant bias is likely to be small, and has in practice been assumed to be negligible, but this has not been tested empirically.

12.2.2 Bias arising from inclusion of OSM(D)s

The selection probabilities of OSM(D)s are described in Section 3.4.2. Potential bias arises if the weighting procedures fail to adequately adjust for variation in selection probabilities. The weighting procedure adopted on BHPS for OSM(D)s is to establish the weight for each of the parents present in the same household as the OSM(D) at the wave in which the birth is identified (typically the first wave after birth, but not always – we shall refer to this as wave t) and to assign the OSM(D) the mean of those weights. Thus the weight for an OSM(D) will depend on how many parents are living with them at wave t and whether that parent/ those parents are themselves OSMs, as follows:

Situation of OSM(D) at wave t	Weight applied to OSM(D)	
1. Two parents present, both OSM	Mean of weights for parents	1.0
2. Two parents present, one OSM	Half the weight of the OSM parent	0.5
3. One parent present, OSM	Weight of the parent	1.0
4. One parent present, not OSM (ie PSM)	Zero	0.0

If we assume for simplicity that all OSMs have a weight of 1 (ie we ignore the slight variation in household selection probabilities described in Section 3.4.1 and also ignore non-response weights), then the weights applied to OSM(D)s can take the values 0, 0.5 or 1, as shown above. However, for unbiased estimation the weights should be inversely proportional to the selection probabilities described in Section 3.4.2. The actual weights are shown alongside the ideal weights in Table 63. There are 8 possible scenarios, corresponding to the cross-classification of the 3 possible selection probability scenarios and the 4 possible weighting scenarios, only 8 of the 12 combinations being possible.

It can be seen the actual weights are not equal to the ideal weights for 4 out of the 8 categories. However, the following observations are relevant:

Where both parents are OSE, but in 2 separate households, and only one parent is present at wave t, the actual weight is 1.0 if the OSM parent is present and 0.0 if the other parent is present. The design weight should be 0.5 in both cases. In practice, the OSM(D) is unlikely to be enumerated if the OSM is not present (ie response rate will be low), so the actual weights may well reflect quite closely the actual inclusion probabilities in such cases.

The other two situations in which the weights are discrepant are situations where only one parent was OSE. If both parents are present at wave t, the OSM(D) will get a weight of 0.5 instead of 1.0; if the OSE parent is absent, the OSM(D) will get a weight of 0.0 instead of 1.0. The practical significance of this is likely to be that children with one immigrant parent (or one parent who was in an institution or north of the Caledonian Canal at wave 1) will be under-represented in weighted BHPS samples.

13. Non-Response Error

13.1 Wave 1 non-response

Table 64 shows household level response at wave 1 by region. Response here includes both complete and incomplete response at the individual level (individual level response is discussed later in this sub-section and detailed in Table 66). The two parts of London stand out as having particularly low response. This pattern is common across a wide range of surveys. In Inner London it is non-contact which is most strikingly high, whilst in Outer London refusal levels are particularly high. Outside London, there is not a straightforward distinction between large urban areas and others. Two conurbations, the West Midlands and Greater Manchester also experience below average response, but other conurbations, such as South Yorkshire and Tyne and Wear experience relatively high response. Response in the South East outside London is also marginally below average.

There is clearly further variation below regional level, but regression analysis of wave 1 household response using Census small area statistics does not suggest major systematic patterns, after taking account of regional variation. Factors such as area deprivation indicators, or their constituent components such as the unemployment rate, or housing tenure, do not provide any additional explanation of non-response. There are however a small number of area characteristics with a significant additional effect. Having a high share black Caribbean or black African population is associated with higher non-response. An increase in the share by one standard deviation lowers the response rate by around 2% in an average region. Housing characteristics were also associated with non-response. Thus a one standard deviation increase in the share of purpose built flats lowered by response by around 1.5%. There was a slightly smaller effect in the same direction for converted flats. Partly related to this, a one standard deviation increase in the share of households made up of one person not of pensionable age lowered response by the same amount.

Table 65 shows household level response at wave 1 by housing type (observed by interviewers). Non-contact rates were higher at flats than at houses, and particularly at converted flats, though these account for less than 4% of the sample of households. Refusal rates varied little over different types of housing, but were slightly lower at flats than at houses. Overall, household response rates were slightly lower at flats than at houses, but there was no difference between types of flats and little difference between types of houses.

Response to the individual interview, conditional upon response to the household interview, is summarised in Table 66. It can be seen that a 95.5% conditional response rate was obtained to the individual interview, consisting of 9,912 personal interviews and 352 proxy interviews. The most common reason for non-response to the individual interview was a personal refusal (4.0%).

13.2 Attrition

Section 6.3 presented basic data on field outcomes and response rates. While this gives some indication of the effectiveness of the survey operation it is not directly helpful in identifying how far non-response error may contribute bias to analysis, or reduce the effective sample size

available for analysis. This is because it is the cumulative effects of response and non-response at a number of waves that influence the sample available for longitudinal analysis, or indeed cross-sectional analysis after wave one. Moreover, in a multi-purpose survey such as BHPS different researchers will be using different subsets of the data subject to different attrition processes. As indicated in Section 11.1, the number of possible subsets, defined by possible combinations of waves, rises very rapidly with the number of waves, being $2^j - 1$ at wave j , so 8,191 by wave 13. This section must focus on a small number of summary measures.

There are also a number of different ways of defining the responding sample, depending on the substantive analysis purpose. Some analyses may require a full response to all instruments, for other purposes proxy responses can also be included, and for yet other purposes all persons enumerated in respondent households can be included. For example labour market behaviour needs to be concerned with full respondents, while the analysis of income mobility may be more concerned with all household members. Much of this section in fact focuses on full respondents. Table 67 contains a number of indicators of outcomes for wave one full respondents. The first column shows the percentage who are not eligible for interview at each wave. This includes those who have died since wave 1 as well as those who have moved abroad. The latter of course are still part of the BHPS sample and may subsequently return and be eligible for interview. The second column shows the proportion of those eligible responding at each wave. Thus 65% of those wave one respondents who were still alive and in Great Britain in 2000 gave an interview at wave thirteen. The next column shows the rather smaller percentage who gave an interview at *all* waves up to the latest wave (55% at wave thirteen), and the final column shows the wave response rate for this core sample.

Table 68 shows equivalent information when the responding sample is defined by being enumerated in a respondent household, including children, and the starting sample is all persons enumerated at wave 1. Columns are defined in a similar way to Table 67.

Analysis of the impacts of attrition can be undertaken in various ways, including comparisons with other external data sources that can be assumed not to be subject to equivalent non-response processes. An alternative is to compare the wave 1 characteristics of those who subsequently drop out with those who remain. Here we adopt the latter approach for a number of reasons:

- Differences can only be due to non-response whereas, with an external comparison, sampling error, coverage error and measurement error can all contribute to observed differences;
- As the same (wave 1) data is used for both elements of the comparison, there is no concern about differences in definitions, data collection methods, or reference points/periods between two data sources;
- A much richer data set is available for the comparison – all data collected at wave 1. (Limited data are available from sources that are not subject to non-response bias – ie non-survey sources.)

For any terminal wave, we can divide the sample into three groups: a) those who respond at all waves, b) those who are non-respondents for at least one wave but still in the sample and c) those who have become ineligible since they entered the sample. Table 69 to Table 71 compare the

distributions of groups a), b) and a) + b) with the original wave 1 sample, for the period up to wave 5 and the period up to wave 13. The contrast distinguishes the effects of rather higher early wave attrition, from that of the lower rates experienced subsequently. Data are weighted by wave 1 cross-sectional respondent weights. In general the most important issue is whether there are substantial differences between the 'Total at wave 1' and the 'respond all waves' distributions. Differences are not large in magnitude, though there are some differences between those who remain respondents and those who drop out. In some cases, these differences are due mainly to those who have died, in other cases they are mainly due to those who have failed to respond on at least one occasion. The latter scenario is of greater concern, as the former does not introduce non-response bias. In some cases, both groups contribute to a difference.

The apparent under-representation of the following groups was mainly due to sample members leaving the eligible population due to death:

- Age 65+
- Retired
- In poor health
- In single-person households
- Single pensioner household

Those who drop out (fail to respond on at least one occasion) include disproportionate numbers of people who had the following characteristics at wave 1:

- Age 16-24,
- Never married
- Unemployed
- No qualifications
- Not active in any organisations
- Resident in Inner London, West Midlands conurbation, Merseyside
- Local authority or housing association tenant
- In the bottom 40% of the income distribution

However, it should be noted that although under-representation of these groups is statistically significant, the actual magnitude of under-representation is generally small. Furthermore, these differences apparent at the data collection stage are largely removed by the application of the weighting (see Section 8.1).

14. Measurement Error

Measurement error is present in all survey data. One of the advantages of a longitudinal survey design is that it provides opportunities to assess the extent of measurement error and the potential effects on substantive analysis of error or 'noise' in the data over time. One source of error is recall error with one of the benefits of an annual panel survey being that the reference period for recalling events is limited to a 12 - 18 month period. Recalling events within this limited time frame is less problematic than recalling events over the life-time. The BHPS, in the early waves of data collection, included various life-time history and 'initial conditions' questions. The recall error on these retrospective data has been assessed, particularly in relation to life-time job histories.

14.1 Unemployment

Elias (1997) reviewed data from the BHPS and the Family and Working Lives Survey and compares these with the Labour Force Survey. He found that recall of periods of unemployment were problematic and that certain groups, women and older workers generally, suffered from worse recall than other respondents. Similar results are found on other surveys collecting retrospective histories. For example, recall problems were found during the design phase of the Social and Economic Life Survey (Elias, 1991) where a significant under-reporting of unemployment occurred. Elias speculates on the causes of such under reporting in social-psychological terms where the experience of unemployment is unpleasant and therefore undesirable knowledge. Periods of unemployment might also be considered as 'non-events' or time where nothing memorable happened, or telescoped resulting in distant events being recalled as though they were more recent than in reality. Significant events may also be remembered with greater clarity than lesser events, relegating non-events to obscurity.

Whichever is the case, unemployment reporting appears more accurate as the respondent approaches the present. However, Elias claims a direct comparability between the BHPS, LFS and the FWL surveys in the calculation of unemployment rates to suggest that a pattern of underreporting of unemployment exist within each survey due to the over reporting of employment. This is summarised within the table reproduced as Figure 2 (from Elias, 1997).

Dex and McCulloch (1997) use BHPS and FWLS (Family and Working Lives Survey) data and find that the retrospective unemployment histories are reasonably reliable within these two surveys. However, they also find that men's histories are more reliable than women's while balancing this against the assumption that the concept of unemployment can be more problematic for women.

Generally, recall data older than two or three years are less accurate while recent data is found considerably more so. The accuracy of spells of unemployment relates to their length, or the precision with which recall is required. Dex and McCulloch cite Mathiowetz and Duncan (1988) who suggest that when asked for dates of one week lay-off periods some 75% of periods were not reported in interview one year later. It appears that it is not that respondents cannot remember that they were unemployed but that they are not sure when and for how long they were unemployed.

Dex and McCulloch suggest that error which may appear in the data as recall error may in fact be due to difficulties respondents have in defining the distinctions between unemployment or inactivity and employment. The Current Population Survey (CPS) in the USA for example has found as many as 10% of the “truly unemployed” were classified as not in the labour force one week or so later. Conversely, a further 3% of the “truly unemployed” were likely to be classified as employed at re-interview. Women are also more likely than men to be classified as not in the labour force when they are actually unemployed, raising concerns about the categories used to collect data regarding women’s employment histories.

14.2 Gross flows

While recall error affects retrospective data particularly beyond a three to five year time horizon, respondents may also vary in the way they report particular aspects of their circumstances year on year resulting in random ‘noise’ in the data. Panel data provide the data needed to measure gross change and the flows into and out of particular states or circumstances. One of the difficulties in measuring gross change in a panel is the potential for measurement errors when using repeated measures for the same individuals. Kalton *et al* (1986) describe measurement error in the Survey of Income and Programme Participation (SIPP) in relation to basic items which would not be expected to change, such as race and sex, which appear to change in inexplicable ways between rounds of the survey. Approximately 0.1% of the sample appeared to change race on the SIPP between each wave of interviewing and 0.1% appeared to change sex. These non-sampling errors were attributed to interviewer error, keying or processing error or a combination of these. Similar apparent changes can be observed in the BHPS data.

Other areas are less easy to attribute to interviewer or processing error. The consistency of occupation and industry codes between waves, where the description of the occupation is collected and coded independently at each wave, provides a test-retest reliability check of the data. In the SIPP, of respondents who had not reported a job change, 37.4% were coded differently at the three digit level between their first and second interview and 19.8% changed three digit industry codes. Overall, only 39.9% of respondents interviewed three times over a 12 month had the same three digit occupation and industry codes, even though they were with the same employer at each interview. Similar findings have been made on UK surveys such as the panel element of the Labour Force Survey.

On the BHPS, occupation and industry were coded post-fieldwork from the verbatim responses recorded by the interviewer. A comparison of the occupation and industry coding on the BHPS waves 1 and 2 for respondents reporting no job change in the intervening year shows that the marginal distributions at each year showed no significant differences. Looking at individual responses, it was found that 32.2% had a different SOC code at the three digit level and in 17.9% of cases this resulted in a change at the major group level. For the industry coding, 28.9% had a different code at the four digit level and for 11.4% this resulted in a change at the major division level.

On examination of a sample of the verbatim responses, response variability on the part of the respondent was found to be a major contributory factor to the coding differences. Possible explanations for the differences included:

- Recall errors in the start date of the current job
- Respondents failure to recognise a within employer job change eg a promotion
- Wave 1 occupations were coded manually. At Wave 2 a computer assisted method (CASOC) was used
- Questionnaire changes - the wave 2 questionnaire did not ask for the type of organisation worked for which made coding difficult in some cases eg nurses working in either the NHS or private sector
- Simple response variability on the part of the respondent
- Different coders at each wave
- Change of interviewer at wave 2
- Coding error at either wave
- Reported difference in managerial status
- Keying errors

Table 72 and Table 73 show the cross-wave comparisons of SOC and SIC for waves 1 and 2. The cross-wave reliability of coding of occupation and industry where other data indicate that the respondent is in the same job as the previous year, is subject to a level of error or difference that can be largely explained by a combination of the various factors listed above. Many of these are difficult to control in an interview setting, even though the use dependent interviewing techniques that becomes possible with CAPI technology may provide one avenue to lessen spurious differences over time in a longitudinal survey. In terms of using BHPS data, analysts should take care to determine whether a real change in occupation has occurred or not. It may be that a respondent's job with their current employer has changed in some significant way due to a promotion, change in hours, managerial responsibility, pay or conditions. Despite this, the start date of that new job may be incorrectly reported by the respondent as being the date the respondent started working for that employer rather than the date they were promoted or changed their job in some other way. Care should therefore be taken to look at a range of variables, not just the occupational code or start date of the current job, when deciding whether a job change has occurred.

It can be seen that coder reliability is high for these BHPS data items. Reliability for SOC is very similar to that typically observed for multiple coding of a single description of a job (Bushnell, 2000). Despite this, analysts should be aware of the implications for analysis of reliability coefficients of less than one (Sarıs and Münnich 1995). (Note that within-item coder reliability was discussed in section 7.4 and some estimates of reliability summarised in Table 57).

Further information about measurement error in BHPS estimates of gross flows comes from the ISMIE study (<http://www.iser.essex.ac.uk/ulsc/methods/research/ismie/>). This study provided evidence of the extent of over-estimation of change in measures of occupation and industry (Sala and Lynn, 2004) and of labour market status (Jäckle and Lynn, 2004).

14.3 Income sources

Another area in which analysts may need to be aware of measurement error in the BHPS survey data concerns measures of the receipt of income from specific sources. BHPS collects information on each of a number of sources from which respondents may have received income during the period since the previous wave, including various state benefits, pensions and other sources such as rent from property. For each source, the respondent is asked to report in which months during the period they received some income, with follow-up questions regarding the amount(s) of income received from each source. These data are primarily used to derive measures of total personal and total household annual income. For these purposes, the data may well be adequate.

But some analysts may be interested in studying the stability of income from particular sources, e.g. flows on and off particular state benefits. For such analyses, measurement error may be important. It is likely that respondents sometimes omit to mention a particular income source. The sources are listed on a series of cards, from which respondents must identify sources relevant to them. With this type of questioning method, it is quite possible that a particular source might sometimes be overlooked, especially if it not greatly significant to the respondent. This can have an impact on measures of change. Suppose, for example, that a particular respondent has received a particular benefit continuously for three years, but that she reports it in the survey interview only in the first and third years and not in the intervening year. In each of the two interviews where she reports the income, she is likely to report having received it in each of the previous twelve months. The result is that the analyst will observe twelve consecutive months of receipt, followed by twelve consecutive months of non-receipt, followed by another twelve consecutive months of receipt. In other words, it appears that this respondent has made two transitions during the three-year period, one off the benefit and one back on to it. But in fact both are spurious. In the presence of this kind of “occasional omission”, rates of transition will be over-estimated, with likely knock-on effects for related analyses.

It should also be noted that a similar effect can arise from incorrect identification of income sources. In this case the respondent may report the income at every interview, but may classify it as a different source on each occasion. This may be particularly likely to occur with, for example, some of the disability-related state benefits: there are several with similar names, and both the names and eligibility criteria have changed since BHPS began.

The ISMIE study sheds some light on the extent of measurement error in BHPS income receipt variables as the sample was matched with administrative records of the Department for Work and Pensions (DWP) so as to provide validation data. For the most commonly-received income sources represented in the DWP data (i.e. those for which the sample sizes of recipients were large enough to provide usefully precise estimates), the *false negative rate* was estimated. The false negative rate is defined as proportion of true recipients (according to the DWP data) who did not report receipt in response to the survey questions. With the standard BHPS questions, this rate varied between 0.00 and 0.50 for the six income sources for which estimates could be made (first column of Table 74). (Though note that the ISMIE study was based not on the main BHPS sample, but on the ECHP “low income” sample that is described above in section 4.1. Consequently the extent of under-reporting on the BHPS may differ from that found in ISMIE.)

14.4 Dependent interviewing

As well as providing estimates of the extent and nature of measurement error, the ISMIE study also concluded that *dependent interviewing* should be able to reduce the extent of measurement error of this kind (Jäckle and Lynn, 2004; Lynn et al, 2004a; Lynn et al, 2004b; Sala and Lynn, 2004). Essentially, this involves asking sample members explicitly about change since the time of the previous interview, either instead of or as well as asking the traditional independent (repeat) questions. For example, Table 74 shows that under-reporting of income receipt can be reduced by either of two forms of dependent interviewing (the forms of dependent interviewing, and the exact questions used, are described in Lynn et al, 2004a). The potential of dependent interviewing to reduce the over-estimation of change in gross flow analysis is demonstrated in Jäckle and Lynn (2004) for labour market transitions and Sala and Lynn (2004) for occupational transitions.

Dependent versions of certain BHPS questions are being developed and tested with a view to implementing dependent interviewing on the BHPS beginning in wave 16. This will affect the current employment, job history and income sources sections of the questionnaire.

15. Confidentiality and Ethics

15.1 Ethical considerations

A household panel survey is subject to the same ethical requirements as any large-scale survey. Primarily these involve ensuring that the confidentiality of respondents' personal details is maintained and that individual respondents cannot be identified in publicly released data. Legislation in the form of the UK Data Protection Act 2000 governs how personal data are handled and the BHPS is required to conform to this legislation. Beyond these legal requirements there are a number of additional ethical considerations to be taken into account when conducting a household panel which will impact on survey and data collection procedures. In some respects there is an inherent tension between the needs of a panel design and the ethics underpinning survey practice. The survey design demands a low attrition rate to maintain the viability and data quality of the panel with the ethics underpinning the implementation of survey procedures having a direct effect on the quality of the survey.

15.2 Informed consent

Some of the key ethical issues are concerned with gaining informed consent and what that should comprise in the context of a panel survey. This is not only informed consent from the respondent taking part but also how one can collect contact details of people not in the survey for tracking and tracing purposes. At Wave 1 of the BHPS extensive consultation and discussion took place to establish the appropriate means of gaining informed consent from respondents when recruiting the panel. Three main options emerged from these discussions:

- (i) Tell respondents that they were being recruited into a panel survey for either an indefinite period or for a fixed number of years.
- (ii) Ask respondents to co-operate at Wave 1 of the survey and make no mention at all of the panel design.
- (iii) Ask respondents to take part at Wave 1, explain the nature of the longitudinal design and tell them that we would like to interview them again the following year.

Option 1 was judged to be problematic in several respects. The funding for the panel, while being secure for the first five years, was not certain beyond that point. The difficulty was arriving at a fixed number of years that we could honestly say the survey would last. As it turns out, even saying that the survey would run for ten years would have been misleading as the panel is now heading into the twelfth year of data collection. It was also considered to be unrealistic and unfair to expect respondents to commit themselves to a panel for an indefinite, unspecified period. The final difficulty with this first option is that if we had specified a period for the panel and the respondent had agreed to take part for that period, it is unclear what the position would be at the end of that period. We would either be obliged to stop interviewing people at that point or gain their consent for an additional period and there was concern about the potential effects on attrition if this strategy were adopted.

Option 2 was seen as being clearly misleading and on the fringes of what would be ethically acceptable. Certainly some argued that we were only asking for co-operation at Wave 1 and

separate co-operation at later waves could either be gained or refused at that point. However, to give no information at all about the longitudinal nature of the survey was judged to be misleading.

Option 3 is the strategy adopted on the BHPS. In many respects this is something of a compromise position between Options 1 and 2. Respondents were asked for their consent to take part at Wave 1. They were not asked to commit themselves for a number of years into the future but were told that we would like to come back to re-interview them the following year. The longitudinal nature of the survey was also described in the purpose leaflet sent in advance of fieldwork. One crucial decision was not to ask respondents for their explicit permission to call on them the following year. Asking for consent one year in advance is a strictly hypothetical question and we know that between ten and fifteen per cent of respondents will refuse this permission. Having had a refusal of this type, ethically it would be very difficult to re-contact these respondents and as a consequence would lead to higher rates of attrition. As a result, respondents were informed we would be coming back the following year but leaving the issue of gaining their consent to the following year when the interviewer made their call. In this way the respondent would be free to either co-operate or refuse at that point.

15.3 Collecting contact details

The BHPS asks all those interviewed to provide a contact name in case they move in the coming year and we are unable to find them. In this case, it is not possible for us to speak to the named contact person to ask for their consent to hold their details. We therefore ask the respondent to tell the contact person so that they are aware we have their details and that they will only be used in the event we cannot find our respondent. On the BHPS we regard this tracking data as an attribute of the individual respondent which is freely provided by them. It is not held in any way which would allow the details of the contact person to be retrieved other than via the indexing of our sample member.

15.4 Confidentiality

As noted earlier, the BHPS is obliged to conform to UK legislation regarding the handling and use of personal data. ISER has systems and procedures in place to ensure that any personal information such as names and addresses is held in strictly confidential conditions. ISER has a set of procedures set out in our Code of Ethics (see Annex A) which forms part of our contract with our survey fieldwork agency. In addition, members of ISER staff handling personal data sign an ethical undertaking to ensure these data are treated correctly. The design of the panel where all household members are interviewed does raise some additional issues as the confidentiality of respondents must be respected within the responding household. Interviewers must not, even unwittingly, tell other members of a household what a given respondent has said. Partners, spouses and parents in particular may not always recognise this and feel they have a right to know what their partner or child has answered. In practice, preserving confidentiality within the household is often difficult especially where other members are present in the room where the interview is taking place. This in turn raises methodological questions about potential contamination effects and how responses to particular questions might vary depending on who is present at the time.

15.5 Feedback to respondents

An aspect of the ethics of social research that can be problematic in any context is the imbalance of information and control of the survey process between respondents and those collecting the data. Respondents provide detailed information about their lives and the control and use of those data effectively passes to the data collector. While this imbalance is an inevitable outcome of the survey process, it is arguable that there is an ethical responsibility incumbent on the survey organisation to provide information to the respondent about how the data are being used and by whom. In a panel where the relationship with the respondent lasts over a period of years, this becomes more important as the aim is to build a relationship of trust and demonstrate the value of continuing within the panel to sample members. On the BHPS, respondents are kept informed via an annual respondent report mailing which gives some key findings from the survey and details of how the data are being used. A named individual contact person at ISER with a Freephone telephone number and Freepost address are included on all correspondence with respondents so that they have a means of contacting us directly if required.

15.6 Refusal conversion

The BHPS has a standard refusal conversion procedure which is used when sample members refuse to take part. This is the area for the panel which presents some of the greatest ethical difficulties and where the tension between survey needs and ethical boundaries becomes most apparent. Participation in the BHPS is voluntary and the right to refuse to take part exists for every sample member. On the other hand, the panel design aims to minimise attrition as far as possible, so there is a reluctance to lose any sample member through a refusal. On the BHPS all refusals are individually reviewed to assess whether there is any chance of a future interview. If it is judged that we might gain an interview, either during the current fieldwork period or at the following year of the survey, respondents are re-contacted. The difficulty with this process is judging when a refusal is a definite and final refusal and when it is a refusal due to a particular situation at the time the interviewer calls. There is evidence that many refusals are situational or temporary in some way. For example, the temporary illness of a family member, a bereavement, a recent house move, a recent separation or divorce, or a particularly busy period at work might all lead people to refuse at a given year. The following year, or even later during the same fieldwork period, the situation which led to the refusal will have disappeared and they will be happy to take part. The main ethical question to which there is no definite answer, is how many times can you re-contact someone when they have refused to take part any further in a panel survey? In practice on the BHPS, households are normally withdrawn from the sample following two consecutive whole household refusals, even though a small number may be re-issued beyond this.

Some discussion of the effectiveness of the BHPS refusal conversion procedures and analysis of the duration of co-operation following successful conversion, appears in Burton et al (2006).

16. External Comparisons

16.1 Socio-demographic comparisons

External comparisons of survey estimates from the BHPS against other sources mainly focus on cross-sectional estimates. This follows from the absence of external longitudinal estimates collected on a sufficiently similar basis, that the source of differences could be identified. An example illustrates some of the issues in longitudinal comparisons. Buck (2000) found a higher estimate of the rate of residential mobility in Great Britain in the BHPS than in the 1991 Population Census. The Census undoubtedly suffers fewer biases in its representation of the GB population than the BHPS. However the Census measure is based on a retrospective question rather than direct recording at two points in time, and it may well underestimate mobility. There are opportunities for further work on longitudinal estimates, for example based on the panel element of the Labour Force Survey, the original UK ECHP and the ONS Longitudinal Study once the 2001 Census is incorporated. However, none of these are ideal, for a variety of reasons.

In terms of cross-sectional estimates, there have been a number of overviews of sample representativeness based on comparison of standard social and demographic indicators. Taylor (1994) compared BHPS wave one data with the 1991 Population Census. He concluded that the unweighted sample showed an under-representation of households in rented accommodation, of large households and of households without access to a car. There was also some over-representation of children and young adults, and an over-representation of older adults. Post-stratification successfully adjusted for these differences. There was also some tendency for BHPS to produce higher estimates of part-time employees and lower estimates of full-time employees, but this may have been a consequence of the use of different questions.

Table 75 to Table 80 update this analysis somewhat, and also show a number of comparisons between the BHPS and the General Household Survey as well as between BHPS 2001 and Census 2001. These attempt as far as possible to use comparable definitions, though this is particularly difficult for ethnic group in 2001. Data are unweighted.

The comparisons with Census data relate to age structure, employment status, socio-economic group and ethnicity. There is no direct possibility of replicating a Census estimate for 1994, but the sequence of BHPS estimates at least indicates whether there is any sign that attrition may be compromising the representativeness of smaller sub-groups. Some of the changes undoubtedly reflect real changes in the population.

Table 75, on age structure, reflects the under-representation of older people referred to above. On the whole the sample evolution over the first four waves does not exacerbate differences from the Census, though the share in the 45-54 age group does increase steadily. This may to some extent reflect real population change. Table 76 suggests that BHPS may have under-represented non-employed persons, especially females, at wave 1. The proportions in many of the non-employed categories had increased by wave 4, but again we cannot tell from these data to what extent this is due to differential non-response and to what extent it may be due to real change in the population. In fact, Table 69 suggests that, if anything, persons in employment at wave 1 may be less likely than others to drop out of the panel. This would suggest that the

increases between waves 1 and 4 in the proportions employed (Table 76) may be caused by real change in the population.

Table 77 and Table 78 suggests that, prior to weighting, BHPS may slightly under-represent semi-skilled and unskilled manual workers and persons of Asian origin.

Table 79 and Table 80 based on a comparison with the GHS, in addition to contributing to an assessment of the representativeness of the initial sample, also allow us to assess whether changes found in repeated cross-sections are also found in the BHPS.

Compared with GHS, fewer BHPS respondents classify themselves as married or widowed, and greater proportions classify themselves as single or co-habiting. However, changes between 1991 and 1994 – a reduction in the proportion married and increases in the proportions co-habiting or divorced – are very similar across the two surveys. In terms of the ownership of consumer durables and cars, housing tenure and household type, both levels and net changes between 1991 and 1994 are generally very similar across the two surveys (Table 80).

16.2 Other specific comparisons

Income has been the area where work on comparisons has been most systematic. The interest here is in the combined impact of unit non-response, item non-response, systematic measurement error, and questionnaire differences in the construction of household income in particular. Jarvis and Jenkins (1995) compared estimates of net household income from Wave One of BHPS with those used by the Department of Social Security for their 'Households Below Average Income' (HBAI) estimates based on the Family Expenditure Survey. They found some tendency to underestimate incomes, in part because BHPS information on investment income is rather incomplete. However a comparison of vingtile ratios suggested that the shapes of the distributions were rather similar, except at the top end. Their analysis also showed that the distribution between main sub-groups of interest to income analysts was very similar to FES based estimates. The composition of the poor population in BHPS tends to over-represent single elderly people and lone parent families, and under-represent couples in poverty compared with HBAI. Overall they concluded: 'In sum, we find that there are some systematic differences between HBAI and BHPS estimates of the overall shape of the income distribution, the composition of the poorest income groups and compositional change. On balance, however, we believe the BHPS distributions are sufficiently close to justify using them to provide a longitudinal component to the HBAI.'

More recently Rigg (2001) has updated these comparison to cover the series of BHPS estimates from 1991 to 1998 with equivalent year HBAI estimates. He found that, although there were some persistent differences at the sub group level, the BHPS continued to reflect the shape of the distribution revealed by the HBAI, and to track broad changes in mean and median income levels. Estimates of poverty rates remained comparable. For example HBAI found a poverty rate (defined as the proportion below 60% of median income) of 18.0% in 1998, while the BHPS poverty rate was 18.7%. BHPS data tended to produce somewhat lower measures of inequality. The BHPS Gini coefficient in 1998 was 31.7, compared with 34.5 for HBAI. This reflects in part the absence of an adjustment for high incomes in the BHPS.

There has been a range of other comparisons of other estimates. For example, Pevalin (2000) investigated whether or not the repeated application of the 12-item General Health Questionnaire (GHQ) results in any discernible retest effects, by comparing it with data from the Health Surveys for England, a series of large cross-sectional studies conducted over the same years. The GHQ has been administered in the BHPS every year since 1991. Other studies have found that the GHQ is liable to retest effects when administered multiple times over a short period, but it is uncertain if a longer time period between applications has similar results. Overall, the results indicate that no retest effects are present in the BHPS data and that the 12-item GHQ is a suitable measure of mental health for use in population based studies with relatively long time periods between applications.

17. References

Banks, R. and Laurie, H. (2000) From PAPI to CAPI: The case of the British Household Panel Survey, *Social Science Computer Review*, Vol. 18 No.4, pp 397 – 406, Sage Publications Inc.

Breeze, E. (1990) *General Household Survey: Report on Sampling Error*. OPCS Series G.H.S. no 18.

Buck, N. (2000) Using panel surveys to study migration and residential mobility, in D. Rose (ed.) *Researching Social and Economic Change*. London: Routledge.

Burton, J., Laurie, H. and Lynn, P. (2006) The Long-Term Effectiveness of Refusal Conversion Procedures on Longitudinal Surveys, *Journal of the Royal Statistical Society Series A (Statistics in Society)*, Vol. 169 No.2.

Bushnell, D. (2000) The impact of coding on data quality, *Survey Methods Newsletter*, 20:1, 16-19.

Butcher, B., Elliot, D. *A Sampling Errors Manual*, OPCS Paper NM13.

Campanelli, P. and Moon, N. (1994) Computer aided occupational class coding, paper presented at the annual meeting of the *World Association for Public Opinion Research*, Boston MA.

Campanelli, P. and O’Muircheartaigh, C. (1999) Interviewers, interviewer continuity and panel survey nonresponse, *Quality and Quantity*, Vol. 33, pp 59 – 76.

Campanelli, P. and O’Muircheartaigh, C. (2002) The importance of experimental control in testing the impact of interviewer continuity on panel survey nonresponse, *Quality and Quantity*, Vol. 36, pp 129 – 144.

Campanelli, P., Thomson, K., Moon, N. and Staples, T. (1997) The quality of occupational coding in the United Kingdom, pp. 139-158 in *Survey Measurement and Process Quality*: Lyberg, L., Biemer, P., Collins, M., deLeeuw, E., Dippo, C., Schwarz, N. and Trewin, D., (Eds.), John Wiley & Sons: New York.

Dex, S. and McCulloch, A. (1997) The reliability of Retrospective Unemployment History Data, *ISER Working Papers*, 97-17

Elias, P. (1997), Who Forgot They Were Unemployed?, *ISER Working Paper, No. 1997-19*, Colchester: University of Essex.

Ernst, L.R. (1989), Weighting issues for longitudinal household and family estimates, pp. 139-158 in *Panel Surveys*: Kasprzyk, D., Duncan, G., Kalton, G. and Singh, M.P., (Eds.), John Wiley & Sons: New York.

Ford, B.L. (1983) An Overview of Hot-Deck Procedures, in W.G. Madow, I. Olkin and D.B. Rubin (eds.) *Incomplete Data in Sample Surveys*, vol. 2, New York: Academic Press, 185-207.

Fowler, F.J. Jr., (1995), *Improving Survey Questions: Design and Evaluation*, Applied Social Research Methods Series Volume 38, Thousand Oaks, CA: Sage Publications

Government Statistical Service (1997) Report of the Task Force on Imputation, *GSS Methodology Series* no.3, London: ONS.

Halpin, B. (1997) Unified BHPS work-life histories : combining multiple sources into a user-friendly format, *Technical Paper 13 of the ESRC Research Centre on Micro-Social Change*, Colchester: University of Essex.

Hox, J. (1999) Item missing data: a software review, *Survey Methods Newsletter*, 19:2, 12-20.

Jäckle, A. and Lynn, P. (2004) Dependent Interviewing and Seam Effects in Work History Data, *ISER Working Paper, No. 2004-24*, Colchester: University of Essex. <http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-24.pdf>

Jarvis, S. and Jenkins, S. (1995) Do the Poor Stay Poor? New Evidence about income Dynamics from the British Household Panel Survey', *Occasional Papers of the ESRC Research centre on Micro-Social Change*. Occasional paper 95-2. Colchester: University of Essex.

Kalton, G. and Kasprzyk, D. (1986) The Treatment of Missing Survey Data, *Survey Methodology*, 12, 1-16.

Laurie, H. Smith, R. and Scott, L. (1999) Strategies for reducing nonresponse in a longitudinal panel survey, *Journal of Official Statistics*, 15:2, 269-282. <http://www.jos.nu/Articles/abstract.asp?article=152269>

Lavallée, P and Hunter, L. (1992) Weighting for the Survey of Labour and Income Dynamics, *Proceedings of Statistics Canada Symposium 92, Design and analysis of longitudinal surveys*, 65-74.

Lee, E. S., Forthofer, R.N., Lorimor, R.J. (1989) *Analyzing Complex Survey Data: Quantitative Application in the Social Sciences No. 71*, Sage Publications: Newbury Park.

Lynn, P. & Lievesley, D. (1991) *Drawing General Population Samples in Great Britain*. London: SCPR.

Lynn, P., Jäckle, A., Jenkins, S. P. and Sala, E. (2004a) The Effects of Dependent Interviewing on Responses to Questions on Income Sources, *ISER Working Paper, No. 2004-16*, Colchester: University of Essex. <http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-16.pdf>

Lynn, P., Jäckle, A., Jenkins, S. P. and Sala, E. (2004b) The Impact of Interviewing Method on Measurement Error in Panel Survey Measures of Benefit Receipt: Evidence from a Validation

Study, *ISER Working Paper, No. 2004-28*, Colchester: University of Essex.
<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-28.pdf>

Martin, J. (1999) An overview of imputation methods and their application to survey data, *Survey Methods Newsletter*, 19:2, 9-11.

O'Muircheartaigh, C. and 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)*, 161:1, 63-77.

O'Muircheartaigh, C., Campanelli, P. and Smith, P.W.F. (1999) A multilevel exploration of the role of interviewers in survey non-response, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 162:3, 437-448

Pevalin, D (2000) 'Multiple applications of the GHQ-12 in a general population sample: an investigation of long-term retest Effects' *Social Psychiatry And Psychiatric Epidemiology* 35: 508-512

Rendtel, U. (1991) Weighting procedures and sampling variance in household panels, *Working Paper 11 of the European Science Foundation Scientific Network on Household Panel Studies*, Colchester, Essex.

Rigg, J. (2001) 'Comparisons Between The BHPS And The HBAI' Unpublished MS, Institute for Social and Economic Research.

Rubin, D. (1996) Multiple imputation after 18+ years, *Journal of the American Statistical Association*, 91, 473-489.

Sala, E. and Lynn, P. (2004) The Effect of Dependent Interviewing on Measures of Employment Characteristics, *ISER Working Paper, No. 2004-26*, Colchester: University of Essex.
<http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-26.pdf>

Saris, W. and Münnich, A. (ed.s) (1995) The multitrait-multimethod approach to evaluate measurement instruments, Eötvös University Press: Budapest.

Scott, J. with Brynin, M. and Smith, R. (1995) Interviewing children in the British Household Panel Survey, pp 259-266, in J. J. Hox et al (eds) *Advances in Family Research*, Amsterdam, Thesis Publishers

Scott, J. (1995) Children as respondents: methods for improving data quality, pp 331-350, in L. Lyberg et al (eds) *Survey Measurement and Process Quality*, Wiley.

Skinner, C.J., Holt, D., Smith, T.M.F. (1989) *Analysis of Complex Surveys*, Wiley: Chichester.

Taylor, A. (1994) Sample characteristics, attrition and weighting in N. Buck, J. Gershuny, D. Rose, J. Scott (eds.) *Changing Households*. Colchester: ESRC Research Centre on Micro-Social Change.

Taylor, M. F. (ed). Brice, J., Buck, N. and Prentice-Lane, E. (2001) *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices*. Colchester: University of Essex.

Table 1: Stratification of Postal Sectors

Region	Major strata per region	Minor strata per major stratum	Stratification factor within minor strata	Number of selected PSUs
Inner London	2	2	SPH	13
Outer London	3	2	SPH	18
Rest of South East	3	2	AGEMP	48
South West	3	2	AGEMP	21
East Anglia	2	2	AGEMP	9
East Midlands	3	2	AGEMP	18
West Midlands Conurbation	2	2	SPH	12
Rest of West Midlands	2	2	AGEMP	11
Greater Manchester	2	2	SPH	12
Merseyside	2	2	SPH	6
Rest of North West	2	2	AGEMP	11
South Yorkshire	2	2	SPH	6
West Yorkshire	2	2	SPH	9
Rest of Yorks and Humberside	2	2	AGEMP	7
Tyne and Wear	2	2	SPH	6
Rest of North England	2	2	AGEMP	9
Wales	2	2	AGEMP	12
Scotland	3	2	AGEMP	22

Table 2: Distribution of selected addresses per PSU

Number of selected addresses (n_i)	Number of PSUs
21	1
24	1
25	1
27	1
29	1
30	1
31	4
32	50
33	177
34	10
35	2
36	1
Total	250

Table 3: Adult population in institutions, by institution type and sex

	Total Residents		
	Male	Female	Total
NHS hospitals/homes	11.6	12.4	12.1
Non-NHS hospitals	1.0	0.9	0.9
LA homes	10.6	16.3	13.9
HA homes and hostels	3.0	2.4	2.7
Nursing homes (Non-NHS/HA/LA)	10.0	21.0	16.5
Residential homes (Non-NHS/HA/LA)	15.0	29.0	23.3
Childrens homes	1.7	1.0	1.3
Prison service establishments	4.2	0.1	1.8
Defence establishments	12.7	1.4	6.1
Educational establishments	6.1	3.1	4.3
Hotels, boarding houses etc.	15.7	7.6	10.9
Hostels and common lodging houses (non-HA)	3.9	1.3	2.4
Other misc. establishments	3.7	3.3	3.4
Civilian ships, boats, barges	0.2	0.0	0.1
Base (total)	310,453	445,288	755,741

Source: 1991 Census (OPCS, 1993)

Table 4: Numbers of achieved individual interviews at each wave

	OSMs	PSM(B)s	PSM(P)s	TSMs	Total
Wave 1	10,264	-	-	-	10,264
Wave 2	9,351	-	10	484	9,845
Wave 3	8,921	-	29	650	9,600
Wave 4	8,609	-	77	795	9,481
Wave 5	8,305	-	120	824	9,249
Wave 6	8,315	-	179	944	9,438
Wave 7	8,155	-	240	1,071	9,466
Wave 8	7,992	-	291	1,032	9,315
Wave 9	7,821	-	359	1,043	9,223
Wave 10	7,600	-	334	1,018	8,952
Wave 11	7,448	-	325	1,110	8,883
Wave 12	7,299	-	310	1,161	8,770
Wave 13	7,120	-	299	1,236	8,655

Note: Figures include both personal and proxy interviews and, from wave 3 onwards, include telephone interviews. 54 persons who were not enumerated at wave 1, but subsequently interviewed on at least one occasion, are excluded from this table

Table 5: Sample status and longitudinal response status of interview respondents at each wave

	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8	Wave 9	Wave 10	Wave 11	Wave 12	Wave 13
Continuing OSM	8934	8335	8187	7899	8019	7895	7714	7545	7348	7163	6942	6801
Of which:												
W1 response, of which:	8568	7839	7577	7183	7132	6900	6647	6396	6143	5914	5694	5481
All waves	8568	7622	7138	6665	6419	6172	5927	5677	5405	5118	4878	4648
Ever wave non-response	0	213	436	508	702	715	705	700	722	778	799	801
Ever out of scope	0	4	3	10	11	13	15	19	16	18	17	15
W1 non-response	185	188	174	169	175	180	169	166	157	146	139	130
W1 aged under 16	153	287	415	526	688	795	877	961	1028	1083	1091	1171
W1 non-contact/non-enumeration	65	53	50	48	52	51	54	52	53	49	45	43
PSM: parent of OSM	9	28	73	116	166	225	272	325	325	314	296	284
TSM	451	608	750	764	900	947	900	898	975	1064	1100	1136
Total	9459	9024	9060	8827	9137	9118	8940	8820	8701	8590	8383	8264

Table 6: Number of sequential waves response starting at Wave 1 by age at Wave 1

	16-24	25-34	35-44	45-59	60+	All
1 or more	1511	2057	1845	2106	2378	9912
2 or more	1237	1816	1640	1833	2029	8568
3 or more	1081	1650	1490	1642	1756	7630
4 or more	997	1549	1417	1543	1621	7138
5 or more	923	1448	1341	1452	1490	6665
6 or more	878	1404	1315	1413	1398	6418
7 or more	840	1354	1285	1378	1306	6171
8 or more	809	1307	1233	1350	1217	5924
9 or more	768	1267	1190	1309	1133	5674
10 or more	735	1223	1141	1268	1032	5404
11 or more	696	1179	1083	1214	951	5123
12 or more	658	1144	1037	1165	879	4883
13	630	1091	993	1131	808	4653

Table 7: Number of sequential waves response starting at Wave 5 by age at Wave 5

	16-24	25-34	35-44	45-59	60+	All
1 or more	1191	1620	1502	1871	1953	8162
2 or more	1128	1557	1454	1817	1836	7817
3 or more	1059	1497	1405	1768	1726	7479
4 or more	1007	1439	1354	1709	1620	7150
5 or more	941	1382	1300	1647	1515	6804
6 or more	875	1319	1249	1588	1395	6445
7 or more	822	1257	1194	1518	1297	6088
8 or more	756	1213	1152	1446	1209	5776
9	702	1156	1103	1393	1127	5481

Table 8: Number of consecutive pairs of waves response by OSM or PSM, by age in 1991

Age in 1991	Number of wave pairs
Under 16	9,907
16-24	16,534
25-34	23,244
35-44	19,960
45-59	22,989
60+	19,618
All	112,252

Note: Includes Scotland and Wales extension samples and Northern Ireland sample, but not ECHP low income sample.

Table 9: Number of observed wave-on-wave employment transitions by OSM or PSM up to wave 13

	Number of transitions	Number of respondents with at least one
Employment to employment	62056	11452
Employment to unemployment	1400	1264
Employment to inactivity	3715	3197
Unemployment to employment	1583	1406
Unemployment to unemployment	1746	862
Inactivity to employment	3896	3326

Note: Includes Scotland and Wales extension samples and Northern Ireland sample, but not ECHP low income sample.

Table 10: Number of observed wave-on-wave demographic transitions by OSM or PSM up to wave 13

	Total events	Full respondent next wave	Number of Individuals
<u>All samples</u>			
Partnership formation	2720	2615	2379
Partnership separation	2067	1536	1830
Child leaving home	2858	2052	2475
			Birth of first child
Birth of child - mother	2631	2519	777
Birth of child - father	2234	1981	759
<u>Original sample only</u>			
Partnership formation	2352	2263	2015
Partnership separation	1761	1330	1533
Child leaving home	2358	1759	1985
			Birth of first child
Birth of child - mother	1851	1780	673
Birth of child - father	1606	1464	653

Table 11: Distribution of within-address selection probabilities of households

Conditional selection probability	Number of households
1.00	5,286
0.75	50
0.60	10
0.50	39
0.429	9
0.375	56
0.333	17
0.30	9
0.273	2
0.25	3

Table 12: Distribution of design weights

Weight	Frequency
.1111	9
.1250	2
.1429	6
.1667	11
.2000	1
.2222	2
.2500	6
.2857	6
.2917	5
.3333	206
.3810	5
.4000	1
.4444	2
.5000	8
1.0000	9883
1.3333	42
1.667	11
2.0000	17
2.3333	2
2.6667	6
3.0000	9
3.3333	16
3.6667	3
4.0000	5
Total	10264

Table 13: Numbers of achieved individual interviews: ECHP sub-sample

	OSMs	PSMs	TSMs	Total
Wave 7	1,727	7	86	1,820
Wave 8	1,591	9	91	1,691
Wave 9	1,487	18	104	1,609
Wave 10	1,425	17	127	1,569
Wave 11	1,349	14	156	1,519

Note: Figures include both personal and proxy interviews and telephone interviews (see Table 19)

Table 14: Stratification of postal sectors (Scotland)

Region	Major strata per region	Minor strata per major stratum	Stratification factor within minor strata	Number of selected PSUs
1.Borders, Dumfries and Galloway	2	1	AGEMP	3
2.City of Glasgow	3	1	SPH	10
3.Remainder of Strathclyde	3	2	AGEMP	22
4.Loathian, Central, Fife	3	2	AGEMP	21
5.Tayside, Grampian	2	2	AGEMP	14
6.Highlands and Islands	2	1	AGEMP	5

Regions consist of the following council areas: 1 Dumfries and Galloway, Scottish Borders; 2 Glasgow City; 3 Inverclyde, Renfrewshire, East Renfrewshire, North Ayrshire, East Ayrshire, South Ayrshire, South Lanarkshire, North Lanarkshire, East Dunbartonshire, West Dunbartonshire; 4 East Lothian, Midlothian, Edinburgh City, West Lothian, Falkirk, Stirling, Clackmannanshire, Fife; 5 Dundee City, Angus, Perth & Kinross, Aberdeenshire, Aberdeen City, Moray; 6 Argyll & Bute, Highland, Western Isles, Orkney, Shetland

Table 15: Stratification of Postal Sectors (Wales)

Region	Major strata per region	Minor strata per major stratum	Stratification factor within minor strata	Number of selected PSUs
1.North Wales	2	2	WELSH	16
2.Mid and West Wales	2	2	WELSH	14
3.South Wales West	2	2	WELSH	13
4.South Central Wales	3	2	WELSH	16
5.South East Wales	3	2	AGEMP	16

Regions consist of the following council areas: 1 Vale of Clwyd, Alyn and Deeside, Wrexham, Delyn, Clwyd South, Clwyd West Conwy, Ynys Mon, Caernarfon; 2 Meirionnydd nant Conwy, Montgomeryshire, Brecon and Radnor, Ceredigion, Carmarthen East and Dinefwr, Carmarthen West and South Pembrokeshire, Preseli Pembrokeshire, Llanelli; 3 Swansea West, Swansea East, Bridgend, Ogmore, Gower, Aberavon, Neath; 4 Vale of Glamorgan Cardiff North, Cardiff Central, Cardiff West, Cardiff South and Penarth, Pontypridd, Rhondda, Cynon Valley; 5 Monmouth, Newport East, Newport West, Torfaen, Islwyn, Blaenau Gwent, Merthyr Tydfil and Rhymney, Caerphilly

Table 16: Numbers of achieved individual interviews: Scottish booster sample

	OSMs	PSMs	TSMs	Total
Wave 9	2,448	-	-	2,448
Wave 10	2,469	-	93	2,562
Wave 11	2,375	-	128	2,503
Wave 12	2,256	-	153	2,409
Wave 13	2,151	-	161	2,312

Note: Figures include both personal and proxy interviews and telephone interviews (see Table 19)

Table 17: Numbers of achieved individual interviews: Welsh booster sample

	OSMs	PSMs	TSMs	Total
Wave 9	2,467	-	-	2,448
Wave 10	2,390	-	78	2,468
Wave 11	2,304	-	145	2,449
Wave 12	2,243	-	151	2,394
Wave 13	2,230	-	159	2,389

Note: Figures include both personal and proxy interviews and telephone interviews (see Table 19)

Table 18: Numbers of achieved individual interviews: British Youth Panel

	OSMs	ECHP	TSMs	Scot/Welsh boosters	Total
Wave 4	751	-	22	-	773
Wave 5	720	-	29	-	749
Wave 6	702	-	46	-	748
Wave 7	671	-	49	-	720
Wave 8	715	167	64	-	946
Wave 9	693	176	69	-	938
Wave 10	688	175	109	442	1,414
Wave 11	698	163	79	396	1,336
Wave 12	727	-	90	365	1,182
Wave 13	553	-	89	379	1,021

Note: Figures include both personal and proxy interviews and telephone interviews (see Table 19)

Table 19: Numbers of personal, proxy and telephone interviews

	In-person	Proxy	Telephone	Total
Wave 1	9,912	352	-	10,264
Wave 2	9,459	386	-	9,845
Wave 3	9,024	324	252	9,600
Wave 4	9,060	309	112	9,481
Wave 5	8,827	286	136	9,249
Wave 6	9,137	249	52	9,438
Wave 7	9,198	236	32	9,466
Wave 7 ECHP	1,630	90	7	1,727
Wave 8	9,030	226	59	9,315
Wave 8 ECHP	1,518	56	17	1,591
Wave 9	8,930	218	75	9,223
Wave 9 ECHP	1,423	52	12	1,487
Wave 9 S&W	4,837	78	-	4,915
Wave 10	8,701	202	103	9,006
Wave 10 ECHP	1,494	48	27	1,569
Wave 10 S&W	4,888	60	82	5,030
Wave 11	8,590	182	164	8,936
Wave 11 ECHP	1,450	47	26	1,523
Wave 11 S&W	4,770	46	136	4,952
Wave 12	8,383	195	240	8,818
Wave 12 S&W	4,408	49	256	4,713
Wave 13	8,264	213	224	8,701
Wave 13 S&W	4,230	85	274	4,589

Note: Figures are for individual interviews, including OSMs, PSMs and TSMs. From wave 7 onwards, figures are presented separately for ECHP sub-sample members. From wave 9 onwards, figures are presented separately for the Scotland and Wales extension samples.

Table 20: Stability of BHPS interviewing panel

	No. of interviewers	No. of new interviewers	No. of wave 1 interviewers	% with same interviewer at both wave 1 and wave <i>x</i> (base= full respondents at both waves)	% with same interviewer at <u>all</u> waves up to and including wave <i>x</i> (base= full respondents at <u>all</u> waves)
Wave 1	243	--	243	100	100
Wave 2	237	35	202	59	59
Wave 3	216	14	181	57	48
Wave 4	217	32	167	50	41
Wave 5	217	18	144	50	37
Wave 6	212	25	139	44	33
Wave 7	218	11	125	39	30
Wave 8	228	23	121	39	28
Wave 9	212	15	118	35	25
Wave 10	212	19	103	32	23
Wave 11	209	18	92	32	23
Wave 12	259	22	83	29	21
Wave 13	261	25	79	27	20

Note: These figures exclude the ECHP sub-sample, the Scottish and Welsh booster samples and the Northern Ireland sample.

Table 21: Distribution of individual interviews by month

	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Missing	Total
W1	1964	5254	2431	243	-	-	-	-	-	20	9912
%	19.8	53.0	24.5	2.5	-	-	-	-	-	.2	100
W2	3672	4243	1181	128	47	65	120	3	-	-	9459
%	38.8	44.9	12.5	1.4	.5	.7	1.3	0	-	-	100
W3	3434	3707	1294	214	122	127	87	39	-	8	9032
%	38.0	41.0	14.3	2.4	2.4	1.4	1.0	.4	-	.1	100
W4	2693	4044	1570	375	193	85	55	40	5	2	9062
%	29.7	44.6	17.3	4.1	2.1	.9	.6	.4	.1	-	100
W5	1720	4078	2225	405	214	86	67	25	7	1	8828
%	19.5	46.2	25.5	4.6	2.4	1.0	.8	.3	.1	-	100
W6	4705	3227	786	207	115	44	27	13	13	2	9139
%	51.5	35.3	8.6	2.3	1.3	.5	.3	.1	.1	-	100
W7	4949	2856	828	229	135	73	29	11	8	-	9118
%	54.3	31.3	9.1	2.5	1.5	.8	.3	.1	.1	-	100
W8	4843	2752	835	284	124	54	28	16	4	-	8940
%	54.2	30.8	9.3	3.2	1.4	.6	.3	.2	0	-	100
W9	4268	2420	1034	215	283	201	189	113	97	-	8820
%	48.8	27.4	11.7	2.4	3.2	2.3	2.1	1.3	1.1	-	100
W10	4574	3082	733	130	107	42	10	11	12	-	8701
%	52.6	35.4	8.4	1.5	1.2	.5	.1	.1	.1	-	100
W11	4579	3022	721	101	109	38	16	4	-	-	8590
%	53.3	35.2	8.4	1.2	1.3	0.4	0.2	0.1	-	-	100
W12	4397	2794	792	184	142	36	22	15	1	-	8383
%	52.5	33.3	9.5	2.2	1.7	0.4	0.3	0.2	0.0	-	100
W13	4661	2567	803	116	58	38	13	7	1	-	8264
%	56.4	31.1	9.7	1.4	0.7	0.5	0.2	0.1	0.0	-	100

Figures are for OSMs and OSM(D)s only – the sub-samples added at later waves are not included. Proxy and telephone interviews are also excluded.

Table 22: Dates of wave 1 pre-tests and pilots

	<u>Date</u>
Pre-test (household & employment)	4/1990
Pre-test (all schedules)	5/1990
Pre-test (all schedules & procedures)	6 - 7/1990
Pre-test (calendar design)	9/1990
Pilot 1 (500 households)	10 - 11/1990
Pilot 2 (500 households)	4 - 5/1991

Table 23: Percentage of respondents giving contact details

	% giving contact name	% mobile telephone	% email address	% both mobile and email
Wave 1	88	--	--	--
Wave 2	93	--	--	--
Wave 3	90	--	--	--
Wave 4	92	--	--	--
Wave 5	91	--	--	--
Wave 6	91	--	--	--
Wave 7	91	--	--	--
Wave 8	92	--	--	--
Wave 9	95	--	--	--
Wave 10	95	38	9	6
Wave 11	92	4	1	1
Wave 12	92	40	11	8
Wave 13	91	45	17	10

Includes ECHP sub-sample from Wave 7 and Scottish and Welsh extension samples from Wave 9.

Table 24: Outcome of field work conversion attempts

	Conversion attempts	Successful conversions						
	Number of households	All eligible persons interviewed	1+ interview, 1+ proxy	1+ interview, 1+ refusal	Telephone interview only	Total converted	next wave - at least 1 interview	next wave - telephone only
Wave 4	294	27 9.2%	12 4.1%	38 12.9%	88 29.9%	165 56.1%	94 57.0%	24 14.5%
Wave 5	305	31 10.2%	7 2.3%	13 4.3%	98 32.1%	149 48.9%	90 60.4%	7 4.7%
Wave 6	126	9 7.1%	--	4 3.2%	40 31.7%	53 42.1%	29 54.7%	2 3.8%
Wave 7	92	12 13.0%	5 5.4%	9 9.8%	25 27.2%	51 55.4%	29 54.9%	4 7.8%
Wave 8	175	14 8.0%	4 2.3%	7 4.0%	59 33.7%	84 48.0%	40 47.6%	5 6.0%
Wave 9	192	27 14.1%	3 1.6%	5 2.6%	55 28.6%	90 46.9%	45 50.0%	18 20.0%
Wave 10	343	15 4.4%	1 0.3%	7 2.0%	154 44.9%	177 51.6%	76 42.9%	38 21.5%
Wave 11	556	27 4.9%	3 0.5%	18 3.2%	246 44.2%	294 52.9%	102 34.7%	87 29.6%
Wave 12	788	12 1.5%	--	9 1.1%	351 44.5%	372 47.2%	124 33.3%	98 26.3%
Wave 13	630	19 3.0%	--	1 0.2%	301 47.8%	321 51.0%	--	--

Notes: Conversion attempts are only made with wholly-refusing households. The next wave figures are based upon the households of a sample of one person per converted household. For example, at wave 4, 165 households were converted. Randomly selecting one person per household provides a sample of 165 wave 5 households, of which 94 (57%) provided at least one interview.

Table 25: Wave 1 field outcomes

Wave One Household Outcomes	Number	%
Addresses issued	8167	
Vacant/non-residential/not main residence	1033	
Multi-households addition to sample	103	
Eligible households	7491	(100%)
Refusal to field agency/research centre	357	(2%)
Household refusal to interviewer	1420	(19%)
Household non-contact	288	(4%)
Language/age/infirmity problems	122	(2%)
Complete household interview	4862	(65%)
Complete household coverage (inc. proxies)	5143	(69%)
Partial household coverage	5538	(74%)

Wave One Individual Outcomes		
Enumerated individuals	13840	(100%)
Ineligible children (Under 16)	3089	(22%)
Eligible adults	10751	(100%)
Refusals	426	(4%)
Non-contact / absent	48	(0%)
Age / infirmity / disability or language difficulty	13	(0%)
Full interviews	9912	(92%)
Proxy interviews	352	(3%)
Total interviews	10264	(95%)

Table 26: Wave 2 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	13840	-	13840
Retiring sample: B	-	-	-
Not issued – inaccessible: C	-	-	-
Issued sample all: D (=E+F)	13840	-	13840
Issued sample adults 16+: E (=G+H)	11006	-	11006
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2834	49	2883
Adult not eligible for interview: G	166	-	166
Assumed eligible for interview: H	10840	-	10840
Extra eligible for interview: I	83	554	637
Field outcomes for eligible sample:			
Refusal: J	1175 (11%)	50 (9%)	1225 (11%)
Non-contact: K	359 (3%)	2 (1%)	361 (3%)
Other NR: L	38 (0%)	8 (1%)	46 (0%)
Response: M	9351 (86%)	494 (89%)	9845 (86%)
Total: N (=H+I)	10923 (100%)	554 (100%)	11477 (100%)

Table 27: Wave 3 Individual outcomes

	OSMs and PSMs		TSMs		ALL	
Total active sample: A (=B+C+D)	14063		591		14500	
Retiring sample: B	345		-		345	
Not issued – inaccessible: C	65		-		65	
Issued sample all: D (=E+F)	13653		591		14244	
Issued sample adults 16+: E (=G+H)	10805		498		11303	
Field outcomes for issued sample:						
Child <16 not eligible interview: F	2848		93		2941	
Adult not eligible for interview: G	230		254		484	
Assumed eligible for interview: H	10575		244		10819	
Extra eligible for interview: I	-		569		569	
Field outcomes for eligible sample:						
Refusal: J	1079	(10%)	125	(15%)	1204	(11%)
Non-contact: K	458	(4%)	7	(1%)	465	(4%)
Other NR: L	88	(1%)	31	(4%)	119	(1%)
Response: M	8950	(85%)	650	(80%)	9600	(84%)
Total: N (=H+I)	10575	(100%)	813	(100%)	11388	(100%)

Table 28: Wave 4 Individual outcomes

	OSMs and PSMs		TSMs		ALL	
Total active sample: A (=B+C+D)	13718		919		14678	
Retiring sample: B	489		9		496	
Not issued – inaccessible: C	564		10		562	
Issued sample all: D (=E+F)	12665		900		13565	
Issued sample adults 16+: E (=G+H)	9966		803		10769	
Field outcomes for issued sample:						
Child <16 not eligible interview: F	2699		97		2796	
Adult not eligible for interview: G	144		298		443	
Assumed eligible for interview: H	9821		505		10229	
Extra eligible for interview: I	-		537		537	
Field outcomes for eligible sample:						
Refusal: J	891	(9%)	166	(16%)	1057	(7%)
Non-contact: K	253	(3%)	23	(2%)	276	(2%)
Other NR: L	37	(0%)	12	(1%)	49	(0%)
Response: M	8640	(88%)	841	(81%)	9481	(91%)
Total: N (=H+I)	9821	(100%)	1042	(100%)	10863	(100%)

Table 29: Wave 5 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	13371	1119	14490
Retiring sample: B	346	17	363
Not issued – inaccessible: C	795	24	819
Issued sample all: D (=E+F)	12228	1078	13306
Issued sample adults 16+: E (=G+H)	9614	946	10572
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2614	120	2734
Adult not eligible for interview: G	131	340	471
Assumed eligible for interview: H	9483	606	10089
Extra eligible for interview: I	-	516	516
Field outcomes for eligible sample:			
Refusal: J	816 (9%)	197 (17%)	1013 (10%)
Non-contact: K	244 (3%)	23 (2%)	266 (2%)
Other NR: L	47 (0%)	29 (2%)	76 (1%)
Response: M	8376 (88%)	873 (78%)	9249 (87%)
Total: N (=H+I)	9483 (100%)	1122 (100%)	10605 (100%)

Table 30: Wave 6 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	13147	1229	14376
Retiring sample: B	296	18	314
Not issued – inaccessible: C	912	37	949
Issued sample all: D (=E+F)	11939	1174	13113
Issued sample adults 16+: E (=G+H)	9426	1004	10430
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2513	170	2683
Adult not eligible for interview: G	115	328	443
Assumed eligible for interview: H	9311	676	9987
Extra eligible for interview: I	-	524	524
Field outcomes for eligible sample:			
Refusal: J	683 (7%)	194 (16%)	877 (8%)
Non-contact: K	172 (2%)	25 (2%)	197 (2%)
Other NR: L	26 (2%)	38 (3%)	64 (1%)
Response: M	8430 (91%)	943 (79%)	9373 (89%)
Total: N (=H+I)	9311 (100%)	1200 (100%)	10511 (100%)

Table 31: Wave 7 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	12996	1417	14413
Retiring sample: B	869	59	928
Not issued – inaccessible: C	569	29	598
Issued sample all: D (=E+F)	11558	1329	12887
Issued sample adults 16+: E (=G+H)	9143	1163	10306
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2433	156	2589
Adult not eligible for interview: G	126	417	543
Assumed eligible for interview: H	9017	746	9763
Extra eligible for interview: I	-	521	521
Field outcomes for eligible sample:			
Refusal: J	513 (6%)	192 (15%)	705 (6%)
Non-contact: K	146 (2%)	19 (1%)	165 (2%)
Other NR: L	28 (0%)	13 (1%)	41 (1%)
Response: M	8330 (92%)	1043 (82%)	9373 (91%)
Total: N (=H+I)	9017 (100%)	1267 (100%)	10284 (100%)

Table 32: Wave 8 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	12287	1405	13692
Retiring sample: B	99	22	121
Not issued – inaccessible: C	732	40	772
Issued sample all: D (=E+F)	11456	1343	12799
Issued sample adults 16+: E (=G+H)	9033	1062	10095
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2423	281	2704
Adult not eligible for interview: G	104	386	490
Assumed eligible for interview: H	8929	676	9605
Extra eligible for interview: I	-	564	564
Field outcomes for eligible sample:			
Refusal: J	547 (7%)	198 (16%)	745 (7%)
Non-contact: K	132 (1%)	23 (2%)	155 (2%)
Other NR: L	34 (0%)	20 (2%)	54 (0%)
Response: M	8216 (92%)	999 (80%)	9215 (91%)
Total: N (=H+I)	8929 (100%)	1240 (100%)	10169 (100%)

Table 33: Wave 9 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	12337	1406	13743
Retiring sample: B	90	12	102
Not issued – inaccessible: C	833	47	880
Issued sample all: D (=E+F)	11414	1347	12761
Issued sample adults 16+: E (=G+H)	9020	1153	10173
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2394	194	2588
Adult not eligible for interview: G	115	325	440
Assumed eligible for interview: H	8905	828	9733
Extra eligible for interview: I	-	464	464
Field outcomes for eligible sample:			
Refusal: J	618 (7%)	209 (16%)	827 (8%)
Non-contact: K	159 (2%)	23 (2%)	182 (2%)
Other NR: L	39 (0%)	48 (4%)	87 (1%)
Response: M	8089 (91%)	1012 (78%)	9101 (89%)
Total: N (=H+I)	8905 (100%)	1292 (100%)	10197 (100%)

Table 34: Wave 10 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	12412	1479	13891
Retiring sample: B	137	16	153
Not issued – inaccessible: C	972	59	1031
Issued sample all: D (=E+F)	11303	1404	12707
Issued sample adults 16+: E (=G+H)	8938	1185	10123
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2365	219	2584
Adult not eligible for interview: G	132	306	438
Assumed eligible for interview: H	8806	897	9703
Extra eligible for interview: I	-	407	407
Field outcomes for eligible sample:			
Refusal: J	615 (7%)	236 (19%)	851 (8%)
Non-contact: K	169 (2%)	20 (1%)	189 (2%)
Other NR: L	34 (0%)	30 (2%)	64 (1%)
Response: M	7988 (91%)	1018 (78%)	9006 (89%)
Total: N (=H+I)	8806 (100%)	1304 (100%)	10110 (100%)

Table 35: Wave 11 Individual outcomes

	OSMs and PSMs		TSMs		ALL	
Total active sample: A (=B+C+D)	12447		1511		14559	
Retiring sample: B	61		3		64	
Not issued – inaccessible: C	1054		66		1120	
Issued sample all: D (=E+F)	11332		1442		13375	
Issued sample adults 16+: E (=G+H)	8963		1269		10232	
Field outcomes for issued sample:						
Child <16 not eligible interview: F	2369		173		2752	
Adult not eligible for interview: G	112		304		416	
Assumed eligible for interview: H	8851		965		9816	
Extra eligible for interview: I			391		391	
Field outcomes for eligible sample:						
Refusal: J	701	(8%)	274	(20%)	975	(10%)
Non-contact: K	106	(1%)	34	(3%)	140	(1%)
Other NR: L	134	(1%)	22	(2%)	156	(2%)
Response: M	7910	(90%)	1026	(75%)	8936	(88%)
Total: N (=H+I)	8851	(100%)	1356	(100%)	10207	(100%)

Table 36: Wave 12 Individual outcomes

	OSMs and PSMs		TSMs		ALL	
Total active sample: A (=B+C+D)	12551		1590		13268	
Retiring sample: B	490		39		529	
Not issued – inaccessible: C	802		39		841	
Issued sample all: D (=E+F)	11259		1512		14638	
Issued sample adults 16+: E (=G+H)	8916		1327		10243	
Field outcomes for issued sample:						
Child <16 not eligible interview: F	2343		185		2528	
Adult not eligible for interview: G	101		296		397	
Assumed eligible for interview: H	8815		1031		9846	
Extra eligible for interview: I			310		310	
Field outcomes for eligible sample:						
Refusal: J	767	(9%)	245	(18%)	1012	(10%)
Non-contact: K	83	(1%)	56	(4%)	139	(1%)
Other NR: L	165	(2%)	24	(2%)	189	(2%)
Response: M	7800	(89%)	1018	(76%)	8818	(87%)
Total: N (=H+I)	8815	(100%)	1343	(100%)	10158	(100%)

Table 37: Wave 13 Individual outcomes

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	12209	1551	14268
Retiring sample: B	250	15	265
Not issued – inaccessible: C	970	63	1033
Issued sample all: D (=E+F)	10989	1473	12970
Issued sample adults 16+: E (=G+H)	8731	1294	10025
Field outcomes for issued sample:			
Child <16 not eligible interview: F	2258	179	2437
Adult not eligible for interview: G	119	268	387
Assumed eligible for interview: H	8612	1026	9638
Extra eligible for interview: I		318	318
Field outcomes for eligible sample:			
Refusal: J	755 (9%)	242 (18%)	997 (10%)
Non-contact: K	70 (1%)	49 (4%)	119 (1%)
Other NR: L	114 (1%)	14 (1%)	128 (1%)
Response: M	7673 (89%)	1039 (77%)	8712 (87%)
Total: N (=H+I)	8612 (100%)	1344 (100%)	9956 (100%)

Table 38: Wave 1 Field outcomes for Scotland and Wales extension samples (BHPS wave 9)

Wave One Household Outcomes	Scotland		Wales	
	Number	%	Number	%
Addresses issued	2475		2475	
Vacant/non-residential/not main residence	302		295	
Multi-households addition to sample	226		11	
Eligible households	2399	(100%)	2191	(100%)
Refusal to field agency/research centre	28	(1%)	33	(1%)
Household refusal to interviewer	668	(28%)	580	(26%)
Household non-contact	189	(8%)	91	(4%)
Language/age/infirmity problems	55	(2%)	59	(3%)
Complete household interview	1241	(52%)	1152	(53%)
Complete household coverage (inc. proxies)	1276	(53%)	1186	(54%)
Partial household coverage	1459	(61%)	1428	(65%)
Wave One Individual Outcomes				
Enumerated individuals	3395		3577	
Ineligible children (Under 16)	724		807	
Eligible adults	2671	(100%)	2770	(100%)
Refusals	139	(5%)	161	(6%)
Non-contact / absent	86	(3%)	142	(5%)
Age / infirmity / disability or language difficulty	-	-	-	-
Full interviews	2405	(90%)	2430	(88%)
Proxy interviews	41	(2%)	37	(1%)
Total interviews	2446	(92%)	2467	(89%)

Table 39: Wave 2 Individual outcomes for Scotland extension sample (BHPS wave 10)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3395	-	3395
Retiring sample: B	-	-	-
Not issued – inaccessible: C	-	-	-
Issued sample all: D (=E+F)	3395	-	3395
Issued sample adults 16+: E (=G+H)	2712	-	2712
Field outcomes for issued sample:			
Child <16 not eligible interview: F	683	-	683
Adult not eligible for interview: G	52	-	52
Assumed eligible for interview: H	2660	-	2660
Extra eligible for interview: I		394	394
Field outcomes for eligible sample:			
Refusal: J	194 (7%)	61 (15%)	255 (8%)
Non-contact: K	71 (3%)	20 (5%)	91 (3%)
Other NR: L	148 (6%)	-	148 (5%)
Response: M	2247 (85%)	313 (80%)	2560 (84%)
Total: N (=H+I)	2660 (100%)	394 (100%)	3054 (100%)

Table 40: Wave 3 Individual outcomes for Scotland extension sample (BHPS wave 11)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3786	151	3937
Retiring sample: B	9	-	9
Not issued – inaccessible: C	173	-	173
Issued sample all: D (=E+F)	3604	151	3755
Issued sample adults 16+: E (=G+H)	2867	128	2995
Field outcomes for issued sample:			
Child <16 not eligible interview: F	737	23	760
Adult not eligible for interview: G	30	44	74
Assumed eligible for interview: H	2837	84	2921
Extra eligible for interview: I		116	116
Field outcomes for eligible sample:			
Refusal: J	301 (11%)	51 (26%)	352 (12%)
Non-contact: K	61 (2%)	27 (14%)	88 (3%)
Other NR: L	92 (3%)	2 (1%)	94 (3%)
Response: M	2383 (84%)	120 (60%)	2503 (82%)
Total: N (=H+I)	2837 (100%)	200 (100%)	3037 (100%)

Table 41: Wave 4 Individual outcomes for Scotland extension sample (BHPS wave 12)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3793	233	4026
Retiring sample: B	24	3	27
Not issued – inaccessible: C	272	2	274
Issued sample all: D (=E+F)	3497	228	3725
Issued sample adults 16+: E (=G+H)	2808	202	3010
Field outcomes for issued sample:			
Child <16 not eligible interview: F	689	26	715
Adult not eligible for interview: G	38	60	98
Assumed eligible for interview: H	2770	142	2912
Extra eligible for interview: I		84	84
Field outcomes for eligible sample:			
Refusal: J	326 (12%)	45 (20%)	371 (12%)
Non-contact: K	58 (2%)	25 (11%)	83 (3%)
Other NR: L	122 (4%)	11 (5%)	133 (4%)
Response: M	2264 (82%)	145 (64%)	2409 (80%)
Total: N (=H+I)	2770 (100%)	226 (100%)	2996 (100%)

Table 42: Wave 5 Individual outcomes for Scotland extension sample (BHPS wave 13)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3786	253	4039
Retiring sample: B	95	1	96
Not issued – inaccessible: C	379	9	388
Issued sample all: D (=E+F)	3312	243	3555
Issued sample adults 16+: E (=G+H)	2679	223	2902
Field outcomes for issued sample:			
Child <16 not eligible interview: F	633	20	653
Adult not eligible for interview: G	33	89	122
Assumed eligible for interview: H	2646	134	2780
Extra eligible for interview: I		91	91
Field outcomes for eligible sample:			
Refusal: J	354 (13%)	53 (24%)	407 (14%)
Non-contact: K	45 (2%)	18 (8%)	63 (2%)
Other NR: L	79 (3%)	7 (3%)	86 (3%)
Response: M	2168 (82%)	147 (65%)	2315 (80%)
Total: N (=H+I)	2646 (100%)	225 (100%)	2871 (100%)

Table 43: Wave 2 Individual outcomes for Wales extension sample (BHPS wave 10)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3577	-	3577
Retiring sample: B	-	-	-
Not issued – inaccessible: C	-	-	-
Issued sample all: D (=E+F)	3577	-	3577
Issued sample adults 16+: E (=G+H)	2812	-	2812
Field outcomes for issued sample:			
Child <16 not eligible interview: F	765	-	765
Adult not eligible for interview: G	69	-	69
Assumed eligible for interview: H	2743	-	2743
Extra eligible for interview: I		331	331
Field outcomes for eligible sample:			
Refusal: J	246 (9%)	49 (15%)	295 (10%)
Non-contact: K	89 (3%)	39 (12%)	128 (4%)
Other NR: L	183 (7%)	-	183 (6%)
Response: M	2225 (81%)	243 (73%)	2468 (80%)
Total: N (=H+I)	2743 (100%)	331 (100%)	3074 (100%)

Table 44: Wave 3 Individual outcomes for Wales extension sample (BHPS wave 11)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3873	122	3995
Retiring sample: B	8	-	8
Not issued – inaccessible: C	193	-	193
Issued sample all: D (=E+F)	3672	122	3794
Issued sample adults 16+: E (=G+H)	2851	111	2962
Field outcomes for issued sample:			
Child <16 not eligible interview: F	821	11	832
Adult not eligible for interview: G	18	41	59
Assumed eligible for interview: H	2833	70	2903
Extra eligible for interview: I		119	119
Field outcomes for eligible sample:			
Refusal: J	361 (13%)	40 (21%)	401 (13%)
Non-contact: K	75 (3%)	8 (4%)	83 (3%)
Other NR: L	83 (3%)	6 (3%)	89 (3%)
Response: M	2314 (82%)	135 (71%)	2449 (81%)
Total: N (=H+I)	2833 (100%)	189 (100%)	3022 (100%)

Table 45: Wave 4 Individual outcomes for Wales extension sample (BHPS wave 12)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3890	198	4088
Retiring sample: B	58	-	58
Not issued – inaccessible: C	287	6	293
Issued sample all: D (=E+F)	3545	192	3737
Issued sample adults 16+: E (=G+H)	2782	178	2960
Field outcomes for issued sample:			
Child <16 not eligible interview: F	763	14	777
Adult not eligible for interview: G	44	75	119
Assumed eligible for interview: H	2738	103	2841
Extra eligible for interview: I		105	105
Field outcomes for eligible sample:			
Refusal: J	322 (12%)	45 (22%)	367 (13%)
Non-contact: K	70 (3%)	29 (14%)	99 (3%)
Other NR: L	84 (3%)	2 (1%)	86 (3%)
Response: M	2262 (83%)	132 (64%)	2394 (81%)
Total: N (=H+I)	2738 (100%)	208 (100%)	2946 (100%)

Table 46: Wave 5 Individual outcomes for Wales extension sample (BHPS wave 13)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	3838	237	4075
Retiring sample: B	95	1	96
Not issued – inaccessible: C	369	8	377
Issued sample all: D (=E+F)	3374	228	3602
Issued sample adults 16+: E (=G+H)	2647	204	2851
Field outcomes for issued sample:			
Child <16 not eligible interview: F	727	24	751
Adult not eligible for interview: G	24	82	106
Assumed eligible for interview: H	2623	122	2745
Extra eligible for interview: I		80	80
Field outcomes for eligible sample:			
Refusal: J	293 (11%)	48 (24%)	341 (12%)
Non-contact: K	27 (1%)	13 (6%)	40 (1%)
Other NR: L	50 (2%)	4 (2%)	54 (2%)
Response: M	2253 (86%)	137 (68%)	2390 (85%)
Total: N (=H+I)	2623 (100%)	202 (100%)	2825 (100%)

Table 47: Wave 1 field outcomes for Northern Ireland sample (BHPS wave 11)

Wave One Household Outcomes	Number	%
Addresses issued		
Vacant/non-residential/not main residence		
Multi-households addition to sample		
Eligible households	2885	(100%)
Refusal to field agency/research centre	123	(4%)
Household refusal to interviewer	500	(17%)
Household non-contact	212	(8%)
Language/age/infirmity problems	71	(2%)
Complete household interview	1492	(52%)
Complete household coverage (inc. proxies)	1656	(57%)
Partial household coverage	1979	(69%)
Wave One Individual Outcomes		
Enumerated individuals	5188	(100%)
Ineligible children (Under 16)	1291	(25%)
Eligible adults	3897	(100%)
Refusals	253	(6%)
Non-contact / absent	186	(5%)
Age / infirmity / disability or language difficulty	-	-
Full interviews	3258	(84%)
Proxy interviews	200	(5%)
Total interviews	3458	(89%)

Table 48: Wave 2 Individual outcomes for Northern Ireland sample (BHPS wave 12)

	OSMs and PSMs		TSMs		ALL	
Total active sample: A (=B+C+D)	5188		-		5188	
Retiring sample: B	-		-		-	
Not issued – inaccessible: C	-		-		-	
Issued sample all: D (=E+F)	5188		-		5188	
Issued sample adults 16+: E (=G+H)	3983		-		3983	
Field outcomes for issued sample:						
Child <16 not eligible interview: F	1205		-		1205	
Adult not eligible for interview: G	110		-		110	
Assumed eligible for interview: H	3873		-		3873	
Extra eligible for interview: I	-		98		98	
Field outcomes for eligible sample:						
Refusal: J	687	(18%)	23	(24%)	710	(18%)
Non-contact: K	146	(4%)	14	(14%)	160	(4%)
Other NR: L	123	(3%)	-	-	123	(3%)
Response: M	2917	(75%)	61	(62%)	2978	(75%)
Total: N (=H+I)	3873	(100%)	98	(100%)	3971	(100%)

Table 49: Wave 3 Individual outcomes for Northern Ireland sample (BHPS wave 13)

	OSMs and PSMs	TSMs	ALL
Total active sample: A (=B+C+D)	5205	148	5353
Retiring sample: B	133	-	133
Not issued – inaccessible: C	204	-	204
Issued sample all: D (=E+F)	4868	148	5016
Issued sample adults 16+: E (=G+H)	3828	104	3932
Field outcomes for issued sample:			
Child <16 not eligible interview: F	1040	44	1084
Adult not eligible for interview: G	397	26	423
Assumed eligible for interview: H	3431	78	3509
Extra eligible for interview: I	-	100	100
Field outcomes for eligible sample:			
Refusal: J	555 (16%)	43 (24%)	598 (17%)
Non-contact: K	152 (4%)	20 (11%)	172 (5%)
Other NR: L	3 (0%)	-	3 (0%)
Response: M	2721 (79%)	15 (65%)	2836 (79%)
Total: N (=H+I)	3431 (100%)	178 (100%)	3609 (100%)

Table 50: Mean levels of item non-response - individual questionnaire

Wave:	1	2	3	4	5	6	7	8	9	10	11	12	13
All	2.60	1.98	1.31	1.92	1.34	1.65	1.22	1.66	1.43	1.57	1.44	1.93	2.49
All n>100	2.59	1.97	1.30	1.90	1.31	1.63	1.22	1.63	1.40	1.51	1.42	1.89	2.46

Note: all substantive variables included; interviewer checks, derived variables and system variables are excluded

Table 51: Mean levels of item non-response - household questionnaire

Wave:	1	2	3	4	5	6	7*	8	9	10	11	12	13
All	2.44	1.81	2.50	2.24	2.28	2.30	5.61	5.69	5.50	5.75	2.20	2.06	1.83
All n>100	2.44	1.78	2.46	2.23	2.24	2.28	5.60	5.67	5.50	5.73	2.20	2.03	1.81

Note: all substantive variables included; interviewer checks, derived variables and system variables are excluded

* At wave 7, respondents reporting the purchase of white goods in the past year were asked for the actual amount spent on each item purchased for the first time. The increase in overall item non-response levels from wave 7 appears to be due to this change in the data collected.

Table 52: Item non-response levels for selected variables

	Var name	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	Sample
Marital status	XMLSTAT	0.15	0.01	0.04	0.06	0.04	0.04	0.02	0.05	0.02	0.13	0.00	0.03	0.00	all respondents ~9000
Health status	XHLSTAT	0.08	0.07	0.05	0.08	0.08	0.03	0.04	0.09	--	0.08	0.04	0.01	0.03	all respondents ~9000
SOC code	XJBSOC	0.05	0.07	0.03	0.02	0.00	0.02	0.02	0.02	0.02	0.00	0.03	0.00	0.02	in employment ~5700
SIC code	XJBSIC	0.75	0.10	0.10	0.11	0.04	0.10	0.10	0.05	0.84	0.57	0.12	--	--	in employment ~5700
gross pay at last payment	XPAYGL	11.7	9.79	8.85	8.77	8.76	9.66	8.96	8.82	9.13	9.86	9.35	8.68	10.71	employees ~4800
net pay at last payment	XPAYNL	2.77	2.54	2.64	2.41	2.17	2.90	2.80	2.61	3.11	--	3.64	3.99	4.19	employees ~4800
usual pay at last payment	XPAYU	15.1	8.35	10.3	10.7	9.33	9.10	8.66	8.98	6.91	8.53	7.29	7.66	11.03	employees ~800
day started current job	XJBBGD	41.8	47.2	44.7	46.3	47.2	47.43	47.45	45.47	46.84	49.49	51.89	51.06	51.08	employees ~4800
Month started current job	XJBBGM	21.7	4.46	3.60	6.11	7.03	5.69	6.32	6.19	7.14	7.54	8.37	8.05	8.10	employees ~4800
year started current job	XJBBGY	3.11	0.23	0.69	0.15	0.26	0.31	0.18	0.20	0.14	0.06	0.39	0.15	0.15	employees ~4800
s/e – net profit	XJSPRF	39.1	35.7	32.8	33.7	33.6	36.62	41.44	43.68	47.16	46.58	45.37	46.78	44.15	self-employed with accounts ~500
s/e – day start present job	XJSBGD	58.8	65.4	62.1	60.2	65.4	64.96	61.86	61.47	63.59	66.19	69.62	66.88	67.09	self-employed ~700
s/e – month start present job	XJSBGM	16.0	10.7	6.95	16.6	17.8	18.84	17.97	18.14	17.51	21.05	21.84	22.94	23.11	self-employed ~700
s/e – year start present job	XJSBGY	0.13	0.14	0.14	0.86	0.59	1.25	0.88	1.35	0.77	0.64	0.16	1.29	0.16	self-employed ~700
Support a particular political party	xVOTE1	0.17	0.07	0.11	0.19	0.14	0.15	0.13	0.10	0.17	0.21	0.26	0.16	0.34	all ~8800
Party would vote for tomorrow	xVOTE3	24.4	--	22.6	23.5	25.8	26.83	20.04	24.54	27.93	27.06	23.20	25.97	26.88	non-identifiers ~3000
Party closest to	xVOTE4	2.07	0.88	1.45	1.53	0.92	0.62	0.58	0.59	0.65	0.64	0.70	0.66	0.83	identifiers ~6000
Amount saved each month	XSAVED	13.8	13.7	11.8	10.8	9.77	10.38	8.78	8.36	8.52	10.60	10.61	11.70	11.59	savers ~3700
Amount borrowed at house purchase	XMGOLD	8.02	17.7	12.5	10.7	10.9	12.81	9.29	7.74	10.45	10.49	9.97	9.27	8.86	house buyers ~300
value of house (owners)	XHSVAL	8.20	44*	8.08	6.36	6.36	5.78	4.90	5.01	5.65	5.15	4.77	5.89	1.28	home owners ~3500
net amount of last rent payment	XRENT	7.12	4.95	5.95	5.44	4.88	5.08	5.08	6.18	7.09	6.59	8.84	16.04	3.74	renters ~1450
gross rent (inc. housing benefit)	XRENTG	17.4	14.6	15.6	15.9	15.4	14.53	14.70	14.02	16.95	15.50	27.86	32.88	9.42	renters ~650

* denotes n<100

Table 53: Item non-response; wealth, assets and debts questions; core BHPS sample

	Saves from current income %	Amount savings current income %	Has non-regular savings %	Amount non-reg savings %	Has investments %	Amount Invested %	Has debts %	Amount owed %
<i>Wave 5</i>								
Missing	0.1		1.0		0.5		0.1	
Refused	0.2		0.1		0.2		0.3	
Don't know	--		0.9		0.3		0.1	
Yes	40.8		50.2		39.9		38.9	
No	58.9		47.8		59.0		60.6	
<i>N cases</i>	8649		5139		8628		8827	
Pre-bands								
Missing		0.7		2.5		2.3		0.3
Refused		6.9		8.3		7.4		2.0
Don't know		13.9		17.0		22.2		5.3
Actual amount		78.5		72.2		68.2		92.5
Post-bands								
Miss/ref/dk		8.7		12.8		12.9		3.1
Actual amount or range		91.2		87.2		87.1		96.9
<i>N valid cases</i>		3501		2629		3490		3434
<i>Wave 10</i>								
Missing	0.0		0.1		0.1		0.0	
Refused	1.0		0.0		0.0		0.5	
Don't know	--		0.0		0.0		0.2	
Yes	42.6		70.7		39.2		39.7	
No	56.4		29.1		60.7		59.6	
<i>N cases</i>	8701		8544		8544		8701	
Pre-bands								
Missing		--		0.0		0.0		0.0
Refused		0.0		8.3		1.1		2.9
Don't know		10.6		14.4		20.9		6.1
Actual amount		89.4		77.3		78.0		91.0
Post-bands								
Miss/ref/dk				11.2		5.8		4.0
Actual amount or range				88.8		94.2		96.0
<i>N valid cases</i>		3709		6044		3088		3456

Note: Only respondents giving a 'don't know' response at the actual amount question were routed through the banded ranges. Respondents who refused at any of the pre-banded questions were routed out of that section. At wave 10, respondents who did not know how much they saved per month from their current income were not asked the banded question.

Table 54: Item non-response; wealth, assets and debts questions; Scottish boost sample

<u>Wave 10 (2nd wave for this sample)</u>	Saves from current income %	Amount savings current income %	Has non- regular savings %	Amount non-reg savings %	Has investments %	Amount Invested %	Has debts %	Amount owed %
Missing	0.0		0.0		0.0		0.0	
Refused	1.4		0.0		0.0		0.6	
Don't know	--		0.0		0.0		0.2	
Yes	43.4		66.2		31.0		40.8	
No	55.2		33.8		69.0		58.5	
<i>N cases</i>	<i>2514</i>		<i>2457</i>		<i>2457</i>		<i>2514</i>	
<i>Pre-bands</i>								
Missing		--		0.0		0.0		0.0
Refused		0.0		19.0		2.9		5.1
Don't know		16.0		14.4		26.4		6.6
Actual amount		84.0		66.6		70.7		88.3
<i>Post-bands</i>								
Miss/ref/dk				21.5		8.6		6.5
Actual amount or range				78.4		91.4		93.6
<i>N valid cases</i>		<i>1092</i>		<i>1626</i>		<i>652</i>		<i>1025</i>

Table 55: Item non-response; wealth, assets and debts questions; Welsh boost sample

<u>Wave 10 (2nd wave for this sample)</u>	Saves from current income %	Amount savings current income %	Has non-regular savings %	Amount non-reg savings %	Has investments %	Amount Invested %	Has debts %	Amount owed %
Missing	0.0		0.0		0.0		0.0	
Refused	3.3		0.0		0.0		1.3	
Don't know	--		0.0		0.0		0.1	
Yes	40.3		64.9		33.9		38.3	
No	56.4		35.1		66.1		60.3	
<i>N cases</i>	<i>2374</i>		<i>2275</i>		<i>2275</i>		<i>2374</i>	
<i>Pre-bands</i>								
Missing		--		0.0		0.0		0.0
Refused		0.0		16.0		2.7		5.2
Don't know		20.3		18.8		25.6		9.1
Actual amount		79.7		65.1		71.7		85.7
<i>Post-bands</i>								
Miss/ref/dk				20.7		9.4		6.2
Actual amount or range				79.4		90.7		93.8
<i>N valid cases</i>		<i>957</i>		<i>1477</i>		<i>676</i>		<i>909</i>

Table 56: Verbatim items coded at each wave

Coversheet and Household Grid	Var Name	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13
Relationship to reference person	HGR2R	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual refusal reason	IVRREF		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Household refusal reason	IVRREFH			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Individual questionnaire														
Place born UK*	PLBORND	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Place born non UK*	PLBORNC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Citizenship (main)*	CITZN1							✓	✓	✓	✓	✓	✓	✓
Citizenship* (secondary)	CITZN2							✓	✓	✓	✓	✓	✓	✓
Advantages cohabitation	COHAD1 COHAD2								✓					✓
Disadvantages cohabitation	COHDS1 COHDS2								✓					✓
Reason wanting to move	LKMOVY	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Other reasons for having moved	MOVY1 MOVY2		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Name of college/university/polytechnic (1 st degree)	UNIB												✓	
Name of college/university/polytechnic (2 nd degree)	UNIM												✓	
Reasons for not going on to further education	FEDNT1 FEDNT2												✓	✓
Job would like once finish education	OCFUT												✓	✓
Baby's weight in grams	BWTGM1									✓	✓	✓	✓	✓
SOC father's job*	PASOC	✓							✓	✓	✓	✓	✓	✓
SOC mother's job*	MASOC	✓							✓	✓	✓	✓	✓	✓
SOC own first job*	J1SOC								✓	✓	✓	✓	✓	✓
SOC current main job	JBSOC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SOC job wanted	JUSOC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SOC second job	J2SOC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SOC all jobs in last year	JHSOC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SOC last job (non-employed)	JLSOC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SOC all jobs over lifetime	LJSOC			✓										
SIC current main job	JBSIC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SIC all jobs in last year	JSBLY	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SIC last job (non-employed)	JLSIC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SIC all jobs over lifetime	LJSIC			✓										
Reason for job change		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Why attracted to current job	JBLKY	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Reason child went to school outside catchment area	SCHOS1 SCHOS2												✓	
Subjective social class	OPCLS2	✓					✓				✓			
Reason/better/ worse off	FISITY			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Reasons for saving	SAVEY1 SAVEY2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Annual earnings	JHPAYS	✓												
Who transfer money to	FTEX (A,B,C,)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Important events	EVENT 1 EVENT1S EVENT2 EVENT2S EVENT3 EVENT3S EVENT4 EVENT4S		✓	✓	✓	✓				✓		✓ LIS LIW NI only		
Generational change	DFWLD1 DFWLD2 DFWLD3 DFWLD4						✓				✓			
Quality of life	QALLIF1 QALLIF2 QALLIF 3 QALLIF4							✓					✓	
Views on neighbourhood	NEIGH1 NEIGH2 NEIGH3 NEIGH4 NEIGH5 NEIGH6								✓					✓
Advantages and disadvantages of being <age>	AGEAD1 AGEAD2 AGEAD3 AGEAD4											✓		
Self-completion														
SOC friend's job	NETSOC			✓		✓		✓		✓	✓	✓	✓	
Relationship to friend/relative	NET1WR NET2WR NET3WR			✓		✓		✓		✓	✓	✓	✓	
Relationship to person respondent can share private feelings with	SSUPR2R													✓
Proxy questionnaire														
Relationship to proxy informant	PRRS21	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SOC current main job	JBSOC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SIC current main job	JBSIC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Youth Questionnaire														
SOC father's job	YPPASC				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SOC future job preferred	DYPSOC				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Main change wants in life	YDDLFA YDDLFB				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Telephone questionnaire														
SOC current main job	JBSOC			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
SIC current main job	JBSIC			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

* not asked if information already collected at earlier wave

Table 57: Validation re-coding results for selected items

	Relationship to HRP			SOC main job			SIC main job			Reason wants move			Reason for saving		
	N	N Diff	% Diff	N	N Diff	% Diff	N	N Diff	% Diff	N	N Diff	% Diff	N	N Diff	% Diff
Wave 1	1201	20	1.6	304	54	17	304	27	9	402	63	16	299	21	7
Wave 2	1756	23	1	152	23	15	299	25	8	245	47	19	258	14	5
Wave 3	2109	10	0.5	268	8	3	268	19	7	263	10	4	270	0	0
Wave 4	2161	12	0.5	261	4	1.6	261	16	6.1	273	11	4	272	2	0.7
Wave 5	2471	22	0.9	295	11	3.7	295	14	4.7	287	28	9.7	297	8	2.7
Wave 6	982	20	2	291	0	0	291	20	6.9	285	35	9.7	291	13	4.5
Wave 7	2214	21	0.9	289	14	4.8	289	19	6.6	298	14	4.7	300	4	1.3
Wave 8	2411	9	0.4	308	9	2.9	308	13	4.2	--	--	--	311	3	0.9
Wave 9*	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
Wave 10	--	--	--	422	80	18.9	--	--	--	258	58	22.5	250	21	8.4
Wave 11	--	--	--	442	101	22.8	442	103	23.3	233	63	27.0	259	24	9.3
Wave 12	--	--	--	395	82	20.8	395	83	21.0	197	49	25.0	279	30	10.8
Wave 13	--	--	--	397	84	21.2	397	85	21.4	197	49	25.0	279	30	10.8

* The re-coding exercise was not undertaken on wave 9

Table 58: Some distribution statistics for weights

Cross-sectional weights (wXRWGHT)					
Wave	n	min	max	mean	St. dev.
1	9912	.202	2.499	1.000	.251
2	9549	.088	2.500	1.000	.338
3	9021	.066	2.500	1.000	.386
4	9057	.070	2.500	1.000	.390
5	8816	.062	2.500	1.000	.423
6	9117	.062	2.500	1.000	.418
7	9088	.066	2.500	1.000	.426
8	8894	.081	2.500	1.000	.403
9	8756	.100	2.500	1.000	.419
10	8626	.086	2.500	1.000	.451
11	8518	.089	2.500	1.000	.439
12	8295	.118	2.500	1.000	.437
13	8162	.111	2.500	1.000	.421

Longitudinal weights (wLRWGHT)					
Wave	n	min	max	mean	St. dev.
2	8721	.201	2.500	1.000	.297
3	7905	.181	2.500	1.000	.321
4	7525	.180	2.500	1.000	.329
5	7169	.183	2.500	1.000	.335
6	7059	.182	2.500	1.000	.339
7	6898	.179	2.500	1.000	.346
8	6719	.182	2.500	1.000	.348
9	6533	.165	2.500	1.000	.354
10	6304	.193	2.500	1.000	.361
11	6042	.200	2.500	1.000	.369
12	5812	.202	2.500	1.000	.375
13	5624	.217	2.500	1.000	.381

Table 59: Analysis of BHPS Usage from Data Archive Records

	Total Data orders by distinct users	Orders by new users	Orders by existing users	Cumulative number of users	Users by wave of last order
Wave 1	77	77	0	77	40
Wave 2	101	73	28	150	43
Wave 3	155	110	45	260	54
Wave 4	164	86	78	346	60
Wave 5	231	124	107	470	133
Wave 6	182	110	72	580	111
Wave 7	186	117	69	697	110
Wave 8	227	140	87	837	123
Wave 9	225	137	88	974	110
Wave 10	392	258	134	1232	214
Wave 11	510	276	234	1508	510
Wave 12	387*				

Notes: The Data Archive user registration system changed in mid-2004, approximately coinciding with the release of the wave 12 data. Users cannot currently be matched between the old and new systems, making it impossible to reproduce columns 2 to 5 from wave 12 onwards. We hope to rectify this in due course. * The data orders for wave 12 relate to a period of approximately 7 months since release.

Table 60: Published papers, books and reports based on BHPS

	No. of publications
To 1994	35
1995	37
1996	44
1997	70
1998	84
1999	106
2000	129
2001	138
2002	128
2003	124
2004 and 2005 (partial)	147
Total	1042

Note: These are publications known to ISER and logged in our data base of publications. Under-coverage is likely.

Table 61: BHPS Registered Users by discipline

	Economics	Sociology	Statistics, Methodology	Health	Demography	Geography	Political Science	Teaching	Total
	%	%	%	%	%	%	%	%	%
Prof.	14.5	7.9	7.9	12.8	16.7	20.0	42.9	0.0	13.4
Dr.	18.8	30.7	34.2	25.6	27.8	13.3	14.3	33.3	22.4
RA/Student	66.7	61.4	57.9	61.5	55.6	66.7	42.9	66.7	64.2
N	415	127	38	39	18	15	9	14	674
Row %	61.6	18.8	5.6	5.8	2.7	2.2	2.1	1.3	

Note: Incomplete data (not all users indicated a discipline) and to 2001 only.

Table 62: Complex Standard Errors, DEFTs and 95% Confidence Intervals for a Range of BHPS Variables

Variable	Proportion	DEFT	Complex se	95% C.I.
Housing Tenure (ATENURE)				
Owner Occupier	66.20	1.46	.93	64.37 - 68.03
LA Rented	20.41	1.57	.85	18.74 - 22.08
Private Rented	9.43	1.65	.65	8.16 - 10.70
Car Available for Private Use (ANCARS)				
None	31.09	1.36	.85	29.42 - 32.75
1	45.94	1.09	.73	44.51 - 47.37
2	19.43	1.26	.67	18.12 - 20.75
Household Type (AHHTYPE)				
Single Non-Elderly	11.61	1.49	.61	10.41 - 12.81
Single Elderly	14.65	1.15	.55	13.57 - 15.73
Couple	17.41	1.11	.57	16.29 - 18.52
Elderly Couple	9.27	1.12	.44	8.41 - 10.12
Couple and Children	28.85	1.12	.67	25.54 - 28.15
Political Party Supported (AVOTE)				
Conservative	39.34	1.57	.81	37.79 - 40.91
Labour	36.04	1.66	.84	34.39 - 37.68
Lib Dem/SDP	11.57	1.36	.46	10.67 - 12.47
Current Employment Status (AHGEST)				
Working	58.30	1.37	.68	56.97 - 59.62
Unemployed	6.31	1.26	.31	5.70 - 6.91
Retired	17.80	1.42	.55	16.72 - 18.87
Family Care	11.17	1.21	.38	10.41 - 11.92
Receiving Unemployment Benefit (AF131)	2.58	1.25	.20	2.19 - 2.96
Receiving Housing Benefit (AF139)	8.15	1.55	.43	7.31 - 8.99
Highest Government Priority (AOPPOL4)				
Living Standards	50.54	1.39	.72	49.14 - 51.95
Protect Environment	46.67	1.34	.69	45.32 - 48.02
School Leaving Age (ASCEND & ASCHOOL)				
Less than or equal to 16	76.47	1.66	.71	75.09 - 77.86
Goldthorpe Social Class (AJBGOLD)				
Service class	32.54	1.38	.84	30.89 - 34.19
Routine Non-Manual	13.75	1.18	.53	12.72 - 14.78
Married Female Employed (Derived)	56.26	1.11	1.01	54.29 - 58.23
Employee Union Member (Derived)	37.45	1.25	.84	35.81 - 39.10
Current Job Spell Began Before 1.9.90	24.07	1.25	.54	23.01 - 25.12
Health Limits Daily Activities (AHLT)	13.18	1.14	.39	12.42 - 13.94
All Health Care Should Be Free (AOPHLA)				
Agree	46.74	1.63	.81	45.15 - 48.33
Disagree	10.20	1.06	.32	9.57 - 10.84
Respondent Smokes (ASMOKER)	29.87	1.37	.63	28.63 - 31.10
Respondent Had In-Patient Stay Since 1.9.90 (AHOSP)	11.50	1.05	.34	10.84 - 12.16

Table 63: Actual and ideal weights for OSM(Ds)

<i>Actual weight</i>	At wave t			
At wave 1	2 present, both OSM	2 present, 1 OSM	1 present, OSM	1 present, not OSM
Both OSEs, 2 hhds	X	0.5	1	0
Both OSEs, 1 hhd	1	X	1	X
One OSE	X	0.5	1	0
<i>Ideal(design) weight</i>				
Both OSEs, 2 hhds	X	0.5	0.5	0.5
Both OSEs, 1 hhd	1	X	1	X
One OSE	X	1	1	1

Table 64: Wave 1 household response status by region (row percentages)

Region	Response	Refusal	Language/ Infirmity	Non-contact
Inner London	60.0	23.8	3.4	12.9
Outer London	64.9	27.5	1.3	6.3
R. of South East	73.2	21.1	2.0	3.7
South West	78.0	17.9	1.4	2.7
East Anglia	79.1	17.5	1.5	1.9
East Midlands	77.1	19.8	0.9	2.1
West Midlands Conurbation	70.4	20.8	3.8	5.0
R. of West Midlands	79.5	17.2	2.4	0.9
Greater Manchester	69.3	23.4	1.7	5.6
Merseyside	74.0	21.5	1.1	3.4
R. of North West	75.5	20.2	0.9	3.4
South Yorkshire	82.5	15.8	0.5	1.1
West Yorkshire	75.7	20.2	1.1	2.9
R. of Yorks & Humberside	76.9	19.7	1.7	1.7
Tyne & Wear	78.8	17.4	1.6	2.2
R. of North	81.8	15.5	1.1	1.5
Wales	77.0	19.7	1.1	2.2
Scotland	73.4	21.5	0.8	4.3
All	73.9	20.6	1.6	3.8

Table 65: Wave 1 household response status by housing type (row percentages)

Region	Response	Refusal	Language/ Infirmity	Non-contact	Base
Detached house/bungalow	79.5	17.8	0.8	1.9	1,324
Semi-detached house/bungalow	80.0	16.7	1.3	1.9	2,172
End terrace house	76.2	20.6	1.0	2.2	505
Terraced house	77.3	18.7	1.2	2.8	1,393
Purpose-built flat/maisonette	74.5	17.7	3.2	4.7	1,075
Converted flat/maisonette	74.9	16.2	1.1	7.7	271
Dwelling with business premises	81.0	14.3	3.2	1.6	63
Bedsit/other	79.5	12.0	2.6	6.0	117
Missing data	20.2	59.4	3.5	16.9	544
All	73.8	20.7	1.6	3.9	7,464

Table 66: Wave 1 response to individual interview; responding households

Outcome	n	%
Full interview	9912	92.2
Proxy interview	352	3.3
Refusal	425	4.0
Absent	12	0.1
Other non-contact	36	0.3
Other non-interview	8	0.1
Total	10,745	100.0

Table 67: Summary outcomes for wave 1 respondents

	% ineligible	% of eligible responding	% of eligible responding at all waves	Wave on wave response rate (interviewed all waves up to previous)
Wave 2	1.4	87.7	87.7	87.7
Wave 3	2.9	81.5	79.1	90.3
Wave 4	4.3	79.9	74.8	94.9
Wave 5	5.6	76.8	70.6	94.8
Wave 6	6.9	77.3	68.7	97.6
Wave 7	8.4	76.0	66.7	97.6
Wave 8	9.5	74.1	64.7	97.4
Wave 9	10.5	72.1	62.4	97.0
Wave 10	12.0	70.4	60.0	96.7
Wave 11	12.8	68.4	59.3	96.1
Wave 12	13.7	66.6	57.1	96.5
Wave 13	14.8	64.9	55.1	96.8

Table 68: Summary outcomes for persons enumerated in wave 1 households

	% ineligible	% of eligible enumerated	% of eligible enumerated at all waves	Wave on wave enumeration rate (enumerated all waves up to previous)
Wave 2	1.3	89.6	89.6	89.6
Wave 3	2.5	86.9	84.3	94.2
Wave 4	3.8	83.5	80.4	95.6
Wave 5	4.8	80.1	76.4	95.2
Wave 6	5.9	79.1	74.1	97.2
Wave 7	7.0	77.3	72.0	97.5
Wave 8	7.9	75.8	70.1	97.6
Wave 9	8.8	73.6	67.8	97.1
Wave 10	10.0	72.0	65.5	97.1
Wave 11	10.8	70.5	64.1	96.7
Wave 12	11.6	69.0	62.2	97.3
Wave 13	12.5	66.4	59.7	96.3

Table 69: Impacts of attrition I

	Total at Wave 1	By Wave 5			By Wave 13		
		Still in Sample	Respond all waves	Ever non- response	Still in Sample	Respond all waves	Ever non- response
<u>Sex</u>							
Male	47.7	47.6	46.2	50.6	47.6	45.5	50.1
Female	52.3	52.4	53.8	49.4	52.4	54.5	49.9
<u>Age</u>							
16-24	15.9	16.4	14.6	20.4	17.5	14.1	21.5
25-34	19.1	19.7	20.0	18.9	21.5	21.7	21.2
35-44	17.5	18.1	19.1	15.8	19.5	20.5	18.3
45-54	14.5	15.0	15.0	15.0	15.7	16.9	14.3
55-64	12.9	12.8	13.3	11.8	12.7	14.6	10.5
65-74	12.1	11.3	11.9	10.0	9.4	10.1	8.5
75+	8.0	6.7	6.1	8.0	3.6	1.9	5.7
<u>Marital status</u>							
Married	58.3	58.9	61.0	54.0	60.1	64.3	55.1
Living as couple	6.3	6.4	6.4	6.3	6.7	6.7	6.7
Widowed	9.1	8.1	8.2	8.0	5.7	5.2	6.2
Divorced/Separated	5.7	5.7	5.8	5.3	5.6	5.7	5.5
Never married	20.7	20.9	18.6	26.3	21.9	18.1	26.5
<u>Current labour force status</u>							
Self employed	7.7	7.9	7.7	8.2	8.2	8.3	8.1
In paid employ	48.9	50.3	51.6	47.2	53.8	56.6	50.6
Unemployed	5.4	5.6	4.6	8.1	6.0	4.2	8.0
Retired	19.3	17.7	18.2	16.4	13.5	14.1	12.8
Family care	11.2	11.2	10.9	12.0	11.2	10.4	12.1
FT student	3.9	4.0	3.7	4.6	4.2	3.5	4.9
Long term sick/disabled	2.7	2.5	2.5	2.5	2.3	2.1	2.5
Other	0.8	0.8	0.7	1.1	0.9	0.8	0.9
<u>Socio economic group: present job</u>							
Not Applicable	41.4	39.7	38.6	42.4	35.8	33.2	38.8
Employers, Managers, Professional	12.3	12.5	13.4	10.5	13.2	14.6	11.6
Int. non-manual	9.8	10.1	10.9	8.2	10.7	12.4	8.8
Junior non-manual	11.6	12.0	12.8	10.4	12.9	14.5	11.1
Skilled manual workers	9.5	9.8	9.2	11.3	10.7	9.7	11.8
Semi- and unskilled manual workers	11.4	11.8	11.3	12.9	12.4	11.4	13.7

Table 70: Impacts of attrition II

	Total Wave 1	By Wave 5			By Wave 13		
		Still in Sample	Respond all waves	Ever non- response	Still in Sample	Respond all waves	Ever non- response
<u>Highest educational qualification</u>							
Degree	7.0	7.2	8.1	5.0	7.5	9.1	5.6
Other Higher QF	16.3	16.4	17.7	13.4	16.9	19.3	14.1
GCE A Levels	9.8	9.9	10.3	9.0	10.4	10.9	9.8
GCE O Level or equiv	20.6	21.0	21.1	20.7	22.3	22.2	22.5
Other QF	11.0	11.2	10.8	12.1	11.5	10.6	12.5
No QF	34.7	33.8	31.4	39.3	31.4	27.8	35.6
<u>Health over last 12 months</u>							
Excellent	28.2	28.7	29.2	27.7	30.0	30.9	28.9
Good	45.0	45.4	45.7	44.8	46.3	46.6	45.9
Fair	18.6	18.3	17.7	19.6	17.5	16.6	18.6
Poor	6.2	5.7	5.6	6.0	4.9	4.7	5.2
Very poor	2.0	1.7	1.7	1.7	1.3	1.2	1.4
<u>GHQ Score</u>							
Missing or wild	3.3	3.0	2.3	4.5	2.6	1.9	3.4
0	49.3	49.7	50.6	47.8	50.8	52.2	49.1
1	14.6	14.6	14.8	14.3	14.5	14.9	14.0
2	8.9	8.9	9.0	8.6	8.7	8.8	8.7
3	5.9	6.0	5.7	6.6	6.1	5.4	6.8
4	4.6	4.6	4.5	4.6	4.5	4.5	4.5
5	3.5	3.4	3.7	2.7	3.3	3.6	2.9
6+	9.9	9.8	9.4	10.8	9.6	8.7	10.6
<u>Number of organisations active in</u>							
None	53.7	53.3	51.4	57.8	52.7	49.6	56.5
1	30.7	30.9	31.2	30.3	31.4	32.1	30.6
2	10.8	10.8	11.8	8.6	11.1	12.3	9.5
3	4.7	4.8	5.6	3.0	4.8	6.0	3.3
<u>Number of people in household</u>							
1	13.7	12.7	13.1	11.6	10.3	10.2	10.4
2	33.4	32.9	33.7	31.1	31.7	33.1	30.0
3	20.4	20.9	20.4	22.1	21.9	21.8	21.9
4	20.7	21.4	20.9	22.3	23.0	22.5	23.6
5	8.9	9.2	9.1	9.3	10.0	9.5	10.5
6	2.9	3.0	2.8	3.6	3.2	2.8	3.6

Table 71: Impacts of attrition III

	Total at Wave 1	By Wave 5			By Wave 13		
		Still in Sample	Respond all waves	Ever non-response	Still in Sample	Respond all waves	Ever non-response
<u>Household Type</u>							
Single Non-Elderly	5.2	5.1	5.2	5.0	5.2	5.3	5.2
Single Elderly	8.5	7.5	7.9	6.6	5.1	4.9	5.2
Couple No Children	29.2	28.7	29.5	26.9	27.5	29.5	25.2
Couple: dep children	30.5	31.6	32.3	30.0	34.2	33.9	34.6
Couple: non-dep child	14.3	14.7	13.6	17.0	15.5	15.5	15.5
Lone par: dep children	3.6	3.7	3.8	3.7	3.9	3.5	4.5
Lone par: non-dep child	4.3	4.2	3.9	4.9	4.2	3.8	4.6
2+ Unrelated adults	2.6	2.5	1.9	3.8	2.5	1.9	3.2
Other Households	1.9	1.9	1.9	2.0	1.9	1.7	2.0
<u>Region / Metropolitan Area</u>							
Inner London	5.1	5.1	4.3	6.7	5.0	3.8	6.4
Outer London	7.7	7.6	7.5	7.9	7.6	7.7	7.4
Rest of South East	19.0	18.9	19.4	17.8	19.3	20.2	18.2
South West	8.6	8.7	9.0	8.0	8.7	9.1	8.2
East Anglia	3.6	3.7	4.1	2.6	3.6	4.5	2.6
East Midlands	7.3	7.4	7.9	6.0	7.3	7.9	6.5
West Midlands Con	4.2	4.2	3.5	5.9	4.4	3.6	5.3
Rest of West Midlands	4.6	4.7	4.8	4.4	4.7	4.7	4.6
Greater Manchester	4.6	4.6	4.4	5.0	4.4	4.1	4.7
Merseyside	2.4	2.4	2.1	2.9	2.5	2.4	2.5
Rest of North West	4.4	4.5	4.7	4.0	4.4	4.7	4.0
South Yorkshire	2.5	2.5	2.6	2.4	2.5	2.6	2.4
West Yorkshire	3.7	3.7	3.5	4.1	3.7	3.3	4.2
R. of Yorks & Hum'side	3.3	3.3	3.4	3.2	3.2	3.3	3.2
Tyne & Wear	2.6	2.5	2.5	2.5	2.5	2.7	2.2
Rest of North	3.5	3.4	3.7	2.9	3.4	3.5	3.3
Wales	5.0	5.0	5.0	5.1	4.9	4.8	5.0
Scotland	7.9	8.0	7.7	8.6	7.9	6.9	9.1
<u>Housing tenure</u>							
Owned Outright	22.8	22.3	23.1	20.5	20.6	22.1	18.8
Owned with mortgage	47.3	48.5	50.3	44.4	51.5	54.1	48.4
Local Authority rent	18.9	18.5	17.1	21.7	17.7	14.9	21.0
Housing Assoc. rent	3.0	2.9	2.7	3.3	2.7	2.3	3.2
Other rented	7.8	7.6	6.8	9.5	7.5	6.6	8.6
<u>Income quintile</u>							
Lowest	20.0	19.4	17.5	23.7	17.5	14.6	21.0
2	20.0	19.5	19.2	20.2	18.2	17.1	19.5
3	20.0	20.3	20.8	19.2	20.3	20.9	19.5
4	20.0	20.5	20.9	19.5	21.8	22.6	20.9
Highest	20.0	20.3	21.6	17.4	22.2	24.8	19.1

Table 72: Wave 1 – 2 comparison of SOC major groups

Wave 2 SOC	Wave 1 SOC Major Group								
	Man/ Admin %	Prof %	Assoc Prof %	Clerical %	Craft %	Personal/ Service %	Sales %	Plant op %	Other %
Man/ Admin	73.9	3.2	4.9	5.2	1.4	4.3	4.9	2.1	--
Prof	2	83.2	5.9	.7		.7	.6	--	--
Assoc prof	4.5	8.8	78.9	4	1.1	1.5	1.4	2.2	.4
Clerical	8.8	1.9	3.9	85.3	2.5	2.1	3.8	4.5	3.4
Craft	2.5	1.5	1.6	1.2	83.6	.3	1	8.6	2.6
Personal/ Service	1.2	.6	2.3	.3	.8	87.1	.9	1	7.9
Sales	5.9	.4	.9	1.7	--	1.1	85.5	1	1.3
Plant op	.8	--	1.6	1.3	7.7	.4	.5	77.9	3.1
Other	.4	.3	--	.4	2.1	2.7	2	2.8	81.4

Table 73: Wave 1 – 2 comparison of SIC major divisions

Wave 2 SIC	Wave 1 SIC major division									
	Agri %	Energy %	Mining %	Engineer %	Manufacture %	Construction %	Hotel/ Cater %	Transport %	Finance %	Other %
Agri	77	--	--	.3	.3	--	.9	--	.2	.2
Energy	--	89.8	.4	.3	--	1	--	.5	.7	--
Mining	--	--	80.1	3.3	1.8	--	.8	--	.3	.1
Engineer	--	1.3	9.5	80.3	3.1	7.6	1.5	.6	1.8	.6
Manufac	--	--	5.1	3.6	84.2	2.4	2.1	1.1	1.1	.3
Construc	2.2	--	1	1.7	1.1	79.9	.4	--	1.7	.8
Hotel	6.8	3.9	2	5.8	6.9	1.2	88.8	1.4	.9	1.2
Transport	2.5	1.4	--	.4	--	1.1	1.2	90.8	1.3	.1
Finance	2.3	1.3	--	2.4	.6	1.1	1.2	1	88.3	2
Other	9.2	2.3	.8	1.3	.9	5.6	2.3	4.1	2.4	94.1

Table 74: False negative rates in the report of income sources: ISMIE study

	Independent Interviewing (standard BHPS questions)	Proactive Dependent Interviewing	Reactive Dependent Interviewing
Retirement Pension	0.00	0.01	0.03
Incapacity Benefit	0.50	0.42	0.35
Income Support	0.11	0.16	0.08
Child Benefit	0.23	0.04	0.08
Family Credit	0.29	0.04	0.25
Housing Benefit	0.18	0.19	0.10

Note: Figures denote the proportion of recipient respondents (according to DWP administrative data) who failed to report receipt in the ISMIE interview. The ISMIE sample did not have the same characteristics as the BHPS sample, as it over-represented low-income households, but the interview was broadly similar to the BHPS interview.

Table 75: Sample distribution by age and sex, Census and BHPS

	Census 1991	BHPS 91	BHPS 92	BHPS 93	BHPS 94	Census 2001	BHPS 01
<u>Male</u>							
16-19	7.1	7.9	7.8	8.1	7.1	6.5	7.4
20-24	9.9	9.1	8.2	8.0	8.5	7.8	7.9
25-34	19.8	21.3	20.4	20.1	19.3	18.2	14.8
35-44	18.3	18.8	18.9	18.6	18.9	19.2	19.2
45-54	15.2	15.7	16.9	17.4	17.5	17.1	18.3
55-64	13.2	11.9	11.7	11.9	11.9	13.6	14.3
65-74	10.5	9.9	10.3	10.4	11.0	10.2	10.6
75+	6.0	5.4	5.7	5.6	5.8	7.3	7.6
<u>Female</u>							
16-19	6.4	6.0	5.8	6.0	6.5	5.8	5.8
20-24	9.3	8.6	8.1	7.6	6.8	7.3	7.5
25-34	18.6	20.4	20.6	20.4	20.0	17.5	14.1
35-44	17.1	18.3	17.9	18.0	18.0	18.1	19.4
45-54	14	15.0	15.9	16.5	17.2	16.1	17.6
55-64	12.8	11.6	11.5	11.4	11.2	12.9	14.5
65-74	11.9	11.8	11.7	11.7	11.8	10.8	10.2
75+	10.0	8.4	8.7	8.5	8.4	11.6	10.9

Table 76: Employment status, Census and BHPS, persons aged under 60

	Census 1991	BHPS 91	BHPS 94	Census 2001	BHPS 01
<u>Males</u>					
Self-employed	12.9	14.0	13.4	13.1	11.8
Employee	64.5	65.2	62.5	63.1	70.4
Unemployed	10.1	9.5	9.3	4.9	3.6
Student	6.1	6.1	8.5	8.6	6.9
Permanently sick	3.7	3.1	4.1	5.1	4.1
Retired	0.8	0.6	1.0	1.2	2.0
Other	2.0	1.2	0.9	3.9	1.2
<u>Females</u>					
Self-employed	3.7	4.8	4.2	4.9	4.8
FT employee	57.5	61.1	59.1	59.1	64.3
Unemployed	4.8	4.1	4.2	3.0	2.4
Student	6.2	4.5	8.3	9.1	7.4
Permanent sick	3.2	2.3	2.8	4.9	4.5
Retired	1	1.6	1.5	1.4	2.1
Other	23.7	21.6	19.8	17.7	14.5

Table 77: Socio-economic group by sex, all current employees and self-employed

	Census 1991	BHPS 91	BHPS 92	BHPS 93	BHPS 94
<u>Males</u>					
Employers and Managers	20.1	19.5	20.2	20.7	21.1
Professional	7.4	8.0	7.9	8.2	8.1
Intermediate Non-Manual	10.1	11.8	12.6	13.2	12.6
Junior Non-Manual	9.0	9.2	9.8	9.4	10.2
Skilled Manual	24.1	26.0	23.9	23.1	22.1
Semi-skilled Manual	14.0	12.9	12.5	12.7	13.1
Unskilled Manual	3.9	2.6	2.7	3.2	2.6
Own Account	10.1	9.5	10.1	9.4	10.2
Armed Forces	1.3	0.6	0.3	0.2	0.3
N		3,295	3,073	2,787	2,966
<u>Females</u>					
Employers and Managers	11.0	11.1	11.9	12.9	13.1
Professional	2.0	2.3	2.2	2.6	2.6
Intermediate Non-Manual	18.4	22.9	21.3	21.6	21.9
Junior Non-Manual	36.7	31.7	32.7	31.4	30.5
Skilled Manual	3.5	3.8	3.6	3.7	3.5
Semi-skilled Manual	18.2	18.5	19.3	19.2	19.1
Unskilled Manual	7.2	6.2	5.9	5.8	6.2
Own Account	3.0	3.4	3.1	2.8	3.1
Armed Forces	0.1	0.0	0.0	0.0	0.0
N		2,854	2,730	2,716	2,740

Table 78: Ethnic group, Census and BHPS, aged 16 and over

	Census 1991	BHPS 91	BHPS 94	Census 2001	BHPS 01 ¹
White	95.4	96.1	96.4	90.9	96.1
Black	1.4	1.4	1.1	2.3	0.8
Indian	1.3	1.1	1.2	2.1	1.1
Pakistani	0.6	0.4	0.4	1.4	0.6
Bangladeshi	0.2	0.1	0.1	0.6	0.1
Chinese	0.3	0.1	0.1	0.4	0.1
Other Asian	n/a	n/a	n/a	0.5	0.2
Mixed	n/a	n/a	n/a	1.3	0.4
Other	0.8	0.8	0.7	0.4	0.6

¹ The ethnicity question used in the 2001 Census differed from that used in the 1991 Census, with more explicit categories, including mixed race categories for the first time. This new question was asked on BHPS for the first time at wave 13 (2003). The proportions for BHPS 2001 are based on wave 13 responses for wave 11 respondents. Where the data were missing, wave 11 responses (the old version of ethnicity) were used.

Table 79: Marital status, BHPS and GHS, aged 16 and over

	1991		1994	
	GHS	BHPS	GHS	BHPS
<u>Men</u>				
Single	23.9	24.8	23.6	25.7
married	62.2	60.7	60.7	57.6
cohabit	5.5	6.8	6.8	8.6
widowed	4.1	3.3	4.1	3.1
divorced	3.2	3.1	3.6	3.5
separated	1.1	1.2	1.3	1.2
<u>Women</u>				
single	17.2	17.3	18.0	19.6
married	56.8	56.3	54.1	53.2
cohabit	5.0	6.2	6.4	7.7
widowed	13.5	12.6	13.2	11.8
divorced	5.5	5.1	5.8	5.9
separated	2.0	2.3	2.5	1.8

Table 80: Household characteristics, GHS and BHPS, 1991 and 1994/5

	1991		1994/5	
	GHS	BHPS	GHS	BHPS
Consumer durables				
Colour TV	94.8	94.7	97	96.5
VCR	68.2	69.4	77	76.6
Freezer	83.1	83.4	88	87.7
Washing machine	87.0	86.5	89	88.5
Tumble Drier	48.2	47.7	50.	51.0
Dishwasher	14.2	14.8	19	20.3
Microwave	54.8	55.1	67	69.5
CD player	26.6	27.1	47	48.7
Home computer	20.9	21.5	24	27.0
Number of cars				
0	32.4	30.7	31	27.8
1	44.1	46.2	45	46.1
2	19.1	19.5	20	21.8
3 or more	4.3	3.6	4	4.2
Tenure				
Owned Outright	24.7	23.3	25	24.2
Owned with mortgage	41.7	43.6	42	45.0
Rented from Local Authority etc.	23.5	20.6	20	17.0
Housing Association	2.8	3.6	4	4.1
Rented from employer	1.4	1.0	1	1.2
Other rent	5.7	8.0	7	8.4
Household type				
1 adult 16-59	10.0	10.2	12	11.4
2 adult 16-59	15.7	16.7	16	17.0
Youngest child 0-4	13.8	14.3	14	13.8
Youngest child 5-15	15.9	16.7	17	16.9
3+ adult (no children)	12.1	12.2	10	12.8
2 adults, at least 1 aged 60+	16.1	14.5	16	13.4
1 adult aged 60+	16.6	15.5	16	14.6
Base N	9,955	5,512	9,668	5,039

Figure 1: BHPS tracking procedures

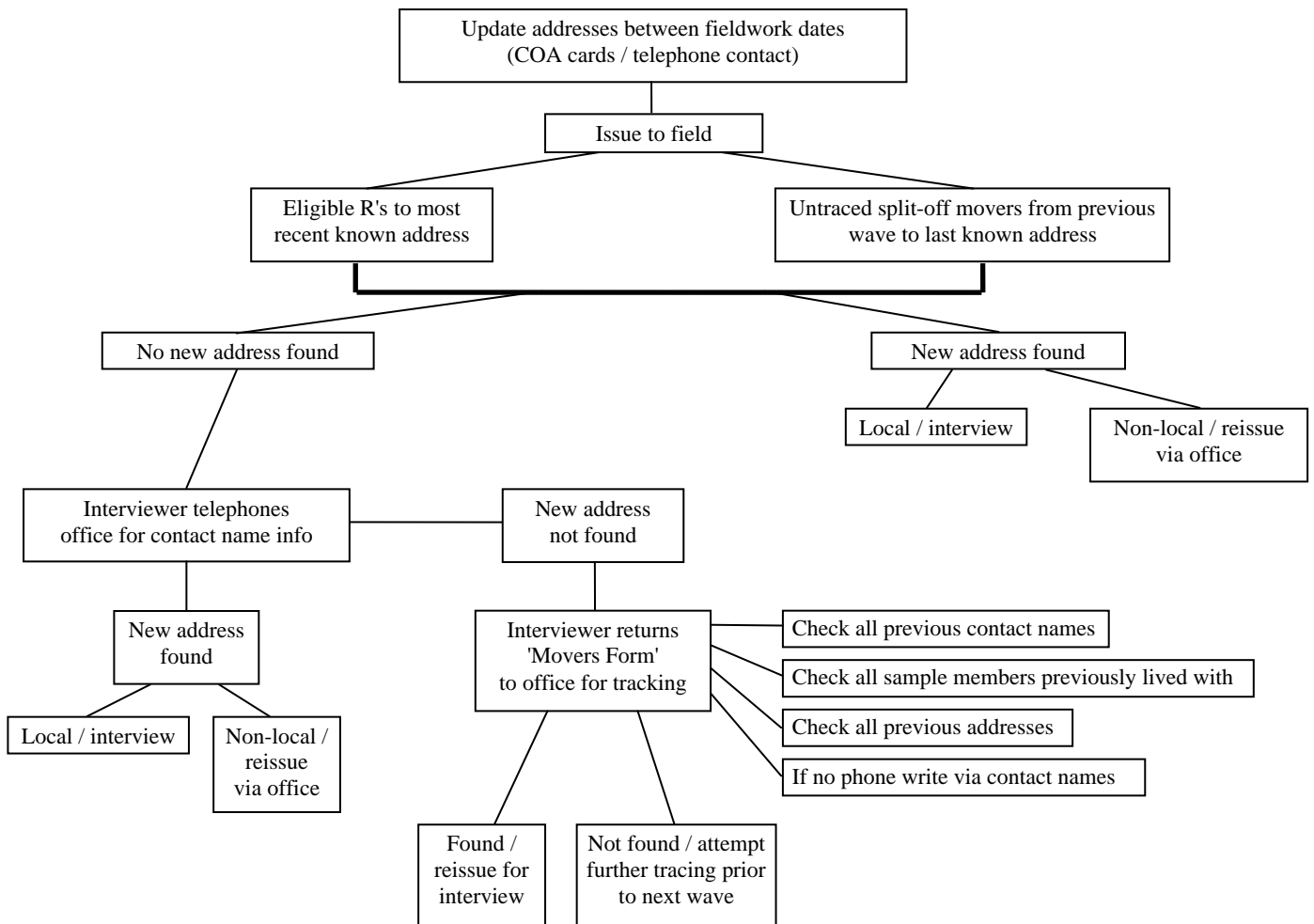


Figure 2: Measurement error regarding unemployment – from Elias (1997)

Table: Summary of Comparison of Unemployment Recorded in BHPS and FWL with LFS unemployment, 1984-92, by age groups and sex

	Males		Females	
	Up to 3 years earlier	More than 3 years earlier	Up to 3 years earlier	More than 3 years earlier
16-24 years	LFS and BHPS agree; FWL <i>underreports</i> unemployment	BHPS <i>severely underreports</i> ; FWL <i>underreports</i> unemployment	LFS, BHPS and FWL in reasonable agreement; BHPS and FWL <i>overreport</i> long term unemployment	FWL and LFS in reasonable agreement; BHPS <i>underreports</i> unemployment
25-34 years	Profile of unemployment growth captured by both FWL and BHPS. Both underreport unemployment slightly	BHPS and FWL <i>severely underreport</i> all unemployment	FWL and LFS in reasonable agreement; FWL <i>underreports</i> long term unemployment; BHPS <i>underreports</i> all unemployment	<i>Chronic underreporting</i> in both BHPS and FWL
35-44 years	BHPS and LFS in good agreement; FWL <i>underreports</i> unemployment	BHPS and particularly, FWL underreport earlier history of unemployment	BHPS <i>severely underreports</i> unemployment; FWL <i>overreports</i> long term unemployment	FWL <i>underreports</i> unemployment; BHPS has <i>chronic underreporting</i> of unemployment
45-54 years	BHPS and FWL in good agreement with LFS	BHPS and FWL underreport earlier history of unemployment	FWL <i>overreports</i> unemployment; BHPS <i>underreports</i> unemployment	BHPS has <i>chronic underreporting</i> of unemployment; FWL has incorrect profile
55+ years	BHPS and FWL in good agreement with LFS; both <i>overreport</i> long term unemployment	BHPS and FWL underreport earlier history of unemployment	BHPS in reasonable agreement with LFS; FWL has <i>chronic overreporting</i> of unemployment	BHPS in reasonable agreement with LFS; FWL has incorrect profile

Annex A: ISER Code of Ethics

CODE OF ETHICS

INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH

UNIVERSITY OF ESSEX

The Ethical Guidelines of the Social Research Association (SRA) and the Institute of Social and Economic Research's (ISER) Ethical Procedures together comprise ISER's Code of Ethics. The SRA Guidelines are not a list of rules of regulations but are intended, as the Introduction states, "to enable the social researcher's individual ethical judgements and decisions to be informed by shared values and experience". ISER's Ethical Procedures relate to issues and practices in ISER's specific work.

The Institute for Social and Economic Research seeks to maintain effective ethical standards in surveys, with respect to both data collection and data usage. ISER has an additional concern in that its activities are long-term with a high public profile. Any breach of ethical standards could have serious implications for ISER's success. Thus, ISER must maintain very high standards in the procedures adopted to guarantee confidentiality and anonymity of data. It must be able to assure not just respondents but users, funders, and the general public that this guarantee is respected at all times.

General Obligations

1. Obligations to society

1.1 Considering conflicting interests

The Institute for Social and Economic Research accepts that social research is predicated on the belief that greater access to well-grounded information is beneficial to society. The fact that statistical information resulting from social research can be misconstrued or misused, or that its impact can be different on different groups, is not in itself a convincing argument against its collection and dissemination. Staff of ISER will, therefore, consider the likely consequences of collecting and disseminating various types of data and should guard against predictable misinterpretations or misuse.

1.2 Widening the scope of social research

ISER shall use the possibilities open to them to extend the scope of their research, and to communicate their findings, for the benefit of the widest possible community.

1.3 Pursuing objectivity

While social researchers operate within the value systems of their societies, ISER will endeavour to uphold their professional integrity without fear or favour. They will also not engage or collude in selecting methods designed to produce misleading results, or in misrepresenting statistical findings by commission or omission.

2. Obligations to Funders and Employers

2.1 Clarifying obligations and role

ISER will clarify in advance the respective obligations of employer or funder and their research programme; they will, for example, refer the employer or funder to the relevant parts of this professional code to which they adhere, Reports of the findings shall (where appropriate) specify their role.

2.2 Assessing alternatives impartially

ISER will consider the available methods and procedures for addressing a proposed inquiry and will provide funders and employers with an impartial assessment of the respective merits and demerits of alternatives.

2.3 Not pre-empting findings

ISER will not accept contractual conditions that are contingent upon specific research findings.

2.4 Guarding privileged information

Even in cases in which ISER has been furnished with information by the funder or employer who may legitimately require it to be kept confidential, the Centre will not keep research methods and procedures that have been utilised to produce published data confidential.

3. Obligations to colleagues

3.1 Maintaining public confidence in social research

Social researchers depend on the confidence of the public. ISER will, in its work attempt to promote and preserve such confidence without exaggerating the accuracy or explanatory power of their data.

3.2 Exposing and reviewing methods and findings

Within the limits of confidentiality requirements, ISER will provide adequate information to colleagues to permit their methods, procedures, techniques and findings to be assessed. Such assessments should be directed at the methods themselves rather than at the individuals who selected or used them.

3.3 Communicating ethical principles

To conduct certain inquiries social researchers need to collaborate with colleagues in other disciplines, as well as with interviewers, clerical staff, students, etc. In these cases, ISER will make their own ethical principles clear and take account of the ethical principles of their collaborators.

4. Obligations to subjects

4.1 Avoiding undue intrusion

ISER will at all times be aware of the intrusive potential of some of their work and recognise that they have no special entitlement to study all phenomena. ISER further recognises that the advancement of knowledge and the pursuit of information are not themselves sufficient justifications for overriding other social and cultural values.

4.2 Obtaining informed consent

Research work involving the active participation of human subjects shall be based as far as practicable on their freely given informed consent. The survey and other research activities of ISER are voluntary inquiries, and subjects shall not be under the impression that they are required to participate; they shall be told of their entitlement to refuse at any stage for whatever reason to give information. Information that would be likely to affect a subject's willingness to participate should not be deliberately withheld.

4.3 Modifications to informed consent

In the unlikely event that technical or practical considerations inhibit the achievement of prior informed consent, the subjects' interests will be safeguarded in other ways. For example:

- a) *Respecting rights in observation studies.* In observation studies, where behaviour patterns are recorded without the subject's knowledge, ISER will take care not to infringe what may be referred to as the 'private space' of an individual or group. This may vary from culture to culture.
- b) *Dealing with proxies.* In cases where a 'proxy' is used to answer questions on behalf of a subject, say because access to the subject is uneconomic or because the subject is too ill or too young to participate directly, care shall be taken not to infringe the 'private space' of the subject or to disturb the relationship between the subject and proxy. Where indications exist or emerge that the subject would object to certain information being disclosed, such information shall not be sought by proxy.
- c) *Secondary use of records.* In cases where ISER has been granted access to, say, administrative or medical records or other research materials for a new or supplementary inquiry, the custodian's permissions to use the records shall not relieve ISER from having to consider the likely reactions, sensitivities and interests of the subjects concerned, including their entitlement to anonymity.
- d) *Misleading potential subjects.* In studies where the measurement objectives preclude the prior disclosure of material information to subjects, ISER shall weigh the likely consequences of any proposed deception. ISER recognises that, to withhold material information from, or to misinform, subjects involves a deceit, whether by omission or commission, temporarily or permanently, which face legitimate censure unless it can be justified.

4.4 Protecting the interests of subjects

Neither consent from subjects nor the legal requirement to participate shall be seen by ISER as absolving it, its staff, Research Associates or collaborators from an obligation to protect the subject as far as possible against potentially harmful effects of participating. ISER shall try to minimise disturbance both the subjects themselves and to the subjects' relationships with their environment.

4.5 Maintaining confidentiality of records

Statistical data are unconcerned with individual identities. Data are generally collected to answer questions such as 'how many?' or 'what proportion?', rather than 'who?'. The identities and records of co-operating (or non-cooperating) subjects will therefore be kept confidential, whether or not confidentiality has been explicitly pledged, and at no time will identifying information be part of machine-readable data

relating to the individual. The confidentiality of the subjects' identity will be maintained at all times during the collection, processing and analysis of statistical data.

4.6 Inhibiting disclosure of identities

ISER will take all appropriate measures to prevent their data from being published or otherwise released in a form that would allow any subject's identity to be disclosed or inferred.

Specific ISER practices and procedures

The Institute for Social and Economic Research's Ethical Procedures relate to issues and practices in specific work. All ISER staff and Research Associates are required to sign an Ethical Undertaking acknowledging that they have read these documents and accept the general obligation the Code of Ethics imposes. These procedures may from time to time be amended in the light of practice.

Informed Consent

5. Voluntary Participation

5.1 Respondents should be aware of the importance of the survey, but the voluntary nature of participation will be emphasised.

“Research work involving the active participation of human subjects shall be based as far as practicable on respondents' freely given informed consent. Respondents should be made fully aware that participation in the survey is voluntary. Information that would be likely to affect a respondent's willingness to participate should not be deliberately withheld. Respondents shall be told of their entitlement to refuse at any stage to give information, for whatever reason, and to withdraw data just supplied.”

(Ethical Guidelines of the Social Research Association)

5.2 An advance letter is sent at the start of fieldwork in each wave, giving respondents the opportunity to make enquiries or to withdraw before interviewers enter the field. Respondents are given the name of the Institute's Survey Liaison Manager to contact in the case of queries.

5.3 The panel survey is based on projections of long-term co-operation. While it is hoped that respondents will come to feel a sense of belonging to the project, participation at each approach is treated as a *single* commitment. No formal commitment to be re-interviewed is made or asked for at any stage.

5.4 At the end of each year's interview, interviewers are instructed to say that the Institute will be approaching respondents about 12 months later for another interview. Information necessary to trace the respondent is then requested. Interviewers are required to give a full explanation of the reasons for asking for this information.

- 5.5 Respondents are provided with a range of material designed to keep them fully informed about the survey. At different times this might include any of the following:
- a) leaflets and brochures about the survey;
 - b) an annual summary of research findings;
 - c) a statement of the Institute's procedures in respect of confidentiality (available upon request).

6. Withdrawal of Participation in the Survey

- 6.1 No interview will be sought with respondents in subsequent waves where the whole household refuses in the first wave. If co-operation is refused after a first interview, the name of the respondent will be retained in the Institute's files unless the respondent demands its removal. An attempt will be made to convert refusals at the time of refusal, but not further requests for an interview will normally be made after two refusals across consecutive waves.
- 6.2 If, after interview, respondents inform the interviewer, the Institute or the fieldwork agency that they are unwilling to be interviewed in subsequent waves, this will be formally recorded. The following rules will then be applied:
- a) If the respondent categorically rejects any further approach she or he will still be sent a thank-you letter, and a summary of findings. The Research Institute will write to the respondent before the next fieldwork period requesting reconsideration. No approach will be made by an interviewer unless the respondent replies accepting a further interview.
 - b) Other respondents who indicate reluctance to continue co-operation but who do not categorically reject any further contact will be approached under standard fieldwork procedures.

7. Withdrawal of Personal Data (names and addresses)

- 7.1 Personal data are currently held in the following forms:
- a) names and addresses of all household members;
 - b) names and addresses of people nominated to help the Institute trace respondents if contact is lost.

Only the former are held on computer. Under the provisions of the Data Protection Act 1998, if requested a photocopy may be provided of such computer data held in respect of that individual, either free-of-charge or for a nominal fee. No such information can be given on behalf of other household members unless jointly requested.

- 7.2 Requests for amendment of names will be accepted. If the respondent demands *removal* of their name from the records, a letter will be sent asking for reconsideration. If permission to retain the name is not forthcoming, it will be removed within 28 days of the request. This applies only to the individual concerned unless joint applications are made. The withdrawal will apply to children who are not yet members of the sample in their own right where no responsible adult remains in the sample.

- 7.3 Withdrawal of a name means that the person's records will be known by ID number only. The respondent will be treated as a refusal in respect of future approaches. Neither rule (A) nor rule (B) will apply.
- 7.4 Names and addresses of contact people will not be held on computer and are therefore not covered by the Data Protection Act 1998. However, requests (in writing) by respondents for removal of the information they have provided will be respected. The record will be destroyed or amended as necessary and the respondent will be informed in writing that this has taken place. If the *contact person* requests (in writing) the removal or alteration of information held on him or her, this will take place automatically, and that person will be informed in writing when the change has been made.
- 7.5 When a respondent requires the removal of *personal* data from a computer record, this will be kept on paper to clarify problems which might subsequently arise, as long as this is in accordance with current legislation. However, an instruction to destroy a questionnaire containing personal data (for example, a cover sheet) will be respected.

8. Withdrawal of Substantive Data

- 8.1 The Data Protection Act 1998 does not provide a right of subject access to non-personal data kept for social research purposes, nor therefore to its amendment or withdrawal.
- 8.2 If a respondent requests the withdrawal of data held on a questionnaire but not yet transferred to computer, the Institute will write to the respondent asking for reconsideration. If permission is not forthcoming by the time data entry begins, the respondent's request will be complied with. This will apply to individuals only, unless joint applications are made. The outcome will be recorded as a refusal. In this case a request for destruction of a questionnaire will be respected.
- 8.3 Withdrawal of data will not be undertaken after data entry except at the Director's discretion, and in no circumstances after final preparation for deposit of the data in the Data Archive (28 days before actual deposit).
- 8.4 If a proxy subject requests the withdrawal of data held on them, a letter will be sent explaining the nature of this information and requesting reconsideration. If permission is not granted, the withdrawal shall proceed within 28 days of the request, as long as the expiry of this period is not later than 28 days before deposit of the data in the Data Archive. Individual data on the proxy subject will be removed from the data set and the outcome recorded as a refusal. Rule (A) will apply for the next wave.

Confidentiality

9 Access and Control

- 9.1 No-one outside the employment of the Institute or of the fieldwork agency will be given access to the schedules or to personal data held on computer.

- 9.2 The fieldwork agency will return all schedules to the Institute after all processing has been completed. If it requires retention of names and addresses beyond the time required for fieldwork and survey-processing functions, whether on paper or on computer, approval will be required from the Research Institute, which should also be informed of security procedures relating to these. The fieldwork agency is separately responsible for any computer holdings under the Data Protection Act 1998.
- 9.3 Separation of personal and substantive data held on computer will be strictly maintained. Temporary linkage will be permissible only in relation to fieldwork operations or to methodological work. Access to personal data within the Institute will be governed by an internal set of procedures which all members of staff are made aware of.
- 9.4 ISER staff, Research Associates and Research Students based in the Institute will be required to sign an Ethical Undertaking accepting the provisions of ISER's Code of Ethics.

10. Release of Data

- 10.1 Sample area identification above a minimum area size (established by the Institute) will be provided. Applications for identification below this criterion, for instance for methodological research on sampling, will be considered by the Director.
- 10.2 The Institute will take all reasonable steps to ensure that data released will not make possible the identification of individuals or households. In particular, personal characteristics such as exact date-of-birth will not be released.

In addition to requirements and obligations stated above ISER staff are advised that it is sensible to realise that respondent anonymity might be more likely to be breached through carelessness than through deliberate action. ISER staff are required to be fully aware of the need to protect sensitive documents. All those with any responsibility for handling questionnaires or other research documents containing confidential information must follow clear rules on confidentiality. Rules exist within ISER for storage and handling of documents. Where these prove inflexible for effective operation, new procedures should be thought through clearly and the relevant manager informed.

All ISER staff and data users are made aware that negligence can be costly. They are advised, for instance, that: on no account is it acceptable to 'loan' data to unauthorised people; that there are simple practical ways in which staff and users can ensure there is no breach; that while access to data which can identify individuals (eg date-of-birth) or groups of individuals (such as geographical indicators) is severely restricted, it is extremely important to take care that data which could potentially identify individual respondents or groups of respondents be closely controlled; and, straightforward measures to protect floppy disks (such as placing them in secure storage boxes), properly exiting from sensitive files before leaving terminals or PCs, and locking doors where necessary, are important safeguards.

Institute for Social and Economic Research
1 October 1999