

Appendix

Technical report on sampling fieldwork and weighting

This section provides a more detailed account of the design of the study, and sets out all the technical information necessary to show how the findings were reached. The questionnaire and opt-out letters are not included in this report but are available on the DWP website at www.dwp.gov.uk/asd

Sample design

A requirement of the research design was to ensure robust samples of respondents in each of the three groups were obtained:

- those currently in employment;
- those in receipt of income replacement benefits (Income Support and Incapacity Benefit); and
- those who consider themselves to be retired.

The sample frame for the study was the Family Resources Survey (April 2001 to March 2002). A certain level of information was already known about respondents from their answers to the FRS and by selecting those agreeing to be re-contacted who were expected to be in the eligible age range in the fieldwork period, this meant there was no need for a sift of the general population.

However, relying solely on the FRS as a sample frame would have resulted in a relatively small number of respondents reliant on income replacement benefits as they would be collected in their natural incidence in this population. In order to satisfy the second research objective identified above, for this reason it was also necessary to boost the sample frame using benefit records supplied by the Department for Work and Pensions.

The FRS sample

Co-operating respondents to the FRS 2001/2 (issued between April and December 2001) were eligible for the this study if they were aged between 49 and 69 (inclusive) and had agreed to re-contact in the FRS interview.

The age range was selected to ensure that all respondents that were aged between 50 and 69 (inclusive) at the time of the fieldwork for this study were included. As there was a potential gap of up to a year between the FRS interview and being selected for this study, it was necessary to include respondents that were 49 at the time of the FRS. Any FRS respondents that were found to be 49 when contacted for this study were excluded.

The number of FRS respondents that satisfied the above criteria was 3,219. Of those, 117 were excluded because their full name was either not given or not recorded correctly at the time of the interview. This, therefore, left a sample of 3,102 FRS respondents, which comprised the FRS sample for this study.

The Benefit Records sample

The Benefit Records sample consisted of 1,000 cases selected from the database of all recipients of Income Support and Incapacity Benefit. In order to improve the efficiency of the fieldwork, the Benefit Records sample was selected from the same postcode sectors as the FRS sample.

The 652 postcode sectors within which the FRS sample was selected (for fieldwork between April and December 2001 inclusive) were identified and people claiming either Income Support (IS) or Incapacity Benefit (IB) in those postcode sectors extracted. From this list, three separate sampling frames were generated – people claiming IS only, IB only and **both** IS and IB.

Each sampling frame was ordered by sex, age-group⁵⁹ and postcode sectors, and simple random samples were selected from each sampling frame (see Table 2.1). The sampling was designed so that the proportions in each type (IS only, IB only and both IS and IB) were approximately the same as the proportions in the sampling frame.

Top-up samples

Because the number of cases that could be issued from the FRS 2001/2 was smaller than had been anticipated, an additional top-up sample of about 450 respondents to the FRS between January and March 2002 (inclusive) were issued. The criteria for inclusion in the sample was the same as those listed above (*The FRS sample* section), with the exception of the change in the FRS time reference period.

In order to increase the efficiency of the fieldwork for this top-up sample, postcode sectors within which there were three or fewer eligible FRS respondents were excluded from the sample. This removed 53 postcode sectors from the 196 within which there were FRS respondents eligible for this study.

The average number of eligible FRS respondents in the remaining 143 postcode sectors was 7.182. Therefore to obtain approximately 450 FRS cases, 63 postcode sectors were sampled ($63 \times 7.182 = 452$). Before selecting the 63 postcode sectors, they were ordered by the grouped number of eligible cases⁶⁰ and postcode.

The 63 selected postcode sectors actually contained 454 FRS respondents. Of those, 22 were excluded because their full name was either not given or not recorded correctly at the time of the interview. This therefore left 432 FRS respondents, which comprised the top-up FRS sample.

⁵⁹ Grouped as 50 to 54, 55 to 59, 60 to 64, 65 to 69.

⁶⁰ Grouped into 4 to 6, 7 to 9, 10 to 12 and 13 to 16.

As the rate of opt-out had been higher than expected, it was also decided to top-up the Benefit Records sample. Therefore an additional sample of 114 was selected from the same 652 postcode sectors as the initial sample (*The Benefit Records sample* section). Again the IS only, IB only and IS and IB samples were selected to be in the same proportions as the original sampling frames (see Table 1).

Table 1 Benefit records sample

	Sampling frame	Initial sample	Top-up sample
IS Only	52,861(32.4%)	320(32.0%)	36(31.6%)
IB Only	73,658(45.1%)	460(46.0%)	53(46.5%)
IS and IB	36,602(22.4%)	220(22.0%)	25(21.9%)
Total	163,121(100%)	1000(100%)	114(100%)

The issued sample

The total size of the sample issued for the study was 4,648 and consisted of:

- 3,102 cases selected from the FRS from April to December 2001;
- 432 cases selected from the FRS from January to March 2002; and
- 1,114 cases selected from Benefit Records.

Development work and piloting

The questionnaire was developed by NatCen in conjunction with DWP. The questionnaire development process was able to draw on a number of other survey questionnaires including the following:

- The Family Resources Survey (FRS).
- The 1994 retirement survey.
- The English Longitudinal Survey of Ageing (ELSA).
- National Adult Learning Survey (NALS) 2001.
- New Deal for Lone Parents Evaluation.
- Workplace Employee Relations Survey (WERS).
- Disabled for Life?: attitudes towards, and experiences of, disability in Britain.

Some questions were taken directly from these surveys and others developed from questions used on these surveys. In addition, questions were developed specifically for use as part of this study.

Pilot study

As part of the development work, the questionnaire was tested in a small scale pilot study. The whole questionnaire was tested, however, specific areas of interest were:

- questionnaire length;
- contact procedures;
- testing the relevance of definitions (fully-retired /semi-retired);
- understanding the relevance of the retirement questions;
- whether questions were understood.

Additionally, the quality of the sample from benefit records was also tested.

In total, 32 interviews were conducted including both FRS respondents and those selected from benefit records. Six trained interviewers were used to conduct pilot interviews and were briefed and de-briefed in person by the research team.

For the main stage of fieldwork, the following changes were made as a result of the feedback from the pilot:

- some cuts to the questionnaire were made due to the length of the pilot interview;
- a number of changes to individual questions were identified and amendments were made for the final questionnaire;
- interviewers were provided with a document explaining the different types of benefits respondents may be claiming to assist them during fieldwork.

Fieldwork procedures

Those selected from benefit records were sent a letter informing them of the study and giving them the opportunity to opt out. Additionally, all potential respondents were sent an advance letter a few days before the interviewer first attempted to contact them.

The interviews were administered using Computer Assisted Personal Interviewing (CAPI). A program was written in Blaise, a computer assisted interviewing software, in order to do this. The fieldwork period was May to August 2002. Details of the questionnaire structure are included in Section 1.3.2.

Response

In total, a sample of 4,647 individuals was issued. All these individuals were expected to be aged 50 to 69 during the fieldwork period. However, as FRS only collects age and not date of birth, it was possible that some respondents may not have reached the age of 50 or may have turned 70 when the interviewer contacted them. Additionally, a small number of the benefit sample also turned out to be out of the age range. These cases were classified as ineligible for the survey.

In total, 2,808 interviews were conducted. As Table A.2 shows, as a proportion of those in scope for the study, this was equivalent to a response rate of 62 per cent. As a proportion of those in scope for the fieldwork, it was equivalent to a response rate of 71 per cent.

Table 2 Main fieldwork – response summary

		<i>Column percentages</i>	
	Total	Population in scope of study	Population in scope of fieldwork
Total sample	4647		
Ineligible:			
Out of age range	122		
In scope for study	4525	100	
Not issued to interviewer			
Refusal to opt-out/advance letter	356		
Invalid cases:			
Duplicate	3		
Address does not exist	2		
Address demolished/derelict	4		
Address vacant/empty	29		
Address occupied – no resident	1		
Communal establishment	5		
Other invalid	4		
Respondent not known at address	2		
Respondent died	28		
Respondent moved, untraceable	111		
Respondent moved out of area	8		
Inaccessible	1		
	198		
In scope for fieldwork	3971	88	100
Refusals			
Refusal before interview	589		
Refusal before interview – proxy	147		
Refusal during interview	8		
Broken appointment – no re-contact	83		
	827	69	79
Unknown eligibility			
Unknown if residential, no contact	1		
Unknown if occupied – refused	3		
Unknown if occupied – no contact	15		
	19	69	79
No contact			
No contact at address	98		
No contact with named respondent	66		
Away/in hospital all survey period	52		
	216	64	73
Other reasons for no interview			
Ill at home during survey period	38		
Physically/mentally unable	32		
Language difficulties	21		
Other unproductive	10		
	101	62	71
Productive interviews			
Full productive interview	2800		
Partial productive interview	8		
Response rate	2808	62	71

Weighting

Because of the complex design of the sample for this study, the weighting strategy was comprised of several stages. First, the weighting strategy was designed to correct for refusal to be re-contacted for the FRS sample and non-response to this study for the full sample. The weighted sample was then adjusted to match estimates of whether claiming IS and/or IB within sex and age group (from the FRS 2001/02) and population estimates within sex and age groups (from the 2001 Census).

Weighting the issued FRS sample

The FRS sample is obtained from an equal probability sample of addresses from the postcode address file (PAF). Therefore, selection weights for the sample of households selected are only required to (a) correct for the number of dwellings at an address where this is different to what is listed on the PAF⁶¹ and (b) correct for the number of households selected at a dwelling. As the impact on the estimates from these weights would be negligible, it was decided not to include them in the weights.

Because the FRS sample is selected from the PAF, there is minimal information available to generate non-response weights from modelling non-response to the FRS. Therefore, the DWP has previously used post-stratification weights to attempt to correct any bias from differential non-response (this also reduces the impact of sampling variance). However, at the time of generating the weights, the post-stratification weights for the 2001/2 FRS had not been generated.

Therefore, neither the selection weights nor post-stratification weights for the FRS were used to calculate the weights for this study. However, not including the FRS selection weight had minimal impact on the estimates, and not including the FRS post-stratification weights was compensated for by generating post-stratification weights for this study at the final stage of the weighting process (see Section 6.5).

Refusal to be re-contacted after the FRS interview

Only people that had agreed to be re-contacted during the FRS interview could be included in the sample (see *Development work and piloting* section). This meant that our issued sample would be biased towards people that had agreed to be re-contacted. Weights were, therefore, generated to try to correct this potential bias.

This weighting was made more complicated because the full FRS dataset was only available for the FRS samples issued between April and November 2001. For the December 2001 to March 2002 sample, only a subset of the full dataset was available. Therefore, separate weighting procedures were carried out for the respondents to this study that had been issued for the FRS between April and November 2001 and those that had been issued between December 2001 and March 2002. The procedure for the sample from the April – November 2001 FRS took advantage of the wider range of variables that were available.

For both procedures, logistic regression models were first fitted for men and women separately on all eligible cases in the FRS, with whether they agreed to be re-contacted or not as the outcome variables. The resultant models were used to estimate the propensity of each FRS respondent to agree to be re-contacted. The weights (w_i) were then calculated as the inverse of these estimates. The only difference between the two approaches was that more variables were available as covariates in the model for the April-November 2001 FRS sample (see *Logistic regression and weights tables* section for the models fitted).

⁶¹ For example, where an address contains more than one dwelling or where the MOI in Scotland is not correct.

Non-response by the FRS sample

The second stage of the non-response process was to generate non-response weights for the sample selected from the FRS. (Note that no selection weights were required for the selection of the issued sample from those eligible in the FRS. This is because we selected all eligible cases from the FRS that consented to be re-contacted.)

A logistic regression was fitted to the issued sample from the FRS with whether people responded or not as the outcome variable and geographic measures (e.g. urbanity) and variables from the FRS as covariates (*Logistic regression and weights tables* section). The resultant model was used to estimate the propensities for respondents to participate in this survey, and the weight for non-response (w_{2a}) calculated as the reciprocal of these propensities.

Non-response by the Benefit Record sample

The parallel second stage of the non-response process was to generate non-response weights for the sample selected from Benefit Records. (Note that no selection weights were required for the selection from the Benefit Records because the sample was selected within strata defined by whether on IS, IB or IS and IB to be in the same proportions as in the population.)

A logistic regression was fitted to the issued sample from the Benefit Records with whether a person responded or not as the outcome variable and geographic measures (e.g. urbanity) and variables from the Benefit Records dataset as covariates (see *Logistic regression and weights tables* section). The resultant model was used to estimate the propensities for respondents to participate in this study, and the weight for non-response (w_{2b}) calculated as the reciprocal of these propensities.

Post-stratification

The sets of weights from each source (FRS sample and Benefit Records) were re-scaled to have mean one, and then merged to form a single set of weights.

Because the achieved sample contained a Benefit Sample boost, it was likely that the proportion in the sample claiming either IB or IS was likely to be greater than in the population. Therefore, post-stratification weights were produced that aimed to adjust the proportion in the following three groups: 'claiming IS or IB: working or not working', 'claiming neither IS nor IB: working' and 'claiming neither IS nor IB: not working'.

Estimates of the proportions in each group within sex and age group⁶² were obtained from the FRS 2001/2. A temporary weight was calculated as $w_t = w_1 \times (w_{2a} \text{ or } w_{2b})$ and the equivalent (weighted) proportions in the sample estimated. Post-stratification weights (w_3) were generated to adjust the (weighted) profile of respondents to this survey so that it matched the profile in the FRS 2001/2. These weights are shown in section.

The final stage of the weighting process adjusted the sample so that the (weighted) age and sex profile matched 2001 Census population estimates⁶³. A temporary weight was calculated as $w_t = w_1 \times (w_{2a} \text{ or } w_{2b}) \times w_3$ and the weighted proportions in this study within age group and sex estimated. Post-stratification weights (w_4) were then calculated that adjusted these weighted estimates to match the 2001 Census estimates (see *Logistic regression and weights tables* section).

⁶² 50 to 54, 55 to 59, 60 to 64, 65 to 69.

⁶³ <http://www.statistics.gov.uk/census2001/pyramids/pages/727.asp> for England and Wales, and <http://www.statistics.gov.uk/census2001/pyramids/pages/179.asp> for Scotland.

The composite weight

The composite weight was calculated as:

$$weight = w_1 \times (w_{2a} \text{ or } w_{2b}) \times w_3 \times w_4$$

This weight was re-scaled so that the weighted total was the same as the sample size (i.e. the mean weight was re-scaled to be equal to 1).

Logistic regression and weights tables

Table 3 Logistic regression to model refusal to re-contact: men (FRS April-November 2001)

	b	s.e.	Wald	d.f.	p	exp(b)
Currently working	-0.566	0.099	33.1	1	0.000	0.568
Married and living with spouse	-0.731	0.124	34.5	1	0.000	0.481
Any educational qualifications	0.242	0.093	6.8	1	0.009	1.273
Any long-standing illness	0.312	0.098	10.2	1	0.001	1.366
Any other investments	0.304	0.099	9.4	1	0.002	1.356
Hold current account	0.501	0.158	10.0	1	0.002	1.650
Currently cohabiting	-0.635	0.230	7.6	1	0.006	0.530
Constant	0.523	0.178	8.6	1	0.003	1.686

Table 4 Logistic regression to model refusal to re-contact: women (FRS April-November 2001)

	b	s.e.	Wald	d.f.	p	exp(b)
Married and living with spouse	-0.616	0.103	35.5	1	0.000	0.540
Any long-standing illness	0.279	0.090	9.6	1	0.002	1.322
Any professional qualifications	0.263	0.087	9.0	1	0.003	1.300
Non-white	-0.434	0.159	7.4	1	0.006	0.648
Currently cohabiting	-0.823	0.251	10.7	1	0.001	0.439
Constant	0.836	0.108	59.9	1	0.000	2.308

Table 5 Logistic regression to model refusal to re-contact: men (FRS December 2001-March 2002)

	b	s.e.	Wald	d.f.	p	exp(b)
Currently working	-0.570	0.088	41.5	1	0.000	0.566
Married and living with spouse	-0.492	0.108	20.9	1	0.000	0.611
Constant	1.090	0.105	108.0	1	0.000	2.973

Table 6 Logistic regression to model refusal to re-contact: women (FRS December 2001-March 2002)

	b	s.e.	Wald	d.f.	p	exp(b)
Married and living with spouse	-0.525	0.096	29.8	1	0.000	0.592
Constant	0.926	0.082	126.5	1	0.000	2.524

Table 7 Non-response model for FRS sample

	b	s.e.	Wald	d.f.	p	exp(b)
Currently married	-0.207	0.082	6.3	1	0.012	0.813
Government Office Region:			18.1	11	0.079	
North East	0.000	(baseline)				
North West	-0.087	0.206	0.2	1	0.675	0.917
Merseyside	-0.087	0.279	0.1	1	0.755	0.917
Yorkshire & Humber	0.093	0.216	0.2	1	0.666	1.098
East Midlands	0.108	0.226	0.2	1	0.633	1.114
West Midlands	-0.088	0.214	0.2	1	0.682	0.916
South West	-0.178	0.230	0.6	1	0.439	0.837
Eastern	-0.169	0.219	0.6	1	0.441	0.845
London	-0.463	0.218	4.5	1	0.033	0.629
South East	-0.302	0.203	2.2	1	0.136	0.739
Wales	-0.431	0.234	3.4	1	0.065	0.650
Scotland	-0.176	0.213	0.7	1	0.409	0.838
Urbanity:			4.2	2	0.124	
Not known	0.000	(baseline)				
Rural	0.084	0.188	0.2	1	0.657	1.087
Urban	-0.099	0.189	0.3	1	0.602	0.906
Constant	1.084	0.257	17.8	1	0.000	2.956

Table 8 Non-response model for Benefit Records sample

	b	s.e.	Wald	d.f.	p	exp(b)
Age group:			8.9	3	0.030	
50 to 54	0.000	(baseline)				
55 to 59	0.041	0.180	0.1	1	0.818	1.042
60 to 64	0.498	0.188	7.0	1	0.008	1.646
65 to 69	0.424	0.270	2.5	1	0.116	1.529
Benefit claimed:			9.6	2	0.008	
IB	0.000	(baseline)				
IS	-0.587	0.191	9.5	1	0.002	0.556
IB and IS	-0.230	0.174	1.8	1	0.185	0.794
Constant	0.410	0.153	7.2	1	0.007	1.507

Table 9 Post-stratification weights for whether claiming IS or IB and working status – men

Age group	IS/IB/Working	Older workers		Weight
		survey ¹	FRS	
50 to 54	Claiming IS or IB: working or not working	23.5%	11.8%	0.503
	Claiming neither IS nor IB: working	68.3%	77.3%	1.131
	Claiming neither IS nor IB: not working	8.2%	10.9%	1.330
55 to 59	Claiming IS or IB: working or not working	32.1%	16.2%	0.504
	Claiming neither IS nor IB: working	53.3%	65.3%	1.225
	Claiming neither IS nor IB: not working	14.6%	18.5%	1.270
60 to 64	Claiming IS or IB: working or not working	53.0%	28.4%	0.535
	Claiming neither IS nor IB: working	24.2%	43.1%	1.781
	Claiming neither IS nor IB: not working	22.7%	28.5%	1.253
65 to 69	Claiming IS or IB: working or not working	15.8%	6.3%	0.401
	Claiming neither IS nor IB: working	11.5%	12.7%	1.102
	Claiming neither IS nor IB: not working	72.6%	80.9%	1.114

¹ Weighted for refusal to re-contact (for FRS sample) and non-response.

Table 10 Post-stratification weights for whether claiming IS or IB and working status – women

Age group	IS/IB/Working	Older workers		Weight
		survey ¹	FRS	
50 to 54	Claiming IS or IB: working or not working	22.1%	8.9%	0.403
	Claiming neither IS nor IB: working	61.3%	67.2%	1.095
	Claiming neither IS nor IB: not working	16.6%	23.9%	1.444
55 to 59	Claiming IS or IB: working or not working	25.6%	11.4%	0.446
	Claiming neither IS nor IB: working	52.3%	50.4%	0.964
	Claiming neither IS nor IB: not working	22.1%	38.2%	1.727
60 to 64	Claiming IS or IB: working or not working	17.9%	7.3%	0.406
	Claiming neither IS nor IB: working	20.7%	20.8%	1.001
	Claiming neither IS nor IB: not working	61.3%	72.0%	1.173
65 to 69	Claiming IS or IB: working or not working	19.2%	8.2%	0.428
	Claiming neither IS nor IB: working	5.5%	7.4%	1.363
	Claiming neither IS nor IB: not working	75.4%	84.4%	1.119

¹ Weighted for refusal to re-contact (for FRS sample) and non-response.

Table 11 Post-stratification weights for age group and sex

Age group	Older workers		Weight
	survey ¹	FRS	
Men:			
50 to 54	13.8%	15.6%	1.134
55 to 59	12.9%	12.8%	0.992
60 to 64	13.7%	11.0%	0.803
65 to 69	9.6%	9.7%	1.007
Women:			
50 to 54	15.1%	15.9%	1.054
55 to 59	15.1%	13.1%	0.870
60 to 64	10.7%	11.4%	1.065
65 to 69	9.2%	10.5%	1.146

¹ Weighted for refusal to re-contact (for FRS sample), non-response and by post-stratification weights (claiming benefit/working status).

