BACKGROUND AND METHODOLOGY

2007

CONTENTS

HISTORY OF THE LFS

Page

SECTION 1 – HISTORY OF THE LFS IN THE UK....1

SECTION 2 – THE LFS IN NORTHERN IRELAND....6

SAMPLE, DESIGN, QUESTIONNAIRE, FIELDWORK AND PROCESSING

SECTION 3 - SAMPLE DESIGN....7

SECTION 4 - THE QUESTIONNAIRE....13

SECTION 5 - FIELDWORK....14

SECTION 6 - CODING AND PROCESSING THE DATA....27

DATA QUALITY

SECTION 7 - NON-SAMPLING ERRORS....29

SECTION 8 - SAMPLING ERRORS AND CONFIDENCE INTERVALS....36

SECTION 9 - NON RESPONSE....42

SECTION 10 - WEIGHTING THE LFS SAMPLE USING POPULATION ESTIMATES....48

SECTION 11 - REPORT ON PROXY RESPONSE STUDY BASED ON LFS QUESTIONS....57

SECTION 12 - IMPUTATION IN THE LFS....67

SECTION 13 - CONTINUITY AND DISCONTINUITY ON THE LFS....73

SECTION 14 - QUALITY....95

SECTION 15 - HARMONISATION....97

SECTION 16- USES OF THE LFS....101

PUBLICATION AND DISSEMINATION

SECTION 17 - LFS DISSEMINATION AND PUBLICATIONS 107

SECTION 18 - LFS DATA FOR SMALL SUB-GROUPS: ANNUAL DATABASES AND AVERAGING OVER SEVERAL QUARTERS 121

ANNEX A: DERIVATION AND EXAMPLES OF STANDARD ERRORS ON THE LFS..... 128

SECTION 1 - THE HISTORY OF THE LFS IN THE UK

The Labour Force Survey (LFS) is a survey of households living at private addresses in the UK. Its purpose is to provide information on the UK labour market which can then be used to develop, manage, evaluate and report on labour market policies. The survey is managed by the Social and Vital Statistics division of the Office for National Statistics (ONS)[1] in Great Britain and by the Central Survey Unit of the Department of Finance and Personnel in Northern Ireland on behalf of the Department of Enterprise, Trade & Investment (DETINI).

For a more detailed description of the LFS and how it has developed, see the August 2006 edition of Labour Market Trends "*Reflections on fifteen years of change in using the LFS: How the UK's labour market statistics were transformed by using the LFS", by Barry Werner* (http://www.statistics.gov.uk/about/platforms/lmt/).

1.1 LFS 1973-1983

The first LFS in the UK was conducted in 1973, under a Regulation derived from the Treaty of Rome. The Statistical Office of the European Union (Eurostat) co-ordinates information from labour force surveys in the member states in order to assist the EC in matters such as the allocation of the European Social Fund. The ONS is responsible for delivering UK data to Eurostat.

The survey was carried out every two years from 1973 to 1983 in the spring quarter (March-May) and was used increasingly by UK Government departments to obtain information which could assist in the framing and monitoring of social and economic policy. By 1983 it was being used by the Employment Department to obtain measures of unemployment on a different basis from the monthly claimant count and to obtain information which was not available from other sources or was only available for census years, for example, estimates of the number of people who were self-employed.

Published LFS estimates for 1973-1983 refer to the spring quarter and are available on a UK basis.

1.2 ANNUAL LFS 1984-1991

Between 1984 and 1991 the survey was carried out annually and consisted of two elements:-

(i) A quarterly survey of approximately 15,000 private households, conducted in Great Britain throughout the year;

(ii) A "boost" survey in the spring quarter between March and May, of over 44,000 private households in Great Britain and 5,200 households in Northern Ireland

Published estimates for 1984-1991 are available for the UK and are based on the combined data from the "boost" surveys and quarterly surveys in the spring quarters (Mar-May). The quarterly component of the 1984 to 1991 surveys were not published because the small sample sizes meant that the results were not robust. However, the quarterly survey proved to be invaluable in developmental terms, and in making early assessments of seasonality. A fuller description of the survey methodology used in this period is available in the annual results published by ONS (previously by OPCS) - see section 18 for details of these publications. [CROSS CHECK REF]

1.3 QUARTERLY LFS FROM SPRING 1992

In 1992 the sample in GB was increased to cover 60,000 households every quarter enabling quarterly publication of LFS estimates. Whilst it built on the annual survey, there were a number of differences which can be summarised as follows:

(i) panel design – from 1992 the GB survey was based on a panel design where a fifth of the sample each quarter is replaced and individuals stay in the sample for 5 consecutive waves or quarters. A shorter fieldwork period was also introduced which together with the panel nature of the survey led to slightly lower response rates.

(ii) sample design - the major difference was the introduction of an unclustered sample of addresses for the whole of Great Britain (the sample for Northern Ireland is similarly unclustered). This improved the precision of estimates particularly when making regional analyses. In the case of Scotland a very small bias arises from partial coverage of the population north of the Caledonian Canal. This area contains about five percent of the total population of Scotland.

(iii) additions to the sample - the inclusion of people resident in two categories of non-private accommodation, namely those in NHS accommodation and students in halls of residence. The students are included through the parental home.

In the winter of 1994/95 a quarterly Labour Force Survey was introduced to Northern Ireland. Each quarter's sample consists of approximately 3,000 household responses spread over five waves - 600 in each wave. A rotational pattern was also adopted, identical to that being operated in the GB LFS. Quarterly UK LFS estimates are available from winter 1994/95.

Estimates of the effects of these, and other changes, between the annual and quarterly changes are included in section 14. [CROSS CHECK REF]

1.4 LFS QUARTERS

The quarterly LFS launched in 1992 in GB and in 1994 in NI operated on a seasonal quarter basis: March-May (Spring), June-August (Summer), September-November (Autumn) and December-February (Winter). The reasons for this were

(i) Many activities associated with the labour market occur seasonally and follow the pattern of the school year. This was more the case when the LFS first started at which point more young people left school at Easter than in the summer;

(ii) Easter can cause difficulty as it varies in timing between March and April - so ensuring that Easter is always covered by the same quarterly survey period avoids this problem.

The first results from the quarterly GB LFS, relating to spring 1992, were published in the LFS Quarterly Bulletin (LFS QB) in September 1992 - that is, about 3½ months after the end of the survey period. From this date, the QB was the main source of LFS data. More timely results were presented in each quarter's ONS 'Labour Force Survey First Release' which provided key results about six weeks after the end of the survey period. Both the QB and the First Release presented GB estimates as Northern Ireland estimates were only available for the Spring quarters until Winter 94/95.

1.5 CALENDAR QUARTERS

In May 2006 the LFS moved to calendar quarters (CQ's). This means the micro data will no longer be available on a seasonal basis (spring – winter). The main reason ONS is moving to CQ's for the LFS is that it is an EU requirement under regulation[2]. Eurostat – the body responsible for the EU LFS – has a target structure for the survey with all Member States providing data on a CQ basis which will promote comparability across countries. In addition

to conforming to the EU regulation, the switch from seasonal to calendar quarters will also enhance the comparability of the LFS with other quarterly surveys which are mostly conducted on a CQ basis. This is particularly relevant with respect to National Accounts.

The following table shows the resultant changes to the quarterly release of micro data.

Seasonal Quarters	**Calendar Quarters (CQ's) – from May 2006**
Winter (December to February)	Q1 = January to March (JM)
Spring (March to May)	Q2 = April to June (AJ)
Summer (June to August)	Q3 = July to September (JS)
Autumn (September to November)	Q4: = October to December (OD)

This means the spring (March-May) questionnaire will move to the April-June questionnaire (Q2) and the June-August questionnaire will move to the July-September (Q3) and so on. Changes were also made to the interview weeks to align them to CQ's.

A note has been published in the June 2006 *Labour Market Trends* (http://www.statistics.gov.uk/about/platforms/lmt/) which looks at the impact of the move to CQ's. There is also a CQ version of the Historical Quarterly Supplement (HQS) that was published on 17 May 2006 to coincide with the move. This will have historical data back to 1997 for certain quarters (mostly Q2 and Q4), so that users can look at trends based on CQ's. A partial series of micro data based on CQs has also been created covering the following periods: Q2 regional datasets 1997, 1999, 2001, and every quarter from then onwards.

It is the intention to complete a full back-series of micro data on a CQ basis.

1.6 EARNINGS FROM EMPLOYMENT QUESTIONS FROM WINTER 1992/93

Whilst questions in the LFS are continually being added, removed or modified, the major change to the early quarterly survey was the introduction of a section of earnings questions in GB from winter 1992/93 onwards. These questions were only asked of respondents receiving their fifth and final interviews, because of concerns that the questions might have an adverse impact on overall response rates. Results from these earnings questions were first published in the summer 1994 QB (in December 1994), and in the December 1994 *Employment Gazette*.

Earnings questions have been asked in the Northern Ireland LFS since the survey went quarterly in Winter 1994/5 but results were not weighted up until early 1998. LFS earnings data on a UK basis are available for each quarter from Winter 1994/5.

1.7 EARNINGS QUESTIONS FROM SPRING 1997

The LFS is an important source of earnings data, particularly for part-time workers. However, because earnings questions were initially only asked in wave 5 interviews, sample sizes were quite small and associated sampling errors tended to be relatively high. Work was done to test whether asking earnings questions in the first wave would lead to higher non-response in later waves, but no evidence was found to support this. So from Spring 1997 earnings questions were asked in both waves 1 and 5 in GB and NI, doubling the sample size and reducing sampling errors by about 30%. For more detail see 'Expanding the coverage of the earnings data in the LFS' in April 1998's *Labour Market Trends*.

1.8 MONTHLY PUBLICATION FROM WINTER 1997/8

A major public consultation on labour market statistics was conducted by ONS during 1997, resulting in a new integrated Labour Market Statistics First Release (LMS FR), first published in April 1998 (see February 1998 *Labour Market Trends* article 'Improved Labour Market Statistics'). The LMS FR, which is published monthly, gives prominence to the ILO measure of unemployment, as measured by the LFS over the administrative claimant count measure and draws together statistics from a range of sources to provide a more coherent picture of the labour market.

LFS results in the LMS FR are published on a UK basis, 6 weeks after the end of the survey period, and relate to the average of the latest three-month period. For the latest release see (http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=1944). For detail on the content of the LMS FR see the ONS "Guide to Labour Market Statistics Releases" which is available on the ONS website (http://www.statistics.gov.uk/about/data/guides/LabourMarket/default.asp)

Another result of the public consultation was that the LFS Quarterly Bulletin was replaced by the LFS Quarterly Supplement to *Labour Market Trends*. The Quarterly Supplement contains UK LFS data, whereas the Quarterly Bulletin contained GB data. From May 2006, the Quarterly Supplement was renamed the Historical Quarterly Supplement, and published only on the National Statistics website, at: http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14365 (for more detail see section 18) (CROSS CHECK REF).

Since April 1998, the Department of Enterprise, Trade and Investment (DETINI) have published a Northern Ireland Labour Market Statistics Release to the same timetable as publication of the Labour Market Statistics First Release

1.9 ENHANCEMENTS TO THE LFS IN ENGLAND, WALES AND SCOTLAND

Since Spring 2000, the Department for Works and Pensions and the Department for Education and Skills have funded an annual enhancement to the sample size of the LFS in England. This enhancement (known as the local LFS or LLFS) involves a boost of approximately 39,000 households per year. It aims to improve labour market information at a local level and to help monitor National Learning Targets for England. Further information on the English boost survey can be found in the January 2002 edition of the Labour market Trends 'Annual Local Area Labour Force Survey data for 2000/2001' (http://www.statistics.gov.uk/about/platforms/lmt/).

During 2001-02 a further expansion of the existing LFS was introduced for Wales. This enhancement is funded by the National Assembly for Wales, and as with the LLFS, it will improve labour market information at a local level. This involved a boost of approximately 14,000 households per year. Fieldwork on the Welsh Labour Force Survey (WLFS) began in March 2001.

Similarly, in Spring 2003 a boost to the Scottish sample, funded by the Scottish Executive, was introduced. This Scottish Labour Force Survey (SLFS) involved a boost of approximately 12,000 households per year.

When results from the LFS are combined with these annual enhancements it is known as the Annual Local Area Labour Force Survey (ALALFS).

Because the databases are larger we can have more confidence in the results produced from them. In practical terms this means it is possible to look at smaller sub-groups of the population that we can using the quarterly LFS databases. The databases are referred to as *Local Area Databases*, reflecting the fact that the databases hold a local area indicator variable and that the main use of the databases is for looking at data for individual local areas.

For further information on local area data please see the *Guide to Regional and Local Labour*

Market Statistics available at the Local Labour Force Survey web page (http://www.statistics.gov.uk/statbase/product.asp?vlnk=4752).

1.10 THE ANNUAL POPULATION SURVEY

In 2004, a further improvement, the Annual Population Survey (APS), was introduced. The APS included all the data of the ALALFS, but also included a further sample boost in more urban areas of England – known as the APS(B) - aimed at achieving a minimum number of economically active respondents, in the sample, in each Local Authority District in England.

This survey interview for the APS(B) is shorter than the main LFS, however it does include many of the core questions on the LFS and its enhancements.

The APS(B) is funded by the Department for Communities and Local Government through the Neighbourhood Statistics programme at ONS.

The first APS covered the calendar year 2004, rather than the ALALFS period of March to February. Also, the ALALFS data were published only once a year, but the APS data are published quarterly, but with each publication including a year's data. The first APS results, covering January to December 2004, were published in July 2005, data will then be published every three months covering a whole year.

Together with other data sources, the survey will provide an annual update on much of the information collected as part of the 2001 Census. Results will be published on the NS Neighbourhood Statistics web site: http://www.neighbourhood.statistics.gov.uk/

A further major advantage of the APS over the ALALFS is that APS micro data are available to users through the Data Archive at Essex University. Access to these data is through a 'Special Licence' scheme, which allows access to detailed data provided that the research use is fully described and strict conditions of access are adhered too. More detail is available on the Data Archive website (http://www.data-archive.ac.uk).

The APS(B) sample element of the APS was stopped at the end of 2005 due to financial constraints. The outputs however still remain albeit without this additional boost in the sample.

[1] Until 5 July 1995, the LFS was the responsibility of the Employment Department (ED). On that date ED was abolished and responsibility for the survey passed to the Central Statistical Office (CSO). On 1 April 1996, the CSO merged with the Office for Population Censuses and Survey (OPCS) to form the ONS which now has responsibility for the LFS.

[2] Council Regulation (EC) No 577/98 and associated revisions.

SECTION 2 - THE LFS IN NORTHERN IRELAND

The Northern Ireland Labour Force Survey is the responsibility of the Department of Enterprise, Trade and Investment (DETINI) and fieldwork is carried out by the Central Survey Unit of the Department of Finance and Personnel.

From 1973 - 1983, as in GB, the survey in Northern Ireland was conducted in alternate spring quarters. From 1984 - 1994 it was carried out annually. This annual survey consisted of 5,200 addresses drawn at random from the Rating and Valuation List - approximately 1% of private addresses in Northern Ireland. Over this period interviewing was conducted only in the spring, with no quarterly element. UK LFS estimates are available for Spring quarters from 1973-1994.

In the winter of 1994/95 a quarterly Labour Force Survey was introduced to Northern Ireland. Each quarter's sample consists of approximately 3,000 household responses spread over five 'waves' - 600 in each wave. A rotational pattern was also adopted, identical to that being operated in the GB LFS. Respondents at 'wave' 1 are interviewed face-to-face with subsequent interviews at 'waves' 2-5 taking place, where possible, by telephone. Computer assisted interviewing has been used in the Northern Ireland Labour Force Survey since 1992. Quarterly UK LFS estimates are available from winter 1994/95.

Income questions have been asked in the Northern Ireland LFS since the survey went quarterly in Winter 1994/5 but results were not weighted up until early 1998. LFS income data on a UK basis is now available for each quarter from Winter 1994/5. From Spring 1997, the income questions in both the GB and NI LFS have been asked of respondents in waves 1 and 5, producing a larger sample size then when previously asked only of wave 1 respondents.

Since April 1998, the Department of Enterprise, Trade and Investment (DETINI) have published a Northern Ireland Labour Market Statistics Release to the same timetable as publication of the Labour Market Statistics First Release.

SECTION 3 - SAMPLE DESIGN

3.1 SURVEY DESIGN OF THE LFS

The LFS is the largest regular household survey in the United Kingdom. The design of the LFS enables estimates of levels, such as the number of people in employment, which are representative of the national population to be produced for any period of three consecutive months - the estimates produced for the monthly Labour Market Statistics First Release are referred to as three-month rolling averages. The figure below illustrates their composition.

	Reference period												
3-month rolling estimate for:	Jan 06	Feb 06	Mar 06	Apr 06	May 06	Jun 06	Jul 06	Aug 06	Sep 06	Oct 06	Nov 06	Dec 06	Publication date
Jan-Mar 06													Mid-May 06
Feb-Apr 06													Mid-Jun 06
Mar-May 06													Mid-Jul 06
Apr-Jun 06													Mid-Aug 06
May-Jul 06													Mid-Sep 06
Jun-Aug 06													Mid-Oct 06
Jul-Sep 06													Mid-Nov 06
Aug-Oct 06													Mid-Dec 06
Sep-Nov 06													Mid-Jan 07
Oct-Dec 06													Mid-Feb 07

The design of the LFS also allows estimates of change to be made. Because of the continuous over lapping nature of the samples used, estimates of change between overlapping periods are volatile and difficult to interpret. Using the figure above, comparisons over time should only be made with the preceding periods shaded in the same pattern, or with earlier periods. For further details on this issue see *Labour Market Trends May 2005 (*http://www.statistics.gov.uk/about/platforms/lmt/).

3.2 SIZE OF THE SELECTED SAMPLE OF ADDRESSES

The LFS is intended to be representative of the whole population of the UK. The sample design of the LFS and its detailed implementation are the responsibility of ONS' Social and Vital Statistics Division. The population covered is all people resident in private households, all persons resident in National Health Service accommodation and young people living away from the parental home in a student hall of residence or similar institution during term time. (These latter groups are included in the LFS sample specifically to improve the coverage of young people).

The sample design currently consists of about 50,000 responding households in Great Britain every quarter, representing about 0.1% of the GB population. A sample of approximately 2,000 responding households in Northern Ireland is added to this, representing 0.1% of the NI population, allowing United Kingdom analyses to be made.

3.3 SURVEY METHODOLOGY

Each quarter's LFS sample of 53,000 UK households is made up of five "waves", each of approximately 11,000 private households. Each wave is interviewed in five successive quarters, such that in any one quarter, one wave will be receiving their first interview, one wave their second, and so on, with one wave receiving their fifth and final interview. Thus there is an 80% overlap in the samples for each successive quarter.

Households are interviewed face to face at their first inclusion[1] in the survey and by telephone, if possible, at quarterly intervals thereafter, and have their fifth and last quarterly interview on the anniversary of the first.

During the build up period (autumn 91, winter 91/2) of the quarterly survey, waves were 'created' in order to provide the required sample structure quickly. The proportion of the sample in each wave from spring 92 was as follows:

	WAVE 1	WAVE 2	WAVE 3	WAVE 4	WAVE 5
Spring 92	30%	23%	47%	0%	0%
Summer 92	20%	30%	23%	27%	0%
Autumn 92	20%	20%	30%	23%	7%
Winter 92/3	20%	20%	20%	30%	10%
Spring 93	20%	20%	20%	20%	20%

Interviewing for the LFS is continuous in that interviews take place to collect information relating to every week of the year. The one exception to this is a “leap week” which was inserted between the winter 2003 and spring 2004 quarters to bring the timing of LFS quarters back into line with the months they refer to. Thus the week from 1 March to 7 March 2004 was not covered by the reference period of either quarter.

3.4 COVERAGE AND SAMPLE SELECTION

The survey results refer to people resident in private households and in NHS accommodation in the UK. For most people residence at an address is unambiguous. People with more than one address are counted as resident at the sample address if they regard that as their main residence. The following are also counted as being resident at an address:

(i) people who normally live there, but are on holiday, away on business, or in hospital, *unless* they have been living away from the address for six months or more;

(ii) children aged 16 and under, even if they are at boarding or other schools;

(iii) students aged 16 and over are counted as resident at their normal term-time address even if it is vacation time and they may be away from it.[42]

A single stage sample of addresses with a random start and constant interval is drawn from the Postcode Address File (PAF) for Great Britain south of the Scottish Caledonian Canal. The PAF is sorted by postcode so the sample is effectively stratified geographically.

The sample is allocated into 212 Interviewer Areas (162 in England, 26 in Wales and 24 in Scotland). The LFS also has the concept of fieldwork "quotas". Each quota is divided into 13 "stints" and then the Interviewer Areas are comprised of one or two quotas (there are 318 quotas in England, 51 in Wales and 43 in Scotland). Some interviewers cover one quota and some

[1] The small proportion of households sampled from North of the Caledonian Canal in Scotland are first approached by telephone.

[2] Students living in halls of residence have been included by changing, (for LFS purposes only), the standard ONS Social and Vital Statistics division definition of a household. The standard definition says that ‘Adult children, that is, those aged 16 and over who live away from home…should not be included at their parental address’. For LFS purposes those living in halls of residence will be included at the parental address. Because of this LFS households are marginally larger than those in other surveys, including for example the General Household Survey (GHS).

cover two. The 13 stint areas are randomly allocated to the 13 weeks of a quarter. The same stint area is covered by an LFS interviewer in the same week each quarter. A systematic sample of addresses is selected for each quarter throughout the country and is distributed between stint areas to provide a list of addresses to be interviewed each week.

Further details about the relationship between the survey design and field management procedures are given in section 5.

3.5 SAMPLING FRAMES

Great Britain

The sampling frame from which most (99%) of the GB sample is taken is the 'small users' sub-file of the Postcode Address File (PAF). The PAF is a computer list, prepared by the Royal Mail, of all the addresses (delivery points) to which mail is delivered. 'Small users' are delivery points which receive fewer than 50 items of mail a day. With the use of special ONS procedures the coverage of the PAF is about 97% of private households.

In addition to private households the PAF contains non-private and non-residential (and therefore ineligible) addresses which cannot be identified as such prior to the interviewer making contact. Interviewers have instructions to exclude such institutions and classify them as ineligible. About 11 per cent of addresses on the PAF are ineligible: allowance is made for this in determining the sample size needed to obtain the required number of interviews.

The sampling frame for NHS accommodation was specially developed for the Labour Force Survey. All district health authorities and NHS trusts were asked to supply a complete list of their accommodation (this accommodation mainly comprises what was once known as 'Nurses Homes', but the coverage is more extensive than that name implies)[3].

Because the area north of the Caledonian Canal is sparsely populated, interviewing a single-stage sample of addresses face to face would be very expensive so the choice lay between a two-stage sample of addresses with interviewing face to face and one drawn from the telephone directory with telephone interviewing. A two-stage sample of addresses interviewed face-to-face would lead to larger sampling errors and would still lead to high travel costs in this area. The disadvantage of a telephone sample is the bias resulting from non-coverage of people not on the telephone and, among those who are on the telephone, non coverage of ex-directory numbers and new numbers not yet in the directory. After investigations of the possible use of random digit dialling it was decided to draw a random sample from the published telephone directory[4]. The sample is drawn with a random start and constant interval. Further details of the fieldwork are provided in section 5.

The sampling interval for the PAF (Postcode Address File), the telephone directory sample and for NHS accommodation was set at 281. This provided a sample of approximately 83,200 addresses from the PAF (introduced over five quarterly, first wave samples of 16,640), 75 eligible addresses North of the Caledonian Canal and 25 units of NHS accommodation.

Northern Ireland

[3] Information was received from 417 out of the 455 authorities, trusts and teaching hospitals and the frame is not therefore complete. If the coverage of the frame is proportional to the coverage of authorities etc, then the frame contains 92 per cent of all NHS accommodation.

[4] The former ONS Social Survey Division conducted a postal survey using a PAF-based sample of people living in the area. The survey found, from a responding sample of 649 households, and a response rate of 86 per cent, that 85 per cent of households in the area are on the telephone and that 84 per cent of these are in a published directory (about 8 per cent are ex-directory and about 6 per cent are yet to be included). Hence, in general, about 71 per cent of the population can be reached through the published directory, although this proportion will obviously tend to be higher when the directory is new and to drop steadily until the new directory is published.

The source of the sample in Northern Ireland is the Valuation List used for rating purposes, excluding commercial units and known institutions, arranged into three geographical strata. These are:

1. Belfast District Council area,
2. Eastern sub-region (most of Antrim, Down and part of Armagh),
3. Western sub-region (remainder of Northern Ireland).

Within each stratum rateable units are selected at random without substitution, to obtain the 650 'new' addresses entering the panel each quarter.

3.6 NON-PRIVATE HOUSEHOLDS (COMMUNAL ESTABLISHMENTS) IN THE LFS

LFS estimates relate almost exclusively to the population living in private households, and exclude most of those living in communal establishments. This section provides more information about the communal establishment population.

Definition of Communal Establishments

The 2001 Census defined communal establishments (CE) as providing managed residential accommodation[5]. Contained within the classification were the following groups:

Type of Communal Establishment, GB	Number of residents (non-staff) aged 16+	
Medical and care establishments		**437,736**
NHS	37,777	
Psychiatric hospital / home	14,316	
Other hospital home	23,461	
Local authority		**45,073**
Children's home	631	
Nursing home	1,597	
Residential care home	41,397	
Other home	1,448	
Housing Association - Homes and Hostels		**10,218**
Other		**344,668**
Nursing home	146,884	
Residential care home	182,757	
Children's home	384	
Psychiatric hospital / home	6,577	
Other hospital	525	
Other medical and care home	7,541	
Other establishments		**451,610**
Defence establishments (including ships)	52,260	
Prison service establishments	50,273	
Probation / bail hostel	1,357	
Education establishments (including halls of residence)	235,399	
Hotel, boarding house, guest house	15,749	
Hostel (non-HA)	26,756	
Civilian ship, boat or barge	26	
Other	69,790	

[5] See Population Definitions for the 2001 Census (Census Advisory & Working Groups), Advisory Group Paper (99)04

Total	**889,346**

Background to the problem

Sampling for the LFS is based on the 'small users' sub-file of the Postal Address File (PAF); this sub-file contains addresses which receive less than 50 items of mail a day. The LFS sampling frame covers about 97% of private households in GB but not most communal establishments. Some provision has been made to sample NHS accommodation using an additional sampling frame; this frame covers about 92% of residents in NHS accommodation (groups 1 and 2 in the above list, about 14% of the total CE population). Students living in halls of residence or boarding schools are also included in the LFS sampling frame as information about them is collected at their parents' address. The LFS, by not sampling from CEs, excludes approximately 1.5% of the total GB population.

LFS is not alone in excluding CEs from its sampling frame; the Expenditure and Food Survey, the Family Resources Survey, the General Household Survey and the Survey of English Housing do not sample from CEs either. Some departments do, however, occasionally conduct samples of sub-sets of the CE population e.g. Department of Health.

Because there is widespread interest in the characteristics of CE residents, a pilot survey looking at the feasibility of a multipurpose survey of CEs was carried out in 2000 (see LMT March 2002). Further feasibility work looking at the need for extending the LFS sample to cover some CEs was conducted in November 2004. However, the findings from this were inconclusive and further investigations are planned in the near future. However, the decennial Population Census is likely to remain the most reliable integrated source of CE population data.

Comparisons between LFS estimates and the Census count of the residents of communal establishments suggest that the main differences are:

- there are proportionately more women in CEs
- the population is generally older in CEs, especially for women
- the economic activity rate is considerably lower amongst CE residents, mainly because of fewer people in employment.

It does seem therefore that residents of CEs are different to the rest of the population in terms of the demographic characteristics.

Implications

In general, the exclusion of most CEs from LFS sampling and population weighting means that the LFS estimates of employment probably omit about 81,000 people who are employed but live in CEs, and about 8,600 who are 'unemployed' (ILO definition)[6].

The International Dimension

One of the most important aspects of the LFS design is that it must meet Eurostat's requirements, as the LFS is conducted under Regulation. Eurostat's requirements are for results for private households only, recognising that "for technical and methodological reasons it is not possible... to include the population living in collective households" (Eurostat, EU LFS Methods and Definitions 2001, p10).

[6] Figures taken from "A pilot survey of people living in communal establishments" (CEPS), Labour Market Trends, March 2002, see http://nswebcopy/cci/article.asp?ID=221&Pos=1&ColRank=1&Rank=374

In the Labour Force Surveys of Australia, Canada and the USA, the sampling frames for the Labour Force Survey are designed to represent the civilian non-institutional population and therefore exclude:

- full-time members of armed forces,
- residents of institutions such as prisons and mental hospitals, and
- patients in hospitals or nursing homes who have been there at least 6 months.

In Australia some effort is made to include non-household residents using a list sample of non-private dwellings such as hotels and motels. The LFS in the US also attempts to include such people; the stratified sampling frame includes a 'group quarter' stratum containing those housing units where residents share common facilities or receive formal care.

SECTION 4 - THE QUESTIONNAIRE

4.1 MANAGEMENT OF THE LFS QUESTIONNAIRE

The questionnaire content is determined by ONS. ONS are responsible for identifying, in conjunction with other government departments, needs for new questions or changes to existing questions (e.g. changes in legislation or new government employment programmes) and for determining priorities, given the constraint of interview length. ONS also have to ensure that European Union data requirements are met.

A number of other Government Departments also sponsor LFS questions, including the Department of Transport (travel to work) and the Health and Safety Executive (accidents at work).

Discussions between ONS and other Government Departments on the questionnaire content for all the four quarters follow an annual cycle. Typically, LMD and other Government Departments would expect to submit to SVS in November an outline for requirements for the survey beginning 14 months from then. Initial discussions would determine by March a package of questions for piloting, to test that the questions were acceptable and understood by respondents. For the 2006 questionnaire development cycle cognitive interviewing is being used to test the new questions at the piloting stage. Pilots usually take place in April / May, following which there would be further discussions on which questions would go to a further round of testing (the Dress Rehearsal). The Dress Rehearsal, which usually takes place in July / August, would test whether potential new questions fit in well with the overall questionnaire. By September the broad content for the following year would be agreed. Final agreement from the LFS Steering Group is normally required in October. The new questionnaires go in the field a few months later, starting with the January to March quarter.

Throughout, the interests and priorities of other government departments are taken into account via the inter-departmental LFS Steering Group, which brings together departments with particular interests in LFS data twice a year.

4.2 QUESTIONNAIRE DESIGN AND STRUCTURE

The questionnaire comprises a "core" of questions which are included in every quarter of the survey, together with "non-core" questions which are not asked every quarter. These "non-core" questions provide information that is needed less frequently. Some “non-core questions are only asked in one or two quarters per year, for example, the majority of the questions on a respondents employment pattern are only asked in the spring quarter. Other “non-core” questions do not appear every year, but are included in the survey every 2 or 3 years. For example, questions on childcare are asked every 2 years, and questions on regional mobility are asked every 3 years.

Some questions in the core are only asked at the first interview as they relate to characteristics that do not change over time (e.g. sex, ethnic group). There is also a section on earnings from employment, which since spring 1997 has been asked in respondents first and fifth interviews (prior to that it was asked only in the fifth interview). The earnings data are processed along with the rest of the data each quarter but are weighted separately.

SECTION 5 - FIELDWORK

5.1 THE CONDUCT OF FIELDWORK

Face-to-face and telephone interviewing

LFS fieldwork is carried out by the Labour Force survey interviewing force which is comprised of both face-to-face interviewers, who work from their homes, and by telephone interviewers, who work in a centralised Telephone Unit in Titchfield, Hampshire, where close supervisory control over the conduct and quality of interviews can be maintained. Field trainers and supervisors regularly accompany face-to-face interviewers to ensure that standard procedures are being implemented and the instructions issued to interviewers on the interpretation and coding of responses are being followed. Many of the interviewers work on a part-time basis and there is some spare capacity to allow for cover for sickness and other absences.

All first interviews (wave 1) at an address are carried out face-to-face, except those North of the Caledonian Canal (see section 3). If the respondent agrees to it, recall interviews are carried out by telephone. Overall, including wave 1, around 62% of interviews are by telephone, and 38% are face-to-face.

Number of interviewers

As mentioned above, the interviewing force for the LFS consists of both face-to-face and telephone interviewers. At the start of 2006, there were approximately 650 interviewers working on the LFS; 450 in the field and 200 in the telephone unit. While numbers of telephone interviewers have remained largely the same over time, the number of field interviewers has increased. This is reflected in declining assignment sizes over time - the number of cases interviewers are allocated within a given fieldwork period – which have decreased by around 40 per cent since 2000.

Timing of interviews

The bulk of the LFS questionnaire requests information about respondents' activities in a seven day period which ends on a Sunday: this is called a reference week. The majority (about 80%) of interviews are carried out in the week following the reference week, although if this is not possible interviewers are given a further week and two days in which to obtain interviews (known as the hangover period). Face-to-face interviewers only interview in the last two days of the hangover period, whilst the telephone unit interview throughout the hangover period. The hangover period is extended during some weeks leading up to and including Christmas in order to minimise non-contact (in addition, during these periods, face-to-face interviewers use the whole of the hangover period).

Fieldwork documents

In advance of a first interview a letter is sent to every address in the selected sample explaining that the address has been selected and that an interviewer will be calling. Additionally, in the advance letter, respondents are assured that the information they give will be treated in the strictest confidence and will not be made available to analysts in any form in which individuals, or their households, can be identified. Respondents are also sent a Purpose Leaflet, giving information on summary results and how the LFS data are used (See Annex A for currently used survey documents).

5.2 FIELD MANAGEMENT AND THE LFS SURVEY DESIGN

Avoiding within quarter bias

In any systematic single stage sample of households spread across 13 weeks there is a need to structure the sample so that fieldwork practice does not inadvertently introduce within-quarter bias. One possibility would be to give up the idea of a quarterly sample and simply take unclustered weekly samples. However, face-to-face interviews for the first wave as well as households needing a face-to-face interview in subsequent waves would amount to a sample of only about 2,400 addresses each week spread over the entire country. The average distance between addresses would then be so great that it would be necessary to train and equip an enormous number of interviewers each of whom would do very few interviews. They would take a very long time to build up useful experience as interviewers, and with such a large number, adequate monitoring and supervision would be difficult. Alternatively with a smaller number of interviewers each would spend most of his/her time travelling between sampled addresses with little or no time to do recalls, leading to heavy non-response bias. Since neither of these options were acceptable to ONS the sample is designed as a series of weekly two stage samples spread over the 13 weeks such that the whole country is covered in the quarter and therefore the quarter as a whole constitutes a single stage sample.

Grouping postcode areas

As noted above, the country is divided up into 110 interview areas each containing an equal number of delivery points working systematically across Great Britain and trying to follow existing regional boundaries as far as possible. Within these 110 areas there is a further sub-division into 13 "stint" areas by grouping postcode sectors. Again the aim is to create weekly stint areas of equal size in terms of their number of delivery points (though geographical size varies considerably). In order to avoid unnecessary travel problems in the weekly areas, ONS attempted to map out areas so as to make a mountain, lake or other geographical obstacles occur on the border of a stint. Inevitably the stints vary in their make up because some of the larger interviewing areas are either very rural or very urban, but where possible the weekly stints are mapped so that they contain a mixture of urban and rural localities.

5.3 DEPENDENT INTERVIEWING AT RECALL WAVES

The LFS uses dependent interviewing, where answers given at the previous wave are available to interviewers. The use of dependent interviewing has been shown to provide more accurate results than asking the questions from scratch each time. Methodological investigations by the US Bureau of the Census have shown the considerable improvements in the quality of data produced from dependent interviewing; this technique was recently introduced on their equivalent of the LFS, the Continuous Population Survey (CPS).

Core questions

For most core questions on the LFS the information from the previous wave is rotated into the next quarter. Interviewers must check this information either by asking the question again or checking that the information given in the last wave is still correct.

There are some core questions which have to be asked each quarter without reference to previous answers. These are as follows:

SCHM04	Whether on Government Training Scheme
WRKING	Whether in paid job
EVEROT	Do you ever work paid or unpaid overtime
TOTUS1	Hours worked in main job
USUHR	Usual hours worked in main job
POTHR	Number of hours paid overtime worked per week in main job
UOTHR	Number of hours unpaid overtime worked per week in main job
TOTUS2	Total usual hours worked per week in main job
TOTAC1	Total actual hours worked in main job
ACTPOT	Actual paid overtime in main job
ACTOUT	Actual unpaid overtime in main job
TOTAC2	Total actual hours worked in main job

DIFJOB	Whether looking for a different or additional paid job
LOOK4	Whether looking for any kind of paid work

5.4 REQUIREMENTS FOR ANSWERS TO QUESTIONS

Whilst every effort is made to obtain answers to all relevant questions from each respondent, it is recognised that there will be some cases when a respondent genuinely does not know the answer to a particular question (particularly in the case of responses by proxy - see below) and cases when a respondent does not wish to give the answer to a particular question. In general ONS would not wish to lose such respondents and a "no answer" or "don't know" will be accepted.

However, there are a number of key questions in the survey, some of which are fundamental in classifying a respondents' economic status, which, if not answered cause that whole record (though not the whole household) to be dropped.

Forced response questions

These 'forced response' questions are currently as follows:

R1-16	Relationship to head of household and to other household members
SEX	Sex of respondent
AGE	Age of respondent
MARSTA[9]	Marital status
MARCHK	Is spouse a member of household
LIVWTH	Whether respondent is living together with someone as a couple
HRPID	Whether accommodation is owned/rented in respondent's name
SCHM04	Whether respondent on a government scheme in the reference week
NEWDEA4	For those respondents on New Deal, the type of option in which they were participating
TECLEC	Whether respondent was on a TEC, LEC or some other scheme
WRKING	Whether respondent did any paid work in the reference week
JBAWAY	Whether respondent was away from a paid job in the reference week
OWNBUS	Whether respondent did any unpaid work in the reference week for a business owned by him/herself
RELBUS	Whether respondent did any unpaid work for a business owned by a relative
STAT	Whether respondent was working as an employee or self-employed
LOOK4	Whether respondent was looking for paid work in the previous 4 weeks
LKYT4	Whether respondent was looking for a place on a Government scheme in the previous 4 weeks
METHMP	Seeking work as an employee
METHSE	Seeking work as self employed
METHAL	Seeking work no preference whether as an employee or self employed
MAINME	Main method of looking for work as an employee
MAINMA	Main method of looking for work as either an employee or self employed
MAINMS	Main method of looking for work as self employed
METHM	Main method of looking for work-combined data from the previous 3 variables

5.5 PROXY INTERVIEWS

Acceptability of proxy responses

The LFS allows interviewers to take answers to questions by proxy if a respondent is unavailable. This is usually from another related adult who is a member of the same household, although there are exceptions to this rule:

(i) a young person, of the same household, may translate for a non-English speaking relative;

(ii) a carer, of the elderly or infirm, although not related, may answer for someone in their care if it can be established that they know the respondent well enough;

(iii) anyone can respond by proxy with the personal permission of the head of household or spouse.

Frequency of proxy responses

About a third of LFS responses are collected by proxy. This figure includes not only people who were unavailable and on whose behalf a proxy response was made, but also two other groups. The first is those who were unavailable and did not have a proxy response made for them this wave, but did have a proxy response made for them the previous wave, which was brought forward to the current wave. The second group comprises economically inactive individuals aged 70 years or more, for whom proxy responses were given.

Hence the $^1/_3$ figure includes all informants for whom proxy data was collected. Information on proxy responses can be obtained using the PRXREL variable. Categories 2 and 3 (defined as 'Spouse/partner proxy' and other proxy' respectively) are combined to give the total number of proxy interviews. Information on proxy responses can also be obtained using the variable IOUTCOME; this identifies the informant's status for a particular variable. However, the main (IOUTCOME=2) category of proxy responses does not include cases where proxy data was imputed from the previous wave, or where proxy responses were given for economically inactive 70+ year olds.

Tables 5.1 and 5.2 show the proxy response rates for different age, sex, ethnic and economic activity categories. The numbers are percentages.

Table 5.1 Proxy response rates, by age, sex, and ethnicity.

per cent

Quarter	All	Age			Sex		Ethcen	
		16-17	18-19	20+	Male	Female	white	non-white
United Kingdom								
Spring 00	*32*	*79*	*63*	*29*	*39*	*26*	*32*	*40*
Summer 00	*33*	*78*	*65*	*30*	*40*	*27*	*33*	*40*
Autumn 00	*32*	*77*	*63*	*30*	*39*	*26*	*32*	*39*
Winter 00-01	*33*	*78*	*64*	*30*	*39*	*27*	*32*	*40*
Spring 01	*32*	*78*	*64*	*29*	*39*	*25*	*31*	*37*
Summer 01	*32*	*78*	*65*	*30*	*40*	*26*	*32*	*40*
Autumn 01	*32*	*80*	*64*	*29*	*39*	*25*	*31*	*39*
Winter 01-02	*32*	*80*	*65*	*30*	*40*	*26*	*32*	*41*
Spring 02	*32*	*79*	*65*	*29*	*40*	*25*	*31*	*40*
Summer 02	*32*	*79*	*65*	*30*	*40*	*25*	*32*	*40*
Autumn 02	*32*	*80*	*63*	*29*	*40*	*25*	*31*	*39*
Winter 02-03	*32*	*79*	*64*	*29*	*39*	*26*	*31*	*39*
Spring 03	*31*	*79*	*63*	*29*	*39*	*25*	*31*	*39*
Summer 03	*32*	*81*	*63*	*30*	*40*	*25*	*32*	*39*
Autumn 03	*32*	*81*	*63*	*30*	*40*	*25*	*32*	*39*
Winter 03–04	*32*	*82*	*66*	*30*	*39*	*26*	*32*	*39*
Spring 04	*32*	*81*	*65*	*29*	*39*	*25*	*31*	*39*

Summer 04	*33*	*84*	*68*	*30*	*40*	*27*	*33*	*40*
Autumn 04	*33*	*85*	*69*	*30*	*40*	*27*	*33*	*40*
Winter 04–05	*34*	*85*	*70*	*31*	*40*	*28*	*33*	*40*
Spring 05	*33*	*84*	*70*	*30*	*40*	*27*	*32*	*40*
Summer 05	*34*	*84*	*72*	*31*	*41*	*28*	*33*	*40*
Autumn 05	*34*	*85*	*70*	*31*	*40*	*28*	*33*	*41*
Winter 05-06	*34*	*85*	*70*	*31*	*40*	*28*	*33*	*41*
Q1: JM06	*33*	*85*	*70*	*30*	*39*	*28*	*33*	*41*
Q2: AJ06	*33*	*86*	*70*	*30*	*39*	*27*	*32*	*40*

Table 5.2 Proxy response rates, by employment status.

per cent

Quarter	All	Economic activity					
		Employees	Self-employed	Govt scheme	Unpaid family	ILO unemp	Inactive
United Kingdom							
Spring 00	*32*	*34*	*37*	*56*	*25*	*32*	*28*
Summer 00	*33*	*36*	*38*	*52*	*24*	*36*	*28*
Autumn 00	*32*	*35*	*37*	*52*	*30*	*34*	*28*
Winter 00-01	*33*	*35*	*38*	*53*	*32*	*34*	*29*
Spring 01	*32*	*34*	*37*	*48*	*30*	*32*	*28*
Summer 01	*32*	*34*	*38*	*53*	*25*	*36*	*28*
Autumn 01	*32*	*34*	*37*	*55*	*27*	*35*	*28*
Winter 01-02	*32*	*34*	*37*	*54*	*25*	*33*	*29*
Spring 02	*32*	*33*	*37*	*56*	*24*	*32*	*29*
Summer 02	*32*	*34*	*37*	*52*	*24*	*35*	*28*
Autumn 02	*32*	*33*	*36*	*54*	*26*	*32*	*28*
Winter 02-03	*32*	*34*	*36*	*55*	*27*	*34*	*28*
Spring 03	*31*	*33*	*35*	*61*	*21*	*33*	*28*
Summer 03	*32*	*34*	*37*	*58*	*25*	*37*	*28*
Autumn 03	*32*	*35*	*35*	*56*	*24*	*35*	*28*
Winter 03-04	*32*	*34*	*37*	*51*	*26*	*35*	*29*
Spring 04	*32*	*34*	*37*	*51*	*19*	*33*	*29*
Summer 04	*33*	*35*	*39*	*52*	*29*	*39*	*29*
Autumn 04	*33*	*35*	*39*	*59*	*24*	*37*	*29*
Winter 04-05	*34*	*36*	*38*	*55*	*30*	*37*	*30*
Spring 05	*33*	*35*	*37*	*58*	*24*	*34*	*30*
Summer 05	*34*	*36*	*39*	*56*	*27*	*40*	*29*
Autumn 05	*34*	*36*	*39*	*56*	*23*	*39*	*30*
Winter 05-06	*34*	*36*	*38*	*59*	*25*	*39*	*30*
Q1: JM06	*33*	*35*	*37*	*59*	*22*	*38*	*29*

Quarter	All	Economic activity					
		Employees	Self-employed	Govt scheme	Unpaid family	ILO unemp	Inactive
Q2: AJ06	*33*	*34*	*36*	*52*	*26*	*35*	*29*

Further information about proxies, especially quality of data, is given in section 11.

5.6 COMPUTER ASSISTED INTERVIEWING (CAI)

The LFS interviews are carried out by face-to-face interviewers using laptop computers and by telephone interviewers using networked desktop microcomputers. The questionnaire and edit instrument that they use is identical in both modes. It is produced using the BLAISE CAI software package, which was created by Statistics Netherlands. SSD has designed a computer system which takes the output from BLAISE and uses it to create derived variables, to weight up population estimates and other processes leading to the production of data files for customers. The system also rotates the data for use at the next wave of interviewing.

Advantages and disadvantages of CAI

The advantages of CAI for the LFS over a paper-and-pencil based system are lower costs, improved speed from fieldwork to analysis and better quality data. Disadvantages such as the initial cost of equipment are outweighed by the advantages for an ongoing, regular survey such as the LFS. Improved quality arises from the completion of editing in the interview, where inconsistencies can be checked with the respondent and the use of automatic routing to ensure that respondents are asked all of the relevant questions. Data capture and editing in the interview, and electronic transmission of the data, are the main contributions to improved speed of delivery results.

5.7 RESPONSE RATES

Panel survey non-response

As the LFS is a panel survey, the calculation of response rates should take the panel design into account. Households may refuse further participation at any of the five quarterly visits they are due to receive.

Households which refuse further participation are not revisited at the next quarter but they remain part of the eligible sample. The response rate for households comprises the ratio of the number of households responding at the current wave to the sum of the number of eligible households found at the same wave at the sampled addresses, plus any households which have refused outright to participate at a previous wave. Outright refusals (as distinct from circumstantial refusals and non-contacts, which are revisited at the next wave) may occur either when the interviewer calls or asks for permission to recall in three months.

The simple model above does not take account of such situations as net addition or subtraction of eligible households at sampled addresses at waves after wave 1, for example by a net increase or decrease in occupation of household spaces. However, the definition of household spaces is fluid and does not provide a firm basis for response rate calculations. LFS response rates which take the panel design into account are based on the simple model.

Figure 5.1 plots wave specific LFS response rates from autumn 1997 to Q2 2006. This includes both face-to-face and telephone interviewing and applies to cases in Great Britain only. Table 5.3 presents the data underlying this figure.

Figure 5.1: LFS Quarterly Survey, wave specific response rates, Great Britain, SN97 to AJ06

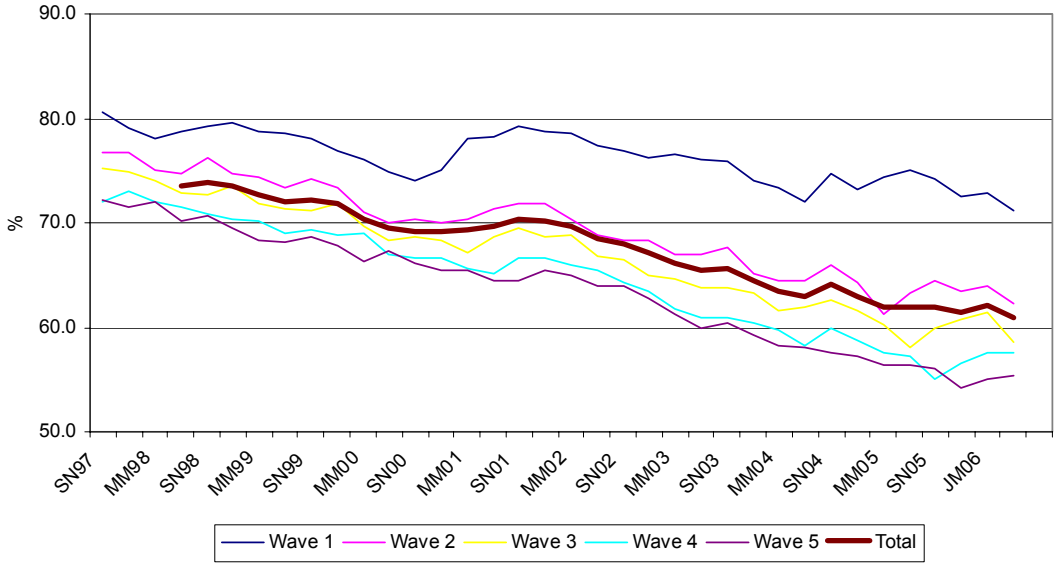


Table 5.3: LFS Quarterly Survey, wave specific response rates, Great Britain, SN97 to AJ06

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Total
SN97	*80.6*	*76.8*	*75.2*	*72.1*	*72.2*	
D97F	*79.1*	*76.7*	*74.9*	*73.0*	*71.5*	
MM98	*78.0*	*75.0*	*74.0*	*72.0*	*72.0*	
JA98	*78.7*	*74.7*	*72.8*	*71.5*	*70.2*	*73.5*
SN98	*79.2*	*76.3*	*72.7*	*70.8*	*70.6*	*73.9*
D98F	*79.6*	*74.7*	*73.6*	*70.3*	*69.5*	*73.5*
MM99	*78.7*	*74.4*	*71.9*	*70.2*	*68.4*	*72.7*
JA99	*78.5*	*73.3*	*71.4*	*69.0*	*68.2*	*72.0*
SN99	*78.1*	*74.2*	*71.1*	*69.4*	*68.6*	*72.2*
D99F	*76.9*	*73.4*	*71.9*	*68.8*	*67.8*	*71.8*
MM00	*76.0*	*71.0*	*69.6*	*69.0*	*66.3*	*70.4*
JA00	*74.8*	*70.0*	*68.4*	*67.0*	*67.3*	*69.5*
SN00	*74.0*	*70.3*	*68.7*	*66.6*	*66.2*	*69.1*
D00F	*75.0*	*70.0*	*68.3*	*66.7*	*65.5*	*69.1*
MM01	*78.1*	*70.4*	*67.1*	*65.6*	*65.5*	*69.4*
JA01	*78.3*	*71.3*	*68.6*	*65.2*	*64.4*	*69.6*
SN01	*79.3*	*71.9*	*69.5*	*66.6*	*64.4*	*70.3*
D01F	*78.7*	*71.9*	*68.7*	*66.7*	*65.4*	*70.2*
MM02	*78.6*	*70.3*	*68.8*	*65.9*	*64.9*	*69.7*
JA02	*77.4*	*68.9*	*66.8*	*65.5*	*64.0*	*68.5*
SN02	*76.9*	*68.4*	*66.4*	*64.3*	*63.9*	*68.0*
D02F	*76.3*	*68.4*	*65.0*	*63.5*	*62.8*	*67.2*
MM03	*76.6*	*67.0*	*64.6*	*61.8*	*61.3*	*66.2*
JA03	*76.0*	*66.9*	*63.7*	*61.0*	*60.0*	*65.4*
SN03	*75.9*	*67.7*	*63.8*	*60.9*	*60.4*	*65.7*
D03F	*74.1*	*65.2*	*63.3*	*60.4*	*59.3*	*64.4*
MM04	*73.3*	*64.4*	*61.6*	*59.7*	*58.3*	*63.4*
JA04	*72.1*	*64.4*	*61.9*	*58.3*	*58.0*	*62.9*
SN04	*74.7*	*65.9*	*62.6*	*60.0*	*57.5*	*64.1*
D04F	*73.2*	*64.3*	*61.6*	*58.7*	*57.2*	*63.0*
MM05	*74.3*	*61.2*	*60.2*	*57.6*	*56.4*	*61.9*
JA05	*75.1*	*63.2*	*58.1*	*57.3*	*56.4*	*62.0*
SN05	*74.2*	*64.5*	*59.9*	*55.1*	*56.0*	*61.9*
D05F	*72.5*	*63.5*	*60.8*	*56.5*	*54.2*	*61.4*
JM06	*72.8*	*63.9*	*61.5*	*57.6*	*55.0*	*62.1*
AJ06	*71.2*	*62.3*	*58.6*	*57.6*	*55.4*	*61.0*

Figure 5.2 and Table 5.4 show the number of responding households over the period SN97 to AJ06 while Figure 5.3 and Table 5.5 show the number of responding persons. Both figures and tables separately identify imputed cases from personal interviews.

Figure 5.2: LFS Quarterly Survey, achieved number of household interviews, GB and UK, SN97 to AJ06

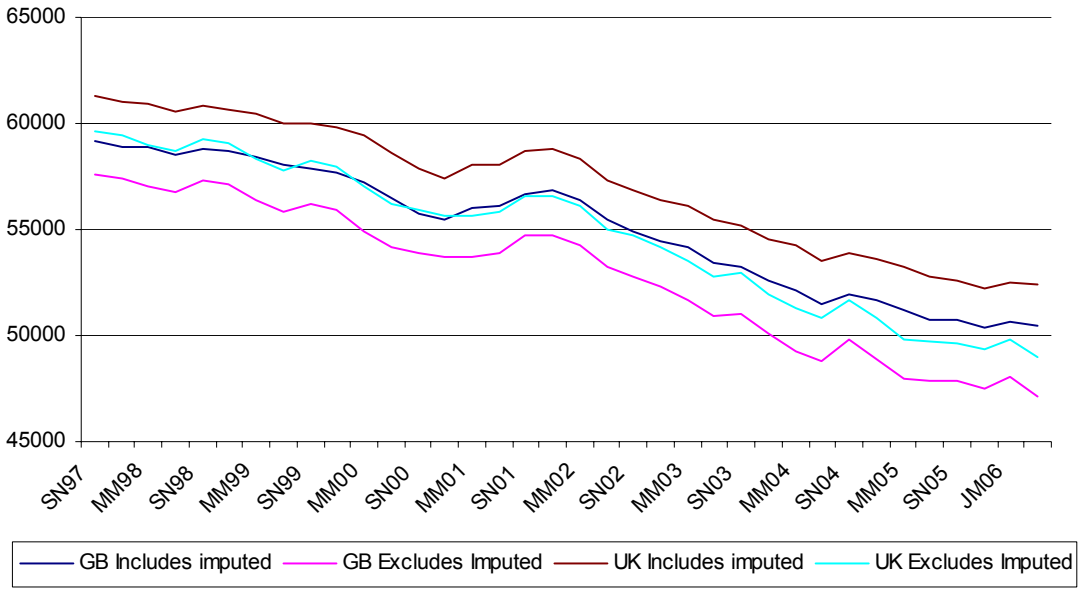


Table 5.4: LFS Quarterly Survey, achieved number of household interviews, GB and UK, SN97 to AJ06

	Total		Total	
	GB Includes imputed	GB Excludes Imputed	UK Includes imputed	UK Excludes Imputed
SN97	*59,122*	*57,597*	*61,273*	*59,659*
D97F	*58,921*	*57,405*	*61,046*	*59,461*
MM98	*58,843*	*57,051*	*60,913*	*59,000*
JA98	*58,525*	*56,785*	*60,581*	*58,741*
SN98	*58,774*	*57,320*	*60,825*	*59,273*
D98F	*58,692*	*57,113*	*60,685*	*59,042*
MM99	*58,406*	*56,380*	*60,418*	*58,321*
JA99	*58,035*	*55,843*	*60,022*	*57,762*
SN99	*57,867*	*56,173*	*60,033*	*58,281*
D99F	*57,653*	*55,969*	*59,781*	*58,007*
MM00	*57,190*	*54,911*	*59,429*	*57,034*
JA00	*56,437*	*54,174*	*58,576*	*56,249*
SN00	*55,773*	*53,877*	*57,895*	*55,936*
D00F	*55,439*	*53,663*	*57,453*	*55,620*
MM01	*55,992*	*53,722*	*58,028*	*55,661*
JA01	*56,119*	*53,914*	*58,077*	*55,805*
SN01	*56,710*	*54,700*	*58,685*	*56,602*
D01F	*56,877*	*54,738*	*58,788*	*56,585*
MM02	*56,410*	*54,223*	*58,354*	*56,118*
JA02	*55,476*	*53,248*	*57,330*	*55,038*
SN02	*54,929*	*52,819*	*56,840*	*54,681*
D02F	*54,465*	*52,273*	*56,389*	*54,152*
MM03	*54,124*	*51,653*	*56,102*	*53,494*
JA03	*53,464*	*50,924*	*55,439*	*52,806*
SN03	*53,261*	*51,033*	*55,229*	*52,921*
D03F	*52,583*	*50,099*	*54,512*	*51,937*
MM04	*52,126*	*49,303*	*54,226*	*51,293*
JA04	*51,453*	*48,809*	*53,547*	*50,794*
SN04	*51,911*	*49,782*	*53,905*	*51,696*
D04F	*51,674*	*48,912*	*53,629*	*50,799*
MM05	*51,187*	*47,967*	*53,195*	*49,840*
JA05	*50,772*	*47,833*	*52,742*	*49,721*
SN05	*50,708*	*47,905*	*52,555*	*49,656*
D05F	*50,394*	*47,535*	*52,253*	*49,335*
JM06	*50,612*	*48,017*	*52,471*	*49,813*
AJ06	*50,447*	*47,146*	*52,398*	*49,001*

Figure 5.3: LFS Quarterly Survey, achieved number of person interviews, GB and UK, SN97 to AJ06

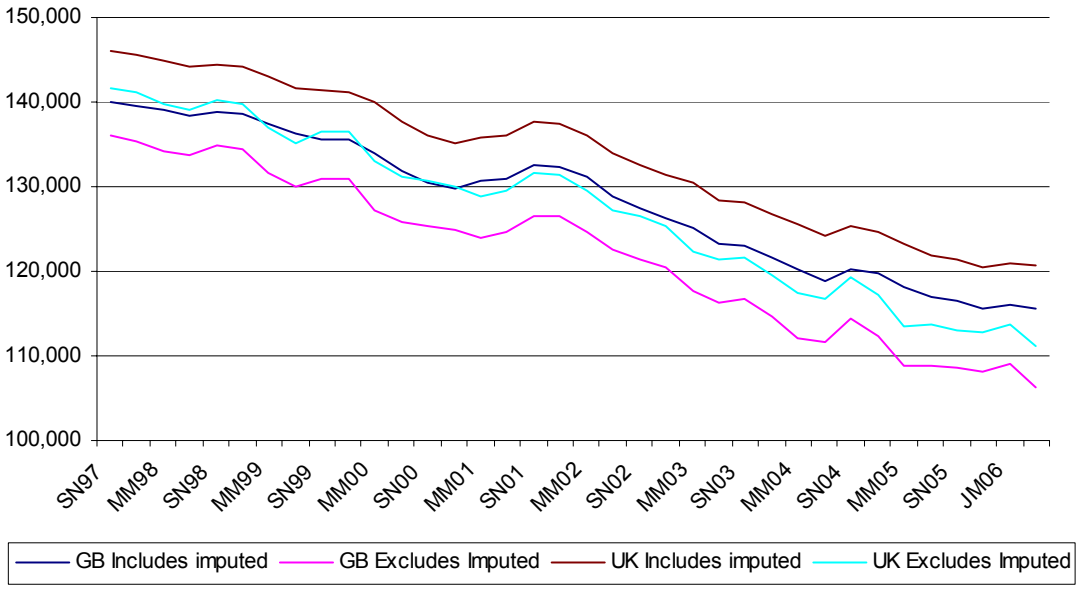
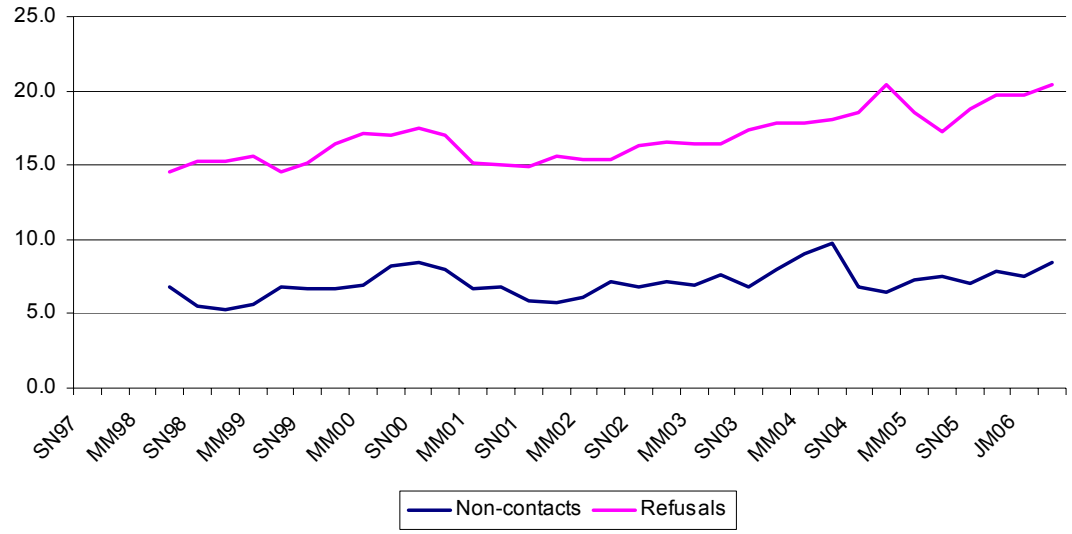


Table 5.5: LFS Quarterly Survey, achieved number of person interviews, GB and UK, SN97 to AJ06

	Great Britain		UK	
	Includes imputed	Excludes Imputed	Includes imputed	Excludes Imputed
SN97	*140,067*	*135,954*	*146,101*	*141,730*
D97F	*139,622*	*135,449*	*145,589*	*141,230*
MM98	*139,185*	*134,262*	*144,980*	*139,690*
JA98	*138,476*	*133,696*	*144,174*	*139,122*
SN98	*138,839*	*134,842*	*144,418*	*140,151*
D98F	*138,653*	*134,444*	*144,099*	*139,744*
MM99	*137,555*	*131,577*	*143,061*	*136,868*
JA99	*136,207*	*129,997*	*141,519*	*135,146*
SN99	*135,672*	*130,858*	*141,441*	*136,501*
D99F	*135,529*	*130,912*	*141,251*	*136,406*
MM00	*134,046*	*127,300*	*140,067*	*133,041*
JA00	*131,942*	*125,718*	*137,644*	*131,260*
SN00	*130,418*	*125,269*	*136,057*	*130,766*
D00F	*129,652*	*124,832*	*135,011*	*130,044*
MM01	*130,613*	*123,867*	*135,889*	*128,905*
JA01	*130,899*	*124,648*	*136,004*	*129,576*
SN01	*132,470*	*126,539*	*137,635*	*131,521*
D01F	*132,388*	*126,494*	*137,460*	*131,395*
MM02	*131,099*	*124,619*	*136,155*	*129,562*
JA02	*128,939*	*122,468*	*133,856*	*127,235*
SN02	*127,516*	*121,477*	*132,614*	*126,457*
D02F	*126,186*	*120,370*	*131,289*	*125,354*
MM03	*125,190*	*117,575*	*130,418*	*122,427*
JA03	*123,179*	*116,274*	*128,484*	*121,344*
SN03	*123,032*	*116,734*	*128,210*	*121,684*
D03F	*121,712*	*114,755*	*126,762*	*119,569*
MM04	*120,168*	*112,196*	*125,612*	*117,338*
JA04	*118,835*	*111,643*	*124,156*	*116,679*
SN04	*120,191*	*114,448*	*125,286*	*119,292*
D04F	*119,668*	*112,283*	*124,683*	*117,099*
MM05	*118,049*	*108,875*	*123,141*	*113,600*
JA05	*116,933*	*108,892*	*121,970*	*113,734*
SN05	*116,605*	*108,697*	*121,323*	*113,126*
D05F	*115,649*	*108,047*	*120,497*	*112,751*
JM06	*116,007*	*109,086*	*120,847*	*113,772*
AJ06	*115,591*	*106,321*	*120,651*	*111,111*

Figure 5.4: LFS Quarterly Survey, non-response, wave 1 interviews, GB, SN97 to AJ06



5.8 QUALITY CONTROL OF FIELDWORK

The LFS is a high quality product. This has been achieved over a period of time with methodological research and continued improvements to both fieldwork and management practices. For example, in order to minimise non-response, interviewers call back at non-contactable addresses a minimum of four times, two of which must be in the evening or at weekends. More recently all interviewers have been trained in Avoidance Refusal Training (ART).

Supervision and training are an important determinant of quality control too; the work of all interviewers is regularly monitored in respect of interviewing technique, dealing with the public, response rates, work efficiency, and the quality of completed work, including the accuracy of coding. Interviewers are provided with both verbal and written feedback on their performance. Where a weakness in performance is identified, additional training and monitoring is carried out.

5.9 NON-ENGLISH SPEAKING RESPONDENTS

Measures to meet the Welsh Language Act

Since 1985, all sample addresses in Wales are sent advance letters in both Welsh and English. A Welsh translation of the Purpose leaflet is also sent. Where a respondent requests that the interview be conducted in Welsh, arrangements can be made to transfer the household to a Welsh speaking interviewer. However, such requests are rare.

Measures to gain response from non-English speakers

All face-to-face interviewers are issued with a language identification card, containing a message written in the eight main foreign languages spoken in Great Britain: Greek, Turkish, Chinese, Bengali, Gujarati, Hindi, Punjabi and Urdu. The card is used to identify a time when an English speaking family member or friend can be contacted to explain the survey's purpose. Where there is no English speaker available, the card also enables interviewers to identify the language spoken so that the interviewer can arrange an interpreter.

Notes

[9] In the Blaise questionnaire this question is XMARSTA.

SECTION 6 - CODING AND PROCESSING THE DATA

6.1 CODING

All coding of data, where required, is carried out by interviewers. At present, coding of Country of Birth, Nationality, Ethnicity, place of residence three months ago and one year ago (if different from present), place of work, (in main and second jobs), and Subject of Qualification is performed using Computer Assisted Coding (CAC) during the interview. In addition, coding of Industry and Occupation (for main, previous and second jobs, job 3 months and one year ago (if different from present), and apprenticeships, is carried out by interviewers after the interview.

The following questions currently require coding and the method currently used is explained:

Variable	Description	Coding method:	
		by CAC during interview	by interviewer after interview
CRYO	Country of Birth	✓	
NATO	Nationality	✓	
ETHO2	Ethnicity	✓	
M3CRYO	Country of residence 3 months ago (if outside UK)	✓	
M3AREA/M3CTY	Place of residence 3 months ago	✓	
OYCRYO	Country of residence 1 year ago (if outside UK)	✓	
OYAREA/OYCTY	Place of residence 1 year ago	✓	
INDD/INDT	Industry in main job		✓
OCCT/OCCD	Occupation in main job		✓
RDINDD/RDINDT	Industry before redundancy		✓
RDOCCT/RDOCCD	Occupation before redundancy		✓
WKTOWN/WKCTY	Workplace of main job	✓	
INDD2/INDT2	Industry in second job		✓
OCCT2/OCCD2	Occupation in second job		✓
WKTOW2/WKCTY2	Workplace in second job	✓	
OYINDD/OYINDT	Industry in job 1 year ago		✓
OYOCCT/OYOCCD	Occupation in job 1 year ago		✓
SUBJCT/SUBJQ/CURSUB	Subject of qualification	✓	
APPD/APPT	Apprenticeship (continuing)		✓
APPIND/APPINT	Apprenticeship industry		✓

6.2 DERIVED VARIABLES

In order to analyse LFS data, a number of derived variables (DVs) are specified. DVs are created variables which combine the answers to two or more questions from the questionnaire. These do not vary significantly from year to year, although only those DVs which relate wholly to core data are created each quarter. DVs relating to non-core data are only created in the quarter(s) those non-core questions are included. All DVs are specified for the UK as they are created after the NI data are merged with GB data. Volume 4 contains the latest set of flow diagrams used to specify the current DVs, which can be split into four groups:

Person: DVs created for each individual record on the database;

Family: DVs created for each family on the database; on a flat file each family member would carry the same value for each family based DV;

Household: DVs created for each household on the database; on a flat file each household member would carry the same value for each household based DV;

Eurostat: DVs created for each record specifically to meet Eurostat requirements. These are not currently available to external customers: they are included here for completeness.

6.3 DATA CHECKING

Whilst some checking is performed in-the-field by the BLAISE survey instrument, other checks are carried out once the data have been received back from interviewers in the field or from the telephone unit.

The principles of the checks are to ensure that the data have no duplication of records etc., that the data have the correct household structure, in terms of persons in the household, and that certain key variables have valid values. These checks are important in maintaining the quality of the data. Examples include:

- checking that families have been correctly assigned within households by referring to the relationship grid
- checking that responses from a previous wave are consistent with the current wave on a number of key variables (eg. age, sex ...). These are known as 'imputation checks'
- country checks to ensure geographies are consistent across variables
- checks to ensure each household has a household reference person (HRP)
- checks to ensure that there is a record for every member of a household – known as the 'full house check'
- checks to ensure that the household type (hhtype) matches the household composition as defined in the relationship grid
- checks on ages for outliers (e.g. very old, old workers etc)
- cross checks between variables indicating activity status (e.g. Inecacar and Statr & Inecacar and Ilodefr)

SECTION 7- NON-SAMPLING ERRORS

The following section is drawn from the report 'Measuring and Improving Data Quality' by Vera Ruddock published as part of the GSS Methodology Series (no. 14).

The provision of accurate, timely data which meets the needs of users at minimal cost is at the heart of government statistics. There are two components to accuracy: sampling errors which occur when data from a sample is used to make inferences about the whole population (see section 8) and the so called 'non-sampling errors' which affect data from sample surveys, as well as administrative and census data. Non-sampling errors should not be viewed as mistakes, rather they are the result of conscious decisions to produce timely, accurate data at minimum cost.

Measuring non-sampling error is much more difficult than measuring sampling error because in many cases the reasons for the non-sampling error are not known, whereas sampling error is a direct result of the survey design and is under the control of the researcher. In some cases it may not be possible to measure non-sampling error or to only give an indication of its possible effect on the survey estimates. Non-sampling errors can also be very expensive to measure.

The rest of this section examines how accuracy of survey estimates can be measured and describes the different types of non-sampling error and their occurrence on the LFS.

7.1 MEASURING ACCURACY

Users of statistics commonly ask the question 'Is the estimate accurate?' The answer to this question influences the value the user attaches to the estimate, and the potential for the estimate to change the user's beliefs about a given subject. Accuracy is one concept, which defines the quality of a survey estimate. Accuracy reflects the difference between the survey estimate and the population parameter being estimated.

The question 'Is the estimate accurate?' is only the first part of the underlying question 'Is the estimate accurate enough for the purpose I want to use it for?' Discussions of the accuracy of estimates must therefore reflect the context in which they are to be used, but this assessment of the suitability of estimates for addressing specific issues requires some measure of the quality of the data. Accuracy is not usually reported; instead the error in an estimate is described by the bias and variance in that estimate, the two components of the total survey error.

Total Survey Error is the inverse of accuracy. A statistic with low accuracy will have high total survey error. Total survey error is measured by the mean square error, which is defined as the sum of all biases and variances:

i.e. MSE = variance + $bias^2$

The biases and variances may be due to sampling error, non-sampling error or both. Reported mean squared errors for survey estimates commonly only include the bias and variance attributable to sampling error. However non-sampling error can lead to biased estimates, for example in surveys people may systematically under report their consumption of alcohol. Similarly slight differences in the way respondents react to different interviewers may lead them to give different answers to different interviewers resulting in interviewer variance.

Bias

Sources of bias can be classified into errors of non observation and errors of observation. Errors of non observation include:

- coverage error. If the register or sample frame used to select the sample does not represent all the target population then the resulting sample estimate may be biased. For example a random sample of people who are in the telephone directory will exclude both those who:

- have no telephone and those who are ex-directory. If the value of a variable being measured in the survey is different for people who are and are not in the telephone directory then the survey estimate will be biased (see Sampling Frames in section 3);

- non response. If the people who do not respond to surveys are different from responders then estimates from the achieved sample may be biased estimates of population values - this can be corrected to some extent by weighting the sample.

Errors of observation include:

- social desirability effects: an unwillingness of respondents to admit to socially undesirable behaviour. This is most obvious in surveys of sexual experience when men commonly over-report and women under-report the number of sexual partners they have had in their lifetime.

Variance

An estimate of the variance of a statistic is based on the variability within the sample, which arises because achieved values differ over the units (e.g. sampled person, interviewers used, questions asked) that are the sources of the error.

Sources of variance include:

- Sampling variance:
  In sample surveys only a proportion of the population has been sampled. The sampling variance reflects the fact that the estimate may have been different if a different sample had been selected.

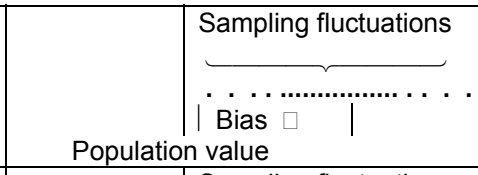
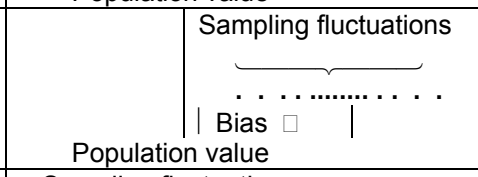
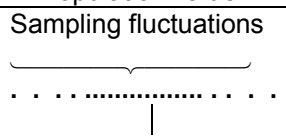
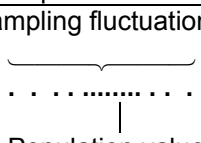
- Non-sampling variance:
  There are a variety of sources of non-sampling variance. For example differences between interviewers may consciously or unconsciously cause variation in the answers given by respondents; this is known as interviewer variance since it is due to differences between interviewers in achieved responses.

In the course of designing a survey many decisions are taken which may influence the relative size of different sources of error. These may reduce the bias in an estimate at the expense of an increase in the variance of the estimate. Alternatively survey designs which minimise one source of survey error may lead to an increase in another source of error. For instance, a common trade-off is the issue of whether to allow proxy responses in household surveys.

In the LFS adult members of a household are allowed to answer questions on behalf of absent members of the household. This minimises the extent of missing data (item non-response), but the quality of the data from proxy respondents is not always as high as data from the actual intended respondent so respondent error is increased. In contrast to the LFS, proxy responses are not accepted on the EFS, because of the very detailed nature of the survey, the result being a much lower response rate.

In some cases the use of proxies may introduce respondent bias into the results if they consistently underestimate variables such as household income, but in others some proxies will underestimate and others overestimate the true value leading to an increase in respondent variance. For more discussion of proxy responses on the LFS, see section 11.

The following diagram (from Moses) illustrates how bias and precision relate to distinct aspects of sampling procedure.

A Large bias, low precision	
B Large bias, higher precision	
C No bias, low precision	
D No bias, higher precision	

The dots in the diagram represent estimates of the population value derived from repeated application of the given survey procedures ie they represent the sampling distribution of the estimates and their mean is the expected value. The population value is what the survey is trying to estimate. The terms ‘large’, ‘low’ etc are, of course, relative.

It can be taken that (D), which is unbiased and relatively precise, is the ideal, whereas (A) is to be avoided. In practice the choice is not as simple as this, and there are circumstances in which a sample designer might be prepared to tolerate some bias if precision could markedly be increased.

7.2 TYPES OF NON-SAMPLING ERROR

Non-sampling errors - bias and variance - can be classified into three broad categories:

1. a) errors of non observation, which can be broken down into:
   - coverage error
   - non-response error

2. b) measurement errors, of which there is:
   - interviewer error
   - respondent error
   - instrument (or questionnaire) error
   - mode error

3. c) processing errors, consisting of:
   - systems error
   - data handling error

Each of these is described below, in relation to the LFS.

7.3 ERRORS OF NON-OBSERVATION

Coverage error

Coverage error is the error which arises because some units are either excluded or duplicated on the sampling frame used to identify members of the population of interest.

A sampling frame has 3 elements :

- a list representing all elements in the target population;
- further characteristics of these elements (auxiliary information);
- the probability of selecting each element on the frame.

The coverage ratio is the proportion of the target population included on the sampling frame. It gives an indication of the level of possible under-coverage, but does not measure the impact of under-coverage on survey estimates. Biased estimates can be caused by under-coverage and duplicate listings while increased variance of estimates can be caused by inclusion of non population elements in the list and errors in auxiliary information.

Coverage bias and variance can be measured by comparing data on the sampling frame with external data and by using special data collection procedures incorporated into the survey. Coverage error is minimised by using accurate up-to-date frames. Out of date lists can have the following impact on data quality:

- bias in survey estimates if new elements differing from elements already on the list have not been added to the list;
- increase in the variance of estimates if auxiliary information used for stratification or estimation is inaccurate, or it is discovered during a survey that sampled elements should not be on the list;
- reduction in survey response rate if elements are untraceable and it is not possible to ascertain that the elements are either old elements which should not be on the sampling frame or elements with inaccurate addresses which cannot therefore be traced.

For more detail on the LFS sampling frame and the way it is kept up-to-date, see section 3 of this volume.

Non Response error

There are two types of non-response error:

- Unit non-response: failure to obtain any of the substantive measurements from the sampled unit (the unit response rate is the proportion of the sampled population responding to a survey);
- Item non-response: failure to obtain specific items of information from an otherwise responding unit.

Non response bias in an estimate has two components:

- the proportion of the sample responding to the particular question;
- the difference between the true answer to a question in respondents and non respondents.

Even if the response rate is high, large differences in the true answer to a question in respondents and non respondents may lead to substantial non response bias. Non response can reduce the precision of survey estimates - this can be pre-empted at the design stage by increasing the size of the survey sample.

Unit non-response

There are four sources of information about non-respondents which can be used to examine the existence of unit non-response bias:

- information on the sampling frame;
- census records for responding and non responding units which can be matched to the sampling frame;
- information collected by interviewers in a follow up survey of non respondents;
- in panel surveys, information collected from respondents in earlier waves of the survey.

Strategies for minimising unit non-response include:

- interviewer training to reduce the number of refusals - interviewer training on the LFS is rigorous and all interviewers work solely on the LFS.
- encouraging interviewers to call on weekday evenings and at weekends - the timings of the LFS face-to-face and telephone interviews are managed in order to maximise the chances of gaining a response from a household, so much of the interviewing is done in the evenings.
- sending an informative well designed advance letter in interview surveys - households chosen for the LFS are sent a letter before their first interview which explains the background to the LFS, that the survey is voluntary and that responses will be treated as confidential, and gives a rough idea of when the interviewer is likely to call.
- reducing the burden on the potential respondent - the length of the questionnaire is reviewed regularly in order to keep the interview length down.
- offering incentives to respondents - while respondents to the EFS are given a monetary reward for completing a diary of their spending, no incentives or rewards are offered to LFS respondents.
- sending follow up reminders for postal questionnaires and making repeat calls in telephone/face-to-face questionnaires. Face-to-face and telephone interviewers will make a number of attempts to contact a household before it is treated as non-response.

For more information on LFS interviewing, see section 5. Weighting is also used on the LFS (see section 10) to compensate for unit non-response. The complicated population weighting allocates a weight to each individual, ensuring that the respondents are representative of the population as a whole, in terms of age, sex and region of residence. It also converts the sample estimates into estimates expressed in terms of the population.

Item non-response
An indication of the level of item non-response bias can be gained by comparing the characteristics of people responding and not responding to a particular question. Strategies for minimising item non-response include :

- clear question design;
- computer assisted modes of administering interviews to reduce routing errors and identify possibly erroneous data in the course of the interview - all LFS data is collected by laptop or PC, allowing a number of data checks to take place during the interview.

On the LFS, imputation (see section 12) is used to estimate missing items on a questionnaire so that the potential bias in estimates due to item non response may be reduced.

7.4 MEASUREMENT ERROR

There are four types of measurement error:

- interviewer error arising from both conscious and unconscious differences in the way interviewers administer a survey, and also from the reactions of respondents to different types of interviewers;

- respondent error arising from the inability or unwillingness of a respondent to produce a correct answer;
- instrument error which reflects the effect of question wording, response categories and form design on responses; and
- mode error which describes the effect of different methods of administering a questionnaire on the recorded responses.

Measurement bias can only be accurately measured in record check studies where the true value of a response is matched to the survey response. An indicator of measurement bias can be obtained from split sample studies where one component of the survey design is varied across subgroups of the sample. Different subgroups may :

- receive different questionnaires to investigate instrument bias; or
- have their interview administered in different ways, for example some may receive a face-to-face interview and others may fill in a self completion questionnaire to investigate mode bias.
- cognitive testing methods which ask respondents to questionnaires why they gave certain answers and attempt to understand the process leading to a response may be used to study respondent and instrument bias.
- measurement variance is important in interviewer surveys. High interviewer variance can have a large effect on the precision of survey estimates.

The different types of measurement error can be minimised using a variety of methods:

- interviewer error is minimised by thorough ongoing interviewer training and the use of small interviewer quotas to reduce the influence of interviewer variance on the precision of survey estimates;
- respondent and instrument error are minimised by careful question testing - new questions for the LFS are generally tested twice and feedback from the interviewers taken into account before the questions become part of the survey.
- mode error is minimised by using appropriate methods to collect data on sensitive questions - it is hoped that response bias on LFS earnings questions can be evaluated and perhaps reduced by asking respondents to check documentary evidence such as a payslip and recording whether such evidence was provided. There are also small but noticeable differences in the information collected by face-to-face interviewers and by telephone interviewers. Although some of the difference can be explained by respondents getting used to the interviewing process with each successive quarter's questioning, some of the difference is also due to the mode effect and it is difficult to disentangle the two causes. Estimates of employment are about 1 per cent lower, on average, in first interviews (face-to-face) than in subsequent interviews (telephone). However, as the survey design has not changed in recent years the estimates are consistent over time, and therefore estimates of change are unaffected by these effects.

7.5 PROCESSING ERROR

There are two types of processing error: systems error and data handling error. Systems errors are errors in the specification or implementation of systems needed to carry out surveys and process results; system errors on the LFS can creep in when derived variables are specified and/or amended . Data handling errors are errors in the processing of survey data.

There are various sources of data handling error:

- *Data capture*

  Information recorded on a paper questionnaire may be inaccurately converted to a format which can be interpreted by a computer. On the LFS, data capture is automatically incorporated into computer assisted interviewing modes of data collection, but interviews themselves may mis-key answers. This type of error on

the LFS is minimised by using mainly computer-assisted data capture with in-built checks.

- *Data transmission*
  Electronic data on interviews may be lost in transit between the field and the head office but this can be minimised by using an effective case management system to track the progress of individual packets of data.

- *Editing*
  Errors may be introduced when raw survey data is transformed into a dataset which can be used for producing estimates. These errors can be minimised by:
    - incorporating survey edits into computer assisted interviews so that the respondent can be asked about suspect responses - the method used on the LFS;
    - involving subject matter specialists so that the edits are appropriate for the data;
    - testing program code used in editing.

- *Coding*
  Coding is the transformation of textual open-ended responses to survey questions into categories to be used in data analysis. Coding systems may be manual, computer assisted - where the computer suggests a list of possible codes to the human coder, or computer automated. The last two of these methods are used on the LFS, particularly for industry and occupation coding.

  Individual coders may unconsciously show preferences for particular codes. The impact of these individual biases in the codes allocated by coders on survey estimates may cancel out, however although the survey estimate may not be biased, the variance of the estimate may be increased. If the individual biases do not cancel out then the coding error will introduce bias into the survey estimate. These types of errors can be minimised by effective training of coders in using the coding system.

- *Weighting and imputation*
  The use of inappropriate methods of weighting and imputation may introduce errors into survey estimates. See section 10 for more detail on LFS weighting and section 12 on imputation.

SECTION 8 - SAMPLING ERRORS AND CONFIDENCE INTERVALS

8.1 NON-SAMPLING ERRORS AND SAMPLING ERRORS

Surveys are prone to errors arising from a number of sources and processes. Frequently a distinction is drawn between *non-sampling errors* and *sampling errors*. Non-sampling errors are covered in section 7 of this volume.

Sampling errors relate to the fact that the sample chosen is only one of a very large number of samples which might have been chosen. It follows from this that an estimate of, say, the number of people in employment, is only one of a large number of such estimates which might have been made. The issue that is of interest to most users of survey data is the *precision* of an estimate - that is to say, the extent or range of the estimates which would (probably) have arisen from the different samples which might have been drawn. Greater precision is associated with a relatively narrow range, or expressed another way, the smaller the group whose size is being estimated, or from which an estimate is derived, the (proportionately) less precise that estimate is.

8.2 MEASURING PRECISION USING STANDARD ERRORS

A measure of the range of different estimates is provided by their *standard*. This is the *standard deviation* (the average amount of variation about the average) of the estimates) which would have arisen from the different samples which might have been selected (see diagram on page 28 cross reference). The smaller the standard error, the more precise is the estimate.

The size of standard errors is determined by a number of factors, including the sample size and the variability of the population from which the sample is drawn. The third important factor in determining the order of standard errors is the sample design. Standard errors calculated from simple random samples will, typically, differ from those calculated from more complicated sample designs, such as clustered or stratified samples.

A useful benchmark to assess the relative magnitude of a standard error is to calculate the ratio of the standard error derived from a particular (complex) sample design with the standard error that would have arisen from a simple random sample of the same size. This ratio (of the standard errors) is the *design factor*. It indicates the gain (or loss) in the estimate of standard error which results from the use of a particular complex sample design compared to a corresponding simple random sample. A design factor (or DEFT) of, say, 1.20 indicates that the standard error of the estimate in question is 20% greater than would have been the case for a simple random sample of the same size. The design factor (DEFT) should not be confused with the design effect (DEFF); the design factor is the square root of the design effect.

In the case of the LFS sample design, there is a clustering effect. This reflects the fact that addresses are sampled, but that results are shown for individuals. For example, ethnicity is particularly clustered, since it is likely that all members of a household living at a particular address will share the same ethnicity. This results in, for example, the design factor for the Pakistani and Indian ethnic groups being 1.71, which is higher than for the other ethnic groups because of the relatively large household sizes for Indians and Pakistanis. The design factor for part-time employees on the other hand is 0.95, reflecting the fact that part-time employee status is not clustered within a household.

By itself clustering would tend to increase the design effect of LFS estimates. However, the LFS sample design employs stratification. Since addresses are stratified by postcode sector there is a reduction in the standard error of estimates related to the factors used in stratification. For the standard errors of change and redundancy rates interviewer areas are used as strata.

The standard errors of the UK LFS estimates shown in Annex A are produced using a linearized jackknife approach by treating paired addresses (sorted by wave, quarter, quota, week and address number) as strata, and the address as a primary sampling unit (PSU).

It is also possible that the seasonal adjustment of LFS estimates will change their standard errors. The standard errors for the seasonally adjusted series should be smaller in the middle of the series but they will be larger at the end of the series because the seasonal adjustment process relies on forecasts.

8.3 STANDARD ERRORS AND CONFIDENCE INTERVALS OF ESTIMATES OF LEVELS

Whilst the standard error and the design effect of estimates are important items of information in their own right - because they indicate the precision of the estimate and the relative efficiency of the sample design in deriving the estimate - they also form the basis for calculating confidence intervals associated with particular estimates. A 95% confidence interval for a population estimate is about ±2 standard errors around the estimate calculated from the sample. The table below shows 95% confidence intervals for quarterly LFS estimates.

Standard errors, relative standard errors and confidence Intervals for Quarterly LFS estimates

Estimate	Standard error	Relative standard error *(%)*	95% Confidence Interval +/-	Confidence Intervals as % of estimate
10,000	2,000	*20*	3,900	*39*
12,000	2,100	*18*	4,200	*35*
15,000	2,400	*16*	4,700	*32*
20,000	2,800	*14*	5,500	*27*
25,000	3,100	*12*	6,100	*24*
30,000	3,400	*11*	6,700	*22*
35,000	3,700	*10*	7,200	*21*
40,000	3,900	*10*	7,700	*19*
45,000	4,200	*9*	8,200	*18*
50,000	4,400	*9*	8,600	*17*
75,000	5,400	*7*	10,600	*14*
100,000	6,200	*6*	12,200	*12*
150,000	7,600	*5*	14,900	*10*
200,000	8,880	*4*	17,200	*9*
250,000	9,800	*4*	19,300	*8*
500,000	13,800	*3*	27,100	*5*
1,000,000	19,500	*2*	38,200	*4*

The tables in annex B to this volume of the LFS User Guide list, for a number of estimates from the December 2004 to February 2005 LFS, the standard errors and other related statistics. The standard errors shown in the Labour Market Statistics First Release, LFS Quarterly Supplement and Annex B are shown for the UK and are calculated using UK design factors. For example, in Winter 2004 63% of people in the UK aged 16 and over were estimated to be economically active. The number of people aged 16 and over in the UK sample was 101,170 and since the design effect was 0.87, the standard error was 0.13%, calculated as:

standard error (given LFS sample design)	=	design factor	*	standard error (assuming simple random sample)
	=	deft	*	$\sqrt{(p(1-p)/n)}$
	=	0.87	*	$\sqrt{((0.63*0.37)/101,170)}$
	=	0.13%.(2dp)		

Hence a 95% confidence interval would be:

63% ±(1.96*0.13%) = 63% ± 0.25%

What this means in practice is that in 19 samples out of 20 we would expect the true economic activity rate to lie within the 95% confidence intervals constructed. Only in exceptional circumstances, in 1 in 20 samples, would we expect the true value of the economic activity rate to be outside the confidence interval around the LFS estimate.

The standard error of the total of the estimate is approximately the standard error of the proportion (or rate) multiplied by the population aged 16 and over:

0.13% * 47,076,144 = 61,199

and so the 95% confidence interval for the number of economically active people aged 16 and over would be:

29,657,970 ± 119,950

We use 95% confidence intervals primarily because they are widely used within ONS and elsewhere; they are something of an industry standard. However, other confidence intervals are equally valid, and might help users appreciate the fact that LFS estimates are always subject to error.

Confidence intervals are multiples of standard errors - the standard error of an estimate multiplied by 1 gives a 68% confidence interval; multiplied by 1.96 gives a 95% confidence interval. For example, the 68% confidence interval around the LFS estimate for the economic activity rate is +/- 0.13%. This means on 2 out of 3 occasions we would expect the true economic activity rate to be within the range 62.87% to 63.13%. The 68% confidence interval is narrower than the 95% confidence interval, but the trade-off is that we are less certain that the true estimate would lie within the band - there is a 1 in 3 chance that it would not lie within 62.87% - 63.13%, compared with a 1 in 20 chance that it will not lie within 62.75% - 63.25%.

Approximate method for standard error estimates of small estimates

It may be that confidence intervals are required for estimates other than those given at Annex A. A useful approximation for the 95% confidence interval of an LFS estimate of M thousand is:

$M \pm \sqrt{(M * 1.92)}$ thousands

This approximation, which works best for relatively small estimates, is derived as follows:

The basis is the equation $s.e.(p) = \sqrt{(pq)} / n$

(where $q = 1-p$ and n = number in sample)

If N = number in population and LFS estimate is M, then $p = M/N$ so $q = (N-M)/N \approx 1$

Then $s.e.(M) \approx N \sqrt{(p/n)}$
$\approx N \sqrt{(M/N) * (1/n)}$

$\approx N/n \sqrt{M/N * n^2/n} = N/n \sqrt{M * n/N}$ where $N/n \approx$ approx. 480 (average weight)

Take $M_T \approx M/1000$ (to give a convenient form for LFS estimates of M thousand)

Then $s.e.(M_T) \approx \sqrt{(M_T * 480/1000)}$

$\approx \sqrt{(M_T * 0.48)}$

Thus 95% Confidence Interval for $M_T \approx M_T \pm \sqrt{(M_T * 1.92)}$

This method may not work for estimates of total employed and total inactive as these tend to have very low defts due to the effect of post-stratification.

8.4 STANDARD ERRORS AND CONFIDENCE INTERVALS OF ESTIMATES OF CHANGES

In the same way that standard errors relating to quarterly estimates of means or proportions can be calculated, so standard errors can be calculated which relate to changes.

Standard errors (and hence confidence intervals) of estimates of changes are calculated as follows:

var $(p_{t1} - p_{t2})$ = var p_{t1} + var p_{t2} - 2kr (var p_{t1} + var p_{t2})/2 = (var p_{t1} + var p_{t2}) (1-rk)

where p_{t1} is the relevant proportion at time t_i

k is the sample overlap (0.8 for successive quarters)

r is the correlation coefficient

Since var p = deff * pq/n and assuming var p_{t1} = var p_{t2}, then, var $(p_{t1} - p_{t2})$ = deff * 2(1 – rk) (pq/n)

and hence s.e. $(p_{t1} - p_{t2})$ = deft * $\sqrt{}$ 2(1 – rk) (pq / n)

A few relevant standard errors are given in the table below:

Autumn 2004 to Winter 2004/05 (UK) (not seasonally adjusted)	Quarterly correlation	Quarterly change	Standard error of change in level
Economically active (working age)	0.92	-36,659	42,587
Employees	0.93	60,854	51,202
Self-employed	0.93	-27,720	30,383
ILO unemployed	0.59	-18,035	27,873
Economically inactive (working age)	0.93	71,970	41,937

Hence a 95% confidence interval for the change in the number of self-employed between Autumn 2004 and Winter 2004/05 would be calculated as:

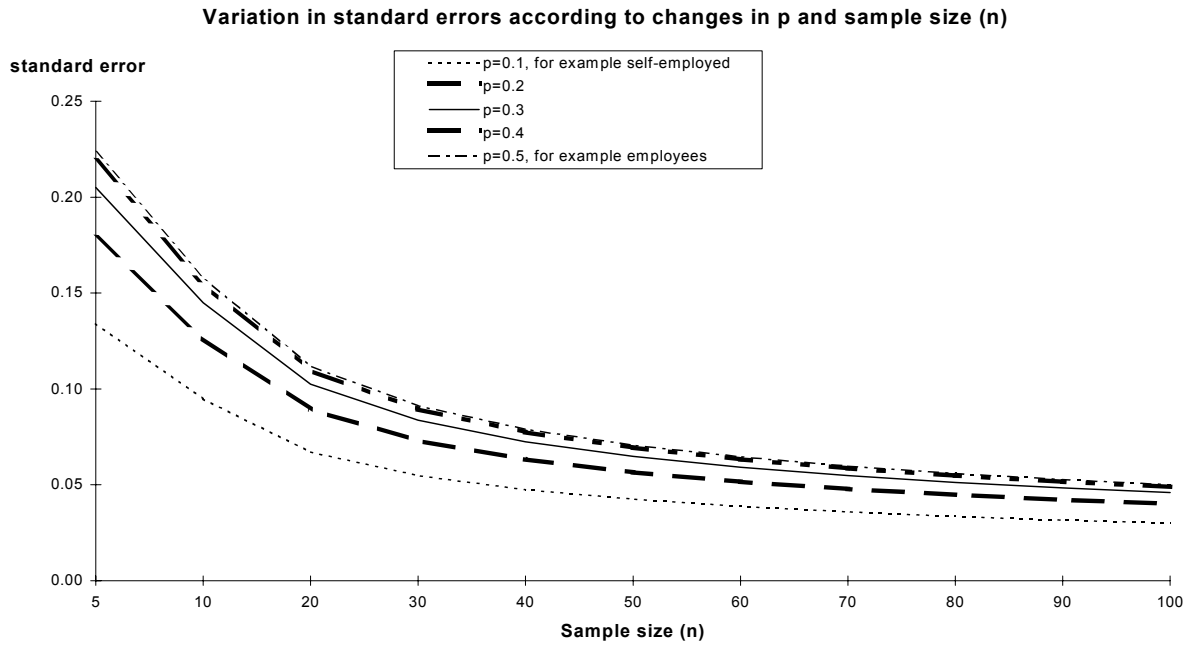
-27,720 ± (1.96 * 30,383) = -27,720 ± (59551)

8.5 RELIABILITY THRESHOLDS

It is the nature of sampling variability that the smaller the group whose size is being estimated, or from which an estimate is being derived, the less precise that estimate is. Put another way, the size of the standard error increases with the level of the estimate, so that the larger the estimate the larger the standard error. But the larger the sample estimate, the smaller will be the standard error in percentage terms (relative standard error being the standard error as a percentage of the estimate). Thus, larger sample estimates will be relatively more reliable than smaller estimates–an estimate of 500,000, while having a standard error of 13,800 will have a relative standard error of 3%, compared with an estimate of 25,000 which has a standard error of 3,100 and a relative standard error of 12%.

Before 2005, quarterly LFS estimates of under 10,000 were not published as they are likely to be unreliable. Although this publication policy changed in 2005 (as a result of the Freedom of Information Act), the unreliability of these LFS estimates did not. It is suggested that the 10,000 threshold is used as a guide to identify which cells will be subject to high sampling variability. Users are advised that estimates below 10,000 are subject to a high degree of sampling variability and should therefore be treated with caution. The 10,000 threshold equates to a sample size of about 25 and a relative standard error of about 20%. The graph below shows how, for different values of p (i.e. for different proportions of the population), the

standard error rises at a much steeper rate when the sample size is less than 30 - very small estimates, those based on less than 25 cases, are subject to such high standard errors (relative to the size of the estimate) as to detract seriously from their value, which is why the reliability threshold is recommended to be around that level. That said, a relative standard error of 20% is a somewhat arbitrary cut-off point, the Australian Bureau of Statistics uses a publication threshold of a relative standard error of 25%. A sample size of less than 3 is potentially disclosive so this information is suppressed even under the Freedom of Information Act.



Estimates of 10,000 have associated 95% confidence intervals of 10,000 ±4,000-so particular care must also be taken in using estimates of change from period to period for these estimates. For example, the estimated size of a particular group may be unchanged at 10,000 in two successive quarters but, because of the unreliability of these estimates, we cannot safely conclude that the true size of the group has not changed between those quarters. Using the results given earlier in this section, such a variable with a quarterly correlation of 0.9, would have a 95% confidence interval for its quarterly change of 0 ± 3000. So, even if the estimate does not change, the true size of the group could have changed by up to 3000 in either direction.

8.6 RELIABILITY THRESHOLDS FOR EARNINGS DATA

Earnings data from the LFS are based on only a part of the sample and so are subject to a different reliability threshold. Earnings is a continuous rather than a discrete variable; other things being equal, estimates of continuous data are more precise than estimates of discrete data. So although earnings estimates are based on a smaller sample size than most other estimates, LFS estimates of gross average hourly or weekly earnings for all employees are usually more precise (in terms of relative standard error) than an estimate of, for example, the number of employees.

However, to set an alternative reliability threshold for estimates of earnings (as opposed to estimates of numbers of people), based on maintaining a relative standard error of less than 20% is problematic for 2 reasons. Firstly, the standard errors of estimates of continuous variables like earnings are not determined by the estimates themselves (unlike estimates of numbers of people), so an appropriate threshold for one continuous variable will not work for another, even if their mean values are the same.

Secondly, standard errors of earnings estimates are dependent on the degree of variation within the sample. Thus, an estimate of earnings of a sub-group whose earnings are similar will have a relatively small standard error, whereas a group with more variable earnings will produce an estimate with a larger standard error, though both estimates may take the same value and be based on similar sample sizes. Estimates derived from discrete variables are affected by variability in a different way since, essentially, there are just two outcomes - whether a respondent is in a specified group or not.

Because of these problems, we recommend adopting a conservative approach by treating these estimates as though they were based on discrete counts. The recommended threshold when using LFS earnings data is set at 60,000 people for estimates based on wave 5 only (for quarters prior to Spring 1997), and at 30,000 for estimates based on combined waves 1 and 5 (for Spring 1997 onwards). As with non-earnings tables, it is suggested that these thresholds are used as a guide to identify those cells which are subject to a higher degree of sampling error, that a similar warning advising of this is added to the cells in question, and that they should be treated with caution. Both of these recommended thresholds equate to a sample size of about 25 respondents, as before. Much output from earnings analyses is in the form of means or proportions, so users need to be aware of the size of the population on which such estimates are based.

8.7 RELIABILITY THRESHOLDS FOR ANNUAL LFS DATA

For Annual LFS data prior to 2000-01 when the survey was enhanced, estimates of fewer than 6,000 are likely to be unreliable. However, since 2000-01, the nature of LFS enhancement has meant that some areas have seen a very large increase in sample size, and others very small increase or none at all. This means that a single threshold for all areas is no longer appropriate.

Following the enhancement in 2000-01, each area in England was allocated to one of three threshold bands – 2,000, 4,000 or 6,000. Annex C of Labour Force Survey User Guide Volume 6 Local Area Data, contains details of how this allocation has been made and Annex D contains details of which Unitary Authorities/Local Authority Districts fall into which band.

For Wales, from 2001-02, each Unitary Authority was allocated to one of four threshold bands - 1,000, 2,000, 3,000, or 4,000. Annex D of Labour Force Survey User Guide Volume 6 Local Area Data, contains details of the allocation.

Similarly for Scotland, from 2003-04 each Unitary Authority was given its own threshold Annex D of Labour Force Survey User Guide Volume 6 Local Area Data, contains details of this allocation.

It has long been known that the design effects for ethnic group and for totals segregated by ethnic group are substantially greater than those for most other groups.

Following an analysis of the variation in design effects between different ethnic groups and different local areas in England, it is recommended that for most ethnic estimates a single multiplier of 2.5 is applied to the general thresholds. A larger multiplier of 3.5 is recommended for totals of individual minority ethnic groups (e.g. the multiplier of 3.5 would apply to the total Indian adults in Birmingham, whereas a multiplier of 2.5 would apply to the total employed Indian adults in Birmingham). A separate analysis for the Welsh Local Labour Force Survey recommended a multiplier of 4.0 in Cardiff and 2.5 for the rest of Wales.

SECTION 9 – NON RESPONSE

This section presents information on the characteristics of non-responding households on the LFS using linked data from the 2001 Census. An earlier Census linked study was conducted after the 1991 Census, results of which appeared in the February 2003 of this User Guide.

9.1 THE 2001 CENSUS LINKED STUDY (BACKGROUND AND METHODOLOGY)

After the 2001 census, a study[1] was carried out to obtain information from census addresses that had been matched with the sampled addresses of the 12 continuous surveys undertaken by the ONS, including the LFS. This was possible because the census and the surveys are both carried out by the ONS so records can be identified and compared within the same organisation without infringing the confidentiality pledge to respondents.

The analysis for the LFS involved carrying out a comparison of the census characteristics of responding and non-responding households, and also separately of non-contacted and refusing households. Multivariate analysis was further carried out to identify those variables that are independently associated with response. In total, a sample of 10,437 addresses taken from April and June 2001 wave 1 LFS sample were included for matching to census records.

Summary of results

The study observed that households that were difficult to contact tend to:
- be located in London;
- contain one person;
- have one adult in employment;
- have no dependent children;
- contain people who did not form a family;
- live in a purpose-built flat or a converted/shared house;
- occupy accommodation containing between one and three rooms;
- rent their accommodation from a private landlord or
- have lived at the sample address for less than a year.

Households who refused to take part in the LFS were most likely to,
- be located in London or the South East or
- have no dependent children.

The Household Reference Person (HRP) of difficult to contact households tend to be,
- aged between 16 and 34 years;
- single (never married), separated or divorced;
- born outside the UK;
- from an ethnic minority group;
- resident at the sample address for less than one year or
- an employee or self-employed.

The HRP of refusing households tend to,
- have no academic qualifications or
- have qualifications other than a degree.

A logistic regression model was used in the study to identify the characteristics most strongly associated with non-response. This revealed that non-contact was most likely to occur in households,
- located in the Midlands, East of England, London and the South East;
- living in a purpose-built flat or in a converted/shared house;
- containing one adult only;
- containing no dependent children or

- whose HRP was single, separated, divorced or widowed, born outside the United Kingdom, an employee or self-employed. Furthermore, the age of the HRP had a significant association with non-contact.

Refusals was most likely to happen in households,
- located in the Midlands, East of England, London, South East, South West and Scotland;
- occupying accommodation that contained four rooms;
- containing no dependent children or
- whose HRP did not have degree level qualifications.

The results of the logistic regression model further suggested that interviewers were most likely to encounter non-response of any kind in households:
- located in the North East, Midlands, East of England, London, South East, South West or Scotland;
- living in accommodation containing fewer than five rooms;
- containing no dependent children or
- where the HRP was single, separated, divorced or widowed or did not have degree level qualifications or was self-employed or unemployed.

Table 1 summarises the results for the final models for the various elements of LFS non-response. It shows there is a close association between the characteristics associated with refusal and total non-response. Four of the six variables included in the model for total non-response are also included in the model for refusals.

Table 9.1 Summary of characteristics associated with elements of LFS non-response

Characteristic	Non-contact	Refusal	Total non-response
Government Office Region	✓	✓	✓
Type of building	✓		
Number of rooms		✓	✓
Number of adults in household	✓		
Number of dependent children in household	✓	✓	✓
Age of Household Reference Person	✓		
Marital status of Household Reference Person	✓		✓
Highest qualification of Household Reference Person		✓	✓
Country of birth of the Household Reference Person	✓		
Economic activity of the Household Reference Person	✓		✓

Further analysis suggested that household characteristics that are significantly associated with total non-response are also related to a number of key LFS measures. This indicates that non-response bias may result in the under-coverage of people in certain economic activity or with certain employment patterns and it may be desirable to re-weight the data to adjust for these effects.

Comparing LFS and census-based data on economic activity

The LFS is the source of internationally comparable measure of economic activity. The census-linked study therefore undertook additional analysis[2], amongst other things, to observe the LFS and census distribution of economic activity. This was achieved by comparing both the gross and net differences.

Gross agreement was considered to be good if 90 per cent or more people were classified to the same category on both the LFS and the Census. The study found a gross agreement of 93.4 per cent when economic activity was grouped into the main categories of employed, unemployed and economically inactive.

The net difference between LFS and census categories were considered large if the absolute difference was one percentage point or more, or the relative difference was 0.1 or more. The study observed a relative difference of less than 0.1 per cent in all but one of the categories – unemployed, which was 0.13.

The above comparison is useful to assess the potential of using the results of the census non-response study as a basis for developing a weighting scheme to compensate for LFS non-response.

LFS economic activity estimates adjusted for non-response

Survey estimates are subject to measurement error, non-response error and sampling error. An adjustment to the estimates is required to provide an alternative means to the Census for identifying likely bias in the economic activity estimates of the LFS. Adjusted estimates can be produced (by weighting and statistical tests) to minimise non-response and sampling error. The adjusted estimate is then compared with the unadjusted estimate to detect if the residual difference can reasonably be viewed as occurring through chance alone. If the residual difference is viewed as occurring by chance, this indicates that the estimates were unlikely to have been affected by non-response and measurement error.

The census linked study found that unweighted and weighted LFS economic activity figures based on the April and June 2001 LFS wave 1 sample were very close to the corresponding unweighted figures. The difference between the categories was not found to be statistically significant. This indicates that the differences detected can reasonably be viewed as occurring through chance alone and the unweighted estimates were unlikely to have been affected much by non-response bias and measurement error.

Changes in non-response on the LFS

The results of the 2001 census-linked study of survey non-response was compared with the results of the 1991 study to examine any changes in non response over a period of time. It showed that there had been no significant change in the overall non-contact rate on the LFS between 1991 (4.4%) and 2001 (5.1%). On the other hand, the refusal rate on the LFS increased significantly from 10.1% for the whole matched sample in 1991 to 14.2% in 2001.

The next table reveals that some of the factors[3] most significantly associated with non-response on the LFS in 2001 and 1991 are the same.

Table 9.2 Changes in the factors related to non-response

Characteristics	2001	1991
Non-contact		
Government Office Region	✓	
Accommodation occupied by the household	✓	✓
Number of adults and children (below 16 yrs) in the household	✓	✓
Number of cars in the household		✓
Household Reference Person/Head of Household		
Age	✓	✓
Marital status	✓	✓
Country of birth	✓	
Economic activity	✓	
Refusal		
Government Office Region	✓	✓
Number of rooms occupied by household	✓	
Number of dependent children in the household	✓	
Household Reference Person/Head of Household		
Sex		✓
Qualifications	✓	✓

Bias in the LFS responding sample

Comparing census data with the LFS sample also showed some bias that could arise as a result of non-response on the LFS. To do this, a correction factor was calculated for each category. The more a correction factor departs from 1.0, the greater the effect of non-response. Categories that are over-represented in the responding sample have factors of less than 1.0 while those with a high correction factors (1.10 or above) tend to have significantly low response rates and hence under-represented. Most of the correction factors for categories defined by single variables were in the range 0.95 to 1.05, indicating that the bias was relatively small. Table 3 shows selected household characteristics that are under or over represented in the LFS responding sample i.e. outside the1.05 to 0.95 range which illustrates the bias in the LFS.

Table 9.3 Household characteristics that are under or over represented in the LFS responding sample

Characteristics of household	Fully responding households (i)	All eligible households in matched sample (ii)	Correction factor (ii) / (i)
Living in a commercial building, caravans or other mobile or temporary structures	0.7	0.8	1.14
Residing in London	11.0	12.1	1.10
Lone parent with non-dependent children only	3.0	3.3	1.10
HRP is unemployed	2.4	2.6	1.08
Residing in purpose-built flat or maisonette	13.1	14.0	1.07
Household with 2 adults aged 16 yrs & over and 1 or 2 children aged under 16 yrs	14.4	13.6	0.94
Residing in North West and Merseyside	12.9	11.8	0.91

9.2 QUESTION SPECIFIC NON-RESPONSE

Sometimes data for a particular question is missing from the results because respondents who are routed to the question do not respond. The figures in Table 9.4 represents the number of people who did not answer particular questions as a percentage of those who were routed to the question. Figures relate to the Spring (March – May/April-June) of each year.

Table 9.4 Question specific non response rates

Variable	1998	1999	2000	2001	2002	2003	2004	2005
ETH01 *(1)* *(Ethnic origin)*	0.02	0.02	0.02	0.02	0.03	0.04	0.04	
EVERWK *(Whether ever had a paid job)*	0.03	0.04	0.02	0.02	0.03	0.04	0.04	
INDM92M *(Industry in main job)*	0.05	0.04	0.07	0.05	0.05	0.04	0.05	
SOC2KM *(2)* *(Occupation in main job)*	0.02	0.01	0.03	0.04	0.04	0.06	0.04	
MANAGE *(Management level)*	0.04	0.04	0.06	0.09	0.09	0.10	0.08	
MPNE02 *(3)* *(Number of employees at place worked)*	0.33	0.30	0.47	0.53	0.00	0.88	0.96	
SOLO *(Working on own or have employees)*	0.04	0.02	0.02	0.02	0.03	0.04	0.04	
FTPTWK *(Whether working FT/PT)*	0.02	0.01	0.02	0.04	0.03	0.04	0.04	
JOBTYP *(Whether permanent or not)*	0.04	0.04	0.04	0.03	0.04	0.06	0.03	
ILLWK *(4)* *(Whether sick in reference week)*	0.02	0.01	0.00	0.00	0.00	0.00	0.00	
EVEROT *(Paid or unpaid overtime)*	0.05	0.06	0.09	0.10	0.10	0.09	0.11	
TOTUS1 *(Total hours usually worked - no overtime)*	1.27	1.32	1.93	2.22	2.00	0.00	0.00	
USUHR *(Number of hours usually worked- ex overtime)*	0.34	0.41	0.81	0.75	0.85	0.00	0.00	
TOTAC1 *(Total hours worked in ref. week - no overtime)*	1.00	1.16	1.67	1.79	1.73	0.00	0.00	
ACTHR *(Actual hours worked in ref. week - ex overtime)*	0.39	0.48	0.55	0.89	0.79	0.00	0.00	
SECJOB *(Second jobs)*	0.02	0.01	0.03	0.03	0.02	0.02	0.03	
DIFJOB *(Whether looking for different or additional job in ref. week)*	0.02	0.03	0.04	0.04	0.02	0.05	0.05	
ED4WK *(Training in last 4 weeks - in employment)*	0.22	0.21	0.31	0.30	0.28	0.41	0.35	
FUTUR4 *(Training in last 4 weeks - not in employment)*	0.04	0.00	0.08	0.08	0.00	0.20	0.00	

Notes to Table 4

1. Prior to MM01 figures relate to variable ETHCEN.
2. Prior to MM01 figures relate to variable SOCMAIN.
3. From MM01 to D01F, figures relate to variable MPNE01. Prior to MM01 figures relate to MPNO.
4. Prior to MM00 figures relate to SICK

To conclude, various strategies are available to weight survey data to take account of non response. A census based weighting scheme is one approach. ONS currently uses another approach, a population weighting procedure, which involves weighting data to sub-regional

population estimates and then adjusting for the estimated age and sex composition by region. When evaluating whether additional information might improve the performance of a population weighting methodology, it is important not to lose sight of issues such as the availability of up-to-date population data at the sub-regional level and the relative transparency of the method.

[1] Freeth, S. (2004) Report of the 2001 Census-Linked Study of Survey Non-Response, ONS unpublished report.
[2] Freeth, S., et al. (2005) 2001 Census-linked Study of Survey Nonresponse on the LFS, Report of follow-up work requested by Labour Market Division, ONS unpublished report.
[3] Because of the changes to definitions, Standard Regions (used in 1991) are not identical to Government Office Regions, HOH is not the same as HRP and dependent children is not the same as the children (aged 15 years and younger).

SECTION 10 - WEIGHTING THE LFS SAMPLE USING POPULATION ESTIMATES

WEIGHTING THE MAIN LFS

Summary

Population weighting serves two purposes. First it enables tables showing population estimates to be produced, second it compensates for differential non-response among different sub groups in the population.

The LFS collects information on a sample of the population. To convert this information to give estimates for the population we must weigh the data. Each case is given a weight which can be thought of as the number of people that case represents. In a perfect world each person in the population would have an equal probability of being selected for the LFS (as in a simple random sample). The weight would simply be the reciprocal of this probability. However, because of differential non-response, some people are more likely to be in the sample than others. People with a lower probability of being in the sample, such as young people in London (see previous section), should have a higher weight.

It is impossible to measure directly what probability each member of the sample had of being selected. Instead, the population is split into sub-groups (or cells) where the number of people in each sub-group is known (based on population estimates). The known population weights are calculated by assigning each case in the sub-group the weight calculated by dividing the population in that subgroup by the number of cases in the sample in that subgroup. As a result the weighted estimates of the total population for that sub-group equals the actual population and the weighted estimate of the total population from the sample will equal the known estimate.

There is one limitation to this method. If the cells are too small, there is a possibility that none of the sample will come from that subgroup. The weighted estimate of the population in that sub-group will be zero and hence the estimate of the total population will be too small. To avoid this (and to make the task of producing the population control totals easier) a multi-stage population weighting procedure is used. Each stage corrects for a different cause of non-response: stage 1 corrects for non response at a local area level, stage 2 corrects for non-response amongst young people by age and sex; and stage 3 corrects for non-response by region, age (banded) and sex.

Description

(i) How the control totals used for current LFS weighting were derived

The LFS weighting methodology requires population figures for each LAD[1], with a five-year age-breakdown by sex, for each region. The starting point in the production of population weighting totals (generally[2]) is projections. All population projections (and estimates) are based, directly or indirectly, on the decennial Census of Population, and use additional information from the NHS Central Register for internal migration, the International Passenger Survey for international flows, and registration data for births and deaths. Projections use a variety of assumptions about the rates at which the components of population change will evolve.

[1] The LFS weighting control totals were calculated at a time, and using underlying population figures, when the administrative geography of the country comprised shire districts and counties, and metropolitan districts (and London boroughs). As mentioned in the article, this geography has been largely replaced following the introduction of unitary authorities.

[2] Population projections are normally the starting point. However, re-weighting exercises give the opportunity to use the most up to date population estimates as part of the LFS weighting. Estimates are also used in weighting the annual local area LFS databases.

A number of adjustments are made to the 'raw' data:

(i) population projections for shire districts are produced by rolling forward the latest growth rate (between estimates) for each LAD, then constraining to the published shire county projections.

(ii) adjustments to reflect the LFS (private household etc.) population are made as follows.

Estimates of communal establishment population have been made by assuming that the percentage of people in communal establishments was the same (by quinary age band, sex, and region) as it was in the results of the 2001 Census. Hence, for example, as the number of old people changes, the number of old people in institutions changes in line. Then the LFS population is calculated by subtracting the estimate of communal establishments from the total population figures.

(iii) monthly estimates are produced from the annual population figures - the mid-year estimate/projection less the communal establishment population - by simple linear interpolation.

(ii) Calculation of weights

The population weights are produced by dividing the population figures by the number of people interviewed in the survey. At its simplest a survey could be used to produce population estimates by giving the same weight to every individual record. Thus, if there were 1,000 respondents and the population from which they were drawn was, say, 5,000,000 then the weight to be applied would be 5,000,000 divided by 1,000, which equals 5,000. This might turn out to be unsatisfactory if, for example, the population contained exactly 2,500,000 women and the same number of men, but the sample contained, say, 600 women and 400 men (Perhaps because, say, the men had proved harder to get hold of, or were more inclined to refuse). If this was the case the tables would show the correct number in the total population, but they would incorrectly, show, 3,000,000 women and 2,000,000 men.

The solution is relatively simple. Instead of weighting all individuals with the same weight men and women can be weighted separately. In the above example the weight for women becomes 2,500,000 (the number of women in the population) divided by 600 (the number of women in the sample) which equals approximately 4,167, and for men it becomes 2,500,000 divided by 400 which equals 6,250 precisely (instead of 5,000 under the simple system). Further sub-divisions may be called for if it proves that the survey population differs from the parent population in terms of the proportions in different age groups, or areas of residence, etc. In principle, as more controls are added (sex, age, region of residence) it is necessary simply to add extra cells to the table containing the population figures.

However, as will be seen, this table increases in size as extra dimensions are added to it. Say that in our imaginary example we could use 2 sex groups, 17 age groups, 10 areas of residence as controls, then our control table would already contain 2x17x10=340 cells.

Were we to include as additional controls household size, simply as those in one person households against those in larger ones, and marital status, then our table will have to increase to 1360 cells, and we already have more cells than we have individuals in the sample. As a result, when we try to carry out our calculations we will find that we have a number of empty cells. If we ignore this we will find, when we run our tables, that the totals are now below the population counts, simply because we have no representatives of some of the population groups we are trying to cover.

The way to avoid the problem of empty cells is not to attempt to construct large matrices but instead to deal with the dimensions one at a time. This is the procedure followed in the Labour Force Survey and it is described in detail below.

(iii) The three-stage process

LFS population weighting involves three distinct stages and, in addition, several iterations of the process.

Stage 1: weighting to population estimates for individual Local Authority Districts;

Stage 2: weighting to national population estimates by sex for the ages, 0-15,16,17, 18, 19, 20, 21, 22, 23, 23, 24 and 25+;

Stage 3: weighting by sex for the following regions and age groups:

	Regions:	Tyne and Wear				Rest of Yorkshire and Humberside			
		Rest of North				West Midlands Metropolitan County			
		South & West Yorkshire				Rest of West Midlands			
		Rest of North West				Greater Manchester and Merseyside			
		East Midlands				South West			
		East Anglia				Wales			
		Inner London				Strathclyde			
		Outer London				Rest of Scotland			
		Rest of South East							
	Age groups:	0-4	5-9	10-15	16-19	20-24	25-29	30-34	35-39
		40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79
					80+				

Thus there are 17 regional groups, 17 age groups and the two sexes, giving a total of 578 cells for stage 3 weighting.

(iv) The iteration process

Stage 1 of the population weighting allocates a weight to each individual according to the LAD in which they live. So all people living in Cleethorpes are given the same weight, all people in Derby are given the same weight, and so on. The weight is derived by dividing the number of people in the sample for each LAD, into the population total for the LAD.

At stage 2, the weighted figures for each individual are read into the single years of age table and the process is repeated. In stage 3, the resulting weights are read into the age by sex by region table.

The problem is that the second and third stages can upset the corrections made in stage one and, of course, stage 3 can upset corrections in stage 2. Were we to stop at stage 3, therefore, the published results might show the "wrong" figures for local authorities and for single years of age. The solution is to iterate, that is to say, after stage 3 go back and do stage 1 again. If this process is repeated a sufficient number of times the resulting corrections gradually approach unity, so that the later stages are no longer upsetting the earlier ones.

The LFS iteration ceases when all the correction factors lie in the range .996 to 1.005 (rounded to three decimal places), so that they are all within the range 1.00 ± 0.01.

(v) Production of population figures

Population figures for the UK are produced by a variety of different organisations, at different periods. The following tables show these responsibilities, and the timing of recent and forthcoming production.

Responsibilities for producing population figures

Organisation	**Mid Year Estimates**	**National projections**	**Sub-national projections**
Office for National Statistics (ONS)	England, Wales		English regions and LADs
General Registrar's Office (Scotland) - GRO(S)	Scotland		Scottish regions and LADs
Government Actuary's Department (GAD)		UK, England, Wales, Scotland, Northern Ireland	
Welsh Assembly			Welsh LADs

Government			
General Registrar's Office (Northern Ireland) - GRO(NI)	Northern Ireland		NI LADs

Timing of the production of recent and forthcoming population figures

Year of release	Month of release	Mid-year estimates	National projections	Sub-national projections
1999	August	1998		
	November		1998-based	
2000	June			1998-based (England)
	March/April			1998-based (Scotland)
	August	1999		
	December			1998-based (Wales)
2001	August	2000		
	November		2000-based	
2002	May			2000-based (England)
	March/April			2000-based (Scotland)
	June/July			2000-based Northern Ireland
	October	2001		
	November		Interim 2001- rebased projections	
2003	August	2002		
	November		2002 based	
2004	August	2003		
	by October/November		2003 based	2002-based England
				2002-based Scotland
2005	August	2004		
	by October/November		2004 based	2003-based England
				2003-based Scotland
2006	August	2005		
	by October/November		2005 based	2004-based England
				2004-based Scotland

NB Quarterly Population Estimates are also produced in September.

Use of Population Figures for LFS weighting

National Statistics has recently carried out a re-weighting of LFS data using the most up-to-date population projections and estimates. Re-weighted LFS data were released in April 2002. An article in the May 2002 edition of LMT described the effects of re-weighting on LFS estimates.

The table below shows the population figures which were used for LFS population weighting at the time of re-weighting. Periods relating to the April 2002 exercise are shown in italics. Where two figures are given, these are interpolated between. For example, the population figures for Autumn 1999 are arrived at by interpolation between the 1999 and 2000 mid-year estimates.

Use of population figures for re-weighting LFS estimates

LFS Quarter	Best data at time of re-weighting/future weighting	
Summer 1993	1993 MYE	
Autumn 1993	1993 MYE	1994 MYE
Winter 1993/4	1993 MYE	1994 MYE
Spring 1994	1993 MYE	1994 MYE
Summer 1994	1994 MYE	
Autumn 1994	1994 MYE	1995 MYE
Winter 1994/5	1994 MYE	1995 MYE
Spring 1995	1994 MYE	1995 MYE
Summer 1995	1995 MYE	
Autumn 1995	1995 MYE	1996 MYE
Winter 1995/6	1995 MYE	1996 MYE
Spring 1996	1995 MYE	1996 MYE
Summer 1996	1996 MYE	
Autumn 1996	1996 MYE	1997 MYE

Winter 1996/7	1996 MYE	1997 MYE
Spring 1997	1996 MYE	1997 MYE
Summer 1997	1997 MYE	
Autumn 1997	1997 MYE	1998 MYE
Winter 1997/8	1997 MYE	1998 MYE
Spring 1998	1997 MYE	1998 MYE
Summer 1998	1998 MYE	
Autumn 1998	*1998 MYE*	*1999 MYE*
Winter 1998/9	*1998 MYE*	*1999 MYE*
Spring 1999	*1998 MYE*	*1999 MYE*
Summer 1999	*1999 MYE*	
Autumn 1999	*1999 MYE*	*2000 MYE*
Winter 1999/00	*1999 MYE*	*2000 MYE*
Spring 2000	*1999 MYE*	*2000 MYE*
Summer 2000	*2000 MYE*	
Autumn 2000	*2000 MYE*	*2001 special estimate*
Winter 2000/01	*2000 MYE*	*2001 special estimate*
Spring 2001	*2000 MYE*	*2001 special estimate*
Summer 2001	*2001 special estimate*	
Autumn 2001	*2001 special estimate*	*2002 special estimate*
Winter 2001/02	*2001 special estimate*	*2002 special estimate*

Notes:

1. Where two numbers are given these are interpolated between.

2. As part of the April 2002 re-weighting exercise, special estimates for 2001 and 2002 were produced to allow LFS results to be weighted to the most up to date population data possible (see article in February 2002 LMT for further details).

Population weighting and re-weighting

Re-weighted LFS data, based on the results of the 2001 census, from summer 2003 were published in 2004. However, Census 2001 data was then republished in Autumn 2003 but the LFS reweighting was then almost completed and it was decided not to begin again with the updated Census 2001 figures but to remain with the weights using the Spring 2003 Census 2001 data and use projections from that data for up to 3 years ahead or 2007, from the time of reweighting, for the current weights rather than MYE's and their projections.

To adjust for the earlier Census derived data being used and to match latest figures, ONS derives interim weights at the approximate GOVTOR level to ensure at least the aggregated estimates at this level match the latest MYE's and NPP's. ONS are have also produced Quarterly Population Estimates which allow the LFS to be weighted to the most up-to-date population MYE data and reduce the scale of revisions when the latest NPP's are released.

All microdata will be reweighted on the republished Census 2001 derived MYE figures back to 1992 at the earliest opportunity in 2007 or beyond and may be annually weighted to the latest MYE's and NPP's if possible so interim reweighting will no longer be required..

WEIGHTING THE EARNINGS DATA

Earnings data, collected at waves 1 and 5 interviews only (before Spring 1997 collected at wave 5 only), is weighted separately. The aim of the earnings weighting exercise is twofold: to weight the cases in the database in such a way that the weight of a sub-group corresponds to that sub-group's size in the population; and to weigh the sample to give estimates of the number of people in certain groups. The weighting exercise is restricted to employees' earnings: other income data are not (yet) weighted. During 1998, income weights for Northern Ireland data were added to existing datasets so that it is now possible to do earnings analysis at UK level from Winter 1994/5 onwards.

The main LFS database is considered to be the best source of data on the size of different groups. The weighting procedure thus attempts as far as possible to replicate the results of the main LFS in the weighted earnings data. However, because the earnings data is based on a sub-sample of the main survey, it is impossible to match the distribution of every variable in the survey in the earnings data. A small number of variables, likely to be important determinants of income (as stated in the December 1994 *Employment Gazette* article 'Income and earnings data from the Labour Force Survey'), was chosen for use in the weighting process. These were sex, age, region, occupation, industry, and whether full or part-time. The variables used are:

SEX	The sex of the respondent;
AGEBAND	Age coded into age bands (16-19, 20-24, 25-29,30-34, 35-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80+);
URESMC	Region of residence, with Metropolitan counties and inner and outer London separately identified;
SOCMAJM	Occupation (single digit SOC) in main job;
INDS92M	Industry (single digit SIC) in main job;
FTPTWK	Whether full or part time in main job.

Even with the limited number of variables, it is not possible to replicate the distribution of all variables simultaneously because there is the possibility that there would be "empty cells" (a certain combination of variables for which there are no matching cases in the earnings sub-sample but where there are cases in the main sample).

As with the main population weighting, a multistage approach is adopted where, for each stage, the distributions of a subset of the variables is matched. Each stage disrupts the distributions for variables matched in earlier stages, so only the variables in the final stage are matched perfectly. Overall the distributions for any of the variables included at some stage should be reasonable.

The five stages are as follows:

Stage 1: SEX, FTPTWK, INDS92M
weighting to reflect distribution of men and women, full-time or part-time, by industry category (drawn from the full LFS);

Stage 2: SEX, FTPTWK, URESMC
weighting to reflect distribution of men and women, full-time or part-time, by usual region of residence;

Stage 3: SEX, FTPTWK, AGEBAND, SOCMAJM
weighting to reflect distribution of men and women, full-time or part-time, by age band, by occupation in main activity;

Stage 4: SEX, FTPTWK, AGEBAND
weighting to reflect distribution of men and women, full-time or part-time, by age band;

Stage 5: SEX, AGEBAND
weighting to reflect distribution of men and women, by age band,

At each stage, the earnings sub-sample is aggregated into groups defined by the relevant variables, weighted by the main weighting factor. For each person in the earnings sub-sample, the count for the appropriate cell is divided into the corresponding cell for the main survey sample and multiplied by that person's weighting factor to obtain a weight to be fed into the next stage of the process.

Finally, as with the main population weighting, an iterative procedure is employed so that the corrections converge.

SECTION 11 - REPORT ON PROXY RESPONSE STUDY BASED ON LFS QUESTIONS

The following report "A study of proxy response in the Labour Force Survey" was an article written by Fiona Dawe and Ian Knight and was published in the Survey Methodology Bulletin (No.40), January 1997.

INTRODUCTION

The British Labour Force Survey (LFS), in common with labour force surveys in other countries, seeks information about all adults living in the households sampled for the survey.[1] In order to maximise response in the short fieldwork period available and to contain the cost of recall interviews, interviewers are allowed to accept information by proxy for those household members not available when the interview takes place. Martin and Butcher (1982)[2] showed that this compromise has a greater effect on some variables than others, but their study was undertaken many years ago and was not specific to LFS. Thus it was decided that Social Survey [has there been a study since 1995] should undertake a special study on key LFS variables in 1995.

Methodology

The basic approach for the study was to interview a respondent from each household who would provide information on other household members, **the proxy.** Subsequently the people on whose behalf the proxy informants had answered, **the subjects**, were themselves interviewed and their answers were then compared with the information given by the proxy informants.

The information collected from the proxy was not made available to the interviewer at the second (subject) interview, so the data collected from subjects was not affected by answers given at the proxy interview. The study used the same questionnaire as the 1996 LFS pilot to create a normal interview context but the analysis of proxy data was restricted to core LFS questions.

At the initial contact the interviewer screened the household to ensure that it contained at least two adult members; households with only one adult member were ineligible because proxy interviews could not be generated. The interviewer then proceeded to interview the first contacted adult and to conduct proxy interviews with that respondent for all the other adult household members. In the majority of cases the interviewer was able to collect data for all the household members at the initial contact. This prevented the proxy discussing the survey content with the subject prior to the proxy interview.

To minimise a time period effect the household was contacted again within ten days and personal interviews were carried out with all adult household members for whom information was collected from a proxy.

Overall, the study was based on the simple model that the reliability of proxy data can be represented by the extent to which the data from the two sources matched.

[1] The LFS fieldwork is carried out every quarter by Social Survey Division. The survey collects information primarily about the informants position within the labour market and their working patterns, as well as other topics such as educational achievements and basic income data. The survey interviews about 150,000 people living at a representative sample of some 60,000 private addresses throughout Great Britain every quarter. Respondents are interviewed five times at quarterly intervals, the last being on the anniversary of the first. The first interview is always conducted face-to-face, and the subsequent interviews (where possible) by telephone.

[2] Martin, J and Butcher, B. (1982): The Quality of Proxy Information - Some Results from a Large-scale Study. *The Statistician, Vol 31, No.4*

Sample selected for the study

A random sample of 500 households, containing at least two people, was drawn from the electoral registers from around Great Britain. In addition, households from the 1996 questionnaire pilot identified as being eligible for this proxy response study (ie. they contained at least two adult members and proxy interviews were carried out for all but one adult member) were included in the sample.

The aim was to carry out as many proxy interviews as possible and to follow up as many of those proxy interviews as possible with subject interviews. In the LFS proxy information is collected from the main respondent as long as he or she feels able to provide that information and is related to the subject, though proxies under 18 years old are not used without the parents permission. (Normally in the LFS about 30% of household member data is provided by proxy.)

However in this study as many proxies as possible were taken and that included people on whom the LFS would not normally rely for proxy information. In terms of their relationship to the subject the distribution of proxies was:

	%
Husband of subject	28
Wife of subject	37
Offspring of subject	11
Parent of subject	21
Other relationship	3

In the event 598 proxy interviews and follow-up interviews with subjects were achieved for 370 respondents.

General design and limitations of the study

Apart from its use of LFS questions, the current study differed from Martin and Butcher in that it was all conducted by telephone. This is because the majority of LFS interviews in Britain are undertaken by phone and the proportion of proxy interviews is a little higher in telephone interviews than it is in face-to-face interviews. The current study was based on a computer assisted telephone interview instrument rather than the conventional paper questionnaire used in the earlier study.

Moreover in the current study, due to organisational constraints and interviewer shift patterns, it was not possible to ensure that the same interviewer conducted the proxy and subject interviews at a household. This introduces the possibility of an interviewer effect, not only in the interviewer-respondent interface but also in the interviewer coding of complex variables like occupation and industry where interviewer interpretation of information is required. (This use of different interviewers to interview the same household on different occasions mirrors the way interviewers are allocated to interviews on the LFS. The probability of an LFS telephone interviewer contacting the same household at subsequent waves by telephone is very small.)
However in other respects this study did follow the Martin and Butcher methodology and was subject to some of the same limitations.

The model assumes that the survey responses would have been totally reliable in the absence of proxy interviewing: ie had the proxy subjects been interviewed twice themselves in respect of the same facts they would have given identical responses. In general this is a reasonable assumption though one might expect a small amount of subject variation causing some mismatch between answers given on two occasions. For that reason the study design included an attempt to measure that mismatch so that it could be subtracted from the total mismatch between proxies and subjects to reveal the true proxy effect. However, this part of the exercise failed because informants were unwilling to repeat the experience of a 30 minute interview that they had already undergone 1-2 weeks earlier. Nevertheless in looking at the

achieved level of match between proxies and subjects we must make some allowance for this **subject variation.**

Missing values caused by proxies

There are two reasons why data might be missing from a proxy:

i. the proxy informant was asked the question but was unable to provide an answer;

ii. the proxy informant answered a previous question incorrectly which affected the routing through the questionnaire and they were not asked the question;

Missing data from either of these two sources are a real loss and could lead to bias from using proxies. In the results which follow, these missing data will be quantified as a percentage of all proxies to whom the relevant question should have applied.

RESULTS: COMPARISON OF PROXY AND PERSONAL INFORMATION

The LFS collects data covering a number of different subject areas, this section uses key variables from each to assess the proxy informant's ability to provide reliable information within that area.

There are three types of variable involved presenting progressively more difficult problems of comparison between proxy and subject:

- relatively straightforward variables where there tends to be equivalent perception of the shared information between proxy and subject (even if the area of shared information is not always as great as we would like)

- more difficult variables where perception of shared information may vary between proxy and subject or where the two may describe the same thing in quite different words

- variables which require significant interpretation and coding by the interviewers, where there will be variation in the recorded codes for proxy and subject due to the actions of different interviewers

The results of the study will be discussed in order of these three levels of complexity.
The results have been analysed by comparing the answers given by proxies with those provided by the subjects at the second interview. The effect of that mismatch, or error, can be presented in two ways.

In comparing the number of proxies and subjects reporting a subject attribute it is quite likely that some of the positive and negative errors will balance out (ie a number of **A**s who answer as **B**s will be matched by a number of **B**s who answer as **A**s). Any residual error after some of these errors have cancelled each other out will be referred to as **net error**. But of course that self cancelling error in the distributions does not remove the individual errors and they are all counted when we calculate the number of matching answers given by the subjects as a percentage of the proxies who gave an answer for those subjects. For simplicity these results are labelled as the percentage of 'correct' proxy responses and the complement of those percentages is a measure of **gross error**.

Table 1 Percentage of proxies giving same answer as subject "(correct") and percentage of missing data

	"Correct" (Proxy same answer as subject)	Lost data
Age	99	-
Marital status	94	-
Economic activity status	93	-
Full-time/part-time work	94	5
Possession of 2nd job	96	4
Usual hours - precise	38	6
Hours last week - precise	21	4
Age of leaving full-time education	62	6
Training in last: 3 months	88	11
4 weeks	82	2
week	70	28
Gross income within 10%	66	29
(precise)	(20)	(29)
Qualifications	63	5
Occupation in major groups	75	4
(at 3 digit level)	(55)	(4)
Industry in summarised divisions	85	4
(at 5 digit level)	(69)	(4)

However, this proxy error can obviously only occur where proxy interviews are taken and if we allow for that, the overall level of error can be seen in perspective. Thus, for example, if gross proxy error for the estimate of training in the last 3 months is 12% and 33% of respondents who should answer this question are interviewed by proxy, then that gross error rate affects 0.12 x 0.33 = 0.04 ie 4%.

Table 2: Percentage of gross error from proxy responses in full LFS dataset

	Est. gross error rate in full LFS	Missing data
Age	0.3	-
Marital status	1.8	-
Economic activity status	2.1	-
Full-time/part-time work	2.0	5
Possession of 2nd job	1.3	4
Usual hours - precise	20.5	6
Hours last week - precise	26.1	4
Age of leaving full-time education	11.1	6
Training in last: 3 months	4.0	11
4 weeks	5.9	2
week	9.9*	28
Gross income within 10%	11.2*	29
(precise)	26.4*	(29)
Qualifications	11.1	5
Occupation in major groups	8.3	4
(at 3 digit level)	(14.9)	(4)
Industry in summarised divisions	5.0	5
(at 5 digit level)	(10.2)	(4)

*Where proportion of missing data was high this estimate should be used with caution because response error could be significant.

DISCUSSION OF FINDINGS

Demographics

Demographic variables like age and sex were no problem for proxies being correctly recorded in over 99% of all cases. However, de facto marital status which allows for actual partner arrangements (as opposed to legal marital status) was incorrectly stated in 6% of cases. Virtually all the errors appear to have been failures to report a household member in what he or she regards as a cohabiting relationship, mostly by proxies who were parents of the subject.

Economic Activity

The results show that 93% of proxies gave the same answer as their subjects for economic activity. Net error for any of the activity categories does not exceed 3% and for the smaller groups it was never greater than 1%.

Table 3 Economic activity: % distribution of proxy and subject responses

	Proxies	Subjects
Employee	62	64
Self-employed	7	8
ILO unemployed	5	6
All economically active	74	77
Economically inactive	26	23
Percentages to nearest whole no.	Base = 376	Base = 376

For employees there was a 5% gross error, ie 95% of proxies gave the same answer as their subjects; and for all people in employment the result was much the same. For self-employment alone the sample in this study was very small and the estimate of proxy error will not be very reliable but the figures show a net difference of 1%. The gross error rate of 25% is mostly accounted for by classification of self-employed subjects as employees. In the absence of a clear single definition of self-employment this is not very surprising.

Table 4 Economic activity: comparison of individual proxy and subject responses

Percentages in bold

Base numbers in brackets	Subjects:			
Proxies	Employee	Self-employed	ILO unemployed	Economically inactive
Employee	**60** (227)	**1** (5)	-	-
Self-employed	**1** (3)	**6** (22)	**1** (2)	**0** (1)
ILO unemployed	**1** (2)	-	**4** (16)	**0** (1)
Economically inactive	**2** (7)	**1** (2)	**1** (5)	**22** (83)

Although there was only a 1% difference in the proportion unemployed reported by proxies, this category contained the most gross error. Unfortunately even with a sample of 376 the number of subjects identified as ILO unemployed was only 23 (6%) which is not sufficient for a very reliable measure of gross error. Nevertheless the figures we have show 1 in 5 were reported by proxies as being economically inactive which suggests they were not aware of the subject's job search activity.

There was only a 3% difference in the proportion classified by proxies as economically inactive. Among employees temporary jobs are a problem area because proxies do not always know when a subject's current work has come to an end. And among the self-employed, the position of freelance workers who have no work at present can also be confusing for proxies. Thus it is not surprising that a small proportion of these groups are misreported as inactive by proxy. Overall the economic activity variable showed a gross error rate of less than 7% among proxies.

Full-time/Part-time, and Hours Worked

The LFS collects information both about the respondent's perception of their working pattern, in terms of whether they classified themselves as working full or part-time, and details about reference week hours and usual number of hours worked.

Subjects were more likely to work full-time than part-time (73% worked full-time). 94% of proxies were able to state accurately their subjects' full or part-time status.
Proxy informants were less able to provide precisely reliable information about the number of hours the subject worked per week. As might be expected, a larger proportion of proxy informants knew the number of hours the subject usually worked per week (38%) than knew the actual number of hours worked by the subject during the reference week (21%). The average number of usual hours worked reported by proxies (35.7) was 1.6 hours greater than that reported by the subjects they represented and the average hours worked last week reported by proxies (33.8) was 2.6 hours greater. In short although proxy informants are generally unable to provide exact estimates of the amount of hours worked per week, there is only a limited amount of net error and when the information is used to produce averages or in a banded form any bias introduced into the results is very small.

Second Job

There was 96% agreement between proxy informants and subject on whether the subject had a second job. Although there was a 4% gross error, the results showed no net error due to proxies.

Income

Questions about individual's income are asked at the end of the last (wave 5) LFS interview for all informants aged under 70. Employees and informants on a government scheme are asked questions about their gross and net income (based on their wage/salary when last paid), which is then used to calculate a weekly amount.

Income can vary from week to week and even the short time gap between proxy and subject interviews could cause some variation between the two. Without reference back to pay slips (which is difficult over the telephone) one can also expect some subject variation. Thus it is unreasonable to expect a precise match between proxy and subject and one should not be concerned that it only existed in 20% of cases. Martin and Butcher found similar results. Comparing the average income reported by proxies with the average reported by their subjects shows an average net error of +£3.25.

In the current study the more realistic but still stringent test of matching within 10% was set and 2 in 3 proxies matched with their subjects at that level. When proxy responses were restricted to spouses or cohabitees, the match was improved to around 80% when husbands answered for their wives and over 70% when wives answered for their husbands. (In the LFS the majority of proxies are women answering for their partners or parents for their offspring.)
Apart from the problem of matching, the proportion of missing data for income through proxy inability/unwillingness to answer is higher than for most other variables. Thus if one is seeking a precise measure of earned income from LFS the results of this study suggest that proxy non-response error is likely to be as important as proxy response error.

Training

Information is collected, within the education section of the questionnaire, about the informant's participation in work-related and educational training over three different retrospective time periods: 3 months, 4 weeks and the week prior to interview. These questions also provide the opportunity to investigate whether or not proxy informants are able to provide retrospective data, and whether the length of time being considered has any bearing on this ability.

One might expect that a brief minor training event would be more easily forgotten by the subject and thus it might only be reported as training in the recent past. Moreover, the person undergoing such training might not change his or her domestic routine to accommodate it, nor bother to

mention it to other members of the household. In these circumstances proxies would be less likely to report such training than subjects and it might only be mentioned by subjects as training within the last week or 4 weeks[3]. Indeed for training in the past 3 months there was very little net error at all between the distributions for subject and proxy. Furthermore it does suggest that the shortfall in proxy reporting relates mainly to minor levels of training.

Education and Qualifications

The LFS questionnaire collects information about the informant's educational achievements and participation. In the current study respondents were asked to list all of their educational qualifications and from this list the highest qualification was determined.

The range of qualifications and levels is such that some classification is necessary before comparing subjects and proxies. The results for highest type of qualification held show nearly two thirds matching with significant net error from proxies understating qualifications.

For subjects with first degrees, over 80% matched compared with around 30% for those with low grade GCSEs (or equivalent school leaving qualifications) and vocational qualifications. These lower qualifications are particularly vulnerable to mis-reporting because they relate to obsolete examinations in many cases and we cannot rely on accurate reporting from the subjects themselves. Other research (Bradley et al, 1996)[4] has shown that respondents often fail to tell other members of the household (or the interviewer without a lot of probing) about low level qualifications which they no longer regard as having any value for them.

Proxy responses on precise age of leaving full-time education were subject to a gross error rate of over a third with those overestimates almost equalling underestimates and comparison of the averages shows a net error of only 0.2%.

Occupation and Industry

Details about informants' occupation and industry are collected using the following questions:

Industry: "What did the firm/organisation you worked for mainly make or do (at the place where you worked)?"

Occupation: "What was your (main) job (in the week ending Sunday the [date])?"
"What did you mainly do in your job?"

Clearly this implies a level of detailed knowledge which proxies will not always have at their command. But the potential for mismatch between subjects and proxies is not simply a matter of proxy knowledge.

The variety of ways in which different tasks and skills can be combined in a job is infinite and occupation is only useful as a statistical variable if it is fitted within a classification system. Not only will any such classification be too complex to cope with the variety of occupations but there will be some occupations which do not fit clearly into any category. Thus there is a possibility that different coders may choose different occupation codes for the same job description. Studies of coders show that **coder variability** of 20% or more occurs in the choice of detailed (3 digit) codes (see Martin et al, 1995 and Dodd, 1985)[5][6] Consequently comparisons of occupations coded at the 3 digit level, as described by subjects and proxies but coded by different interviewers, could not show better than an 80% match even if both respondents had described the job in the same words. The description of jobs is also liable to **subject variation** in that a person may describe their (same) job in different ways on two separate occasions, laying stress

[3] The proportion recorded as having had training in the past week was noticeably lower than for other time periods but the question was only put to a small sub-sample and the result may not be representative.
[4] Bradley M, Knight I and Kelly M (1996) Collecting qualifications data in sample surveys -a review of methods used in government surveys. *London, HMSO* (forthcoming)
[5] Martin, J., Bushnell, D., Thomas, R. and Campanelli, P. (1995): A Comparison of Interviewer and Office Coding of Occupations. *Presented to the American Association of Public Opinion Research, Fort Lauderdale, May 1995*
[6] Dodd, P. (1985): An assessment of the efficiency of the coding of occupation and industry by interviewers. *New Methodology Series, No. 14, Office of Population Censuses and Surveys, London*

on different activities, and thereby implying different detailed occupation codes are appropriate. Interviewers are trained to probe for full details in order to minimise this problem but paradoxically the more detail we gather the more difficult it can be to fit some jobs unequivocally into one of the detailed categories[7].

For all these reasons therefore, the variation in coded occupation between subjects and proxies of 45% at the 3 digit level was substantially due to other factors. Martin et al (1995) estimated that interviewer coder agreement at the 3 digit level is 74% so gross error at that level after allowing for that coder effect is: 0.74 x 0.45 = 0.33 = 33%

In fact LFS occupation data is rarely used at this level of detail because it produces many sub groups which are too small for analysis. At the aggregated **major group** (MG) level the variation between subjects and proxies fell to 25%. Again using the Martin estimate for interviewer coder agreement at MG level (86%), gross error after allowing for that coder variability is 21.5%.

Many of the problems with coded occupation data apply equally to coded industry data. There is less room for variable description of what the organisation makes or does than of the individuals job, but there is room for ignorance on the part of both subjects and proxies. For example the main activity of the organisation might not be reflected in the main activity of the workplace or unit in which the subject works. Even employers can get confused by this[8].

Dodd (1985) produced an estimate for agreement between two coders for detailed industry codes at 74%. Thus the gross error of 31% at the detailed (5 digit) level of coding can be adjusted in the same way as occupation codes (see above): 0.74 x 0.31 = 0.23 = 23%

Again however, when the detailed codes are collapsed to the industry 'divisions' commonly used for analysis the gross error is reduced to only 15% and after allowing for Dodd's estimate of interviewer coder variability (86%) that gross error rate drops to 13%.

Variation in proxy accuracy with relationship to subject

The study looked at the relationship of proxies to their subjects to see if there was an ideal across all variables. As Table 5 shows, there was not. Parents were the poorest proxies for *Economic Activity Status* but the best for *Highest Qualification;* spouses were much better than other proxies for income.

[7] This problem could lead to panel surveys showing a high level of spurious occupational mobility. To avoid that, the occupation questions in the LFS use **dependent interviewing** whereby respondents are told which occupation was coded from their response in the previous wave interview and asked whether their occupation has changed since then.

[8] Even if the respondent is a senior manager in an organisation and knows a lot about the goods and services it provides, mistakes can still arise. Particular sub-sections of the organisation may be engaged in an activity quite different from the main activity of the enterprise. If such a sub-section is the main activity within a particular establishment then that local unit may well be classified according to its own main activity though it is not the main activity of the organisation.

Table 5: Percentage of correct answers from proxies by relationship of proxy to subject

		Relationship of proxy to subject*			
	All proxies	Husband	Wife	Offspring	Parent
Econ Activity status	93	94	95	93	86
Occupation in major groups	75	80	75	68	71
(at 3 digit level)	(55)	(62)	(55)	(50)	(51)
Industry - summarised divisions	85	89	84	82	85
(at 5 digit level)	(69)	(80)	(65)	(68)	(65)
Full/part-time work	94	93	97	100	89
Usual hours (precise)	38	44	35	43	35
Possession of 2nd job	96	100	95	93	94
Age of leaving f/t education	62	56	58	74	68
Highest qualification	63	62	62	56	92
Training in last 3 months	88	90	88	82	90
Training in last 4 weeks	82	87	93	82	79
Gross income from employment					
(within 10%)	66	79	72	42	50
Base	224-376*	57-105	86-138	25-43	52-79

*There were 11 proxies in a relationship to the subject other than those listed above but this group was too small for separate analysis.

Estimating the overall effect of proxy error

This study can usefully be extended to take account of the overall effect on the LFS by standardising the results to take account of likely proxy error in the sub groups where the PRS showed it to have occurred, following the method used by Martin and Butcher (1982).

This can be done with the simple model:

No of subjects (from reporting as themselves as wrongly having attribute X having	+	No. of proxy responses classifying respondents as having attribute X	+	proportion (from PRS) of those which proxies wrongly reported having alternative attributes when they really had attribute X	-	proportion PRS) of those proxies reported as attribute X

For example, using this form of standardisation the economic activity variable from the Winter 1996 LFS (unweighted) can be adjusted as follows in Table 6.

Table 6: Net effect of proxy error on economic activity variable

	Unweighted LFS results unadjusted for proxy error	**Unweighted LFS results removing proxy error**
	%	%
Employees	57.5	58.1
Self-employed	8.3	8.3
ILO unemployed	5.7	6.0
Economically Inactive	28.6	27.6

CONCLUSION

The agreement levels between information given by proxy informants and the same information given by the subjects themselves were found for many key variables to be above 80% and several were above 90%. Taking account of the proportion of LFS responses given by proxy, this implies overall gross error rates of around 1-5% for these variables.

Those variables requiring less straight-forward information (such as training in the last week and highest qualification obtained) and those requiring very detailed numerical information (such as hours worked and income) showed a less satisfactory match between proxy and subject responses which means higher gross error rates. Indeed where information is sought at a level of detail that people do not normally carry in their head it is unreasonable to expect a precise match and it is unnecessary for most purposes. For example, for gross weekly income, 66% of proxies were able to give an answer within 10% of the subject's answer and if such data were taken from spouse proxies only, the level of matching would have been much higher. A bigger problem for such variables is the high level of “don’t knows.”

In short, the reliability of proxy data, for areas where the proxy informant is required to provide precise numerical answers (e.g. hours worked and income), increases with a decrease in the required level of detail. If the data is used in a banded form, or to calculate averages the match of proxy and subject data is greatly increased.

The study also investigated whether it was possible to identify a key relationship of proxy to subject that would guarantee a lower rate of proxy error. However, whilst spouse proxies were better for some variables like occupation they were worse for others like qualifications. Rather than attempting to find an "ideal" proxy informant in terms of personal characteristics, it would seem that the best placed household member to provide proxy information is the person most affected by the subject's actions. In general, no single type of household member is able to supply reliable proxy information for all questions, though as suggested earlier the income variables would be more reliable if restricted to spouse proxies.

Proxy error for coded occupation and industry data looked higher than for most variables but after making allowance for variation due to different interviewers coding the data, the error was as low as most other variables.

SECTION 12 - IMPUTATION IN THE LFS

In the earlier section regarding dependent interviews it was noted that for many quarters, responses may be rolled forward (for one quarter only) if a respondent is unavailable. This is referred to as 'imputation'.

The following examines some of the implications of this. The first part deals with the situation that arises in the case of non-core questions (which are not asked in every quarter). If a respondent is unavailable in the latest quarter, then the variable will be coded as DNA (Does Not Apply - there will be no data to 'roll forwards'). A procedure has been established to separate these 'non-responding' DNAs from 'genuine' DNAs.

The second part of this section looks at the imputation methodology used when new ethnicity questions were introduced in spring 2001.

The final part reports on work conducted to examine the extent to which the use of imputed data on the LFS leads to estimates which depress estimates of change.

IMPUTATION AND NON-RESPONDING DNAS

When running LFS tables the DNA ('Does Not Apply') category may be unexpectedly large. This is because certain questions are not asked every quarter (see list overleaf) and some respondents are not contacted in successive waves.

If respondents from one quarter are non-respondents in a subsequent quarter (for wave 2 to wave 5 interviews) then data is carried forward from previous quarters. However, if the question was not asked in the previous quarter there is no data to bring forward, so the response to the question is coded as DNA. As no current data is available for these non-respondents, one way to treat them would be to leave them in the population distribution as effectively “Not known” in the same way as the “Not answered” category is used. Alternatively, if the best estimates for the whole population are required, then by assuming that these cases with missing data have the same distributions as the respondents, they can be eliminated from the survey estimates. To achieve this, an additional weight is required.

To check whether there are non-responding DNAs the variable concerned should be cross tabulated with a variable called IOUTCOME to differentiate between 'genuine' DNAs and non-responding DNAs, which will be shown in code 6 of IOUTCOME (data brought forward from the previous quarter).

The process of imputing the non-responding DNAs is as follows:

(i) calculate:
$$\frac{\text{Valid response total}}{\text{(Valid response total) - (non-responding DNAs)}}$$

(ii) multiply each of the valid responses by this factor (exclude DNA)

Example of imputation of non-responding DNAs

The LFS questionnaire explains which groups of people should be asked each question. For example, in the case of SHFTWK99 the people asked the question consist of those who are:

IF	WRKING=1	did paid work in the reference week
or	JBAWAY=1	temporarily away from a job in reference week
or	OWNBUS=1	Unpaid work for own business in reference week
or	RELBUS=1	Unpaid work for relative's business in reference week
or	YTETMP=1	on Govt scheme with employer providing training
or	YTETMP=2	on Govt scheme on a project providing training
or	YTETMP=4	on Govt scheme and temporarily away from an employer or project
or	NEWDEA4=3 or 4 or 5 or 7	Employee in public / private sector, voluntary sector, environmental task force, assisted self employment
or	YTETJB=1 AND NEWDEA4 = 1, 6, 8, 9, or 19	Work done in addition to that done on New Deal Scheme

By filtering on these groups it is possible to produce the following table of SHFTWK99 by IOUTCOME.

UK - Spring 2004

SHFTWK99	**Base**	**Personal response**	**Proxy response**	**Data brought forward**
Base	27,954,666	16,951,266	8,697,633	2,305,767
Most of the time	3,655,725	2,466,805	1,188,920	0
Occasionally	854,045	566,634	287,411	0
Never	21,120,466	13,911,095	7,209,371	0
NA	18,663	6,732	11,931	0
DNA	2,305,767	0	0	2,305,767

As we would expect, having filtered on only those groups that are actually asked the question, the only DNAs that are picked up are those where the data has been brought forward due to non contact. This can be used as a check to see that no-one else (who should not be asked the question) is being inadvertently asked the question.

It is quite simple to calculate the weight required to adjust the estimates of the non-missing categories and eliminate the non-responding DNAs. The factor is:

$$\frac{27{,}954{,}666}{27{,}954{,}666 - 2{,}305{,}767} = 1.08989$$

This weight can then be used to multiply the frequencies of the valid codes as follows.

SHFTWK99

Base 27,954,666
Most of the time 3,655,725 x 1.08989 = 3,984,338
Occasionally 854,045 x 1.08989 = 930,815
Never 21,120,466 x 1.08989 = 23,018,985
NA 18,663 x 1.08989 = 20,341
DNA 0

Variables that may be, or may have been, affected by non-responding DNAs are:

ACCDAY4	CTRM	HSTQUL	OYMNGE	TELQA2	VCQPLO
ACCDNT	DAYSPZ	ILLWRK	OYMPR02	TELQB2	VOCQPL

ATFRM2	DEGQUL	LANG	OYSIND	TEMLEN	WCHDAY[1-7]
ATFROM	DIFQUL4	LANGD1	OYSOCC	TMEOFF	WCHFR
AWARE	DLTYP	LANGD2	OYSOLO	TMPPAY	WCHJB
AXFA	DRFP	LSSOTH	OYSTAT	TRATIR	WCHJB3
AXFB	DRIVL	M3CRY	OYSUPVI	TRDSAME	WCHMO
AXPA	EDINS	M3CRYO	PRIVEH[01-10]	TREFEMP	WCHSA
AXPB	ETHC	MAINDRV[1-10]	PTNCRE[1-2]	TREFT[01-11]	WCHSU
BANK	EVDAY	MATLVE	QALPL99	TRHR93	WCHTH
BHNOTA	EVEVE	MOVED	QLPLO99	TRLEAVE	WCHTU
BHNOTB	EVHM98	NEWQUL	QULADV[1-9]	TRNDAY	WCHWE
BHNOTC	EVNGHT	NOCUST	QULFUT	TRNFEE[1-5]	WKFRI
BHPAID	EVSAT	NOLWF	ROAD	TRNLEN	WKMON
BNKHOL[01-11]	EVSUN	NUMILL	RSAQUL	TRONJB	WKSAT
BNKHOLF	FEEIR[1-5]	NVQHI	SAMQUL	TRSITE	WKSUN
CARADV	FLED9D	NVQQUL	SCQUL	TRVDRV	WKTHU
CARADV2	FLEX9D[1-3]	NVQSAM	SHFTWK99	TRVMTH	WKTUE
CASHFUL	FLEXW[1-9]	NWNCRE[1-2]	SHFTYP	TRVTME	WKWED
CASHTIM	GNVQUL5	OMCONT	SMESIT	TUCOV	YNOTFT
CGQUL	GOBACK	OMROLE	SMEST2	TUPRES	YPTCIA
CHATT5[01-10]	HGHNOW	OYCIRC	TDIFEMP	TYPILL	YSTRTF
CHINF[1-2]	HGHQUL	OYCRY	TDIFT[1-6]	TYPVEH[01-10]	
CHPEO[1-9]	HOLS	OYCRYO	TECQUL	UNION	
CHPRI[1-2]	HOMED[1-3]	OYEQM3	TELEQA	USEVEH	
COTH	HOMED2[1-3]	OYFTPT	TELEQB	USUWRK[1-3]	

IMPUTATION OF ETHNICITY IN SPRING 2001

The recommended output classification of ethnic groups for National Statistics data sources was changed in 2001 to be broadly in line with the 2001 Census (see LFS User Guide Volume 5 for full details of the classification). From spring 2001, the LFS introduced new questions on ethnicity in line with this new classification.

In normal circumstances, where information about the respondent does not change between quarters, e.g. date of birth, or in situations where respondents could not be contacted in a later quarter, information from the previous quarter is rolled forward. With the introduction of new ethnicity questions to the LFS, there were no data to roll forward for respondents who could not be contacted. An analysis of respondents showed that they represented 6 per cent of the total population aged 16 and over. An examination of their known characteristics (using data from the winter 2000/01 quarter) showed that the imputed group contained a smaller proportion of white people, a higher proportion of men and a higher proportion of people aged under 25; they were more likely to be employed and less likely to be economically active than the rest of the respondents in that quarter.

Without treating missing values, analysis by ethnic group for this quarter would be misleading. For this reason, additional imputation procedures were adopted to ensure the greatest possible number of cases had the new ethnicity information present for the spring 2001 quarter.

Methodology

The work to correct for the quality issues in the spring 2001 files can be separated into four stages: augmentation, re-coding, model development and imputation.

Augmentation

Spring 2001 ethnicity data (questions Eth01 to Ethbl) were augmented with data collected in summer 2001. That is, cases where ethnicity was missing in spring, but for which a response was recorded in summer, data was fed-back to repopulate the spring 2001 data set.

Re-coding

'Other' type responses recorded verbatim were re-coded according to a provisional census coding schema. This code was used in conjunction with the response at the first question (Eth01) to derive a new six-point classification for each case. Some adjustment was needed to the outcomes to reflect the differences in questionnaire design between the Census and the LFS.

Modelling

Using adult cases where both new (spring 2001) and old (winter 2000/01) ethnicity was present, a predictive model for new ethnicity was devised. Taking old ethnicity as the best predictor of new ethnicity, an exhaustive 'chaid' analysis (using AnswerTree® software) further identified tenure, age and number of children in the family unit, as variables to be included in the model for some of the old ethnic groups. These breakdowns determined the imputation classes to be used in the imputation process.

Imputation

The remaining cases of missing new adult ethnicity were imputed using the computer package Stata®. A method of hot-decking imputation which randomly selects a donor case from within an imputation class was employed to populate the missing ethnic group values. This process was repeated five times to produce five replicate datasets to investigate the amount the final distribution of ethnic groups varies according to the imputation process.

The table below shows ethnicity of respondents (numbering 75,118) and imputed cases (numbering 3,129) for each imputation. It demonstrates that the amount of variation due to the imputation process was very small.

Ethnicity of respondents and imputed cases for each imputation

per cent

	First replicate	Second replicate	Third replicate	Fourth replicate	Fifth replicate
Ethnic Group					
White	94.82	94.81	94.81	94.82	94.82
Mixed	0.4	0.41	0.41	0.42	0.4
Asian	2.71	2.71	2.71	2.7	2.71
Black	1.5	1.5	1.49	1.49	1.5
Chinese	0.25	0.25	0.25	0.25	0.25
Other	0.32	0.32	0.33	0.33	0.32

IMPUTATION AND ESTIMATES OF CHANGE

The practice of imputation in the LFS - rolling forwards information from the previous quarter for non-respondents in the current quarter – can be criticised for depressing measures of change. In order to investigate this issue effectively it is necessary to use linked LFS databases.

Consider two consecutive quarters. Then we have full information for respondents contacted in both quarters. However for cases interviewed in the first quarter who failed to respond in the second, imputed values are substituted by rolling forward their answers from the first quarter. About 3% of cases have their values imputed in this way in each quarter. For some individuals, these imputations will be correct and for others they will be incorrect. If a large proportion is correct, then including them will lead to an improvement in the quality of the current quarter's estimates at the cost of only a small bias in the estimates of change from the previous quarter. Alternatively, if a large proportion of them is incorrect then the quality of both the current estimates and the change estimates will suffer.

Although we cannot be certain what the correct value is for a particular non-respondent, we can look at the speed of change among those who respond in successive quarters and at the answers given by these temporary non-respondents in subsequent quarters. It is also important to compare these temporary non-respondents with current respondents in order to assess whether dropping them from the survey, rather than imputing values for them, would create any larger non-response bias.

If non-respondents whose values are imputed resemble respondents to the survey (in terms of their employment status characteristics and propensity to change this status, for example), then it would be appropriate to weight for these non-respondents on the basis of the values and patterns of change observed amongst the respondent population – i.e. there would be nothing to gain by imputing values for them.

However, if non-respondents and respondents are sufficiently different from each other (on non-demographic factors), then information derived from the respondent population is unlikely to be successful in estimating the characteristics of the whole population.

Empirical Evidence

The rather limited evidence we have on this derives from a study undertaken using data from three quarters in 1992/93. The data examined were the numbers in the main economic status categories (mainly those in employment) for respondents in winter 1992/3, spring 1993 and summer 1993; and for winter 1992/3 and summer 1993 for the separate group who did not respond in spring 1993 (and whose data for this quarter were imputed).

The increase in employment for respondents was from 32,174 in winter 1992/3 (55.2% of respondents) to 32,312 in spring 1993 (55.4%) - an increase of 0.4%. For those non-respondents

in spring 1993, who had their values imputed from the previous quarter, there was obviously no (recorded) change – 1290 were recorded as employed in both quarters.
As one indicator of the level of real change in this latter group, the answers given in winter 1992/3 and summer 1993 were compared, the implication being that if little change is recorded over the longer 6 month period, then it is unlikely that such changes occurred in the two 3 month periods.

The change in employment for respondents was from 32,174 (in winter) to 32,487 (in summer) - up 0.97%. For those people whose data were imputed in spring, the corresponding figures were 1,290 (66.5% of imputed cases) in winter and 1,289 (66.4%) in summer. That is to say, there was virtually no change - which is, of course, what the imputation process assumes for the previous quarter.

So as there was little change in the numbers in employment between winter and summer for those whose spring data were imputed, and the change between winter and spring and between spring and summer for respondents were of the same orders of magnitude (increases of 0.4% and 0.6% respectively), then it seems unlikely that there were substantial counter-balancing moves between employment and the other states between winter and spring and between spring and summer for those whose spring data were imputed.

Although the effect of the imputation on the change in the sample numbers who were recorded as employed was trivial with these data, there is still an argument for avoiding *any* increased risk of bias. Rolling forward data from the previous quarter is only one method of dealing with non-response. In a situation like this, where non-response means that no data is available for the current quarter, the only realistic alternative to imputation is to rely on population weighting. This assumes implicitly that the characteristics of non-respondents are broadly similar to those of respondents with respect to economic status etc.

For these data, the economic status distribution (in winter) of those interviewed in spring and of those not interviewed and imputed in the spring are different, as shown below:

	Winter characteristics of:	
	those who responded in spring	those whose data were imputed in spring
Employees	47.5%	57.6%
Self-employed	6.7%	7.8%
ILO unemployed	5.8%	8.5%
Inactive	39.0%	25.0%

So, the group who were not interviewed in the spring and whose data were imputed from their winter responses had a substantially higher proportion of economically active individuals than the group who responded to the survey in the spring.

The implication of this finding is that to drop these non-responding cases and to rely solely on the population weighting used on the survey to deal with this type of non-response would lose valuable additional information from the survey and hence would probably reduce the quality of the current survey estimates slightly.

SECTION 13 - CONTINUITY AND DISCONTINUITY ON THE LFS

MINIMISING THE RISK AND IMPACT OF LOSING LFS CONTINUITY

Background

As a biennial or annual survey up until 1991, the LFS was principally valuable for the in-depth cross-sectional analyses of the labour market which it provided. Since its switch from annual to quarterly frequency in 1992, however, a wide range of users of the LFS have increasingly looked towards the survey as a source of time series as well as cross-sectional data. This change in emphasis in the analytical capability of the LFS has increased users' awareness of, and sensitivity to, loss of continuity.

In addition to the use of the LFS for monitoring changes in key labour market activity variables - employment, ILO unemployment, total hours worked etc - from quarter to quarter (using seasonally adjusted data), the survey is used for monitoring changes over time in a number of other aspects of people's behaviour which are of interest in various fields of government policy. Examples are: the progress made towards the achievement of four of the six National Targets for Education and Training; the education, employment and training of young people; the extent of job-related training provided by employers; the relative situations in the labour market of men and women, different ethnic minority groups, lone parents and of older workers; the impact of the Disability Discrimination Act; and the extent of Trade Union membership and recognition.

Against this background, the continuity of LFS time series has increasingly been of importance to users of the survey. This section describes a set of guidelines for seeking to ensure the continuity of LFS data - the guidelines themselves are in annex 1. Relevant issues are discussed under the following headings:

(a) possible causes of discontinuities in LFS data;

(b) the circumstances in which discontinuities in LFS time series may be justified

This background information supports the overall strategy for minimising the risk and impact of a loss of LFS continuity, contained in annex 1.

POSSIBLE CAUSES OF DISCONTINUITIES IN LFS DATA

It is important, at the outset, to emphasise the point that the central aim of the LFS is to categorise the adult population according to the main categories of - in employment, unemployed and economically inactive - and sub-divisions of these, defined according to the guidelines promulgated by the International Labour Organisation (ILO). These variables are to be regarded as the LFS "core" and accorded particular care in respect of their continuity.

Changes in the administrative arrangements for eligibility for unemployment-related or other social security benefits, in so far as they have an impact on people's labour market behaviour, may in principle, have an impact on the LFS measures of employment, unemployment or economic inactivity. For example, the switch from Invalidity to Incapacity Benefit, accompanied by the introduction of a more stringent qualifying medical test, might over a period induce a greater degree of job seeking activity in the labour market. However, such changes can never cause discontinuities in the LFS series, as long as the basis of the survey in terms of the ILO definitions remains constant.

ONS will, subject to resource constraints, investigate the impact of administrative changes on LFS estimates (such as those arising from the switch from Invalidity to Incapacity Benefit and the introduction of the Job Seeker's Allowance). However, it should be recognised that it will almost always be difficult to disentangle such effects from the impact of the general economic

or social factors which affect the LFS measures, and that it may not be practically possible to generate useful estimates of the impact of such administrative changes on the LFS estimates. In the case of the introduction of JSA, for example, a hypothetical impact over a six month period on the LFS measure of unemployment of the order of 35,000 (which was predicted to have been the approximate impact on the claimant count) would be undetectable in the context of estimates of quarterly changes in ILO unemployment for which the 95% Confidence Limits are ±58,000.

Discontinuities in LFS series have arisen in the past, or could arise in future, because of the following:

(i) Definitional changes

While changes in the benefit system do not cause discontinuities in LFS time series, any changes in the underlying definitions on which the estimates are based, clearly do. Prominent examples are: the switch to the current ILO definition of unemployment in 1984; the inclusion of unpaid family workers among the employed population in 1992; and the LFS re-definition, on DfES (now DfES) advice, of people in full-time education. In each of these cases, statistics have been published by ONS describing the impact of the changes on the LFS estimates.

(ii) Impact of switch from annual to quarterly LFS design

The revised 1996 LFS Historical Supplement describes and assesses the impact of a number of changes that were made in the LFS design, sampling frame and methodology, when the LFS was switched from annual to quarterly frequency. As far as possible, estimates are made of the magnitude of the discrepancies between both the annual and established quarterly surveys and between the introductory and established quarterly surveys.

(iii) Data processing effects

A number of processes need to be implemented in order to convert the raw returns from LFS interviews into the published estimates: data editing and imputation for missing values; changes in coding frames or classifications; sample weighting to known population controls; and seasonal adjustment. Major changes in the methods or external data used in these processes may in some, but not all, circumstances cause discontinuities in LFS series.

In the case of data editing, a number of minor improvements in LFS editing procedures were introduced at the interviewer stage and this enabled a greater degree of cross-checking of the validity of data with the respondent. The aim was to increase the quality of the LFS data relating to households and families, and no major discontinuities occurred.

In the case of imputation a particular feature of the LFS - which has been shown to be beneficial for the quality of the data - is that missing responses for people still resident in the sampled household are substituted by values carried forward from the responses made for the same person in the previous quarter. Where new questions are introduced, or amendments are made, however, this process may not function and a discontinuity may, potentially, arise because of an increased level of question non-response.

In the case of changes in coding frames or classifications the introduction of a new standard nomenclature, such as the 1990 Standard Occupational Classification or the 1992 Standard Industrial Classification, can cause disruption to time series. In the latter case, ONS was able to create a key series retrospectively on a consistent basis. SOC 2000 was introduced in March 2001.

In the case of sample weighting the impact on LFS estimates caused by changes to the population controls used for the survey, which were introduced to take advantage of the improved series of post-1991 Census population estimates, was catered for by ONS by re-issuing a complete new set of LFS estimates and databases from 1984 to 1993, and to take advantage of more-up-to-date population projections, a re-weighting of LFS estimates and databases from Autumn 1993 to Autumn 1999. Then in 2001, LFS estimates back to Autumn 1998 were re-weighted. Recently, all LFS datasets from Autumn 2003 back to 1984 were reweighted, incorporating the population estimates based on information from the census carried out in 2001. As part of the normal production round, data from winter 2003 onwards was brought in line with census 2001 based population estimates and therefore consistent with the back series for 1984 to autumn 2003.

Future changes in the methodology for sample weighting are a potential risk to continuity. For example, it may be possible to increase the quality of LFS estimates by enhancing the weighting process to allow for household structure by constraining individual weights within households to be equal. A discontinuity in LFS series need not arise; the new procedure may simply reduce the sampling errors of the LFS estimates.

A review of LFS seasonally adjusted data is conducted in the spring of each year and a full series of revised estimates is published back to 1992. A continuing programme of improvement to the methodology used for LFS seasonal adjustment is underway but no discontinuity effects are anticipated over and above the normal annual review process.

(iv) Questionnaire changes

Discontinuities in LFS series can arise, sometimes in unexpected ways, if the questionnaire is changed in order to collect new data or to improve the quality of existing items. While such effects may most obviously occur in the time series from an existing question if it is changed, more subtle side-effects may occur in the time series for other, related questions. Alterations in question routing may also, in some circumstances, have an impact.

Prominent examples of these sort of cases are: the inclusion of a 13-week job-related training question which, almost certainly, improved the quality of, but caused a discontinuity in, the existing 4-week job-related training information (by filtering out wrongly included training occurring outside the 4 week reference period); and changes to the LFS qualifications questions which improved the quality of the data collected but also introduced discontinuities. In neither case, does an obvious method of estimating the extent of the discontinuity exist.

CIRCUMSTANCES IN WHICH THE INTRODUCTION OF DISCONTINUITIES MAY BE JUSTIFIED

Clearly, the introduction of discontinuities to LFS time series (see annex 2) is, in itself, usually undesirable because of the potential disruption which may be caused to users of the data. However, there are circumstances where the advantages of making changes which may cause discontinuities over-ride the disadvantages, or there are external factors outside the control of ONS. The guidelines for dealing with discontinuities envisage ONS, in consultation with OGDs and, where appropriate, other LFS users, assessing the balance between the benefits and disadvantages in each case so that a view of how to proceed can be formed.

One over-arching issue in considering whether a discontinuity might be justified concerns the importance of the series affected. Arguments exist to support the view that virtually all LFS series are "important" to one user or another, but the series from the survey relating to the ILO-defined estimates - employment, ILO unemployment and economic inactivity - yield the “core” information which defines the primary reason for the existence of the LFS.

While the benefits of changing any LFS procedures or any part of the LFS questionnaire, therefore, need to outweigh the disadvantages of possible discontinuities, the balance is strongly in favour of the status quo in the case of the part of the questionnaire covering the "core" series, as para 4 above notes. The strategy outlined below explicitly recognises the distinction between core and other series, but also recognises that certain users will regard particular series as vital to their interests.

Some of the circumstances are as follows.

(i) Major survey re-design

An obvious case where some impact on the continuity of some LFS series was a price worth paying in order to gain other benefits, was the major up-grade of the LFS from an annual to a quarterly survey. The fact that the previous annual LFS was not seen by many users as a source of time series data, meant that the issues arising from any discontinuities have only relatively recently been raised. Nevertheless, considerable retrospective efforts have been made by ONS to determine the impact of this change on the estimates.

(ii) Change of contractor

Should a change of contractor occur, there would inevitably be some impact on the continuity of LFS data, even in the core series, resulting from the well-documented "contractor effect".

(iii) Quality improvements resulting from change

Examples of cases where the benefits of quality improvement have been seen by users to outweigh the problems caused by discontinuities are as follows:

- the LFS qualifications questions - where refinements to the questions have been designed to monitor the National Targets more precisely.

- the LFS disability questions - where the changes have been designed to bring the LFS estimates closer to the concepts of the Disability Discrimination Act.

- the method of determining family and household structure in the LFS - where changes have been made to harmonise the LFS methodology with that used for other household surveys, and hence to improve the quality, and comparability, of the LFS household and family data.

In each of the three examples quoted in the previous paragraph, ONS worked closely with DfES (DfES) to explore the extent of the discontinuities caused and, where possible, to make allowance for them in LFS time series.

(iv) External factors outside ONS control

A potential external source of impact on the LFS questionnaire, and hence on the continuity of LFS series, is a change in the EU Regulation covering the conduct of the LFS. A new Regulation for a continuous LFS was introduced in 1998 which introduces some changes to the LFS questionnaire requirement.

ONS have consistently pursued a vigorous defence of the existing UK LFS methodology and questionnaire in the discussions of the Eurostat Working Party which led up to the development of the new Regulation. As a result, changes to the existing UK LFS questionnaire needed to conform to the new Regulation will be minimal and will certainly not affect the "core" LFS series.

The above discussion has indicated the importance of trying to maintain continuity of LFS series, particularly those which are considered “core”. It identified some of the ways in which discontinuities have occurred in recent years. And finally, it looked at some of the circumstances in which the introduction of a discontinuity might be justified. The annex contains the Strategy and Guidelines for minimising the risk and impact of a loss of LFS continuity. It is intended to be forward-looking, to highlight relevant actions and behaviour, and to identify both potential sources of risk and guidelines for dealing with each source.

ANNEX 1: STRATEGY AND GUIDELINES FOR MINIMISING THE RISK AND IMPACT OF A LOSS OF LFS CONTINUITY

STRATEGY FOR CONSIDERATION OF LFS CONTINUITY

The issues relating to the maintenance of continuous time series from a household survey, such as the LFS, are very complex. No overall prescriptive basis exists for dealing with all the circumstances that may arise[1] but the following strategy - which also serves the purpose of putting into context the detailed guidelines (see below) - should be valuable as a basis for future consideration of the issues:

(i) to attempt to recognise the risks of discontinuity in advance

The most efficient use of resources in the present context is to attempt to ensure that discontinuities do not arise in the first instance. If changes to the LFS (survey method, questionnaire and so on) are planned, then part of the planning process should specifically include an assessment of the potential for discontinuities (and an assessment of the ways in which, and the extent to which, the possible discontinuity can be quantified). If such changes are implemented despite the likelihood of the introduction of a discontinuity, or if unpredicted discontinuities occur, due consideration should be given to their seriousness, reflecting (a) the over-riding importance of "core" variables, and (b) the fact that even "non-core" variables are of considerable interest to different users[2]. This should help determine the appropriate response, and by implication the resources required.

(ii) to assess the benefits and disadvantages of changes to the LFS

Any benefits of changes to the LFS will be assessed along with the impact these changes may have on the continuity of LFS time series and, where possible, decisions on the implementation of such changes will be taken by ONS in consultation with LFS users, in the light of all the relevant factors.

(iii) to consult with users

ONS will attempt to resolve discontinuities, subject to resource constraints, in consultation with appropriate LFS users, by applying one or more of the approaches outlined below. Input from subject-matter specialists within OGDs will be actively encouraged.

(iv) to determine the appropriate response to the discontinuity or risk of discontinuity

A number of different ONS responses to discontinuities in the LFS series have already been mentioned in this paper. In summary, these may be grouped into the following categories. Although it is difficult to generalise, responses (b) and (e) below are likely to provide most accurate information about the extent of a discontinuity. But (b) only applies in specific instances, whilst (e) is highly resource intensive. On the other hand, response (c) is likely to provide sufficiently accurate information for most users, and is applicable in the context of most potential sources of discontinuity.

(a) ... by fully revising LFS historical series onto a consistent basis

Examples where this has been done by ONS include: the revision, back to 1984, of all LFS series and sample weights following the re-basing to post-1981 Census population controls; the revision, of seasonally adjusted series

[1] Indeed, whilst changes in the external environment, such as changes in the administrative system for assessing eligibility for social security benefits, will not be considered to cause discontinuities in LFS series where the underlying LFS definitions do not change, ONS will still consult with users about estimating the effects of such changes – see para 6.

[2] see para 9.

following each annual review; the revision of key series by industrial sector following the introduction of the SIC 1992 classification (by means of using the longitudinal nature of the LFS to generate a "splice" in the series coded on both old and new classifications); and the revision of LFS series of persons in full-time education following the re-definition of this concept on DfES advice.

(b) ... by publishing dual estimates for one or more benchmark quarters

Examples where this has been done include: the publication of dual sets of unemployment estimates for 1984 on the pre-1984 LFS definition and on the current ILO definition; the publication, for the 1991 LFS, of occupational analyses on both the old and new classifications; and the publication of estimates of employment, ILO unemployment and economic inactivity with and without the inclusion of unpaid family workers as in employment.

(c) ... by indirectly estimating the size of the discontinuity

Unless a basis for benchmark estimation exists, such as a dual set of questions, (which, in general, is not practical), the options for the indirect estimation of any discontinuity effects are limited. Nevertheless, ONS has made considerable efforts to quantify the impact of the switch of the LFS from annual to quarterly frequency (as illustrated by the annex to the revised 1996 LFS Historical Supplement). Other work (carried out jointly with DfES) related to changes in the LFS questions relating to qualifications and to disability.

(d) ... by bringing the discontinuity to users' notice, for example by indicating the discontinuity in LFS tables and user guides

In many instances, data users will be the first to raise the possibility of (or indeed, to notice the existence of) a discontinuity, but in general it is ONS's responsibility to make users aware of discontinuities (including via LFS Data Service, Nomis®, and the Data Archive). This should include reporting to the LFS Steering Group, as appropriate, and relevant entries in the LFS User Guide. In some instances there may even be merit in including a relevant piece in *Labour Market Trends*. Discontinuities should certainly be highlighted in publications concentrating on time series, such as the LFS Historical Supplement. And the practice of changing the names of LFS variables (on the databases) when a discontinuity is likely to occur, or is observed, should continue.

In some circumstances, such as the 4-week training data series, all that can obviously be done is to indicate the existence of a discontinuity in LFS publications and user guides. Volume 3 of the LFS User Guide gives such details for each LFS variable. In addition, a new section of Volume 1 of the User Guide has been compiled. This brings together information about LFS discontinuities to help minimise the risk that users misuse or misinterpret LFS data, and will mention alternative sources to the LFS where these will help avoid, or alleviate, problems of discontinuities.

(e) ... by conducting an enlarged pilot

In general, LFS piloting has always been concerned with testing face validity - that is to say, ensuring that new or revised questions can be readily understood - and with ensuring that questions work satisfactorily within the context of the LFS. Both aspects of testing can be satisfied by relatively small pilots and are not intended to collect data *per se*.

But in particular circumstances, a case could be made for conducting an enlarged pilot with the specific intention of collecting LFS data in order to assess, for example, the implications of a different survey design. Clearly such enlarged pilots will tend to be exceptional, because of the very great costs involved, but they should be considered part of the armoury - a last resort - in assessing potential discontinuities.

(v) to monitor the success of this strategy in minimising discontinuities, over time

It is worth clarifying that all aspects of dealing with discontinuities - from initial efforts to minimising them through to assessing their effects - have a dynamic element. For example, it will be important to review periodically how well the strategy, and the raised awareness of the importance of continuity, are helping to minimise unexpected discontinuities, and to the assessment of all known discontinuities. The LFS Steering Group is best-placed to assess the way in which the strategy is working, and will consider "LFS continuity" as necessary.

It is also important to realise that an assessment of the effect of a discontinuity can change, as more information becomes available. For example, if a revised question is introduced, and the resulting series changes, then the initial assessment of the effect of the revised question is likely to be that the revision led to the changed data. But this assessment should be reviewed once more data are available, because the additional data might help refine the assessment.

(vi) to establish a relevant dialogue with other National Statistical Institutes

ONS should establish a dialogue with other National Statistical Institutes, as a means of sharing knowledge about the potential risks to data continuity and means of dealing with such losses of continuity. This should be taken forward by ONS, in consultation with other Departments, via the LFS Steering Group, and the implications considered by the Steering Group on an ad hoc basis.

POTENTIAL SOURCES OF A LOSS OF CONTINUITY, AND GUIDELINES FOR ADDRESSING THEM

In the context of the strategy outlined above, and taking account of the discussion in para 7 of the main part of this section regarding where discontinuities have arisen in the past, the following guidelines are intended to minimise the risk of discontinuities arising in LFS data. Also, recognising that discontinuities *may* arise - sometimes by design - they are also intended to offer maximum assistance to LFS users. The guidelines are structured according to possible sources of loss of continuity. In the case of each guideline, separate consideration is given to the case of the "core" LFS variables (see para 9 above). Each element of the strategy potentially applies to each source of loss of continuity.

(i) Externally-imposed changes to the LFS

This would include changing Eurostat requirements (both data and survey design), changes in (ILO) definitions, changes resulting from Government policy, and so on. With respect to both core and non-core series ONS should seek to influence the decision-making process, taking into account users' views about the balance between the desire for continuity and that for change. It is important that ONS and other departments should keep each other informed about relevant developments in Eurostat meetings, for example.

(ii) Survey redesign and changes in data processing practices

This would include, for example, any proposal to alter the design of the LFS, the possibility of introducing imputation for item non-response to the LFS, and so on. In general such proposals are high profile, and are likely to pose significant threats to

continuity in all data series. ONS should seek to ensure that the issue of continuity is one criterion used to assess the benefits and disadvantages of the redesign/change, and should attempt to evaluate the likely effects using the sorts of approaches referred to above, with the emphasis, in terms of resolution of discontinuities, being on the core series. (Note that in the case of an enlarged pilot study, it is likely that both core and non-core variables could equally easily be assessed for discontinuities).

(iii) Questionnaire changes

All proposals for change to the LFS questionnaire (and all other causes of potential discontinuities) should be carefully evaluated by the (inter-departmental) LFS Steering Group; this evaluation should consider potential discontinuities not only in the series most obviously affected, but elsewhere within the survey. Where it is considered that discontinuities are likely to arise, the Group should consider whether the need for the proposed change outweighs the likely disadvantages of a loss of continuity. This will inform an assessment of the expected net effect of making a particular change.

If such a loss of continuity is accepted, the Group should consider whether an assessment of the impact of the discontinuity is appropriate - subject to resource constraints - and which of the methods is most appropriate.

Independent of this, the effects of introducing the change to the questionnaire should be monitored by representatives of the LFS Steering Group. Any such discontinuities should also be publicised. If such discontinuities can be corrected, and users wish them to be corrected, and the resources are available, then they should be corrected.

(iv) Other quality improvements to the LFS

ONS are continuously seeking to improve the quality of all aspects of the LFS, from data collection - for example, interviewer training, increased response rates - to data processing - such as improvements to the weighting methodology, the treatment of data from proxy respondents, and so on. ONS should seek to ensure that the issue of continuity is one criterion used to assess the benefits and advantages of the quality improvement, and should attempt to evaluate the likely effects using the sorts of approaches referred to above with the emphasis, in terms of resolution of discontinuities, being on the core series

(v) Mistakes

Mistakes (leading to potential discontinuities) in the questionnaire specification and implementation, or in different aspects of data processing, can occur. ONS should guard against this by sharing knowledge, involving data users, and quality control procedures. But if mistakes do occur, then ONS should report them to data users, and to the Steering Group, with an assessment of (i) the likely impact of the mistake; (ii) whether the mistake can be corrected within existing resources; (iii) how long the correction would take. ONS would take into account users' views in reacting to the mistake.

ANNEX 2: DISCONTINUITIES ON THE LABOUR FORCE SURVEY

Summary

Topic	Time of discontinuity	Section
Employment	1983 Spring 1992	1 2
Unemployment	1984 Spring/Summer 1992 Spring 1993	3 4 5
When left last job Redundancies in the last three months	Spring 1992	6
When started with current employer	Spring 1992/Summer 1993	6
Redundancies in the last three months	Spring 1995	7
Reasons for economic inactivity	Spring 1992/Summer 1993	8
Long term health problem	Summer 1993/4 Spring 1996	9 9
Qualifications	Spring 1996	10
Numbers of graduates	1991-1993	11
Coding of occupations	1991/2001	12
Coding of industry	Winter 1993-94	13
Household and family data	Spring 1992	14
Job-related training	Summer 1994	15
Ethnic origin and nationality	Spring 1992/Spring 2001	16
Irish nationality	Winter 1994-95 Autumn 1995	17 17
Temporary employees	Spring 1992	18
Northern Ireland qualifications	Spring 1996	19
Education courses	Spring 1997	20
Disability data	Spring 1997	21
Benefits questions	October 1999	22
Sickness absence	October 1999	23
Enchancement reference period	Spring 2004	24
Number of O-level/GCSE etc passes held	Spring 2004	25

1. Employment before and after 1983

1.1 Since (spring 1983), people in full-time education who also did some work in the survey reference week, and people on employment and training programmes, have been classified as being in employment.

1.2 In 1983, the former group was estimated to be about 320 thousand, whilst the latter group was 355 thousand. However, users should note in considering people on government training schemes, that some may consider themselves to be employees or self-employed, so the figure of 355 thousand is likely to be an under-estimate of the true number of people on such schemes (but a useful indication of the size of this element of the discontinuity, as this recording problem is likely to have affected results from earlier LFS's similarly).

2. Employment before and after spring 1992

2.1 The LFS was conducted annually from 1984 to 1991, and quarterly from spring (March to May) 1992 thereafter. There were considerable differences between the ways in which the annual and quarterly surveys were conducted, in terms of design, method, sampling frame and definitions. The main effects were as follows:

- the annual LFS understated employment by about 210 to 220 thousand, relative to the established quarterly survey, mainly because of the addition of about 160 thousand unpaid family workers, and an extension to the sampling frame adding in about 40 to 50 thousand;
- the introductory quarters understated employment by 43 thousand (spring 1992) then 16 thousand (summer 1992), relative to the established quarterly survey, because of the way in which the sample was developed during these early quarters;
- the net effect from 1991 to spring 1992 was a discrepancy of about 170 thousand, plus/minus the effects of certain unquantifiable elements;

2.2 More details are included in the table below

Summary of differences between employment and ILO unemployment estimates from annual and quarterly LFS

Difference between annual and established quarterly LFS	**Effect on employment estimate**	**Effect on ILO unemployment estimate**
Sample clustering	*	*
Rotation patterns	*	*
Interviewing modes	Annual estimates were understated relative to the quarterly LFS by about 12 thousand	Marginal reduction in quality in quarterly LFS
Type of questionnaires	*	*
Use of PAPI/CAI	Quality improvement in quarterly LFS	
Levels of proxy responses	Annual LFS slightly overstated employment relative to the quarterly LFS. Marginal reduction in quality in quarterly LFS.	
Sampling frame enhancements	Annual estimates understated relative to quarterly by 40-50 thousand	No effect
Definitional differences - UFWs	Annual estimates understated relative to quarterly by 160 thousand	No effect
Seasonal adjustment	*	*
Difference between introductory and established quarters		
Quarterly build-up	Spring 92 understated by about 43 thousand, summer 92 by 16 thousand, relative to the established quarterly LFS	Spring 92 overstated by about 23 thousand, summer 92 by 12 thousand, relative to established quarterly LFS
Combined	Annual LFS understated employment by about 210 to 220 thousand, relative to the quarterly. Introductory quarters understated the fully established quarterly LFS by 43 thousand, then 16 thousand. Net effect from 1991 to spring 1992: discrepancy of about 170 thousand, plus/minus unquantifiable.	No observable discontinuity as a result of moving from annual to quarterly LFS. Spring 92 overstated (relative to established LFS) by about 23 thousand, summer 92 by 12 thousand.

* effect could not be readily quantified

2.3 The variables affected are: *ecara* (in 1991) and *inecaca* (spring 1992 onwards).

3. Unemployment before and after 1984

3.1 In 1984, the International Labour Organisation (ILO) definition of unemployment (out of work, available to start within the two weeks following their LFS interview, and who had either looked for work in the four weeks prior to interview or who were waiting to start a job they had already obtained) was adopted in the UK LFS. Prior to that the

Labour Force (LF) definition of unemployment was used. This was based on a one-week job search period, and the availability criterion applied only to full-time students.

3.2 Estimates of unemployment (and economic inactivity) are published on both definitions for spring 1984, to aid interpretation.

3.3 The variables affected are: *ecarax* (in 1983) and *ecara* (in 1984).

4. ILO unemployment before and after spring/summer 1992

4.1 A routing mistake in the 1992 introductory quarterly surveys (spring, summer) resulted in most people who were waiting for a new job to start not being asked whether they were currently available for work.

4.2 This is estimated to have understated ILO unemployment in spring 1992 by 25 to 30 thousand, and in summer 1992 by between 40 and 45 thousand.

4.3 The variables affected are: *inecaca* and *start*.

4.4 In the derived variable for economic activity, to be classified as ILO unemployed a respondent must answer *start* = yes (available to start work in two weeks). Therefore anyone who is not routed to *start* in the questionnaire cannot be classified as ILO unemployed. Since autumn 1992, the routing to *start* in the questionnaire has included (*inter alia*) those waiting to take up jobs (*jbaway* = waiting to take up new job; *wait* = yes: waiting to take up job); in spring and summer 1992 these two categories were not included in the routing to *start*. The discontinuity arising from *jbaway* is up to 5 thousand in both spring and summer 1992; that arising from *wait* is larger - about 25 thousand in spring 1992, and about 40 thousand in summer 1992.

4.5 There will be corresponding discontinuities in the economically inactive series, where the people incorrectly routed in these two quarters would have been classified.

5. ILO unemployment before and after spring 1993

5.1 A routing mistake in the spring 1992 to winter 1992-93 surveys had the effect of excluding men aged 60-69, and women aged 60-64, from being asked whether they were waiting to take up a job they had already obtained. The routing (of the questionnaire variable *wait*) was corrected in spring 1993.

5.2 The discontinuity was likely to be between 1 and 2 thousand.

6. When started with current employer/When left last job/Redundancies in the last three months before and after spring 1992

6.1 Before spring 1992 the period since leaving last job or since starting a new job was recorded in the interview in pre-defined categories such as “less than three months”. From spring 1992 onwards, the LFS does not collect information about the duration since starting or leaving a job, only the month and the year of the event. This change is likely to have affected the quality of recall, although we do not know in what way.

6.2 However, it is possible to estimate the effect of deriving duration bands from the event dates recorded by the LFS. The date of the LFS interview is not used in the derivation of any variables - only the month and year. Consequently when determining whether someone left/started their job "in the last three months" the options are to take those who left/started in:

(A) the reference month and the THREE previous calendar months
OR
(B) the reference month and the TWO previous calendar months.

WNLEFT (when left last job) and EMPLEN (length of time with current employer – SuperCross only from summer 1993) take the current month and the previous TWO months for the category "less than three months" (option (B)). The measure of redundancies in the last three months (REDUND) took the current month and the previous THREE months (option (A)) until the re-weighting in April 2000, since when it has been made consistent with the other variables and uses option (B).

6.3 ONS has estimated the effect of taking the first of these two options compared with the ideal redundancies measure, i.e. people made redundant in the 13 weeks before interview. This was reported in the Technical Note of the May 1999 *Labour Market Trends* article on Redundancies in the UK. The effect of moving from option (A) to option (B) is described in the May 2000 LMT article on Redundancies estimates.

6.4 Difference between the basis of the pre 2000 re-weighting redundancies measure (option (A)) and the ideal measure of redundancies in the last three months.
The question design implies that since spring 1992, the LFS includes a few people who were made redundant between three and four months prior to their interview. In the main, this is people who were not in employment in the reference week for which their response to when they left their last job is used. For example, the spring 1998 estimate would include a person who was not in employment during the reference week ending on 31 March 1998 and who had been made redundant from their last job on 1 December 1997.

The proportion of people included though they were made redundant more than three months prior to their interview can be estimated using a theoretical model. This relies on the assumption that the distribution of the dates of the redundancies is uniform for each reference week, and that the distribution of the reference weeks is uniform across the quarter. This is a fairly reasonable assumption, as further analysis of the data has shown.

6.5 Of those who were not in employment and who had had no job since their redundancy, those among them who were made redundant truly *within three months prior to the interview* form about 6/7 (86%) of the estimate of people made redundant and not in employment using option (A). Those who were in employment in the reference week and who had been made redundant in the previous three months are very slightly affected (see May 2000 LMT article) but the effect has not been quantified.

6.6 According to the LFS estimates, those who had been made redundant from their last job and who were not in employment during the reference week represent approximately 60% of all those made redundant. So in total only about (6/7*60%) + 40% = 91% of the LFS estimate (prior to the April 2000 re-weighting) were made redundant truly within the three months prior to their interview (see the May 1999 article for illustration). The overestimation is consistent over time, staying between 11% and 9% of the LFS estimate. Hence comparisons over time since 1992 are not significantly affected by including these people who were made redundant more than three months prior to their interview but within the three calendar months preceding it.

6.7 At the same time as the LFS re-weighting, the derivation of REDUND was made consistent with other duration variables, i.e. using option (B) above. All duration variables including REDUND now UNDER-ESTIMATE the true numbers of occurrences in the less than three months category. The effect is again about half a month but this is a larger proportion of the total – see below.

WNLEFT
The only significant differences between WNLEFT as derived since spring 1992 and the way it would ideally be derived are:

- The category *"less than three months"* is underestimated; for spring 1998 the "correct" estimate would be 13 per cent above that found in WNLEFT.
- The category *"1 year but less than 2 years"* is overestimated; for spring 1998 the "correct" estimate would be 91 per cent of that found in WNLEFT.
- For key categories *"less than 6 months", "less than 1 year"*, and *"less than 2 years"*, the bias is smaller than 7 per cent.

Information on the method of making these estimates (which could be applied to EMPLEN and EMPMON) may be obtained from Lester Browne in Labour Market Division, 020 7533 6143.

6.8 The main variables affected are:

REDUND since spring 1992,
WNLEFT since spring 1992,
EMPLEN from spring 1992 to spring 1993, (ONS datasets only each quarter from Summer 1993)

Use of EMPMON (which replaced EMPLEN in all formats except SuperCross from summer 1993) to derive employment duration bands would similarly produce different results to the pre 1992 variable EMPLEN since it does not take into account the dates in the month of either the job start or the reference week.

7. Redundancies: changes in the questionnaire in spring 1995

7.1 Three main changes to the derivation of the redundancies measure were made in spring 1995:

(i) a question was added on whether the person made redundant had been made redundant from any other job in the last three months.
(ii) from 1995 the redundancy total includes those reporting that they were made redundant from a job in which they were self-employed, as well as those who were employees.
(iii) in 1992-94 people who were in employment during the reference week could be included only if their previous employer was either closing down or cutting back on staff. In spring 1995 this condition was removed.

In the process of revising the coverage, the ordering and formulation of the questions was changed. The LFS may (or may not) have recorded a higher or lower number of redundancies as a consequence, but this impact is not quantifiable. It is however expected to be small. Apart from the unknown effect of the revised question wording and their ordering, ONS estimates, very approximately, that in 1995 the LFS counted some 20,000 more redundancies than it would have under the 1992-94 criteria. The nature of the discontinuity is such that comparisons between sub-groups are not affected.

7.2 The variables affected are REDUND, and indirectly, REDINDY.

8. Reasons for economic inactivity before and after the 1992 quarterly surveys

8.1 A number of factors affect the continuity of the data for reasons for economic activity. These are described below.

8.2 The first issue concerns the comparability of 1984 to 1991 data with data from spring 1992 onwards. Unlike the questionnaire for the quarterly period (spring 1992 onwards), between 1984 and 1991 in the questions used to derive the main variable for economic activity there were categories for 'No answer'/'Not stated'. Because of this, data on reasons for inactivity have not previously been published alongside those for spring 1992 onwards, in the LFS Historical Supplement, for example. But

the evidence supports the idea that these NAs can fairly safely be included in the relevant "no" categories, in order to produce a (reasonably) consistent time series back to 1984.

8.3 The following section provides more detail about the numbers and characteristics of the NAs to each of the relevant questions, in support of the proposal.

8.4 The questions used to derive the main economic activity status variable ECARA are:

Whether available to work in the next 2 weeks
Whether looked for work in the last 4 weeks

If the answer is YES to both of these, the person is, of course, ILO unemployed.

There were between 76 and 96 thousand people in each year (84-91) who were *available* but did not state whether they had *looked for work*; since they would be unemployed if they had, we must treat them as 'not looked' or else revise the ILO unemployment series. This is a possible source of discontinuity in 1992, although it seems unlikely that many had actively looked for work.

There are between 48 and 89 thousand who were *not available* and did not state whether they had *looked*. Most who responded (two-thirds to three-quarters of men and 80-90 percent of women) had not looked for work and it seems reasonable to assume that the non-respondents had not.

There are a small number who did not state whether they were *available* (19,000 in 1984, 10,000 in 1985 and less than 8,000 in other years), of whom the majority had *looked for work*. This probably results in a small discontinuity in the ILO unemployment estimates between 1984 and 1986, but even if all those who had looked were also available, the addition would range from 7,000 in 1984 to 2,000 in 1991. The effect on the various permutations of inactivity would be miniscule.

Whether would like to work:

The 'no answer' category for this question was 74,000 in 1984, 29,000 in 1985 and between 13,000 and 8,000 in the other years. Two-thirds of the 1984 'no answers' were women, although the division is more equal in other years. There is thus a possible discontinuity between 1984 and later years. Including these larger numbers of 'no answers' with 'would not like work' may mean that the 'would like work' estimates, particularly in 1984, are understated. The size of this group in 1984 may be the effect of the high unemployment and represent a type of discouraged worker who, if pressed, would say they would like to work.

8.5 It seems reasonable to publish estimates for economic inactivity which are consistent with the published estimates of ILO unemployment for 1984 to 1991, especially as there is little hard evidence to revise estimates of ILO unemployment. The estimates for 1984 and 1985 may be of poorer quality than the later years' estimates (as with many other LFS estimates for these years). ONS are currently consulting on this proposal.

8.6 A second "reasons for inactivity" data problem concerns 1992 alone. From spring 1992 to winter 1992/93 there was a mistake in the routing of the questionnaire, which resulted in the "reasons for not seeking work" questions only being asked of those under the age of 60, missing out women aged 60-64 and men aged 60-69.

8.7 This affects estimates of discouraged workers (who are economically inactive, and who are not looking for work because they believe there are no jobs available). Assuming constant growth rates in the number of discouraged workers aged over 60 between spring 1991 and spring 1993, it is likely that the spring 1992 survey missed

about 38 thousand male discouraged workers aged 60-69, and about 15 thousand women aged 60-64.

8.8 The routing problem also affects estimates of the number of retired people, and those who gave no reason for being economically inactive. The number of retired people fell by nearly 300 thousand between 1991 and spring 1992, and increased by almost 250 thousand between spring 1992 and spring 1993, whilst the number who gave no reason fell from about 700 thousand in spring 1992 to about 70 thousand in spring 1993.

8.9 Finally, changes to the questionnaire between the annual and quarterly surveys led to a discontinuity in the 'not want or need job' reason for inactivity. Before 1992 this had been of the order of half a million, but after that fell to about 150 thousand.

8.10 The variables affected are: *inecaca*, *ecara*, and *ilodefa*.

9. Long term health problem/disability before and after summer 1993/94, and from spring 1996

9.1 The specific reasons for these discontinuities are complex, but they all stem from the fact that the frequency with which the relevant questions were asked was changed from every quarter to every other quarter (summer 1993-94 - that is to say, spring 1993 was the first quarter to be missed) and then back to every quarter (spring 1996).

9.2 The complexity relates to two specific aspects of the LFS design. Firstly, in cases where respondents are not contactable or do not want to take part in the LFS but wish to remain in the survey (circumstantial refusals), their data are imputed from their previous survey interview (for one quarter only). Hence when questions are not asked in a quarter, there is no response to impute forward in this way. Secondly, respondents in the first wave of the LFS are interviewed face-to-face; those receiving their second to fifth interviews tend to be interviewed by telephone.

9.3 The combined effect is that when telephone interviewers do not have access to previous information which will enable them to probe, respondents are less likely to say that they have health problems. In face-to-face interviews, certain health problems may be visible and respondents may be more open and willing to discuss such a topic in a personal interview, where rapport can be developed more easily than over the telephone.

9.4 A further issue, though not a discontinuity in itself, is that a question on long term health problems/disabilities was only introduced in winter 1993-94.

9.5 Hence data from winter 1993-94 to winter 1995-96 (alternate quarters) are consistent, but understate the numbers relative to data collected every quarter by approximately 15%-20%. The combined effect of asking the questions half-yearly cannot be quantified easily. Data for spring 1996 onwards are not consistent with the earlier quarters. The numbers gradually moved towards "1992 comparability" between spring 1996 and winter 1996-97, by which time all respondents had been asked the questions on a quarterly basis.

9.6 The variables affected are: *limitt* and *lnglim*.

10. Qualifications before and after spring 1996

10.1 A number of changes were made to the section on qualifications in spring 1996. Whilst these greatly improved the flow of the interviews, and enabled the LFS to collect more coherent and accurate data on qualifications - for example, by asking about types of qualifications, and levels (and in some cases the numbers of such qualifications) in separate questions.

10.2 Inevitably such improvements lead to discontinuities. ONS and DfES have assessed the effect by comparing responses in the winter 1995-96 and spring 1996 surveys, for individuals interviewed in both quarters. The outcome of this work will be reported in an article in the January 1998 issue of *LMT*.

10.3 The main variable affected is *hiquap* (which has been renamed to *hiqual* from spring 1996 to help draw attention to the discontinuity). Hiqual was replaced by hiqual4 in 2004 and hiqual5 in 2005 mainly as new categories of qualifications were added e.g Welsh Baccalaureates.

11. Numbers of graduates 1991-1993

11.1 There is a step change in the LFS estimate of the number of people with a highest qualification at degree level or higher (those with NVQ level 4 or better), between 1991 and spring 1992, and a possible further discrepancy compared with administrative records between 1992 and 1993.

11.2 ONS have looked closely at this apparent discontinuity, in terms of the expansion of the sampling frame, the questionnaire, the interviewer instructions, editing procedures, levels of proxy responses, and the build up of the LFS wave structure. But these factors cannot account fully for the discontinuity.

11.3 In most cases the effects of these changes were similar for all qualifications or were so small that the effect on graduates would have been negligible. Minor effects resulted from the differential response of proxies and the expansion of the sampling frame. An analysis of the distributions of people with NVQ4 or more by age, sex, region, ethnic group and employment status showed no one category with exceptionally large increases sustained over the years from 1992 onwards (sampling variability has quite marked effects in individual quarters). It is interesting to note, however, that largest increases between 1991 and 1993 are not for first degrees but for higher and “other” degrees and for BTEC/HND qualifications. The “other” degree category includes professional qualifications such as membership of a professional institute or chartered accountant and higher degree includes Post Graduate Certificate in Education (PGCE).

11.4 Nevertheless, DfES and ONS judgement is that the LFS estimates of the stock of people with NVQ4 level and higher qualifications, including degrees and equivalents, are not consistent between 1991 and 1993. The discrepancy between 1992 and 1993 is smaller and may not be significant, but users are advised not to make comparisons of these estimates for 1991 and earlier years with those for 1992 and subsequently.

11.5 A measure of new graduates has recently been defined (see March 1997 *LMT* - LFS Helpline). An article on LFS qualifications data is planned for January 1998 issue of *LMT*.

11.6 The main variable affected is *hiquap*.

12. Coding of occupations before and after 1991

12.1 From 1984 to 1990 the occupations of employed LFS respondents were coded using a classification system called CODOT (the Classification of Occupations and Directory of Occupational Titles). Corresponding data for 1991 were dual-coded to CODOT and to the Standard Occupational Classification (SOC) code, and from 1992 onwards only SOC has been used.

12.2 This means that it is not possible to produce a consistent time series of occupations spanning 1991. However, table 13 in the April 1992 issue of the Employment Gazette cross-classifies the two codings for the 1991 LFS.

12.3 From March 2001, the new Standard Occupational Classification (SOC2000) was introduced. The new classification does not directly map to the old one and certain related variables have disappeared. From 2001 there will no longer be any variables that show a manual/non manual split, social class, or socio-economic group. In place of Socio-economic Group and Social Class, the new National Statistics Socio-economic Classification (NS-SEC) has introduced in new variables NSECM and NSECMMJ. Further details of the SOC2000 and the new NS-SEC classification are available from the ONS website: hhtp://www.statistics.gov.uk/methods_quality/classifications.asp

13. Coding of industry before and after winter 1993-94

13.1 In winter 1993-94, details of the industries in which LFS respondents worked were coded to SIC(92), replacing the earlier classification to SIC(80). Simply mapping detailed categories of SIC(80) to SIC(92) proved insufficient as a means of producing a continuous time series, and a more sophisticated exercise, linking respondents' mapped SIC(92) data for autumn 1993 to the coded SIC(92) data for winter 1993-94. This enabled a relatively smooth series to be produced with no evident discontinuities.

13.2 Further details of this exercise, including a description of the limited data available on the databases using the linked approach, is included in volume 5 of the LFS User Guide.

14. Household and family data, and marital status

14.1 Because the LFS was designed and developed as a survey focusing on individuals, little attention was given in the early stages to the information on households and families which were included by virtue of the household-based survey design. Over time users began to draw on this information, and a number of inconsistencies and discontinuities were identified in the household and family data, caused by changes over time in the definitions of the variables underlying household and family type, and changes or anomalies in applying these definitions.

14.2 Further details of problems with these data from spring 1992 to spring 1996 are given in User Guide volume 8. There are also a number of relevant Labour Market Trends articles. Pam Tate's article in the March 1997 issue of Labour Market Trends summarises the problems with the quarterly data, and provides similar information in respect of the period 1984 to 1991. This article also demonstrates a constructed series for the proportion of lone parent families, a statistic which suffered a discontinuity when the quarterly LFS was introduced. David Hastings' article in June 1997's LMT describes the introduction of a consistent approach to collecting household data from all household members (the "household matrix") from spring 1996. Subsequent LMT articles, (September 1997, August 1998) describe the problems with deriving a consistent series on the economic activity of working age households between spring 1992 and spring 1996, and the adjustments made to compensate for these inconsistencies on the household datasets.

15. Job-related training before and after summer 1994

15.1 In summer 1994, a new question asking whether employees had received job-related training (JRT) in the previous thirteen weeks was added to the LFS, immediately prior to the long-standing questions about JRT in the previous four weeks. This appears to have introduced a significant discontinuity (a fall of 1-2% points in the participation rate) in the four-week series.

15.2 At first sight this appears to be because respondents might previously have reported training undertaken just over four weeks previously in response to the four week question, whereas once asked the 13 week question they would tend to report it more

accurately. But this would suggest a change in the relationship between those receiving JRT in the previous four weeks, and the previous week, and such a change did not occur.

15.3 ONS are not proposing to undertake further work on this issue; the trends before and after the discontinuity are relatively similar, and it is straightforward to splice the two series together - although it is not clear whether the 'previous' or the 'current' form of the questions gives rise to the better measure of JRT.

15.4 The variables affected are *jobtrn* and *trnopp*.

16. Ethnic origin and nationality, and foreign workers etc., before and after spring 1992.

16.1 Two aspects of the move from the annual to quarterly LFS in spring 1992 led to discontinuities in these data series. First, the changes to the LFS design and methodology resulted in significant reductions in non-response, producing a discontinuity in recorded responses. The variables this affected were *ethcen* and *ethnica* (in 1992).

16.2 Second, the removal of the clustering from the LFS sample design when the LFS became quarterly meant that each quarter's estimates were "free-standing", in contrast to estimates produced from the LFS from 1984 to 1991, when nationally representative results were only available for three years taken together because of the geographically clustered sample.

16.3 Further details are given in the May 1994 *Employment Gazette* (pp147-159) on ethnic groups, and January 1995 *Employment Gazette* (pp11-19) on foreign workers. The May 1994 article also describes how the ethnic classification used before 1992 compares with the classification based on the 1991 Census definition.

16.4 The recommended output classification of ethnic groups for National Statistics data sources was changed to be broadly in line with the 2001 Census. From Spring 2001, LFS introduced new questions on ethnicity in line with this new classification. More information on the National Statistics interim standard classification of ethnic groups can be found on the National
Statistics website:
hhtp://www.statistics.gov.uk/nsbase/themes/compendia_reference/articles/ns_ethnic_classification.asp.

17. Irish nationality before winter 1994-95 and after autumn 1995

17.1 This problem stemmed from the fact that nationality is regarded as a contentious issue in Northern Ireland, such that country of birth may be a better measure of legal nationality (in the NI context).

17.2 Before winter 1994-95 (spring quarters only, effectively, as the NI LFS only became quarterly in winter 1994-95) an edit check automatically recoded values of *nation* = 6 (nationality = Irish Republic) to 1 (UK/British), if the response to *cry* is 1 (country of birth = UK/Britain). But this recode was not applied to the more detailed nationality variable *nato*, so those coded to Irish (Republic), Irish (Southern) and Irish (part not stated) have all remained as code 6 (Irish Republic), regardless of their country of birth.

17.3 This has led to over-inflated estimates of the number of Irish nationals resident in Northern Ireland for winter 1994-95, spring 1995, summer 1995, and autumn 1995. From winter 1995-96 the editing procedure was reinstated. For the problematic quarters, users should cross-tabulate *natox* = 6 on the basis of *cry*.

18. Temporary employees before and after spring 1992

18.1 There are two separate discontinuities affecting data for temporary employees before and after spring 1992.

18.2 First, before 1992 the self-employed were asked about their permanent/temporary status, and the nature of their temporary work. After 1992, these questions were only asked of employees.

18.3 By itself this discontinuity could easily be countered by filtering out the self-employed from the results prior to 1992. But the use of the Census-edit makes the issue more complicated. This edit checks that occupation codes are compatible with employment status; certain occupations (such as police officers) cannot be self-employed, and the employment status of anyone claiming so on the LFS would be recoded from self-employed to employee. (Note that there are no recodes from employee to self-employed). Because some people reporting (temporary) self-employment will have been recoded as (temporary) employees prior to 1992, whilst subsequently the series of temporary employees will comprise only those who say in their LFS interview that they are employees, a (small) discontinuity will have arisen.

18.4 Second, there is a discontinuity in the responses to the reasons why a job is non permanent. From 1992 onwards, respondents have a choice of one of five answers (excluding no answer) as to why a job was temporary - seasonal work, casual work, agency temping (where the employee is leased out by a temporary workers' bureau and remains the employee of the bureau rather than of the company with which they are placed), on a fixed-term contract or task or some other reason for the job being temporary. Prior to 1992, the first three categories (seasonal, casual and agency temps) were grouped as one response. Hence data on the totals of temporary workers remain relatively consistent from spring 1991 to spring 1992 despite the change of question and processing, but the totals of the responses given for reasons why a job is temporary prior to spring 1992 are not comparable with estimates after this date.

19. Northern Ireland Qualifications Variables

19.1 Following the changes to the qualification section in spring 1996, a routing problem was introduced to the questionnaire which created a discontinuity in two of the Northern Ireland qualification variables.

19.2 From spring 1996 to winter 1996/97, those with GCSE's or CSE grade 1's were not asked the question about how many O levels or GCSEs they had. Consequently, over one quarter of valid cases are missing from the NUMOL variable for these four quarters and it is not possible to obtain data retrospectively.

19.3 In spring and summer 1997, those with CSEs were incorrectly routed to the GCSE question. This had the effect of increasing by 102 the sample number in the "No" category of this variable (no GCSEs at grade C or above). In fact, these individuals have no GCSEs at any level and analysis of the variable (for Northern Ireland) should first select on QUALS=18 (i.e. GCSE). The problem will be rectified from autumn 1997 onwards.

20. Education courses (including full-time students) from spring 1997

20.1 From spring 1997, a change to the LFS questionnaire led to more people on education courses being identified, which created a discontinuity in the full-time students series. Full-time students are identified in the LFS by means of questions asking about enrolment and attendance at school/college and type of course and educational institution.

20.2 The question about attendance, called ATTEND, was changed in spring 1997 when a new response category, "waiting for term to (re)start", was introduced. Previously there were just two categories ("still attending" and "stopped going"), and interviewers were instructed that anyone who said that they were waiting for term to (re)start, and asked the interviewer for guidance, should be classified as "still attending". It was, however, expected that making this more explicit by including it in the question would increase the number of students recorded by the LFS (see *Labour Market Trends* July 1996 pp. 334-336).

20.3 Respondents who answer that they are "still attending" or "waiting for term to start" to ATTEND are routed to the questions on type of educational course (COURSE), type of educational institution (EDINS) and whether training undertaken was part of an education course previously mentioned (JOBED). It is the routing to COURSE which is of most interest because COURSE is the key variable in the derivation of type of current education received (CURED), and the definition of a full-time student is in terms of CURED. COURSE is also involved in the routing to some of the job-related training questions, but any discontinuity effects from the change to ATTEND are unlikely to be significant.

20.4 The discontinuity in the full-time students series is most obvious in summer because of the long vacation when existing students are likely to respond that they are waiting for term to re-start. The estimated discontinuity effects in 1997 were:

- In spring 1997, the discontinuity between the actual figure and the figure which would have been expected in the absence of any change is up to about 60 thousand (2.3%)
- In summer 1997, the discontinuity in the number of full-time students is considerably larger than spring, between 300 and 400 thousand
- The discontinuity in autumn 1997 is more difficult to assess, but is likely to be between 150 and 250 thousand
- The discontinuity in winter 1997/98 is in the range of 0 to 60 thousand

20.5 It will be possible to make more accurate estimates of the discontinuities when administrative data for 1997 is available. In time series tables of full-time students, a discontinuity at spring 1997 should be indicated, and the discontinuity ranges quoted. See *Labour Market Trends*, June 1998, pp. 337-340 for further details about this discontinuity.

21 Disability data from spring 1997

21.1 The focus and number of questions in the Health and Disability Module of the LFS questionnaire changed in spring 1997 to reflect the provisions of the Disability Discrimination Act 1995. In particular there were new questions which were concerned with all health problems, whilst until spring 1997 the emphasis had been on problems which affected respondents' work.

21.2 The disability module was changed in spring 1997 to an extent that comparison with previous quarters is not straightforward. Conceptually, it is possible to compare, under both sets of questions, numbers of people with a long term health problem which affected their work.

21.3 However, subtle changes to the question wording and ordering have led to a significant fall in the number of disabled people recorded by the LFS. Previously the module asked whether the disability (i) would affect (ii) any kind of paid work..., whereas in spring 1997 the module asked (i) does it affect (ii) the kind of paid work. The more focused approach of the new module appears to have reduced the number of people who provide hypothetical responses. For example, an office worker with a back problem would have been justified in saying that it did affect any kind of work he/she might do, in that it would tend to rule out any job involving hard physical labour. But faced with the more robust new question set, such people appear to have

thought to themselves that their back problem does not affect their ability to work in an office, and so have concluded that they were not disabled.

21.4 ONS considers that data for summer, autumn and winter 1997 are of better quality than spring 1997 because it was not possible to impute for non-response. The total number of people in summer 1997 with a long-term disability which affected the kind of paid work they might do was 4.7 million - 14 per cent lower than winter 1996/97. The total number in autumn 1997 was 11 per cent lower, and in winter 1997/98 it was 9 percent lower than winter 1996/97. Therefore ONS estimates that the introduction of the new questions in spring 1997 reduced reported levels of disability by about 10 per cent. However, it is not yet possible to adjust for the discontinuity.

21.5 Further details of this discontinuity can be found in *Labour Market Trends*, December 1997, pp. 494-5, and June 1998, pp. 321-335.

22. Changes to the benefits questions from October 1999

22.1 From March-May questions on benefits were changed to take account of the new tax-credits that were introduced in 1999/2000. Thus, the questionnaire now asks respondents if they received Disabled Person's Tax Credit or Working Families Tax Credit

23. New questions on sickness absence from Spring 2000

23.1 The LFS asked respondents about days off work due to illness or injury and then went on to ask about the number of days they were too ill to work (including non-working days). As the sickness questions did not differentiate between days that respondents were scheduled to work, and days that they were not meant to work, it was impossible to calculate the number of working days lost to illness or injury.

23.2 The Cabinet Office and HM Treasury felt that the existing questions about sickness absence were inadequate. The Cabinet Office required better information about sickness in the whole economy and wanted to monitor sick absence in the public and private sectors. The new questions enabled 'days lost by sickness absence out of the number of days on which an employee was expected to work' to be calculated for the first time. The new questions, introduced from Spring 2000, now make it possible to also identify on which day(s) of the week the respondent was absent.

24. Enhancement survey questions reference period amended from Spring 2004

From March to May, the reference period used for respondents partaking in the enhancements, changed from three years to one year. Thus, respondents are now asked questions with reference to their situation within the last year instead of the last three years.

25. Education questions regarding number of O-level, GCSE etc passes already held from Spring 2005

Prior to Spring 2004, the LFS provided 3 broad categories to respondents in terms of how many O-level, or GCSE etc passes that they held. From Spring 2004, as requested from the Department for Education and Skills, respondents were given a more detailed set of response options with the hopes of gaining more specific data. It was found that this produced a larger than expected increase in the number of don't know responses, due to respondents not being able to provide the exact number of passes that they had. The end result showed that there was an increase of about 2% in each category, which could only be explained due to the change in the question. As a result, from Spring 2005, DfES requested that the LFS return to the response options used prior to Spring 2004. There are now follow on questions in place in order to gain the more specific detail.

SECTION 14 – QUALITY

The LFS Performance & Quality Monitoring Report (PQM)

The LFS reports quality issues in the LFS Performance and Quality Monitoring (PQM) Report, which is published quarterly on the NS website. The LFS PQM was first produced in 1992 and contains detailed information on aspects of survey quality, including achieved sample size and response rates. The PQM was revised in 2006 to coincide with the first release of LFS data on a calendar quarter basis beginning with the January - March 2006 quarter. There were a number of reasons for revising the PQM at that time. The first was that the format of the PQM had remained largely unchanged since its launch in the early 1990s and its contents were considered to be in need of review. A number of key indicators of data quality were identified as missing from the report, and some of the items being reported on were no longer thought to be of interest. The process of identifying and agreeing changes to the PQM was done in consultation with customers (internal and external) to ensure that it met their requirements. In addition, the format and content of the LFS PQM was brought it into line with the corporate approach to quality reporting which the ONS implemented with the support of the Statistics Commission.

PQM Contents

Overall, the PQM has adopted the ONS corporate approach to quality reporting in which quality is reported against the six European Statistical System (ESS) dimensions of quality, namely: relevance; accuracy; timeliness & punctuality; accessibility and clarity; comparability; and coherence. Each dimension of quality is defined in the PQM.

The PQM contains a 2-page 'Executive Summary' that briefly describes the current status of the survey in terms of:

- achieved sample size
- response rates
- delivery dates of data
- quarter to quarter changes, including any changes to the survey, such as new or amended questions, and
- fieldwork issues, particularly those which are likely to have an impact on data quality.

Specific quality issues reported in detail in the PQM include:

- the relevance of the LFS including its primary purpose, users and uses, strengths and limitations, and key definitions.
- The accuracy of the LFS including
    - a time series of achieved number of household and person interviews for GB & UK.
    - sampling variability estimates for a number of key LFS variables for the UK
    - wave-specific response rates and an overall response rate for the quarter for both GB & UK.
    - a time series of wave-specific response rates for GB
    - the composition of non-response and how it has changed over time
    - wave-specific response rates by Government Office Region for the quarter
    - proxy response rates for the quarter
    - income response rates by NS-SEC for the quarter, and
    - attrition rates by key person level characteristics for the quarter
- information on timeliness and punctuality including delivery dates of data
- information on accessibility and clarity, including various access points
- information on comparability, including definitions, quarter to quarter changes and fieldwork issues

- information on coherence with other sources of data on the labour market, including the strengths and limitations of the LFS
- a summary of methods used in the LFS
- technical definitions, and
- website references

PQM Publication Date

The PQM is released to coincide with the release of the quarterly LFS data to which it relates. As the PQM contains a small amount of market sensitive data in the form of estimates for key variables, it is released in accordance with the published timetable for the Labour Market Statistics Integrated First Release. This is a requirement of the National Statistics Code of Practice and Protocol on Release Practices.

Other Quality Reports relevant to the LFS

In addition to the PQM, the LFS also has a Summary Quality Report (SQR) which also provides users with information on fitness for purpose of the LFS, and contains qualitative information covering the six ESS dimensions of quality covered by the PQM, and a summary of methods used to compile the output. What it does not contain, however, is quantitative information relevant to each quarterly release of LFS data, as this is contained in the PQM. In addition to the LFS SQR, quality issues relevant to the LFS can be found in the Labour Market Statistics (LMS) SQR, along with quality issues relating to the other components of labour market statistics. The LFS and LMS SQRs are published on the NS website and are updated only when there is a change to the qualitative information they contain.

Website Links

Labour Force Survey Performance & Quality Monitoring Report (PQM):

http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=10675

Labour Market Statistics Summary Quality Report (SQR):

http://www.statistics.gov.uk/cci/article.asp?id=1354

SECTION 15 - HARMONISATION

Background

The United Kingdom conducts a wide range of Government surveys of persons and households, which provide sources of social and economic statistics. These surveys were designed at different times, to meet different needs, and have been commissioned by a range of departments. Consequently, the surveys were developed to a significant degree in isolation from each other. This resulted in a lack of cohesion, with differences arising in concepts and definitions, in design, in fieldwork and processing practices and in outputs.

In an attempt to overcome these shortcomings the Social Survey Division of the (then) Office of Population Censuses and Surveys agreed to undertake work to introduce common classifications, definitions and standards for social survey questions, and to improve comparability between social statistics. All this with a view to 'harmonising' the surveys as far as possible without compromising or jeopardising their objectives, and to provide a robust methodological structure within which future developments to these surveys could be framed.

Harmonisation concentrated initially on standardising the inputs to surveys and a differentiation was established between 'primary' questions and concepts (i.e. relevant to all surveys) and those of a 'secondary' nature (relevant to a subset of surveys).

Primary topics reviewed by the harmonisation project team included the definition of household response unit, household composition (sex, age, marital status and co-habitation, etc), ethnic group, tenure, economic activity, industry, occupation, employment status, socio-economic classifications, full/part-time work and income classification. Secondary topics covered were social security benefits and allowances, consumer variables, income from main job as an employee, income from self-employment, accommodation type, household costs and benefits, vehicles, and period of residence at the current address.

Clearly not all of these topics are relevant to the LFS but a number of harmonised alternatives to existing LFS questions were successfully tested in the pilot and dress rehearsal for the 1996/7 surveys in the following subject areas:

- demographic data (especially collecting details of household composition about all members of a household whether or not they respond to the main body of the survey);

- period of residence at current address;

- housing tenure;

- collection of comprehensive data on the relationships between all members of each household; and

- ethnicity.

These harmonised questions were incorporated into the main LFS from spring 1996.

Certain LFS questions were considered suitable for adoption on other surveys - for example, LFS questions on marital status and economic activity. A comprehensive description of the harmonisation process and the initial outcome from the project has been published by the (then) OPCS in the booklet *'Harmonised Questions for Government Social Surveys'*, published in September 1995.

During 1996 work continued on the harmonisation of survey *inputs* - question wording and response categories, edits, interviewer instructions - and *outputs* - categories for publication or analysis, for example.

An updated and expanded version of the GSS booklet on harmonised questions, renamed *Harmonised Concepts and Questions for Government Social Surveys*, was published in November 1996. The revised booklet adds interviewer instructions and edit checks to the questions and harmonised output concepts for several topics. It also made a number of changes to some of the questions. Most recently, a December 1997 update to the Harmonised Concepts and Questions booklet was published: this extended the range of harmonised inputs and output concepts. An up-to-date version of this publication is included in the National Statistics website: http://www.statistics.gov.uk/about/data/harmonisation/defualt.asp.

ONS is increasingly positioning itself to make greater use of administrative data, for example, to support Neighbourhood Statistics. In order to exploit such data it will be important to extend the principles of harmonisation beyond surveys and the Census. This is likely to be the focus of new harmonisation activities over the next few years. Other factors influencing harmonisation activities include:

- Eurostat requirements – developing harmonised Key Social Indicators
- Emerging social topics such as social capital, e-society and cultural identity.

Harmonisation and the LFS - Inputs - Potential for Discontinuities

Whilst the benefits of harmonisation are clear, it is also the case that changing questions or interviewing practices risks the possible introduction of discontinuities. Analysis of responses to the harmonised questions on ethnic origin and housing tenure questions in 1996 showed that they had not caused significant discontinuities. A further change to the ethnic origin categories was incorporated into the LFS in March 1997[1]. Again, there is no evidence that this led to a discontinuity. From Spring 2001, the Labour Force Survey introduced new questions on ethnicity based on recommended output classification of ethnic groups from the 2001 Census. The new classification has two levels. Level 1 is a broad classification into 5 main ethnic groups. Level 2 nests within Level 1 and provides a finer classification. No comparison should be made between the old and new ethnic classifications in the LFS, because not only are the categories different but, the questions and coding of answers underlying the data are also very different.
More information on the classification of ethnic groups can be found on the National Statistics website: http://www.statistics.gov.uk/about/ethnic_group_statistics/

In addition to this new classification, ONS recommended the collection of national group information in recognition of user requirements. The national categories tested and introduced are English; Scottish; Welsh; Irish; British and Other. In addition to presenting the overall results from the national group question, analyses of national group data of particular ethnic groups could also be presented according to user needs.

Harmonisation and the LFS - Outputs

The harmonised output categories for economic status are consistent with those used in the LFS. Their development resulted in an improvement in the routing of the harmonised question on reasons for economic inactivity to allow the classification of all such persons, consistently with the LFS.

[1] Up to February 1996, LFS respondents who considered themselves “Black”, but not “Black – Caribbean” or “Black – African” could be recorded as “Black – Other”. Between March 1996 and February 1997 this latter category was harmonised to “Black – neither Caribbean nor African”. From March 1997, it has been “Black – Other Black Groups”.

Harmonised Questions for Government Social Surveys - LFS (as at May 2005)

Variable	Whether harmonised in the LFS	Comments
1. Primary set		
Household response unit	No	See (i)
Gender	Yes	
Date of birth	Yes	
Age	Yes	
Legal marital status	Yes	
Living arrangements	Yes	
Who owns or rents accommodation/Tenure	Yes	
Household Reference Person (HRP)	Yes	
Relationship to HRP	Yes	
Ethnic origin	Yes	
National Identity	Yes	
Economic status	No	See (ii)
Employment status	No	See (ii)
Industry – SIC code	No	See (ii)
Occupation – SOC code	No	See (ii)
Social class	No	See (ii)
Socio-economic group	No	See (ii)
Full-time/part-time work	No	See (ii)
Geography – use of GORs	Yes	
2. Secondary set		
Social security/benefits	No	Equivalent questions are asked, but they are not harmonised
Consumer Durables	No	Not in survey
Accommodation type	No	Not in survey
Length of Residence	Yes	
Household motor vehicles	Yes	Introduced in 2001/2
General health – limited activities	No	
Educational attainment	No	Though can be derived from input variable
Qualifications	Yes	
Length of time since last did paid work	No	See (ii)
Time in present job	No	See (ii)
Usual hours worked in main job	No	See (ii)

Notes:

(i) The harmonised definition of the household response unit is “one person or a group of people who have the accommodation as their only or main residence and (for a group) either share at least one meal a day or share the living accommodation, that is, a living room or sitting room”. The LFS definition differs slightly. The LFS adds students who live in halls of residence in term-time and residents in National Health Service accommodation to the coverage allowed in the harmonised definition, but these are clearly identified and the harmonised definition can be derived.

(ii) The LFS uses the International Labour Office (ILO) definition of economic status. The harmonised input is *based* on this definition, but differs in minor respects since the ILO standards depend on more complex questions than are possible for a harmonised question for general social surveys. The minor differences with the LFS affect people who were on a government supported training scheme and the classification between full-time and part-time work. This departure from the harmonised question affects several topics, including economic

activity and usual hours in main job. The result of this is while the questions relating to economic activity are mostly harmonised; the outputs differ from the harmonised outputs.

SECTION 16 - USES OF THE LFS

Introduction

The Labour Force Survey (LFS) began as a condition of UK membership of the European Community and was carried out biennially from 1973 - 1981 and annually from 1984 - 1991. Over this time Government departments, especially the Employment Department, found the information collected in the LFS increasingly valuable in the framing of social and economic policy. In 1990, the Secretary of State for Employment announced the development of a quarterly LFS which began in spring 1992.

The main purpose of the quarterly LFS is to provide information needed to develop, manage, evaluate and report on labour market policies. Currently, interviewing takes place in approximately 53,000 households a quarter, yielding labour market and demographic information about some 120,000 adults. Its main strengths are that it provides a self-contained, integrated source of information about the Labour market activity (or inactivity) of the whole (household) population, based on a large sample size, and that it uses the internationally standard definitions of employment and unemployment recommended by the International Labour Organisation (ILO).

Topics covered by the LFS

The LFS provides regular information relating to the following topics:

- demographic characteristics of the population;
- employment, unemployment and inactivity;
- qualifications held and in the process of being attained;
- job-related training;
- trade union membership and the coverage of collective bargaining,
- industrial accidents and their causes;
- work related illnesses;
- earnings and sources of income

Macro-economic monitoring

The quarterly LFS is highly valuable in helping to assess changes in the labour market. First key results are now published one and a half months after the survey period ends, with full results available two months later. Main indicators regularly published from the LFS include -

- ILO unemployment, total employment, ILO unemployment rate and economic activity rate (employment and unemployment as a percentage of the total population), by age group;

- employees and self-employed people, full- and part-time workers, second jobs and temporary workers, by industry and occupation;

- average actual working hours and total hours worked in the economy;

- redundancies;

- reasons why people are economically inactive (not employed or unemployed) and whether they would like to work, including groups such as:

    - discouraged workers - those who say they would like to work but have not looked for work recently because they believe no jobs are available and therefore are excluded from measures of unemployment;
    - people (usually women) looking after the family or home;

- students;
- retired people;
- people unable to work because they are sick or disabled.

The LFS is useful as an alternative source of information, relying, on a different collection method, with which to compare the trends shown by the claimant count of unemployment and the surveys of employers about employees. Each source has its own strengths and weaknesses[2] in particular, the articulated nature of the LFS means that it can provide important information to explain such unexpected (to the casual observer) phenomena as a fall in unemployment at the same time as a fall, or a smaller rise, in employment. The LFS may be able to show that the difference is explainable, for instance, by an increase in the number of people in full-time education, information which is not available from unemployment or employment records. The LFS also provides estimates for sections of the labour force who are not covered by the employer surveys, such as the self-employed and temporary employees, or the claimant count of unemployment such as those ineligible for unemployment-related benefits (e.g. most under 15 year olds), and those with a low propensity to claim (such as married women).

The LFS provides the basis for labour force projections which provide an assessment of the likely chances in the composition of the labour force over the next 10-15 years. These projections assist in the formulation of policies which will take account of predictable chances in the economically active population.

The "flexible" labour market

The LFS collects a wide range of information about people's employment, such as type of employment, industry and occupation, identifying separately their main job and any second job they might have. The LFS is the only source of quarterly statistics on self-employment, temporary workers and the type of contract they have - fixed period/task, agency work, casual etc, none of which is available as frequently from any other source. Because it is a survey of people not employers, the LFS can show the mix of employment types varying from full-time to part-time and temporary, self-employed and unpaid working for a family business. This basic information can be linked to more in depth results such as the reasons why people work part-time, such as the proportion who do so because they could not get a full-time job. The LFS is also the only regular source of estimates of the extent of homeworking.

The survey collects information on usual and actual working hours, including separate figures for overtime, used to show, for example, that this country has the most varied pattern of working hours in Europe. Questions are also asked about evening, Saturday and Sunday work. A research feature in the January 2000 *Labour Market Trends* explored the data available from the LFS on working patterns, and describes the characteristics of people who work flexibly.

Another aspect of the flexibility of the labour market is labour mobility and the LFS helps to monitor this by means of questions asking people about the job they were doing one year earlier, and whether they moved to find work. The survey also asks how long employees have been working with their current employer and if they have left a job recently, people are asked the reason why. It is also possible to identify people who have returned to the labour market since the previous year, such as women returning to work after a break to bring up a family.

Regional statistics

Regional data have always been available from the LFS and now a limited number of key - variables on employment and training are provided for local authority districts and Training and Enterprise Council (TEC) areas. This helps TECs and the Government Offices for the regions to assess local labour markets to inform their planning processes and to advise local people and businesses. Although small area data are not the LFS's strong point, the estimates which are available go some way towards meeting the need for information about

[2] For a comparison between LFS and claimant count estimates of unemployment see *Labour Market Trends,* February 2004.

areas such as inner cities and rural areas, whose special needs are considered on an interdepartmental basis.

The characteristics of the unemployed

The information about the characteristics of unemployed people which is available from the LFS, such as marital status and qualifications, complements the information collected about benefit claimants. The LFS is able to identify groups of interest such as disabled people and lone parents who may face particular problems in getting work, and people from ethnic minorities. It also provides information about the duration of unemployment, and the occupations and industries where the unemployed previously worked.

The Department for Work and Pensions (DWP) uses information from the LFS to help devise and assess services to help people not in work. They are interested in the reasons why people do or do not seek work, and the methods they use, both to judge the effectiveness of their policies and to encourage active and effective job search. The LFS is the main source for monitoring redundancies. A recent article in *Labour Market Trends*[1] describes the characteristics of redundant workers and this information also helps the DWP to improve their understanding of this group and the influences on their chances of returning to work.

The LFS uses the internationally standard ILO definition of unemployment[2]. Respondents are also asked whether they were claiming unemployment related benefits. The LFS helped to inform Government about the number of people who were likely to be affected by the changes from Invalidity Benefit and Sickness Benefit to Incapacity Benefit (introduced in April 1995) and from Unemployment Benefit to the Jobseeker's Allowance in 1996. It helped to monitor the effects of the introduction of Incapacity Benefit and JSA both on the claimant count of unemployment and on the ILO measure from the LFS.

Training and qualifications

A number of the Department for Education and Skills' (DfES) publications, including Trends in Education and Skills[3], make extensive use of the LFS. The survey is a key source of information about the amount and type of training done (particularly job related training). Information on the qualifications and employment status achieved by people in different categories - women and ethnic minority groups, in particular - helps to inform policy on further action in the area of training which may be beneficial in promoting equality of opportunity in the labour market.

The Learning and Skills Council is responsible for planning and funding vocational education and training in England and it uses the Labour Force Survey as a major source of information when evaluating their effectiveness[4]. Such information also forms part of a range of indicators used to assess the effectiveness of the Training and Enterprise Councils contracted to manage the provision of training for young people and unemployed adults around the country.

Work relating to policies and programmes aimed at increasing adult commitment to learning, (eg, Career Development Loans, Small Firms Training Loans) requires information from the LFS as comparative background information when monitoring the performance of such programmes in terms of participation rates of groups including women, people with disabilities and those from ethnic minorities.

The youth labour market

The LFS is an important source of information about the youth labour market. In particular, it provides up-to-date, quarterly, information about whether young people are in education,

[1] "Redundancies in the UK", *Labour Market Trends,* May 2004.
[2] The ILO definition of unemployment covers persons: without a job, available to start work in the next fortnight and had actively looked for work in the last four weeks or had found a job and were waiting to start.

[3] See http://www.dfes.gov.uk/trends/index.cfm
[4] http://readingroom.lsc.gov.uk/lsc/2005/research/commissioned/skills-in-england-2004-vol-2.pdf

which can be combined with information about their economic activity to reflect the multiple activities that they are often engaged in. The LFS is also the primary source of statistics on apprenticeships.

Working conditions

The LFS helps to monitor the coverage of the provisions of the employment protection legislation and to assess the number of people who might be affected by proposed changes. The survey provides estimates of the numbers of employees who qualify for the right to go to an Employment Tribunal if they feel they have been dismissed unfairly (i.e. having completed one years' service). This helps to forecast the number of cases likely to come to the Employment Tribunals. The LFS also provides information on the number of people in small workplaces, where legislation may create a different burden. The LFS also is the only regular source of information on the holiday entitlements of full- and part-time employees which is of interest in relation to the EU directive on working time.

Trade union membership

The LFS is an important source of information about the level of trade union membership, filling gaps in other sources. The demographic and employment data collected by the LFS is useful in analysing the extent of trade union membership among different groups in the population (e.g. ethnic minorities), sectors of industry, small workplaces, the public sector etc. The LFS also provides a measure of the extent to which employees' pay and conditions are determined by collective bargaining arrangements. These data provide a useful adjunct to workplace based estimates of collective bargaining[5]. An article used to appear every year in *Labour Market Trends* on this topic.

Incomes

Since winter 1992/93, the LFS in Great Britain has included questions on employees' earnings and other household income. After careful evaluation, these data were released for public use in December 1994 and described in an article in *Employment Gazette.* Income questions were included in the LFS in Northern Ireland from Winter 1994/5. There are other sources of earnings data (e.g. the Annual Survey of Hours and Earnings (ASHE)), but the LFS data is largely unique in that it covers groups such as temporary employees, part-timers and the low-paid, who are not necessarily covered by employers' records. For this reason the LFS is a key source of data for the Low Pay Commission when setting the National Minimum Wage[6]. The LFS has been used extensively to explore the relationship between pay and qualifications[7]. It has also been used to provide data for the European Union survey on the Structure of Earnings.

Equal opportunities at work for women, people from ethnic minorities, people with disabilities and older workers.

The LFS is a key source of statistics on the characteristics and labour market status of people from different ethnic groups, women, people with health problems and disabilities and older workers. This information is used in monitoring and promoting equal opportunities regardless of race, sex, disability or age, both in the workplace and in other fields covered by government.

The information available from the LFS assists in taking into account relevant factors such as levels of qualification and age when considering the position of particular groups in the labour force, and possible reasons for differences in employment and unemployment levels between them. For example, LFS results contribute to the monitoring of the industrial and occupational segregation of ethnic minority people and women, and their progress in achieving managerial

[5] For example, see Inside the Workplace: First Findings from the 2004 Workplace Employment Relations Survey, DTI
[6] The Annual Report of the Low Pay Commission includes extensive analysis based on LFS earnings data (see http://www.lowpay.gov.uk/lowpay/report/pdf/DTi-Min_Wage.pdf)
[7] For example, Walker and Zhu, "Education, Earnings and Productivity: recent UK evidence", *Labour Market Trends,* March 2003.

positions. The survey also provides information about the types of work done by people with health problems and disabilities, and the types of work previously done by those who have given up employment.

The Equal Opportunities Commission makes great use of the LFS, including many items based on the survey, in their annual publication *Facts about Women and Men in Britain*[8].

Households and families

The LFS records information about all members of a household so it is possible to look at family and household characteristics. This aspect of the data has most commonly been used to monitor labour market participation in households, a recent update of which appeared in the November 2004 edition of *Labour Market Trends.* Information about women with dependent children, including lone mothers, is available from the LFS and is used to monitor their participation in the labour market, and to help assess the support needed by working mothers, through childcare provision and other policies. There is also interest in questions such as whether unemployed people tend to have unemployed partners and whether people on low earnings are often in low income households. Rather more work on employment and earnings at the household or family level has been done by academic researchers, using the LFS over a 10 year period, for example[4]. Data from the LFS on incomes has contributed to the debate on low income households and the concept of a minimum wage.

Work-related accidents and illness

The Health and Safety Executive (HSE) requires a benchmark against which to interpret the information on workplace accidents reported by employers, which is known to be incomplete. They are interested to know both the level and trends in workplace accidents and the variation in risks between the main sectors of industry. The LFS results were a major input to a recent review of the reporting regulations and will be used to judge whether or not the revised regulations are working. Data are used to inform the allocation of inspectors based on the level of risks at a detailed industry level, and the Annual Report to the Health and Safety Commission regularly features data from the LFS.

Longitudinal analysis

The design of the LFS makes it possible to conduct longitudinal analysis. Datasets are produced linking respectively two and five consecutive waves of data, including all people of working age who respond at each of the waves. To date, analyses of these data has been largely restricted to the study of labour market flows[9].

Links with International Organisations

The LFS is an European Union survey. The UK LFS includes all the questions required by the EU and the data are sent to Eurostat (the Statistical Office of the European Communities) each quarter. The ILO recommended definitions of employment and unemployment, which are used in the LFS are also used in similar surveys in other major nations of the world such as the USA, Canada and Australia. LFS data are also widely used by organisations such as the Council of Europe, the United Nations, ILO and OECD for international comparisons.

LFS data are used to work out the cost to the UK of various proposed EU directives relating to employment conditions. Information about foreign nationals living and working in the UK and about corporate transfers required by the European Union to monitor the freedom of movement of workers within the EU is obtained through the LFS. The LFS also asks about people who have worked, or applied for a job abroad in the last five years.

[8] See http://www.eoc.org.uk/cseng/research/facts_about_GB_2005.pdf
[4] For example, Harrop and Moss, "Working parents: trends in the 1980's", *Employment Gazette,* October 1994.
[9] For example, "People leaving economic inactivity: characteristics and flows", *Labour Market Trends*, April 2002.

Other Government Departments and Agencies

Many government departments also use the LFS for purposes not directly related to the labour market. Some Departments sponsor a limited number of questions in the survey. For example, the Department for Transport sponsor questions on place of work, mode of travel and time taken to travel to work to supplement those collected in the decennial population census. The Home Office makes significant use of the questions on ethnicity and country of birth asked in the LFS to support policy on race and immigration issues.

Other Departments who do not sponsor questions in the LFS, use data collected in the survey primarily for different purposes. The LFS is the only statistical source of information between decennial population censuses which gives estimates of the size of the different ethnic minority populations in Great Britain and this information is used by ONS. ONS also uses the LFS to obtain estimates of the numbers, and characteristics, of households and families, especially of one-parent families. The Home Office uses estimates from the LFS as a benchmark against which to monitor different ethnic groups in the criminal justice system. The Department for Education and Skills uses the LFS for assessing the educational participation and qualifications of the population as a whole. HM Treasury is interested in the potential of the LFS to provide information of labour inputs (as hours worked) for calculations of industrial productivity. The Monetary Policy Committee who advise the Treasury on the economy, and the Bank of England, also regularly use information from the LFS. Various other departments such as Inland Revenue, Welsh Assembly Government and Scottish Executive are also regular users of the LFS.

Dissemination to the wider public

The LFS is widely used by Government Departments for analysis of the labour market and to develop government policies in this field. The Government Statistical Service is also committed to providing statistical information as a basis for informing the wider public debate. LFS results are made publicly available to provide the basis for research, analysis and debate about the labour market in the following ways:

- *Labour Market Statistics First release*
- *Labour Force Survey Historical Quarterly Supplement*
- LFS data for TEC/LEC areas, counties and local authority districts available on Nomis® (National Online Manpower Information System);
- databases at *ESRC Data Archive* at Essex University, for academic research;
- tables, feature articles and the LFS Help-Line feature in *Labour Market Trend*s;
- LFS estimates are included in *Economic Trends, Social Trends, Monthly Digest, Regional Trends, Social Focus, and Regional Profile* publications produced regularly by ONS.
- The LFS is widely used by local authorities, the CBI and other employer organisations, the TUC and individual trade unions, by labour market analysts in the City, economics correspondents in the broadsheet newspapers, and researchers in a wide variety of other organisations ranging from the Unemployment Unit to the Institute for Employment Studies, from major retailers to solicitors. The House of Commons Library has access to the LFS through Nomis® and the dial up service provided by the LFS Data Service.

This article has described many, but not all, of the enormous range of uses to which Labour Force Survey results are put. If you would like more information about the Labour Force Survey or how to access it, please telephone the Labour Market Helpline on 020 7533 6094.

SECTION 17 - LFS DISSEMINATION AND PUBLICATIONS

AVAILABILITY OF ELECTRONIC DATA

The following LFS data exist electronically:

- annual LFS data for each survey from 1984-91
- annual "time series" database covering period 1984 to 1991
- each quarter's results since spring 1992 on both a regional and sub-regional basis
- household databases for Spring 1990, Spring quarters 1992-95 and Spring and Autumn quarters from 1995 onwards
- local area (counties, LADS, TECS, LECs) tabulations from spring 92 onwards (limited set of variables).
- 2 quarter and 5 quarter longitudinal datasets from winter 92/93. (available as portable SPSS files with a limited set of variables).

1. Services available from ONS

The LFS Data Service can provide clients with tabulations from the LFS. Tables can be provided in Excel.

☎ LFS Data Service, 01633 655732
✉ LFS.dataservice@ons.gov.uk

Customers may have full LFS databases sent to them regularly or on an ad hoc basis for use on PCs at their own site. Databases are currently available in SPSS, SAS and SuperCROSS formats.

☎ LFS Data Dissemination, 01633 655732
✉ LFS.dataservice@ons.gov.uk

Provision of these services will attract a charge, which will be detailed at the time of application.

2. SuperCROSS Service

The LFS now allows users to access data in SuperCROSS format. Training is provided by ONS in how to use this system, which can be accessed online by external users. Subscription prices are as follows:

Database	Annual Subscription
Regional Quarterly Datasets	£250 for all 4 quarters
Household Quarterly Datasets	£100 for 2 quarters
LADB's (1992-1999)	£100 for all years
Annual datasets (1984-1991)	£100 for all years
Annual datasets (1979-1983)	£50 for all years
APS Jan-Dec Datasets	£100 for 2 years-more to be added each August

☎ LFS Data Service, 01633 655732
✉ LFS.dataservice@ons.gov.uk

3. The Data Archive

Formerly known as the ESRC Data Archive, based at the University of Essex, hold copies of all LFS databases. Academic users can access the data at specially agreed rates.

☎ The Data Archive, 01206 872017/873574
✉ Archive-Userservices@essex.ac.uk
www.data-archive.ac.uk

4. Nomis®

The National On-line Manpower Information System holds local area LFS data.

☎ Nomis®, 0191-374-2468
www.nomisweb.co.uk

5. Statbase

This is an on-line database containing up to date *seasonally adjusted* rolling monthly and calendar quarter data, as well as historical data (back to 1992). *Non-seasonally adjusted* data is also available for Spring quarters from 1984 to 2000. Statbase is only available via the National Statistics website (www.statistics.gov.uk). If you do not have access to the internet, printed copies or Excel copies of this data is available from the ONS.

☎ Labour Market Statistics Helpline, 020-7533-6094
✉ Labour.Market@ons.gov.uk

6. The Labour Market Statistics First Release Historical Supplement

The Historical Supplement to the Labour Market Statistics First Release is a series of data tables on the National Statistics website relating to various aspects of the labour market. Where possible, the tables aim to give access to the full historical data series, which cannot be shown in the First Release.
http://www.statistics.gov.uk/Onlineproducts/LMS_FR_HS.asp

PAPER PUBLICATIONS OF LFS RESULTS

A number of publications, either in part or in total, provide LFS results. The main ones are:

1. Labour Market Statistics First Release

In April 1998 most of the existing Labour Force Survey First Release became subsumed in the new integrated Labour Market First Release (LMS FR). The LMS FR contains LFS data for the most recent 3-month period, alongside other labour market statistics, giving a coherent picture of the labour market within the UK. The LMS FR is published about 6 weeks after the end of the LFS reference period.

☎ The ONS Press Office on 020 7533 5707

2. LFS Quarterly Supplement and LFS Historical Quarterly Supplement

In April 1998 Labour Market Trends was also re-launched; the LFS Quarterly Bulletin was redesigned and is now issued as a supplement to Labour Market Trends. The Supplement, which is available separately from Labour Market Trends, is issued in May, August, November and February. It contains all of the most recent quarterly data which has been previously been published in the LMS FR, along with further tables, charts, maps and commentary. It is the most comprehensive source of published LFS data. This publication is currently available on the National Statistics Website.

Launched in 2006 the LFS Historical Quarterly Supplement replaced the LFS Quarterly Supplement and consists of 44 tables in the form of excel spreadsheets. Supplements can be found at the following web address: http://www.statistics.gov.uk/StatBase/Product.asp?vlnk=14365.

3. Labour Market Trends

Labour Market Trends contains tables of historical LFS data, along with other labour market data. In addition it contains articles containing data from the Labour Force Survey. Some of these are produced annually - for example, Analysis of the Labour Market, Characteristics of Ethnic Minorities. The articles which are available from 1995 are shown below:-

- Foreign workers in the UK Jan 1995
- Redundancies in Great Britain Jan 1995
- Apprentices and other long-term trainees Feb 1995
- Older workers Apr 1995
- Trade Union membership and recognition May 1995
- An analysis of working time, 1979-94 May 1995
- Revisions to the Quarterly LFS; re-weighting and seasonal adjustment review May 1995
- Ethnic groups and the labour market Jun 1995
- Disability and the labour market: findings from the LFS Dec 1995
- Re-weighting of the annual LFS results 1984-91 Dec 1995
- The new Hours Worked Series Dec 1995
- Measuring employment: comparison of official sources Jan 1996
- Redundancies in Great Britain: results from the LFS Feb 1996
- Membership of trade unions in 1994: an analysis based on information from the Certification Officer Feb 1996
- What happens to men and women with SET degrees? Feb 1996
- Options for producing monthly estimates of unemployment according to the ILO definition Mar 1996
- Women in the labour market: results from the spring 1995 LFS Mar 1996
- Earnings data from the LFS and the New Earnings Survey Apr 1996
- Longitudinal data from the LFS Apr 1996
- Trade union membership and recognition: an analysis of data from the 1995 LFS May 1996
- The new LFS local area database May 1996
- Ethnic minority participation in the labour market: trends from the LFS 1984-95 Jun 1996
- The identification of full-time students in the LFS Jul 1996
- Annualised redundancy data calculated from the LFS Jul 1996
- Disability and the labour market Sep 1996
- The Labour Force Survey in Northern Ireland Nov 1996
- Parental employment in the European Union Dec 1996
- Labour market recoveries in the UK and other OECD countries Dec 1996
- Spotlight on the South West Jan 1997
- British labour force projections 1997-2006 Feb 1997
- Women in the labour market: results from the spring 1996 LFS Mar 1997
- Data on household and families from the LFS Mar 1997
- Redundancies in Great Britain: results from the LFS Apr 1997
- Spotlight on Yorkshire and the Humber May 1997
- Household and Family data from the LFS (illustrative data) Jun 1997
- Trade union membership and recognition Jun 1997
- Working Fathers Jul 1997
- Trends in labour market participation of ethnic groups: 84-96 Aug 1997
- Spotlight on the West Midlands Sep 1997
- Economic Activity Sep 1997
- Temporary Workers in Great Britain Sep 1997
- Measuring labour market attachment using the LFS Oct 1997
- LFS estimates of claimants of unemployment-related benefits: results of an ONS record linkage study Nov 1997

- Assessment of the quality of data from the spring 1997 LFS disability module Dec 1997
- Spotlight on Wales Dec 1997
- Comparison of sources of employment data Dec 1997
- Averaging LFS data over several quarters Jan 1998
- Qualifications data in the LFS Jan 1998
- Monthly publication of up-to-date quarterly data from the LFS Feb 1998
- Improved ONS labour market statistics Feb 1998
- Benefits data from the LFS Feb 1998
- Women in the labour market Mar 1998
- Self-employment in the 1990's Mar 1998
- Expanding the coverage of earnings data in the LFS Apr 1998
- Towards reconciliation of NES and LFS earnings data May 1998
- The new presentation of labour market statistics: guidance for users about sources May 1998
- The annual seasonal adjustment review Jun 1988
- Disability data from the LFS: comparing 1997-98 with the past Jun 1988
- Discontinuity in LFS data on education courses Jun 1988
- Analysis of household data from the Labour Force Survey Aug 1998
- Production of household data sets from the Labour Force Survey Aug 1998
- Additions to Labour Force Survey household tables Sep 1998
- Industry comparisons of employment estimates Oct 1998
- LFS grossing: the management of change Nov 1998
- Labour market participation of ethnic groups Dec 1998
- LFS Household data: spring 1998 analyses Jan 1999
- Spotlight on the East Midlands Feb 1999
- Women in the labour market Mar 1999
- The 1997/8 LFS Annual Local Area database Apr 1999
- Changes to the LFS seasonal adjustment methodology Apr 1999
- Spotlight on London May 1999
- Redundancies in the United Kingdom May 1999
- Update on Labour Force Survey household datasets May 1999
- Owner-managers in the LFS Jun 1999
- Trade union membership and recognition 1997-98 Jul 1999
- Survey of users of ONS labour market statistics products and services Jul 1999
- Review of the labour market statistics First Releases Jul 1999
- Longitudinal data from the Labour Force Survey Jul 1999
- Seasonal adjustment review of the Labour Force Survey Jul 1999
- Sickness absence in Great Britain Aug 1999
- Using the LFS to estimate time-related underemployment Aug 1999
- Estimating ILO unemployment before 1984 Aug 1999
- Trend estimation of LFS data Aug 1999
- Disability and the labour market Sep 1999
- Evaluation of new benefits data from the LFS Sep 1999
- Spotlight on the North East Oct 1999
- Lone parents and the labour market revisited Nov 1999
- Trends in the labour market participation of ethnic groups Dec 1999
- Spotlight on the South East Jan 2000
- LFS household data: Spring 1999 analyses Jan 2000
- Employment rates 1959-1999 Jan 2000
- LFS questions on working patterns Jan 2000
- Improvements to LFS estimates: weighting and seasonal adjustments Feb 2000
- The 1998/9 Labour Force Survey annual Local Area Database Apr 2000
- The Local Labour Force Survey for England May 2000
- LFS: Regrossing and seasonal adjustment May 2000
- Redundancies: enhancing the coherence of LFS estimates May 2000
- Developments in the local area LFS data May 2000
- Spotlight on the North West Jun 2000
- Trade Union Membership 1998-99 Jul 2000
- Developing improved estimates of ILO unemployment for unitary authorities and local authority districts Sep 2000

- Review of unemployment rate measures Sep 2000
- Spotlight on the Regions Nov 2000
- Estimating job entry and exit from the Labour Force Survey Nov 2000
- Improved data on sickness absence from the LFS Dec 2000
- The introduction of additional questions on employment status to the LFS Dec 2000
- Labour Market participation of ethnic groups Jan 2001
- LFS household data: Spring 2000 analyses Jan 2001
- Measuring low pay using the New Earnings Survey and the LFS Jan 2001
- Women in the labour market: results from the Spring 2000 LFS Feb 2001
- Trends in female employment Feb 2001
- Report on the consultation on developments in local area LFS data Mar 2001
- Seasonal adjustment of ILO unemployed aged 18 to 24 series Mar 2001
- The 1999/2000 Labour Force Survey annual Local Area Database Apr 2001
- Disability and the labour market: results from the summer 2000 LFS May 2001
- The launch of the Annual Business Inquiry May 2001
- Seasonal adjustment review of the Labour Force Survey May 2001
- Redundancies in the United Kingdom Jun 2001
- Estimating participation in education, training and employment Jun 2001
- Changing to SOC 2000 – dual coding on the Labour Force Survey Jul 2001
- Time series analyses of the LFS two-quarter Longitudinal datasets Aug 2001
- Trade Union Membership 1999-2000; and analysis of data from the CO and the LFS Sep 2001
- Some labour market implications of employment legislation Sep 2001
- Foreign labour in the UK; patterns and trends Oct 2001
- People leaving employment; characteristics and flows Nov 2001
- The relationship between work-rich and workless household Dec 2001
- People and jobs; comparing sources of employment data Jan 2002
- Annual local area Labour Force Survey data for 2000/2001 Jan 2002
- Economic inactivity and the labour market Feb 2002
- Labour Force Survey Regrossing April 2002 Feb 2002
- Women in the labour market: results from the spring 2001 LFS Mar 2002
- A pilot survey of people living in communal establishments Mar 2002
- Labour market and demography project Mar 2002
- Trends and sources of data on sickness absence Apr 2002
- People leaving economic inactivity: characteristics and flows Apr 2002
- People with second jobs May 2002
- Regional labour market performances May 2002
- Labour Force Survey: regrossing and seasonal adjustment May 2002
- Patterns of economic inactivity among older men Jun 2002
- Teleworking in the UK Jun 2002
- Who trains? Employers' commitment to workforce development Jun 2002
- Trade union membership: an analysis of data from the autumn 2001 LFS Jul 2002
- Measuring jobs: levels, short-term changes and industry classification Jul 2002
- Analysis of the claimant count by age and duration including clerical claims Jul 2002
- Underemployment and overemployment in the UK Aug 2002
- Labour market experiences of people with disabilities Aug 2002
- Hours worked: a comparison of estimates from the Labour Force and New Earnings Surveys Aug 2002
- Managerial qualifications and organisational performance Aug 2002
- Migrant workers in the UK Sep 2002
- Effect of the introduction of SOC2000 on employment estimates Sep 2002
- Review of the framework for labour market statistics Sep 2002
- Volume of underemployment and overemployment in the UK Oct 2002
- Housing tenure and the labour market Oct 2002
- A new survey of job vacancies Oct 2002
- Review of the Labour Force Survey Oct 2002
- The economically inactive who look after the family or home Nov 2002
- Labour disputes in 2001 Nov 2002
- Trends in female employment 2002 Nov 2002
- Review of statistics on distribution of earnings Nov 2002

- Patterns of Pay Dec 2002
- The new ethnicity classification in the LFS Dec 2002
- The impact of bonus payments on the Average Earnings Index Dec 2002
- Interim LFS estimates consistent with the 2001 Census Dec 2002
- International comparisons of labour disputes in 2000 Jan 2003
- Methodology for the 2001/02 annual local area Labour Force Survey data Jan 2003

- Development of improved estimation methods for local area unemployment levels and rates Jan 2003
- Implications of population ageing for the labour market Feb 3003
- Earnings growth 1997 to 2002 Feb 2003
- Revisions to workforce jobs and comparison with Labour Force Survey jobs Feb 2003
- Job seperations Mar 2003
- A century of labour market change 1900-2000 Mar 2003
- Education, earnings and productivity Mar 2003
- Patterns of low pay Apr 2003
- International comparions of labour disputes in 2001 Apr 2003
- Migration within Britain for job reasons Apr 2003
- Modernising China's labour market statistics Apr 2003
- Job relocations and employer-assisted migration May 2003
- Linking together economic and social data May 2003
- Labour disputes in 2002 Jun 2003
- Economic Inactivity in selected countries Jun 2003
- The role of working-age benefits data in understanding labour market supply Jul 2003
- The Vacancy Survey: a new series of National Statistics Jul 2003
- Jobcentre Plus notified vacancy series Jul 2003
- Analysis of the claimant count by age and duration including clerical claims Jul 2003
- Composition of pay Aug 2003
- Job densities for local areas: a new indicator Aug 2003
- Self-employment in the UK labour market Sep 2003
- Jobs in the public sector Sep 2003
- Life stages of economic activity Oct 2003
- Key indicators of women's position in Britain Oct 2003
- How people answer Labour Force Survey questions about economic inactivity Oct 2003
- Job mobility and job tenure in the UK Nov 2003
- Labour market turning points Nov 2003
- Work permits and foreign labour in the UK Nov 2003
- Patterns of pay Dec 2003
- Economic inactivity among students Dec 2003
- Changes to self-employment in the UK: 2003-2003 Dec 2003
- Measuring working time arrangements Jan 2004
- Recent changes in hours worked, summer 2003 Jan 2004
- Changes in working trends over the past decade Jan 2004
- Improving labour market statistics in Ukraine Jan 2004
- Comparisons between unemployment and the claimant count Feb 2004
- The UK Time Use Survey from a labour market perspective Feb 2004
- Trade Union membership Mar 2004
- Skills shortages in skilled construction and metal trade occupations Mar 2004
- Working time patterns in the UK, France, Denmark and Sweden Mar 2004
- Characteristics of the short-term and long-term unemployed Apr 2004
- International comparisons of labour disputes in 2002 Apr 2004
- Ethnic differences in women's demographic, family characteristics and economic activity profiles, 1992 to 2002 Apr 2004
- Public sector pay growth by industry May 2004
- Redundancies in the UK May 2004
- Employment by industry and occupation Jun 2004
- Labour disputes in 2003 Jun 2004
- Jobs in the public sector mid-2003 Jul 2004
- Employment and unemployment in the new EU member countries Jul 2004
- Analysis of the claimant count by age and duration including clerical claims Jul 2004
- The effects of bonuses on earnings growth in 2004 Aug 2004
- The demand for data in the UK Aug 2004
- Local area job densities: 2002 Aug 2004
- The increase in employment in Wales during 2002 and 2003 Sep 2004
- Growth in self-employment in the UK Oct 2004
- Labour market data for local areas by ethnicity Oct 2004

- Workless households: results from the spring 2004 LFS Nov 2004
- Labour productivity Nov 2004
- Low pay estimates for 2004 Dec 2004
- International comparisons of labour market sources Dec 2004
- Employment and unemployment estimates for 1971 to 1991 Jan 2005
- Annual local area Labour Force Survey 2003/04 Jan 2005
- The difference between pay settlements and earnings growth Feb 2005
- The employment rate of older workers Feb 2005

Supplied by Palgrave Macmillan
✉ ons@palgrave.com

UK Annual Subscription (12 issues) £100, Overseas Subscription £126.

4. LFS Annual Reports

Between 1973 and 1991, OPCS produced reports on each of the surveys, which were published by HMSO (now TSO). The following reports are available from TSO on request:

OPCS, Labour Force Survey 1973, 1975 and 1977, (published 1980)
OPCS, Labour Force Survey 1979, (1982)
OPCS, Labour Force Survey 1981, (1982)
OPCS, Labour Force Survey 1983 and 1984, (1986)
OPCS, Labour Force Survey 1985, (1987)
OPCS, Labour Force Survey 1986, (1988)
OPCS, Labour Force Survey 1987, (1989)
OPCS, Labour Force Survey 1988 and 1989, (1991)

☎ The Stationery Office on 0870 6005 522
✉ Book.Orders@tso.co.uk

5. Training Statistics

Annually the Department for Education and Skills produces Education and Training Statistics for the UK as a publication of the Government Statistical Service.

☎ DfES Statistics on 0114-2591012

6. Social Trends

ONS publishes Social Trends annually. This publication includes data from the Labour Force Survey. It is also available electronically via the National Statistics website (www.statistics.gov.uk).

✉ Supplied by Palgrave Macmillan
✉ ons@palgrave.com

☎ ONS Direct on 01633 812078 (for copies on CD)

7. Regional Trends

ONS publishes Regional Trends annually. This publication includes data from the Labour Force Survey. It is also available electronically via the National Statistics website (www.statistics.gov.uk).

✉ Supplied by Palgrave Macmillan
✉ ons@palgrave.com

ARTICLES ON LFS METHODOLOGY IN OTHER PUBLICATIONS SINCE 1989

In addition, methodological issues and developments have been reported in a number of published articles:

- 'An evaluation of telephone interviewing on the British Labour Force Survey', Norris P, Blackshaw N and Wilson P, *Journal of Official Statistics*, 1989, Statistics Sweden

- 'The development of computer-assisted interviewing (CAI) for Household Surveys: The case of the British Labour Force Survey', Manners T, *Survey Methodology Bulletin* No. 27, July 1990, OPCS

- 'Developing Computer Assisted Interviewing on the Labour Force Survey: A Field Branch Perspective', Blackshaw N, Trembath D and Birnie A, *Survey Methodology Bulletin* No. 27, July 1990, OPCS

- 'The Enhancement of the Labour Force Survey in Great Britain'; Chamberlain E, Manners T and Bradley M, *Statistical News*, Winter 1991, HMSO

- 'Computer Assisted Interviewing for the Labour Force Survey'; Manners T, *Statistical News*, Winter 1991, HMSO

- 'The quarterly labour force survey - a new dimension to labour market statistics'; Chamberlain E, Purdie E, *Employment Gazette* October 1992.

- 'Estimating employment: a comparison of household and employer-based surveys'; Watson M, Spence A, *Employment Gazette,* October 1993

- 'Changes in the sample design for the LFS'; Knight I, *Survey Methodology Bulletin* No. 34, January 1994, OPCS

- 'Economic activity in local areas: some new results from the LFS'; Ellison R, Newman P, *Employment Gazette*, June 1994

- 'Income and earnings data from the LFS'; Laux R, *Employment Gazette*, December 1994

- ‘Revisions to the Quarterly LFS; re-weighting and seasonal adjustment review’, Emma Tonks*, Employment Gazette*, May 1995

- ‘Re-weighting of the annual LFS results 1984-91’, Richard Laux, *Labour Market Trends*, December 1995

- 'Longitudinal data from the Labour Force Survey'; Laux R and Tonks E, *Methods and Quality Papers*, May 1996, ONS

- ‘The New Labour Force Survey local area database’; Ian Wood, *Labour Market Trends*, May 1996

- ‘Revisions to the quarterly Labour Force Survey’; Sheena Gordon*, Labour Market Trends*, May 1997

- ‘LFS estimates of claimants of unemployment-related benefits: results of an OND record linkage study’, Penny Pease, *Labour Market Trends*, November 1997

- Comparison of sources of employment data; Penny Pease, *Labour Market Trends,* December 1997

- ‘Averaging LFS data over several quarters’; Frances Sly, *Labour Market Trends*, January 1998

- 'Monthly publication of up-to-date quarterly data from the LFS'; Richard Laux, *Labour Market Trends*, February 1998

- 'Improved ONS labour market statistics'; Labour Market Division, *Labour Market Trends*, February 1998

- 'Expanding the coverage of earnings data in the LFS', James Jenkins*, Labour Market Trends*, April 1998

- 'The new presentation of labour market statistics: guidance for users about sources'; Richard Laux, *Labour Market Trends*, May 1998

- 'Towards reconciliation of NES and LFS earnings data', David Wilkinson, *Labour Market Trends*, May 1998

- 'The annual seasonal adjustment reviews'; Sheena Gordon and Jenny Myall*, Labour Market Trends*, June 1998

- 'Production of household datasets from the LFS'; Pam Tate, *Labour Market Trends,* August 1998

- 'LFS grossing: the management of change', Richard Laux, *Labour Market Trends,* November 1998

- 'Changes to the LFS Seasonal adjustment methodology', Rachel Hill, *Labour Market Trends,* April 1999

- 'Longitudinal data from the Labour Force Survey', Pam Tate, *Labour Market Trends,* July 1999

- 'Update on Labour Force Survey household datasets', Emma-Jane Cooper-Green, *Labour Market Trends,* May 1999

- 'Seasonal adjustment review of the Labour Force Survey', Emmanuelle Modica, *Labour Market Trends,* July 1999

- 'Trend estimation of Labour Force Survey Data', David Hastings, *Labour Market Trends,* August 1999

- 'Evaluation of new benefits data from the Labour Force Survey', James Jenkins and Richard Laux, *Labour Market Trends,* September 1999

- 'Methodological Issues in the production and analysis of longitudinal data from the Labour Force Survey', P.S. Clarke and P.F. Tate, *GSS Methodology Series*, No. 17, 1999

- 'Labour Force Survey questions on working patterns', Rachel Hill, *Labour Market Trends,* January 2000

- 'Improvements to the LFS estimates: weighting and seasonal adjustment', Mehdi Hussain, *Labour Market Trends,* February 2000

- 'Labour Force Survey: regrossing and seasonal adjustment', Mehdi Hussain and Adrian Jones, *Labour Market Trends,* May 2000

- 'Developments in Local Area LFS data', Don Burke and Tricia Williams, *Labour Market Trends*, May 2000

- 'Changes to the design of the Labour Force Survey', Dave Elliot, *Survey Methodology Bulletin*, No 47, July 2000, ONS

- 'Investigating mismatches in data between waves on the Labour Force Survey using qualitative methods' Kirsty Deacon, *Survey Methodology Bulletin*, No 47, July 2000,ONS
- 'Variance estimation for Labour Force estimates of Level and change', D.J Holmes and C.J Skinner, *GSS Methodology Series*, No 21, 2000

- 'Report on the consultation on developments in local area Labour Force Survey data', Ann Blake, Don Burke and Tricia Williams, *Labour Market Trends*, March 2001

- 'Seasonal adjustment review of the Labour Force Survey', Lara Curran, *Labour Market Trends*, May 2001

- Changing to SOC2000 – Dual coding on the Labour Force Survey', Roeland Beerten, Laura Rainford and Adrian Jones, *Labour Market Trends*, March 2001

- 'Labour Force Survey regrossing April 2002', Allan Smith, *Labour Market Trends*, February 2002

- 'LFS Seasonal adjustment review and regrossing', Allan Smith and Lara Curran, *Labour Market Trends*, July 2002

- 'Effect of the Introduction of SOC2000 on employment estimates', Catherine Gibbins, *Labour Market Trends* September 2002

- 'Review of the Labour Force Survey', Elaine Chamberlain, *Labour Market Trends*, October 2002

- 'Trade union membership: an analysis of data from the autumn 2001 Labour Force Survey', Keith Brook, *Labour Market Trends*, July 2002

- 'Measuring Jobs: levels, short-term changes, and industry classification', Helen Ganson, *Labour Market Trends*, July 2002

- 'Labour market experiences of people with disabilities', Allan Smith and Breda Twomey, *Labour Market Trends*, August 2002

- 'Hours worked: a comparison of estimates from the Labour Force and New Earnings Surveys', Richard D. Williams, *Labour Market Trends*, August 2002

- 'Effect of the introduction of SOC2000 on employment estimates', Catherine Gibbins and Richard Laux, *Labour Market Trends*, September 2002

- 'Review of the Labour Force Survey', Elaine Chamberlain, *Labour Market Trends*, October 2002

- 'Trends in female employment 2002', Melanie Duffield, *Labour Market Trends*, November 2002

- 'The new ethnicity classification in the LFS', Allan Smith. *Labour Market Trends*, December 2002

- 'Seasonal Adjustment review of the claimant count series', Helen Treasure, *Labour Market Trends*, May 2003

- 'Developing a quarterly labour costs index', Derek Bird and Nerissa Wood, *Labour Market Trends*, June 2003

- 'Comprehensive manual for labour market statistics', Keith Tyrrell, *Labour Market Trends*, August 2003

- 'Experimental consistent time series of historical LFS data', Craig Lindsay and Paul Doyle, *Labour Market Trends*, September 2003

- 'Seasonally adjusting the Average Earnings Index excluding bonuses series', Helen Spyrou, *Labour Market Trends*, November 2003

- 'Investigating hours worked measurements', Richard D. Williams, *Labour Market Trends*, February 2004

- 'Labour Force Survey reweighting and seasonal adjustment review', Alex Clifton-Fearnside and Alyson Whitmarsh, *Labour Market Trends*, April 2004

- 'Seasonal adjustment review of the claimant count series', Helen Treasure, *Labour Market Trends,* May 2004

- 'Methodology for 2002/03 annual local area Labour Force Survey data', David Hastings and Joe Traynor, *Labour Market Trends*, June 2004

- 'A guide to interim reweighting and using Labour Force Survey microdata' by Trish McOrmond and Stephen Hicks, *Labour Market Trends*, September 2004

- 'Sources of data for measuring labour demand', Richard D. Williams, *Labour Market Trends*, September 2004

- 'Methodology for the 2004 Annual Survey of Hours and Earnings', Derek Bird, *Labour Market Trends*, November 2004

- 'An analysis of historical ASHE data 1998 to 2003', Chris Daffin, *Labour Market Trends*, December 2004

- 'Seasonal adjustment of the Vacancy Survey data', Helen Treasure, *Labour Market Trends*, December 2004

- 'Comparison of 2001 Census and Labour Force Survey labour market indicators', Daniel Heap, *Labour Market Trends,* January 2005

☎ The Stationery Office on 08706 005522
✉ Book.Orders@tso.co.uk

For any publications from the former OPCS contact the Labour Force Survey Unit, Social Survey Division, ONS, 1 Drummond Gate, London, SW1V 2QQ.

ALTERNATIVE SOURCES OF LABOUR MARKET DATA

Employment

The Workforce in Employment quarterly surveys of employers provides figures for employees in employment. They count jobs not people and are particularly valuable for providing estimates for specific industry sectors. A comparison of jobs from the two surveys - combining main and second employee jobs for the LFS - shows that the workforce series gives lower estimates of the number of employee jobs in the economy than the LFS, and that the gap has widened in recent years. For more details see 'Comparison of sources of employment data', *Labour Market Trends*, December 1997.

Unemployment

Figures for claimant count are produced every month from records of people claiming unemployment related benefits. These figures are available quickly and can be provided for small areas. The claimant count measure of unemployment and ILO unemployment are defined in different ways, and whilst there is some overlap between them, they could not be expected to correspond exactly. ILO unemployment tends to be higher than the claimant count with the smallest gap (72,000) occurring in spring 1993 when both measures were near a peak, and the largest (422,000) in spring 1990 when both measures were near a trough. For more details see 'LFS estimates of claimants of unemployment-related benefits: results of an ONS record linkage study'; Penny Pease, *Labour Market Trends*, November 1997.

Earnings & Hours

Statsitics from the New Earnings Survey (NES) were published for the last time in 2003. A new survey, the Annual Survey of Hours and Earnings (ASHE) has been developed as a replacement. The ASHE survey is based on the same coding frame as NES. The new methodology includes improved coverage of employees and weighting of earnings estimates. The data and variables collected remains broadly the same. An article will be published on the web providing a summary of the ASHE design.

More information is available at the ONS website at the following address:
http://www.statistics.gov.uk/about/data/guides/LabourMarket/.

SECTION 18 - LFS DATA FOR SMALL SUB-GROUPS: ANNUAL DATABASES AND AVERAGING OVER SEVERAL QUARTERS

INTRODUCTION

The quarterly Labour Force Survey (LFS) contains information based on about 53,000 households in the United Kingdom and for many analyses this is sufficient. For small groups in the population such as ethnic minorities and young people, and for small area analyses in general, a single LFS quarter does not provide a large enough sample to give reliable estimates when cross-tabulations of several variables are required. The average of a larger sample contacted over a longer period will provide estimates of greater precision. Users may also wish to calculate annual average values for their own sake, or in order to smooth out seasonal or other variation.

METHODS OF COMBINING LFS SAMPLES

A single LFS quarter includes information about some 150,000 individuals. This is sufficient to allow a wide range of labour market, educational and demographic analyses, but does not provide a large enough sample to give reliable results in detailed cross-tabulations for small population sub-groups or for local areas. The nature of sampling variability means that the smaller the group whose size is being estimated, the (proportionately) less precise that estimate is. Basing estimates on data for a larger sample can increase precision.

In order to meet the demand for more data at local authority district level, ONS released the first annual LFS Local Area Database (LADB) in May 1996 which covered March 1994 to February 1995. These databases contain a range of key variables together with a unitary authority local authority district (UA/LAD) identifier, which allows users to carry out cross-tabulations at local level for the first time.

Annual databases
To explain the concept of an annual database, it is first necessary to describe the panel design of the LFS, usually described as the 'wave' structure of the sample. Each quarter's sample of 60,000 households is made up of five 'waves' each of approximately 12,000 households. Each quarter a new group (wave) of households is selected and its members interviewed for the first time. The same people are interviewed four more times at quarterly intervals. In any one quarter, one wave will be receiving their first interview, one wave their second and so on. Up to 1999/2000, the annual LADB was created by taking waves 1 and 5 from each of four consecutive quarters.

Selecting waves 1 and 5 ensures that each respondent is included once and only once - a database of ‘distinct cases’. The selected records are weighted according to definitive mid-year population estimates. As a result of the larger sample size (60 per cent more than a quarterly LFS database), the standard errors of estimates (the usual measure of sample variation) from the local area database are about 80 per cent of those for estimates for a single quarter. This resulted in the minimum publication threshold being reduced from 10,000 for a quarter, to 6,000 for annual data.

From March 2000, there was a boost to the sample in England - a partnership project between the ONS, the Department for Work and Pensions (DWP) and the Department for Education and Skills (DfES). More details including a summary publication, fact sheets for UA/LADs and counties are available on the National Statistics website – www.statistics.gov/llfs. Respondents in the boost are interviewed once a year for four years. The increase in sample meant that for some UA/LADs, the publication threshold was reduced to 2,000 or 4,000.

From March 2001, there was a similar boost to the LFS sample for Wales – a partnership project between ONS and Welsh Assembly Government. Again, the increased sample resulted in lower publication thresholds for all the UAs in Wales, some as low as 1,000.

More information on local area LFS data is contained in Volume 6 of the LFS User Guide.

From March 2003 there was a similar boost to the LFS sample in Scotland – a partnership project between ONS and Scottish Executive.

From January 2004 there was an additional boost in England and the LADB (or ALALFS) was renamed the Annual Population Survey (APS). The boost was subsequently discontinued in January 2006 but the survey name remained as the APS. The APS is published quarterly.

Annual averages
An alternative method is simply to add together the weighted estimates for four consecutive quarters (from the quarterly LFS databases or the time-series database) and divide by four. Such estimates will be averaging both the numerator and denominator. Due to the panel design of the LFS, databases for consecutive quarters have approximately 80 per cent of their samples in common. Hence the calculation of the precision (standard errors) of annual averages is not straightforward. The improvement in accuracy or precision, (i.e. lower standard errors) arising from averaging estimates over a year is illustrated by the reduced thresholds for estimates regarded as sufficiently reliable for publication shown below.

Minimum publication thresholds

	Minimum publication level	95 per cent confidence interval
One quarter	10,000	+/- 4,000
Four quarters	6,000	+/- 2,640
Eight quarters	4,000	+/- 1,600
12 Quarters	3,000	+/- 1,200
16 Quarters	2,000	+/- 800

RELIABILITY OF AVERAGED QUARTERLY ESTIMATES

As a consequence of the overlap in samples between quarters, the precision of annual averages varies according to the correlation between responses from the same individuals in different quarters. Where responses are likely to change between quarters, as with economic activity variables, the gain from averaging is greater than with an annual database of distinct cases because the latter excludes additional information which has been collected about respondents in waves 2 to 4 each quarter. For variables which cannot change between quarters, such as ethnic origin, there is still a clear advantage, in terms of sampling error, in averaging data over a year compared with using a single quarter's data, although an annual database of distinct cases would produce somewhat more precise estimates than simple annual averages. In practice, users will `often be combining ethnic origin with labour market or education variables, so the difference between annual averages and annual database estimates would not be as great as this suggests.

Social Survey Division of ONS has provided some guidance on the level of estimates which would have the same relative precision as estimates of 10,000 in single quarter's LFS database. In other words, these are the smallest estimates which are considered reliable enough to use, based on a criterion of a maximum of 20 per cent coefficient of variation (the standard error as a percentage of the estimate). The recommended lower limits for reliable data for averages of successive quarters are shown in on the previous page.

The recommended thresholds are based on the quarter-on-quarter correlation exhibited by total employment and unemployment. They have been rounded up to the next 1,000, partly for simplicity and partly to allow for the additional loss of precision in variables which have virtually 100 per cent correlation. The calculation incorporates some design effects, to reflect the fact that the sample is clustered within addresses. However, these design effects vary widely for different ethnic groups and estimates up to two times these value may, for some groups, have confidence intervals as wide as those given on the previous page.

OTHER BENEFITS AND DRAWBACKS OF AVERAGING

Calculating the average of estimates from existing databases has advantages over the setting up of special annual databases. This method allows complete flexibility in the data that can be combined. Annual averages can be calculated for any consecutive four-quarter period

(combining numbers of quarters other than multiples of four is not recommended unless the data are seasonally adjusted).

Annual databases have to be created specially and therefore cannot be available immediately the last quarterly database of the year is released. The number of variables included is also likely to have to be limited due to the need to check for comparability between quarters (although, of course, it is not advisable to calculate averages over periods where the questionnaire has changed).

There is however, a cost to annual averaging in terms of the time taken to extract data and calculate averages over four (or more) quarters which has to be done each time averages are required.

For small area estimates there are other considerations than sampling errors when selecting a method of increasing the available sample size on which to base estimates. Carrying out the weighting of sample estimates to population totals as a last stage, as is done for the UA/LAD, is likely to produce more robust estimates for small areas than the averages of estimates from independently weighted-up quarterly databases. The UA/LADs also have the benefit of being weighted to the definitive mid-year population estimates rather than to projections as used for the quarterly databases. Hence, for the variables included, they may still be the better source for county estimates, although these are available from the quarterly databases and hence could be averaged.

Data for small sub-groups and areas always need to be interpreted with great caution. Differences between annual averages for different periods present a special problem since again there is some overlap in the samples for consecutive periods. In particular, drawing comparisons between annual averages for very small groups for periods ending one quarter apart (e.g. spring 2001 - winter 2001/02 compared with summer 2001- spring 2002) is not recommended.

For examining detailed characteristics of special groups, such as ethnic minorities, data could also be combined over two, three or more years. However, such estimates would be centred on a point a year or more in the past, making this less suitable for rapidly changing variables.

Whilst there are advantages to using annual averages, flexibility and timeliness, from 2000/01, the increased annual sample has resulted in more reliable estimates. As a result, ONS dramatically increased the amount of annual LFS data available free for local areas on Nomis®. Due to confidentiality constraints, ONS had to withdraw the publicly – available LADBs. However, annual LFS data are available, unrounded and unsuppressed with a disclaimer, from the Sub-National Data Service – contact LFS.dataservice@ons.gov.uk . A charge may be made for this service. The APS datasets also became available from January 2004.

NON-STANDARD VARIABLES

Earnings information is only available up to winter 1996/97 from fifth wave interviews, from spring 1997 it is available from first and fifth wave interviews. It is recommended that single-quarter estimates up to winter 1996/7 based on weighted-up estimates of less than 60,000 employees should not be used. From spring 1997 the threshold is 30,000 employees. When four quarters' earnings data are pooled, the threshold is unchanged as there is no overlap in the data. For annual databases, the thresholds were lower and from 1997/98, the thresholds for earnings data from the annual database are the same as for all other variables.

Some LFS variables are only available in two quarters each year (for example, flexible working). The comparative standard errors and reliability thresholds for multi-quarter averages quoted here do not apply to such variables. If there is demand, ONS would consider the appropriate guidance for such variables.

For variables which are only available once a year (for example, trade union membership), or indeed for other variables, users may wish to calculate averages of quarters a year apart. Here, the overlap (for the quarterly LFS conducted from 1992 onwards) is about 16 per cent.

The appropriate thresholds for publication in this case are: two years: 6,000; three years: 4,000; four or five years: 3,000.

CALCULATION METHODS

Annual averages at national, regional and larger local authority level can be calculated by extracting tabulations of the same data for four quarters and placing them in a spreadsheet for manipulation. Care needs to be taken to check that the variables are consistently defined in all the quarters to be averaged. The data extraction and manipulation can be quite time-consuming for detailed cross-tabulations, and where variables are already on the LFS Supercross database the process can be simplified by using the functions available in Supercross (averages can be calculated directly but care should be taken to check that the correct data have been included). At present the variables available are limited to about 60 of the most frequently used variables. It is hoped to extend the range of variables available on the time-series database during 1998, subject to continuity in the variables and availability of resources. A list of variables currently on the quarterly time-series database and those under consideration for addition is available on request.

USES OF MULTI-QUARTER AVERAGES

The value of averaging LFS estimates over four or more quarters is that the larger number of responses on which the estimate is based results in a more reliable estimate. This in turn means that smaller estimates become sufficiently reliable to use. One particular topic where annual averages are already proving to be of great value for this reason is ethnic minority groups. By means of annual averaging, users are able to analyse smaller ethnic minorities and/or smaller sub-groups, such as age groups.

One statistic of great interest among policy makers and ethnic minority special interest groups is the youth unemployment rate for different ethnic groups. Annual averages can also make more regional estimates for ethnic minorities available. Aggregation over several years would allow analysis of the ethnic minority populations and other groups and regions in more depth than is possible from four quarters. Such estimates, while not fully up-to-date, would nevertheless be more recent than the 1991 population census. An example of information which is only available from the LFS is people who do unpaid work for a family business. The number of these is small, and up to now no analysis of the types of occupations they are employed in has been published by *ONS.*

CONCLUSION

One of the aims of ONS is to make better use of the statistics collected. Averaging LFS data over several quarters enables reliable estimates to be obtained for smaller groups than are available from the individual quarterly databases. Labour Market Division has already made use of the advice in this article to provide more data for publication in the LFS Help-Line, *Social Trends and Regional Trends.* The advice given in this article should enable a wider range of users to obtain useful data from the LFS to meet their requirements.

In addition, the increased annual samples have improved the reliability of LFS data for small areas. A wider range of data are now available free on the National Statistics and Nomis® websites and also from the Sub National Data Service. The timeliness of annual data has also been improved.

ANNEX A: DERIVATION AND EXAMPLES OF STANDARD ERRORS ON THE LFS

The sampling variance in the survey was estimated by treating each household as a primary sampling unit (PSU). The method used to estimate the standard error of a ratio R=Y/X (where X might, for example, be the number of men aged 16 and over who are economically active and Y might be the number of them who are unemployed) was that of successive differences. The formulae used were:

$$Variance(Y) = \sum_s \frac{N_s}{2(N_s - 1)} \sum_{p=2}^{N_s} (Y_{s,p} - Y_{s,p-1})^2$$

$$Variance(X) = \sum_s \frac{N_s}{2(N_s - 1)} \sum_{p=2}^{N_s} (X_{s,p} - X_{s,p-1})^2$$

$$Covariance(X,Y) = \sum_s \frac{N_s}{2(N_s - 1)} \sum_{p=2}^{N_s} (X_{s,p} - X_{s,p-1})(Y_{s,p} - Y_{s,p-1})$$

$$StandardError(R) = \frac{1}{X} \sqrt{(VAR(Y) - 2RxCOV(X,Y) + R^2 xVAR(X))}$$

where s represents the stratum, p the PSU, i the individual case, $X_{s,p}$ and $Y_{s,p}$ the PSU totals, and N_s the number of PSUs in stratum s.

LABOUR FORCE SURVEY STANDARD ERRORS: March - May 1999, United Kingdom

Economic activity status

Characteristic	Percent	Standard Error	Design Factor	Base
All people 16+ by Economic Activity:				
Economically active	62	0.17	1.17	111,437
All in employment	58	0.17	1.17	111,437
Employees	51	0.17	1.16	111,437
Self-employed	7	0.08	1.07	111,437
ILO unemployment	4	0.06	1.04	111,437
Economically inactive	38	0.17	1.17	111,437
Men 16+ by Economic Activity :				
Economically active	70	0.20	1.02	52,817
All in employment	65	0.21	1.02	52,817
Employees	54	0.22	1.02	52,817
Self-employed	10	0.13	1.02	52,817
ILO employment	5	0.09	1.01	52,817
Economically inactive	30	0.20	1.02	52,817
Women 16+ by Economic Activity:				
Economically active	55	0.21	1.02	58,620
All in employment	52	0.21	1.02	58,620
Employees	48	0.21	1.02	58,620
Self-employed	3	0.08	1.00	58,620
ILO unemployment	3	0.07	1.02	58,620
Economically inactive	45	0.21	1.02	58,620

People aged 16+ in employment by type of employment

Characteristic	Percent	Standard Error	Design Factor	Base
All people 16+ in employment by type of employment				
Full-time employees	64	0.19	1.02	64,909
Part-time employees	23	0.16	0.96	64,909
Full-time self-employed	9	0.12	1.03	64,909
Part-time self-employed	3	0.06	1.02	64,909
Government employment and training programmes	1	0.03	1.00	64,909
Unpaid family worker	0	0.03	1.04	64,909
Men 16+ in employment by type of employment				
Full-time employees	76	0.23	1.02	34,471
Part-time employees	7	0.14	1.01	34,471
Full-time self-employed	14	0.19	1.01	34,471
Part-time self-employed	2	0.07	1.01	34,471
Government employment and training programmes	1	0.05	1.00	34,471
Unpaid family worker	0	0.03	1.00	34,471
Women 16+ in employment by type of employment				
Full-time employees	51	0.29	1.01	30,438
Part-time employees	41	0.29	1.01	30,438
Full-time self-employed	3	0.10	1.01	30,438
Part-time self-employed	4	0.11	1.00	30,438
Government employment and training programmes	0	0.04	1.00	30,438
Unpaid family worker	1	0.04	1.01	30,438

People aged 16+ in employment by industry sectors

Characteristic	Percent	Standard Error	Design Factor	Base
All aged 16+ in employment by:				
Agriculture & Fisheries	2	0.06	1.16	64,909
Energy & Water	1	0.04	1.02	64,909
Manufacturing	18	0.16	1.05	64,909
Construction	7	0.11	1.03	64,909
Distribution	20	0.17	1.06	64,909
Transport and Communication	6	0.10	1.04	64,909
Banking, Finance, Insurance	15	0.15	1.06	64,909
Public Admin, Health, Education	26	0.18	1.04	64,909
Other Services	5	0.09	1.05	64,909
Men aged 16+ in employment by:				
Agriculture & Fisheries	2	0.09	1.05	34,471
Energy & Water	2	0.07	1.00	34,471
Manufacturing	24	0.23	1.01	34,471
Construction	12	0.18	1.02	34,471
Distribution	17	0.21	1.02	34,471
Transport & Communication	9	0.16	1.01	34,471
Banking, Finance, Insurance	15	0.19	1.01	34,471
Public Admin, Health, Education	14	0.19	1.00	34,471
Other Services	4	0.11	1.01	34,471
Women aged 16+ in employment by:				
Agriculture	1	0.05	1.01	30,438
Energy & Water	1	0.04	0.99	30,438
Manufacturing	11	0.18	1.01	30,438
Construction	1	0.07	1.00	30,438
Distribution	23	0.24	1.01	30,438
Transport & Communication	4	0.11	1.01	30,438
Banking, Finance, Insurance	15	0.20	1.01	30,438
Public Admin, Health, Education	38	0.28	1.01	30,438
Other Services	6	0.14	1.01	30,438

Economically active people aged 16+ by sex who were ILO unemployed

Characteristic	Percent	Standard Error	Design Factor	Base
Economically active people 16+ who were ILO unemployed:				
All persons 16+	6	0.09	1.04	68,960
Men 16+	7	0.13	1.01	36,923
Women 16+	5	0.12	1.02	32,037

Economically active people aged 16-17 who were ILO unemployed

Characteristic	Percent	Standard Error	Design Factor	Base
Economically active people 16-17 who were ILO unemployed:				
All persons 16-17	18	0.87	1.02	1,993
Men 16-17	21	1.28	1.02	1,058
Women 16-17	14	1.14	1.00	935

Economically active people aged 18-24 who were ILO unemployed

Characteristic	Percent	Standard Error	Design Factor	Base
Economically active people 18-24 who were ILO unemployed:				
All persons 18-24	16	0.57	1.03	4,338
Men 18-24	18	0.83	1.03	2,288
Women 18-24	13	0.75	1.02	2,050

Economically active people aged 25-34 who were ILO unemployed

Characteristic	Percent	Standard Error	Design Factor	Base
Economically active people 25-34 who were ILO unemployed:				
All persons 25-34	6	0.18	1.02	17,079
Men 25-34	6	0.25	1.00	8,961
Women 25-34	6	0.25	1.00	8,118

Economically active people aged 35-49 who were ILO unemployed

Characteristic	Percent	Standard Error	Design Factor	Base
Economically active people 35-49 who were ILO unemployed:				
All persons 35-49	4	0.13	1.01	25,437
Men 35-49	5	0.19	1.00	13,381
Women 35-49	4	0.17	1.00	12,056

Economically active people aged 50-59/64 who were ILO unemployed

Characteristic	Percent	Standard Error	Design Factor	Base
Economically active people 50-59/64 who were ILO unemployed:				
All persons 50-59/64	5	0.18	1.02	14,504
Men 50-64	6	0.25	1.00	8,655
Women 50-59	3	0.23	1.00	5,849

Economically active people aged 60+/65+ who were ILO unemployed

Characteristic	Percent	Standard Error	Design Factor	Base
Economically active people 60+/65+ who were ILO unemployed:				
All persons 60+/65+	2	0.31	1.00	2,118
Men 65+	3	0.57	1.00	768
Women 60+	2	0.37	1.00	1,350

People aged 16+ in employment by region of usual residence (GOR)

Characteristic	Percent	Standard Error	Design Factor	Base
All persons 16+ in employment				
North East	51	0.85	1.20	5,033
North West	58	0.58	1.19	10,170
Merseyside	47	1.21	1.22	2,521
Yorkshire & Humberside	57	0.59	1.17	9,672
East Midlands	60	0.66	1.19	7,876
West Midlands	58	0.58	1.19	10,408
Eastern	61	0.57	1.17	10,113
London	59	0.53	1.16	11,646
South East	64	0.46	1.16	14,691
South West	60	0.59	1.17	9,299
Wales	53	0.81	1.18	5,279
Scotland	56	0.55	1.15	10,598
Northern Ireland	54	0.87	1.13	4,131

People aged 16+ who were ILO unemployed by region of usual residence (GOR)

Characteristic	Percent	Standard Error	Design Factor	Base
All persons 16+ ILO unemployed				
North East	6	0.34	1.04	5,033
North West	3	0.18	1.02	10,170
Merseyside	5	0.43	0.99	2,521
Yorkshire & Humberside	4	0.21	1.05	9,672
East Midlands	3	0.20	1.00	7,876
West Midlands	4	0.20	1.05	10,408
Eastern	3	0.17	1.04	10,113
London	5	0.20	1.04	11,646
South East	2	0.13	1.06	14,691
South West	3	0.18	1.04	9,299
Wales	4	0.27	1.02	5,279
Scotland	4	0.20	1.03	10,598
Northern Ireland	4	0.34	1.09	4,131

People aged 16+ who were economically inactive by region of usual residence (GOR)

	Percent	Standard Error	Design Factor	Base
All persons 16+ economically inactive				
North East	43	0.83	1.19	5,033
North West	39	0.58	1.20	10,170
Merseyside	48	1.21	1.22	2,521
Yorkshire & Humberside	39	0.58	1.17	9,672
East Midlands	37	0.65	1.20	7,876
West Midlands	38	0.56	1.18	10,408
Eastern	36	0.56	1.18	10,113
London	36	0.52	1.16	11,646
South East	34	0.46	1.17	14,691
South West	37	0.59	1.17	9,299
Wales	44	0.80	1.17	5,279
Scotland	40	0.55	1.15	10,598
Northern Ireland	42	0.86	1.12	4,131

People aged 16+ by ethnic origin: White

Characteristic	Percent	Standard Error	Design Factor	Base
People aged 16+ by ethnic origin:				
All	95	0.09	1.38	111,437
Men	95	0.11	1.13	52,817
Women	95	0.10	1.10	58,620

People aged 16+ by ethnic origin: Black - Caribbean

Characteristic	Percent	Standard Error	Design Factor	Base
People aged 16+ by ethnic origin:				
All	1	0.03	1.25	111,437
Men	1	0.04	1.09	52,817
Women	1	0.04	1.09	58,620

People aged 16+ by sex and ethnic origin: Indian

Characteristic	Percent	Standard Error	Design Factor	Base
People aged 16+ by sex and ethnic origin:				
All	1	0.05	1.53	111,437
Men	1	0.06	1.15	52,817
Women	1	0.06	1.20	58,620

People aged 16+ by sex and ethnic origin: Pakistani

Characteristic	Percent	Standard Error	Design Factor	Base
People aged 16+ by sex and ethnic origin:				
All	1	0.05	1.65	111,437
Men	1	0.05	1.27	52,817
Women	1	0.05	1.25	58,620