THE IMPACT OF INTERVIEWING METHOD ON MEASUREMENT ERROR IN PANEL SURVEY MEASURES OF BENEFIT RECEIPT: EVIDENCE FROM A VALIDATION STUDY

Peter Lynn, Annette Jäckle, Stephen P. Jenkins and Emanuela Sala

ISER Working Papers
Number 2004-28

Institute for Social and Economic Research

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

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

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

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

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

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

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

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

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

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

Acknowledgement: This paper derives from the project, "Improving Survey Measurement of Income and Employment" (ISMIE), funded under the Economic and Social Research Council (ESRC) Research Methods Programme, grant number H333250031. We also benefit from the core funding of the UK Longitudinal Studies Centre (ULSC) at ISER, by the ESRC (award no. H562255004) and the University of Essex. We are grateful to our ISER colleagues for their assistance in producing the ISMIE data set, especially Nick Buck, Jon Burton, John Fildes, Heather Laurie, Mike Merrett and Fran Williams. NOP Research programmed the ISMIE CAPI script and carried out the field work.

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

> Lynn, Peter, Jäckle, Annette, Jenkins, Stephen P., and Sala, Emanuela (December 2004) 'The impact of interviewing method on measurement error in panel survey measures of benefit receipt: evidence from a validation study', *Working Papers of the Institute for Social and Economic Research*, paper 2004-28. Colchester: University of Essex.

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

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

ABSTRACT

This article is concerned with measurement error in survey reports of social security benefit receipt. Survey respondents may under-report benefit receipt or, less likely, over-report. Our aims are three-fold. First, we attempt to quantify the extent of measurement error. Second, we assess the extent to which this varies according to the questioning method used. Specifically, dependent interviewing has been proposed as a way to reduce under-reporting in some circumstances (Mathiowetz and McGonagle, 2000) and we compare two versions of dependent interviewing (DI) with traditional independent interviewing in an experimental design. Third, we seek to identify why measurement error arises and to identify new ways of reducing it. We use data from a large-scale UK household panel survey, though some of our findings are applicable also to cross-sectional surveys. To assess measurement error, a validation exercise was conducted, with administrative data on benefit receipt matched at the individual level to the survey micro data.

Key words: dependent interviewing, income, over-reporting, under-reporting

1. Introduction

This article is concerned with measurement error in survey reports of social security benefit receipt. Benefit receipt is an important component of income for many households in the UK and consequently survey measures of benefit receipt are important for studies of income, poverty and related issues. For example, in May 2004, 4.9 million adults of working age (14% of the working age population) and 10.6 million adults of retirement age (99.9% of the retirement age population) were claiming at least one key benefit, while 2.5 million children aged under 16 (22% of the population) were living in a household claiming a key benefit (Department for Work and Pensions; 2004a, 2004b, 2004c). Amongst the poorest fifth of households, ranked by net disposable household income, around 55% of all income is accounted for by benefits (Department for Work and Pensions; 2004d, Table 2.2).

Survey measures of benefit receipt are subject to measurement error (Bound *et al*., 2001, pp. 3770-3779). Some survey respondents may under-report benefit receipt. This could be due to simple forgetting (many households will receive income from several different benefits as well as other sources and it is not always straightforward to remember all sources in an interview situation), due to misplacement in time or misclassification, or due to conscious suppression (e.g. caused by social desirability or sensitivity effects). Over-reporting is also possible, perhaps due to misclassification or misplacement in time.

Our aims are three-fold. First, we attempt to quantify the extent of measurement error. Second, we assess the extent to which this varies according to the questioning method used. Specifically, dependent interviewing has been proposed as a way to

reduce under-reporting in some circumstances (Mathiowetz and McGonagle, 2000) and we compare two versions of dependent interviewing (DI) with traditional independent interviewing in an experimental design. Third, we seek to identify why measurement error arises and to identify new ways of reducing it. We use data from a large-scale UK household panel survey, though some of our findings are applicable also to cross-sectional surveys. To assess measurement error, a validation exercise was conducted, with administrative data on benefit receipt matched at the individual level to the survey micro data.

Earlier studies present evidence that reported levels of benefit receipt are greater with DI (Dibbs *et al*., 1995; Lynn *et al*., 2004). To interpret that as a reduction in measurement error requires an assumption that measurement error consists primarily of under-reporting. In this article, after describing our data (Section 2), we directly assess that assumption as well as assessing what proportion of the measurement error in prevalence levels is eliminated (Section 3). We discuss the possible explanations for the small amount of over-reporting we find (Section 4) and we explore the role of errors in recalled dates as a factor contributing to both over-reporting and under-reporting (Section 5). We then propose and investigate ways in which dependent interviewing for panel surveys, or filtered questioning for cross-sectional surveys, could be extended to further reduce under-reporting (Section 6). Section 7 summarises our findings and draws conclusions.

2. The Data

2.1 Survey data

Our data are from the 'Improving Survey Measurement of Income and Employment" (ISMIE) project, funded by the Research Methods Programme of the UK Economic and Social Research Council. A sample of respondents to an existing panel survey which had come to an end were interviewed one more time for purely methodological purposes. The sample was the GB "low income supplemental sample" of the European Community Household Panel Survey (ECHP). This sample was selected in 1997 from respondents to the 1994-96 UK ECHP who exhibited characteristics associated with an increased likelihood of low household income (e.g. elderly, single parents, in receipt of income support, etc.). A description of the sample design appears in Lynn (2003). Though the sample was not designed to represent the general population, it covers a broad range of characteristics and is in important respects not dissimilar to the total population (Jäckle *et al.*, 2004). For experimentation with questions about income sources, it is an advantage that benefit recipients are over-represented in the sample.

The ECHP involved interviewing all adult members of sample households 8 times at annual intervals, the last wave of interviewing having taken place between September 2001 and February 2002. The 1,163 sample members (in 700 households) who had provided full interviews at wave 8 (2001-02) of the ECHP were included in the ISMIE study. They were randomly assigned to one of three treatment groups, which we refer to as the "independent interviewing" (INDI), "reactive dependent interviewing" (RDI) and "proactive dependent interviewing" (PDI) groups.

Assignment to groups was random within strata defined by sex, age and whether or not income from employment was reported at wave 8. Consequently, sample members within the same household were not necessarily allocated to the same group.

In each household containing at least one sample member, a household interview was carried out (median interview length 5 minutes), plus an individual interview with each sample member (median interview length 24 minutes) using Computer-Assisted Personal Interviewing (CAPI). A total of 1033 interviews were achieved, representing a response rate of around 89%. We refer to these 1033 persons as the "ISMIE respondents". Field work was carried out between February and April 2003, constituting an interval of between 13 and 18 months since the previous interview. The two dependent interviewing versions of the instrument called upon data from the previous interview ("wave 8"). For further details of the ISMIE survey, see Jäckle *et al*. (2004).

The questions regarding benefit receipt were part of a module on non-employment income, which worked as follows. Respondents were asked to look in turn at four cards, each of which contained a list of possible sources of income. The first card listed 6 types of pension, the second listed 10 state benefits related to disability or injury, the third listed 9 other state benefits and the fourth listed 8 other miscellaneous income sources, plus a catch-all category, "any other regular payment." The respondent was asked whether he or she had received any of the types of income or payments shown since September 2001. The interviewer noted each source reported. Then, for each reported source, a series of questions asked in which months (since the previous interview) income was received from that source,

whether income was still being received currently, the amount of the most recent payment, the period covered by that payment, and whether the income was received solely or jointly. The questions are reproduced in the Appendix.

A question requesting consent to link DWP administrative data to the survey data was asked at the end of the ISMIE individual interview. If respondents answered that they didn't know whether to give consent, or queried why the information was required, the interviewer provided more information, and then repeated the consent question. Respondents who gave oral consent also signed a form confirming consent. Of the 1033 ISMIE respondents, 799 (77.3%) gave consent to the data linkage. There were some differences between subgroups in consent propensity; it had a U-shaped relationship with age and was lower amongst respondents who lived alone. For further details see Jenkins *et al.* (2004a).

2.2 Administrative data

The DWP data were linked to the ISMIE survey data using non-hierarchical pooled matching based on 5 criteria. The first match criterion was National Insurance Number (which ISMIE respondents were asked to supply immediately after the consent question). The other 4 criteria were combinations of sex with two or three out of date of birth, forename, family name, postcode, and first line of address. In cases where an ISMIE respondent was matched to more than one person in the DWP data, the modal match was accepted as the correct one, provided that the match involved at least three of the five criteria. All other cases where an ISMIE respondent was matched to more than one person in the DWP data, of which there were few, were inspected visually to determine which match appeared to be correct.

Amongst ISMIE respondents for whom no match was made, it is not possible to distinguish between those who were genuinely not represented in the DWP data (because they were not benefit recipients) and those for whom the matching variables were inaccurate, though it seems likely that the latter group is small. For further details of the matching process, see Jenkins *et al.* (2004b).

In this article we are concerned with survey measures of receipt of each of six benefits: state retirement pension, child benefit, income support, incapacity benefit, working family tax credit and housing benefit. State (contributory) retirement pension is paid to persons who have reached State pension age (presently 65 for men, 60 for women) and have also met the contribution conditions (specified levels of National Insurance contributions paid by either the claimant or their spouse). Child Benefit is a fixed-amount entitlement paid for children up to the age of 16 and those aged 17 or 18 in full-time non-advanced education at a recognised educational establishment. Income support (IS) is intended to help people on low incomes who do not have to be available for employment. The main types of people who receive it are pensioners, lone parents, the long and short-term sick, people with disabilities and other special groups. Incapacity Benefit is paid to people who are assessed as being incapable of work and who meet certain contribution conditions. Working Families Tax Credit (WFTC) was designed to supplement the income of low income families with at least one person undertaking at least 16 hours of paid employment per week, thereby increasing the incentive to accept low-paid jobs. (It was replaced in April 2003 – around the end of the ISMIE field work period - by Working Tax Credit.) Housing Benefit (HB) is designed to help people on low-income pay their rent. Three of these six benefits (IS, WFTC, HB) are means tested, based on income received by the family unit. Numbers of recipients in the UK population ranged from about 1.5

million for incapacity benefit to 11.1 million for retirement pension, as of February 2003 (Department for Work and Pensions, 2004e, table C1).

3. The Effect of Interviewing Method on Measurement Error

3.1 Estimation of measurement error

Of the social security benefits represented in both data sources, we restrict our analysis to the six described in the previous section as these were the most prevalent amongst the ISMIE sample. For these six income sources between 61 respondents (incapacity benefit) and 256 (retirement pension) reported receipt in the survey interview and between 78 (family credit) and 255 (retirement pension) were recipients according to the administrative data – though these were not necessarily the *same* respondents, as we shall see.

For each benefit, we constructed a dichotomous measure of whether or not the DWP data indicated receipt in at least one month during the survey reference period. The survey reference period is from September 2001 until the month of the ISMIE interview (inclusive) for the INDI and RDI groups (mean length 18 months), and from the wave 8 month of interview until the month of the ISMIE interview (inclusive) for the PDI group (mean length 17 months). This is the period about which ISMIE respondents were asked. An equivalent indicator was constructed based upon the survey reports. We are interested in the relationship between these two measures. Specifically, we want to assess whether under-reporting is reduced with either form of DI, and also whether over-reporting is affected. As indicators of under-reporting and over-reporting, we analyse “false negatives” (cases where receipt is indicated by

the DWP measure but not by the survey measure) and "false positives" (cases where receipt is indicated only by the survey measure) respectively. If the DWP measure is taken to be accurate, then false negatives can be interpreted as cases of survey under-reporting and false positives as cases of over-reporting. However, these interpretations should be made with caution, as there may be other explanations for false positives (see section 4 below).

We should also take into account that the survey questions ask about receipt "either yourself or jointly." (Note that 3 of the 6 benefits are means-tested at the level of the family unit – see section 2.) In order to minimise the risk of erroneously counting a case as a false positive, we have counted it as a "true positive" if the survey measure indicates receipt and the DWP measure indicates receipt for *any* household member (not necessarily the respondent). This does not completely eliminate the possibility of erroneous false positives, however, as there may still be other recipient household members who were not interviewed, did not give consent for the matching, or were not successfully matched. The definition of our derived variable indicating the match between the survey and DWP measures is summarised in Table 1.

Table 1: Definition of derived indicator of correspondence between survey and DWP data

Respondent is recipient according to DWP data	Other household member is recipient according to DWP data	Survey report of receipt ("either yourself or jointly")	Derived variable
No	Yes or No	No	True negative
Yes	Yes or No	Yes	True positive
No	Yes	Yes	True positive
Yes	Yes or No	No	False negative
No	No	Yes	False positive

We will denote the sample proportion in each category of our match indicator by p^c_{ab}, where *a* = 1 if receipt is indicated by the administrative data, 0 if not; *b* = 1 if receipt is indicated by the survey response, 0 if not; c indicates treatment group. Thus, for example, p^{PDI}_{10} indicates the proportion of the PDI group classified as false negatives. Additionally, we will indicate marginal totals of the 2 x 2 table for each treatment group (where the rows and columns are defined by *a* and *b*) as follows: $p^c_{a\bullet} = p^c_{a0} + p^c_{a1}$; and $p^c_{\bullet a} = p^c_{0a} + p^c_{1a}$. So, for example, $p^{PDI}_{1\bullet}$ indicates the proportion of the PDI group classified as recipients according to the administrative data (the sum of true positives and false negatives). Several of our hypotheses of interest concern not the total sample proportion in a particular cell of the table a x b, but rather a row or column proportion. Specifically, only respondents classified as recipients according to the administrative data are at risk of being false negatives, so we define the false negative *rate* for treatment group *c* as $p^c_{10} / p^c_{1\bullet}$. Similarly, we define the false positive rate as $p^c_{01} / p^c_{0\bullet}$.

Our hypotheses are as follows, where H_1 indicates the hypothesis in which we are interested; H_0 the corresponding null hypothesis:

(1) DI should reduce under-reporting. If true, we would expect to observe lower false negative rates with each of the DI treatments than with INDI.

$$H_0 : \left(p^c_{10} / p^c_{1\bullet}\right) = \left(p^{INDI}_{10} / p^{INDI}_{1\bullet}\right),\ c = PDI, RDI$$

$$H_1 : \left(p^c_{10} / p^c_{1\bullet}\right) < \left(p^{INDI}_{10} / p^{INDI}_{1\bullet}\right),\ c = PDI, RDI$$

(2) DI may increase over-reporting. If true, we would expect to observe higher false positive rates with each of the DI treatments than with INDI.

$$H_0 : \left(p_{01}^c / p_{0\bullet}^c\right) = \left(p_{01}^{INDI} / p_{0\bullet}^{INDI}\right),\ c = PDI, RDI$$

$$H_1 : \left(p_{01}^c / p_{0\bullet}^c\right) > \left(p_{01}^{INDI} / p_{0\bullet}^{INDI}\right),\ c = PDI, RDI$$

(3) Under-reporting is the dominant component of measurement error. If true, we would expect to observe a higher false negative rate than false positive rate with INDI.

$$H_0 : \left(p_{10}^{INDI} / p_{1\bullet}^{INDI}\right) = \left(p_{01}^{INDI} / p_{0\bullet}^{INDI}\right)$$

$$H_1 : \left(p_{10}^{INDI} / p_{1\bullet}^{INDI}\right) > \left(p_{01}^{INDI} / p_{0\bullet}^{INDI}\right)$$

(4) Overall measurement error for benefit receipt prevalence rates should be less with DI. If true, we would expect to observe a smaller magnitude of error with each of the DI treatments than with INDI.

$$H_0 : \left|p_{01}^c - p_{10}^c\right| = \left|p_{01}^{INDI} - p_{10}^{INDI}\right|,\ c = PDI, RDI$$

$$H_1 : \left|p_{01}^c - p_{10}^c\right| < \left|p_{01}^{INDI} - p_{10}^{INDI}\right|,\ c = PDI, RDI$$

(Note that the observed error on the prevalence rate, $p_{\bullet 1}^c - p_{1\bullet}^c$, can be rewritten as $p_{01}^c - p_{10}^c$.)

The distribution of our derived indicator, for each benefit and each treatment group, is presented in Table 2 (analysis is restricted to the 77% of ISMIE respondents who gave consent for the DWP match – see section 2). These observations will

subsequently be used to estimate the false positive rates, false negative rates and differences in observed error relevant to our hypotheses.

Table 2: Income receipt indicators from administrative and survey data (row proportions)

		True negative p_{00}	True positive p_{11}	False negative p_{10}	False positive p_{01}	Admin $p_{1\bullet}$	Survey $p_{\bullet 1}$	Diff $p_{01} - p_{10}$
Retirement Pension	INDI	.698	.298	-	.004	.298	.302	.004
	PDI	.663	.330	.004	.004	.333	.333	.000
	RDI	.664	.321	.011	.004	.332	.325	-.007
Incapacity Benefit	INDI	.882	.057	.057	.004	.115	.061	-.054
	PDI	.897	.058	.042	.004	.100	.062	-.038
	RDI	.869	.073	.040	.018	.113	.091	-.022
Income Support	INDI	.790	.179	.023	.008	.202	.187	-.015
	PDI	.785	.180	.035	-	.215	.180	-.034
	RDI	.818	.168	.015	-	.183	.168	-.015
Child Benefit	INDI	.767	.172	.050	.012	.221	.183	-.038
	PDI	.774	.192	.008	.027	.199	.218	.019
	RDI	.770	.208	.018	.004	.226	.212	-.014
Family Credit	INDI	.901	.057	.023	.019	.080	.076	-.004
	PDI	.877	.092	.004	.027	.096	.119	.023
	RDI	.894	.077	.026	.004	.102	.080	-.022
Housing Benefit	INDI	.767	.179	.038	.015	.218	.195	-.023
	PDI	.644	.264	.061	.031	.326	.295	-.031
	RDI	.668	.274	.029	.029	.303	.303	.000

Notes: For the definition of true negative, true positive, false negative and false positive, see Table 1. The columns headed "Admin" and "Survey" show prevalence rates for receipt estimated from the administrative and survey data respectively. Bases are 262 INDI cases, 261 PDI and 274 RDI, consisting of all ISMIE respondents who gave consent for DWP matching.

Table 3: Results of hypothesis tests

	H(1)		H(2)		H(3)	H(4)	
	PDI	RDI	PDI	RDI		PDI	RDI
Retirement Pension							
Incapacity Benefit					***		*
Income Support					***		
Child Benefit	**	*			***	*	*
Family Credit	*				***		
Housing Benefit					***		**

Each hypothesis is tested using a Pearson χ^2 test. The survey design is taken into account by specifying households as PSUs. *** P≤0.001; ** 0.001<P≤0.01; * 0.01<P≤0.05. Blank entries indicate no significant difference (P>0.05).

3.2 Under-reporting

Amongst the INDI group, false negatives depress the survey estimate of the proportion in receipt of the benefit by between zero (retirement pension) and six percentage points (child benefit) (Table 2). This translates to a false negative rate ($p_{10} / p_{1\bullet}$) of between 0% (retirement pension) and 50% (incapacity benefit) (Figure 1). Hypothesis 1 was tested by comparing the false negative rates for each form of DI with that for INDI, separately for each benefit (Table 3). Dependent interviewing significantly (P<0.05) reduces the prevalence of false negatives for child benefit (both RDI and PDI) and family credit (PDI only). In the case of child benefit, this represents a reduction in the false negative rate from 22% with INDI to 4% (PDI) or 8% (RDI). There is also a suggestion that the false negative rate is reduced for incapacity benefit, but these reductions do not reach statistical significance (incapacity benefit is the least prevalent of the six benefits included in our analysis and consequently the tests have least power). These findings provide some support for hypothesis 1.

3.3 Over-reporting

False positives appear to inflate the survey estimate of the proportion in receipt of benefit amongst the INDI group (Table 2) by between 0.4 percentage points (retirement pension and incapacity benefit) and 1.9 percentage points (family credit). This translates to a false positive rate ($p_{01} / p_{0\bullet}$) of between 0.4% (incapacity benefit) and 2.1% (family credit) (Figure 2). Neither method of dependent interviewing has a significant impact (at the 0.05 level) on the prevalence of false positives for any of the benefits (Table 3). We therefore find no evidence to support hypothesis 2.

3.4 Measurement error

Overall, false positive rates are much lower, for all three interviewing methods, than false negative rates (Figure 1 and Figure 2). With INDI, false positive rates are significantly lower for five out of the six benefits (Table 3). This supports hypothesis 3 and is consistent with the widely-held belief that, with respect to income data, under-reporting is the major form of measurement error with which researchers should be concerned.

Hypothesis 4 was tested by comparing the estimated measurement error associated with the prevalence estimate (final column of Table 2) between INDI and each form of DI, separately for each benefit. For child benefit, measurement error was significantly less ($P<0.05$) with both forms of DI; for both incapacity benefit and housing benefit measurement error was less with RDI than with INDI.

4. Why does Over-reporting Occur?

There are at least three possible explanations for false positives, other than actual over-reporting. In this section, we explore the likely extent of each, to understand better the extent to which observed false positives represent genuine over-reports by survey respondents.

4.1 Failure of the Matching Process

The DWP measure may be incorrect in some cases due to a failure in the matching process. This could cause a false positive if a correct record for a particular benefit was present in the DWP data (respondent is a recipient) but was not matched to the

Figure 1: False negative rates for six benefits and three interviewing methods

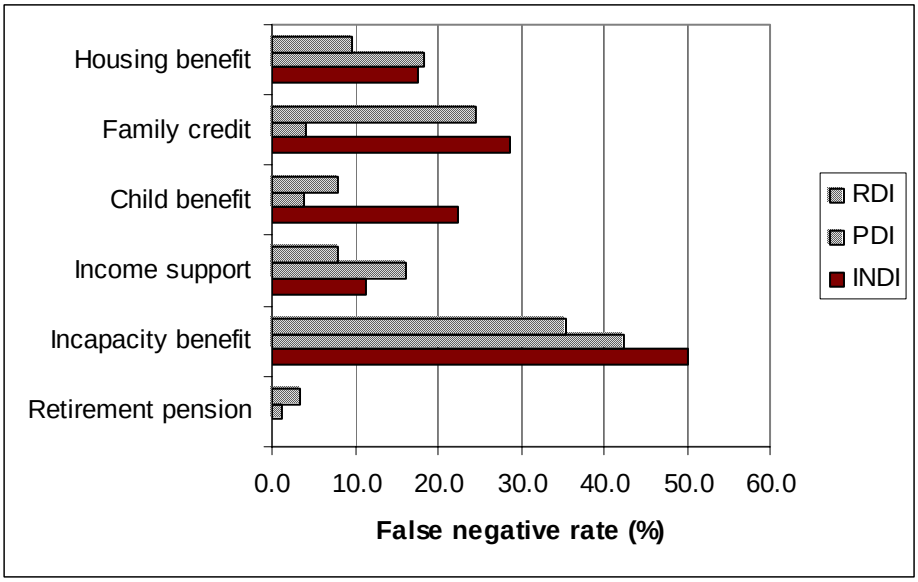
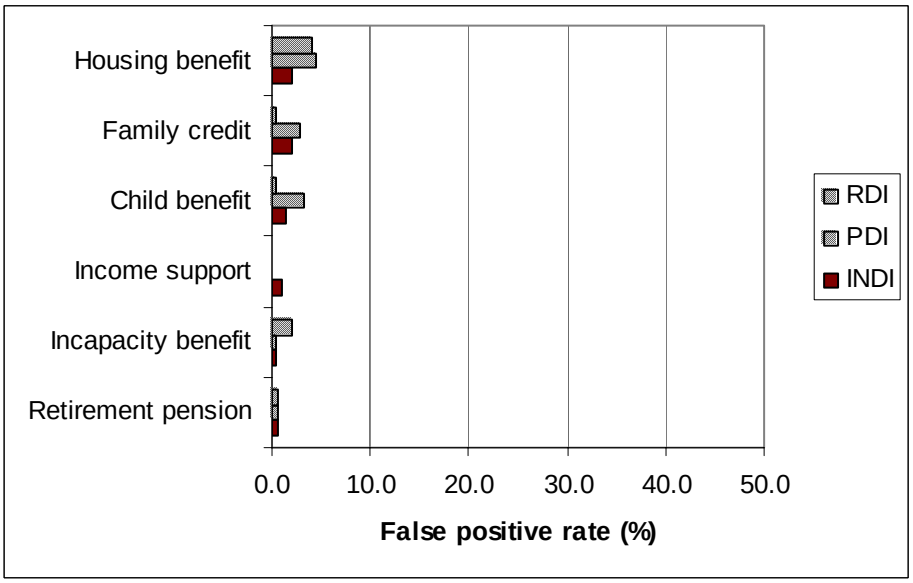


Figure 2: False positive rates for six benefits and three interviewing methods



survey data, either because no record was matched for the respondent or because an incorrect record (pertaining to a different person) was matched. However, we can rule out the possibility of match failures of the first sort (no match at all) for some respondents, where a match was successfully achieved to other DWP data. This is

because the matching process involved first matching to a unique personal identifier on the DWP data and using this identifier to obtain the records for each benefit.

We found that around two-thirds of the cases classified as a false positive on a particular benefit *had* been successfully matched to the DWP data (i.e. for a different benefit or for the same benefit in a different time period). Although based on fairly small numbers, this suggests that linkage failures are unlikely to explain a large part of the apparent over-reporting.

4.2 Receipt by Other Household Members

False positives could also occur due to survey reports relating to receipt by a partner for whom DWP data is absent (recall that the questions ask about receipt "either yourself or jointly"). In addition to the above reasons for matching failure, this could happen due to the partner having not responded to the survey or not given permission for the matching. To investigate this possibility, we repeated the analysis of Table 2, restricting it to households containing one adult (or one pensioner in the case of retirement pension). Amongst these sub-samples, the false positive rates are similar to those for the whole sample. This suggests that receipt by partners does not explain a large part of the apparent over-reporting.

4.3 Errors in the DWP Data

Even if the correct DWP record is linked to a survey respondent, the record may contain errors of a sort that cause the respondent to be classified as a non-recipient of a particular benefit within the reference period, even though he or she was in fact a recipient. An example would be the incorrect entry of dates of the beginning or end

of a claim. We cannot assess this possibility, though we believe that such errors are likely to be of a very low prevalence.

4.4. Genuine Over-Reporting

As the three possible explanations for false positives put forward above do not find much support in the data, it may be concluded that there is *some* over-reporting in the survey data. Some respondents report receipt of a benefit that they have not in fact received during the reference period. Some of this may be due to respondents getting dates wrong (but see section 5.2 below). Also, some over-reporting could be caused by confusion on the part of respondents between different benefits. We find a few cases in the data of respondents whose responses constitute a false positive for one benefit but a false negative for another.

5. Measurement Error in Recalled Dates and Transitions

As already suggested, some of the errors in dichotomous indicators of whether a particular benefit was received at any time during a reference period may be caused by misplacement of dates when receipt either started or ended. This relates to the suggestion of Bound *et al.* (2001, pp.3770-3771) that measurement error in benefit income ("transfer program income" in American terminology) is more likely to occur when recipiency status is volatile rather than stable. In this section we examine explicitly measurement error in recalled dates. We relate our findings to the discussion in sections 3 and 4 above.

5.1 Mis-recalled dates as an explanation for under-reporting

Under-reporting might be particularly likely to occur when a respondent had received a benefit only during the early part of the survey reference period. We will refer to sample members who had received the benefit (according to the administrative data) at some point during the survey reference period but not since January 2003 (and therefore not currently at the time of the ISMIE interview) as "past recipients" and those who had received it since January 2003 as "recent recipients." (The modal reference period is September 2001 to February 2003, so the period since January 2003 can reasonably be thought of as representing recent or current receipt.) Under-reporting by past recipients would be consistent with the idea of 'constant wave response bias' (Young, 1989), whereby "respondents may give an answer for earlier months in an interview period, identical with the answer they give for the most recent month or their current state" (Young, 1989 p. 395). Kalton and Miller (1991) provide a possible explanation for this phenomenon: "Respondents may give the same answers for each month because they have forgotten that a change occurred during the … reference period or simply because repeating the same answer requires less effort" (Kalton and Miller, 1991 pp. 243-244).

We find that almost half of the false negatives observed (58 out of 128 cases, aggregated across the six benefits) were past recipients. Given that, overall, the proportions of recipients who were past recipients were much lower (8.9% overall across all instances of receipt of any of the six benefits: from 0.0% for retirement pension to 27.0% for family credit), this suggests that cessation of receipt during the reference period is associated with an increased risk of under-reporting. Indeed, the overall false negative rates are about nine times greater amongst past recipients

than amongst recent recipients (Table 4). It is also apparent that DI was disproportionately successful at reducing the odds of under-reporting amongst past recipients, as indicated by the lower odds ratios.

Table 4: False negative rates amongst recent and past recipients by treatment group

False negative rate	INDI	PDI	RDI
Past recipients	.781	.654	.571
Recent recipients	.094	.075	.070
Odds ratio	34.3	23.2	17.7
Base (past recipients)	*32*	*26*	*28*
Base (recent recipients)	*265*	*305*	*314*

Note: Bases are all cases indicated as recipients by the DWP data; cases defined as in Table 7.

5.2 Mis-recalled dates as an explanation for over-reporting

Over-reporting may occur if a respondent had received a benefit during the period immediately prior to the survey reference period, but not during the survey reference period. To test this hypothesis, we constructed two indicators of receipt in the immediate prior period. The first defined the prior period as March 2001 to August 2001; the second defined it as September 2000 to August 2001. Amongst the 56 cases of false positives in our data, only one was classified as a past recipient (under both definitions). It therefore seems that there is no association between transition off benefit during a period immediately before the survey reference period and false positives. Recall error in the dates of transitions does not therefore seem to contribute to the observed over-reporting.

6. Modifying DI Designs to Further Reduce Under-Reporting

Although dependent interviewing appears to reduce the extent of under-reporting, at least for two of the benefits, it does not eliminate it. Indeed, for all five benefits where there is some under-reporting with INDI, under-reporting remains with DI. This is mainly because DI can only have an impact on respondents who are actually asked the DI question. Many of the under-reporters in the DI treatment group were not asked the DI question as they did not report the benefit at wave 8. Amongst respondents who *did* report receipt of a particular benefit at wave 8, the effect of DI is clear (Table 5). For each of the five benefits for which there was under-reporting with INDI, the rate of under-reporting was lower with both PDI and RDI. Only six of these ten reductions in error rate are significant (P<0.05), but this may largely be due to the small sample sizes within each benefit x treatment group combination.

Table 5: False negative rate by treatment group amongst wave 8 reporters

	False negative rate ($p_{10}/p_{1\cdot}$)			*Base (n)*		
	INDI	PDI	RDI	*INDI*	*PDI*	*RDI*
Retirement Pension	.00	.00	.00	*(73)*	*(81)*	*(85)*
Incapacity Benefit	.29	.00*	.07	*(14)*	*(12)*	*(14)*
Income Support	.12	.03*	.03	*(41)*	*(39)*	*(30)*
Child Benefit	.25	.04**	.07**	*(49)*	*(52)*	*(59)*
Family Credit	.21	.00*	.20	*(14)*	*(17)*	*(25)*
Housing Benefit	.10	.05	.02*	*(41)*	*(57)*	*(58)*

PDI and RDI are each compared with INDI using a Pearson χ^2 test. The survey design is taken into account by specifying households as PSUs. *** P≤0.001; ** 0.001<P≤0.01; * 0.01<P≤0.05.

It therefore seems likely that overall under-reporting rates could be further reduced if the DI questions could be extended to other sample members (other than those who reported receipt of the benefit at the previous wave) who have a high propensity to

under-report, provided that this can be done without excessively increasing the proportion of the sample who would need to be asked the DI questions. Indeed, given that propensity to under-report is likely to be associated with some fixed characteristics of the survey respondent, those who under-report at the current wave could be expected to have an increased propensity to have under-reported also at the previous wave, so it is *a priori* likely that limiting the DI questions to those who reported receipt at the previous wave will exclude some under-reporting recipients from the DI treatment.

An obvious extension would be to ask the DI question of all sample members who reported receipt at any of the previous i waves, $i > 1$. In the following we refer to such a design as the $n = i$ design.

In Table 6 we present for each of the six benefits the numbers of false negative cases who had reported receipt of the benefit at one or more of waves 4 to 7 of the ECHP (i.e. receipt at some point during the reference period covered by that interview). The analysis is limited to cases where the benefit in question was *not* reported at wave 8, as our focus here is on the impact of extending a DI question *beyond* respondents who had reported receipt at the previous wave (wave 8). All three treatment groups are combined in the analysis, as the treatment was essentially identical if receipt had not been reported in the previous wave: only the standard independent question was asked.

Of the 79 cases of under-reporting by respondents who had also not reported receipt at wave 8, 34 (43%) would be asked a DI question with the $n = 5$ design (ask the question of all respondents who had reported receipt at any of waves 4 to 7). The

n = 3 design is almost as effective, capturing 31 (39%) of the cases of under-reporting. Of course, we cannot expect that all of these cases would then report their receipt in response to the DI question, but we would expect a high proportion to do so. By extrapolating the false negative rates amongst ISMIE respondents who were actually asked the DI question (i.e. those in the PDI and RDI groups who had reported receipt at wave 8), we would predict that around 28 of the 31 cases might be expected to report receipt in response to the DI question. The 31 cases are of course distributed over the benefits (Table 6), so it would be unwise to present empirical estimates of expected error rates for each benefit. But on average, we would expect that around one-third of the under-reporting that remains with the *n* = 1 DI design would be removed with the *n* = 3 design (28 out of 79 cases in our data).

This further reduction in measurement error comes at a cost, namely the need to ask the additional DI questions of more respondents. For example, the *n* = 3 design would have resulted in our study in an extra 227 questions being asked across the six benefits relative to the *n* = 1 design, i.e. an extra 0.22 per sample member on average. In Table 7 we report the mean number of DI questions per sample member that would have been asked for each of the *n* = *i* designs, *i* = 1, … , 5. For *i* = 2, … , 5 we additionally show the proportions of under reporters with the *n* = 1 design, and others, who would be asked the DI question. We would like to maximise the former while keeping the latter as small as possible.

Using the *n* = 1 design, as for the ISMIE PDI group, would result in a mean of 1.03 extra questions per respondent. The *n* = 5 design would increase the number of extra questions to 1.43 per respondent, a fairly modest increase (note that this compares with an extra 6.00 questions per respondent if an explicit question were

asked of each respondent about each benefit). In terms of sample coverage of the DI questions, the *n* = 3 design seems to be optimal. Any further extension of the questioning to *n* = 4 or *n* = 5 brings only a very small increase in the coverage of under reporters, but a much larger increase in the coverage of other respondents (for whom the DI questions have no benefit). For example, with the *n* = 5 design, 7.5% of "other" sample members (those who would not be asked the DI question with the *n* = 1 design and would not be under-reporters) would be asked the DI question. As these constitute 82% of the total sample, this is not a trivial increase in the questioning effort. With the *n* = 3 design, only half this number of the "other" respondents would be asked the DI question. Note, however, that this assumes PDI. RDI would greatly reduce the mean number of questions asked per respondent.

Table 6: Numbers of under-reporters who had reported receipt at past waves

	Reported receipt at wave…				
	7	6 or 7	5, 6 or 7	4, 5, 6 or 7	*Base*
Retirement Pension	0	0	0	0	*2*
Incapacity Benefit	2	5	5	7	*29*
Income Support	5	5	5	5	*12*
Child Benefit	2	2	2	2	*2*
Family Credit	1	2	2	2	*7*
Housing Benefit	8	17	18	18	*27*
Total	18	31	32	34	*79*

Note: Base is ISMIE respondents who were deemed a "false negative" for the specified benefit and did *not* report receipt of that benefit at wave 8.

Table 7: Mean numbers of DI questions per respondent under five alternative designs

i	1	2	3	4	5
Mean DI questions per respondent	1.03	1.16	1.25	1.31	1.43
Coverage of rep(*t*-1) (%)	100.0	100.0	100.0	100.0	100.0
Coverage of nonrep(*t*-1) false negatives (%)	-	22.8	38.0	39.2	41.8
Coverage of other nonrep(*t*-1) cases (%)	-	2.3	3.8	5.1	7.5

Notes: The $n = i$ designs are described in the text; rep(*t*-1) indicates reporters of the benefit at wave *t*-1; nonrep(*t*-1) indicates non-reporters; false negatives are recipients who do not report receipt at wave *t*. In the ISMIE sample, 17.15% of cases were rep(*t*-1), 1.28% were nonrep(*t*-1) false negatives, and 81.57% were other nonrep(*t*-1). A case is defined as a respondent-benefit combination, so there are 6,192 cases in this analysis (1,032 respondents x 6 benefits)

Aside from previous reports of receipt of the benefit in question, there may be other survey items from previous waves that could be used to identify respondents eligible for a DI question. For some benefits, there exist items which match closely (though not perfectly) the eligibility criteria for a particular benefit. For retirement pension, the DI question could be asked of all persons of retirement age. In our sample, this would capture the remaining two under-reporters in Table 6, without greatly increasing the number of other respondents who would be asked the DI question (14 other respondents in our sample; all classified as "true negatives" though it is possible that some of these are under-reporters for whom a successful match was not made). For child benefit, the DI question could be asked of all women[1] with dependent children in the household (aged 0-16 or 17-18 and in full-time education). Again, this would capture the remaining two under-reporters in Table 6, though it would also capture 69 other respondents with no record of child benefit receipt in the administrative data. This number could no doubt be reduced by restricting the question to women who were the mother of at least one child in the household.

[1] Child benefit is usually paid to the mother. Asking the DI question of both men and women with children in the household would triple the number of respondents asked the question unnecessarily to 207.

For the other four benefits, it may be possible in principle to identify other survey items that predict under-reporting and could be used as filters for DI questions. For each benefit, receipt of related benefits, for example, might be a useful indicator. There may also be other items of relevance to specific benefits, such as health or disability items for incapacity benefit. We do not explore these possibilities further here, as the remaining numbers of under-reporters in our sample are small (only 75 in Table 6). However, we do suggest that the approach of filtering DI questions on predictors of receipt may be promising.

Although our focus here is on panel surveys, we would note that some of the question filtering approaches suggested here, such as those based on age and gender for retirement pension, or gender and presence of children for child benefit, could be applied also in cross-section surveys provided that the demographic details were collected earlier in the interview.

7. Summary and Conclusions

We have found that DI reduces the extent of under-reporting of benefit receipt, at least for two important benefits (section 3.2). There is no evidence that this comes at the cost of an increase in over-reporting (section 3.3). For five out of the six benefits examined, under-reporting is by far the dominant component of measurement error under INDI, suggesting that DI is likely to reduce measurement error (section 3.4). However, some net under-reporting remains even with DI.

We believe that DI has the potential to reduce under-reporting even further. It could achieve this if ways could be found of targetting a DI question at respondents most

at risk of under-reporting, provided that this did not result in excessively large proportions of other respondents also being asked the DI question. We have explored two strategies that seem to be promising in terms of meeting these criteria (section 6). One is to filter the DI questions based on the responses to other survey questions that indicate likely eligibility for the benefit in question. For retirement pension, the question could be asked of all respondents who meet the age eligibility criterion. For child benefit, it could be asked of all mothers of dependent children. The second strategy is to ask the DI question of all respondents who reported the benefit, not just at the previous interview but at any of the previous n interviews. In our study, $n = 3$ appears to be optimal, corresponding to reported receipt of the benefit at any time in the previous 3.5 years. We show that this strategy is likely to bring about a further worthwhile reduction in measurement error (in addition to that brought about by asking the DI question of those who reported receipt in the previous interview). We therefore suggest that it is a design worth pursuing. Of course, the optimum value of *n* could differ if the interval between waves was something other than one year.

A possible third strategy is to identify other survey variables that predict a tendency to under-report. We have identified mis-remembering of dates as an important factor contributing to under-reporting that remains even with the DI design we tested (section 5.1). Misclassification and simple forgetting are also likely to be important. Good candidate variables to trigger a DI question would therefore be those related to the tendency to mis-classify, to forget or to misremember dates of receipt. These might include reported receipt of other benefits, a tendency to move on and off the benefit, age, level of education and so on. We were unable to pursue this strategy further in our study as sample numbers were insufficient to permit modelling of the

propensity to under-report amongst those who were not asked the DI question under our DI design. This warrants further research on a larger sample for which validation data are available.

The first and third strategies described here could also be applied in cross-sectional surveys (or the first wave of panel surveys) provided that the relevant indicator variables (age, gender, education, etc) are collected earlier in the interview.

References

Bound, J., Brown, C. and Mathiowetz, N. (2001) Measurement error in survey data, 3705-3843 in Heckman J.J. and Leamer E. (ed.s) *Handbook of Econometrics vol.5*. Amsterdam: Elsevier Science.

Department for Work and Pensions (2004a) *Quarterly Bulletin on the Population of Working Age on Key Benefits*. Newcastle-upon-Tyne: DWP. http://www.dwp.gov.uk/asd/asd1/cga_wa/CGA_WA_May04_bulletin.pdf

Department for Work and Pensions (2004b) *Client Group Analysis of the Population over State Pension Age*. Newcastle-upon-Tyne: DWP. http://www.dwp.gov.uk/asd/asd1/state_pension/cga/CGA_Pen_Bulletin_May_2004.pdf

Department for Work and Pensions (2004c) *Quarterly Bulletin on Families with Children on Key Benefits*. Newcastle-upon-Tyne: DWP. http://www.dwp.gov.uk/asd/asd1/cga_famchild/CGA_FC_May04.pdf

Department for Work and Pensions (2004d) *Households Below Average Income 2002/03*. London: The Stationery Office. http://www.dwp.gov.uk/asd/hbai/hbai2003/chapters.asp

Department for Work and Pensions (2004e) *Benefit Expenditure Tables*. http://www.dwp.gov.uk/asd/asd4/medium_term.asp

Dibbs, R., Hale, A., Lovelock, R. and Michaud, S. (1995) Some effects of computer-assisted interviewing on the quality of the Survey of Labor and Income Dynamics, *SLID Research Paper 95-07*, Statistics Canada, Ottawa.

Jäckle, A., Sala, E., Jenkins, S. and Lynn, P. (2004) Validation of survey data on income and employment: the ISMIE experience. *ISER Working Paper 2004-14*. Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-14.pdf

Jenkins, S.P., Cappellari, L., Lynn, P., Jäckle, A. and Sala, E. (2004a) Patterns of consent: evidence from a general household survey, *ISER Working Paper 2004-27*. Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-27.pdf

Jenkins, S.P., Lynn, P., Jäckle, A. and Sala, E. (2004b) Linking household survey and administrative record data: what should the matching variables be?, *ISER Working Paper 2004-23*. Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-23.pdf

Kalton, G. and Miller, M.E. (1991) The seam effect with social security income in the Survey of Income and Programe Participation, *Journal of Official Statistics*, 7:2, 235-245.

Lynn, P. (ed.) (2003) *Quality Profile: British Household Panel Survey Waves 1-10: 1991-2000*, Colchester, University of Essex. http://iserwww.essex.ac.uk/ulsc/bhps/

Lynn, P., Jäckle, A., Jenkins, S.P. and Sala, E. (2004) The effects of dependent interviewing on responses to questions on income sources, *ISER Working Paper*

2004-16. Colchester: University of Essex. http://www.iser.essex.ac.uk/pubs/workpaps/pdf/2004-16.pdf

Mathiowetz, N.A. and McGonagle, K.A. (2000) An assessment of the current state of dependent interviewing in household surveys, *Journal of Official Statistics*, 16:4, 401-418.

Young, N. (1989) Wave-seam effects in the SIPP, *Proceedings of the Survey Research Methods Section, American Statistical Association*, 393-398.

Appendix: Income Source Questions

Independent Interviewing

I am going to show you four cards listing different types of income and payments. Please look at this card and tell me if, since September 1st 2001, you have received any of the types of income or payments shown, either just yourself or jointly?

If yes: "*Which ones?*" Probe: "*Any others?*" Until final "no"

Code entered for each that applies. Question repeated for each card in turn.

CARD 1

N.I. Retirement (Old Age) Pension 01
A Pension from a previous employer 02
A Pension from a spouse's previous employer 03
A Private Pension/Annuity 04
A Widow's or War Widow's Pension 05
A Widowed mother's allowance 06

CARD 2

Severe Disablement Allowance 16
Industrial Injury or Disablement Allowance 18
Disability Living Allowance/ Care Component 19
Disability Living Allowance/ Mobility Component 20
Disability Living Allowance/ Components not known 21
Disabled Person's Tax Credit 22
(Formerly Disability Working Allowance)
Attendance Allowance 23
Invalid Care Allowance 24
War Disability Pension 25
Incapacity Benefit 26
(Formerly invalidity benefit/NI Sickness benefit)

CARD 3

Income Support 32
Job Seeker's Allowance 34
Child Benefit 35
Child Benefit (Lone Parent) 36
Working Family Tax Credit 37
(Formerly Family Credit)
Maternity Allowance 38
Housing Benefit/Rent rebate or allowance 39
Council Tax Benefit 40
Any other state benefit 41

CARD 4

Educational Grant (not Student Loan) 51
Trade Union/Friendly Society Payments 52
Maintenance or Alimony 53
Payments from a family member not living here 54
Rent from Boarders or lodgers (not family members) living here with you 55
Rent from any other property 56
Foster Allowance 57
Sickness or accident insurance 58
Any other regular payment
(PLEASE GIVE DETAILS) 59

For each code entered: *And for which months since September 1st 2001 have you received…?*

Reactive Dependent Interviewing

Independent questions, as above, followed by:

For each income source reported at wave 8 but not wave 9:

> *Can I just check, according to our records you have in the past received <SOURCE>. Have you received < SOURCE > at any time since <INTDATE>?*
>
> *For which months since <INTMON> have you received < SOURCE >?*

Proactive Dependent Interviewing

For each income source from card 1 reported at wave 8 (i.e. received in one or more month between September 2000 and the wave 8 interview, September 2001-February 2002):

> *According to our records, when we last interviewed you, on <INTDATE>, you were receiving <SOURCE>, either yourself or jointly. For which months since <INTMON> have you received < SOURCE >?*

Then:

> **CARD 1:** *I am going to show you four cards listing different types of income and payments. Please look at this card and tell me if, since <INTDATE>, you have received any other of the types of income or payments shown, either just yourself or jointly?*

Then equivalent questioning for each of cards 2, 3 and 4 in turn (excluding income sources 41 and 59 from the initial proactive question).