# The impact of questioning method on measurement error in panel survey measures of benefit receipt: evidence from a validation study

Peter Lynn and Annette Jäckle,

*University of Essex, Colchester, UK*

Stephen P. Jenkins

*London School of Economics and Political Science, UK*

and Emanuela Sala

*University of Milano Bicocca, Milan, Italy*

**Summary.** We assess measurement error in panel survey reports of social security benefit receipt, drawing on a unique validation study. Our aims are threefold. First, we quantify the incidence of measurement errors (under- and over-reporting). Second, we assess the extent to which this varies according to the questioning method that is used. Specifically, dependent interviewing has been proposed as a way to reduce under-reporting in some circumstances. We compare two versions of dependent interviewing with traditional independent interviewing in an experimental design. Third, we identify and assess new ways of reducing measurement error in panel surveys. We use data from a large-scale UK household panel survey and we consider six benefits. To assess the measurement error, a validation exercise was conducted, with administrative data on benefit receipt matched at the individual level to the survey microdata.

## 1. Introduction

Survey measures of benefit receipt are important for studies of income, poverty and related issues since social security benefit receipt ('transfer program income' in American terminology) is an important component of income for many households in the UK. 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, and 2.5 million children aged under 16 years (22% of the population) were living in a household claiming a key benefit (Department for Work and Pensions, 2004a, b, c). Among the poorest fifth of households, around 53% of gross household income is accounted for by benefits (Department for Work and Pensions (2009), Table 2.2).

Survey measures of benefit receipt are subject to measurement error (Bound *et al.*, 2001). Some survey respondents may under-report benefit receipt. This could be due to simple forgetting since for instance 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. It could also be due to misplacement in time or misclassification, or due to conscious suppression caused, for example, by social desirability (DeMaio, 1984) or by an unwillingness to reveal sensitive information (Tourangeau and Smith, 1996). Over-reporting is also possible, perhaps due to misclassification or misplacement in time. One form of misplacement in time that is often observed in surveys is 'telescoping', whereby events are recalled as being more recent than they actually are (Bradburn *et al.*, 1994). In a panel survey, this may be reduced by the effects of 'bounding' (Neter and Waksberg, 1964), depending on the question design and data editing procedures that are adopted. Bounding involves comparing events reported at the first two interviews and then ignoring second-interview reports of events that were already mentioned at the first interview. Estimates are based solely on (remaining) events reported at the second or later interviews. Thus, the date of the first interview serves as a 'bound' on the reports, as events that took place before that date should be excluded from estimates.

Suppression due to social desirability effects (Tourangeau *et al.*, 2000) may occur for benefit data, especially among people who consider benefit receipt to be stigmatizing. A possible reason for misclassification to occur in our data is respondents' confusion between different benefits. Hancock and Barker (2005) reported confusion between certain benefits among respondents to the UK Family Resources Survey. One other possible source of error is the fact that some benefits are assessed at the family unit level. This leaves room for confusion among survey respondents about which one of two partners in a family should report certain benefits.

In longitudinal surveys, dependent interviewing (DI) is a method that was designed primarily to reduce overestimation of changes in status (Mathiowetz and McGonagle, 2000). DI can take various forms, but the essential elements are that it involves reminding a survey respondent of relevant responses that they gave at a previous interview or asking a different form of question depending on responses previously given (Jäckle, 2009). DI differs from traditional independent interviewing (INDI), in which all respondents are asked an identical question without reference to any answers given in previous interviews. The forms of DI can be classified as either proactive or reactive. Proactive DI (PDI) involves using information from a previous interview to form the question, in place of an INDI question, whereas reactive DI (RDI) involves asking an INDI question and then a follow-up question, the nature of which may be determined both by responses in an earlier interview and the response to the INDI question. An example of a PDI question is 'Last time we interviewed you, you said you were receiving housing benefit; is that still the case?'. An RDI approach might involve first asking an INDI question, 'Are you currently receiving housing benefit?', and then a follow-up question if the response differed from the response given in the previous interview, e.g. 'So you have stopped receiving housing benefit since we last interviewed you, is that right?'. Lynn *et al.* (2006) further described the two types of DI and how they differ from INDI. Appendix A illustrates the differences between the three approaches. DI and INDI may have different implications for measurement error in survey measures of benefit receipt.

In this paper, we focus on measurement error in estimates of prevalence levels of benefit receipt. By prevalence level we mean the proportion of the population in receipt of a particular benefit or type of benefit. Our aims are threefold. First, we attempt to quantify the incidence of measurement errors (under- and over-reporting). Second, we assess the extent to which this varies according to the questioning method used. We compare two versions of DI with INDI 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 data. This is the first study of DI to use validation data.

Earlier studies presented evidence that reported levels of benefit receipt are greater with DI (Dibbs *et al.*, 1995; Lynn *et al.*, 2006). However, those studies—unlike ours—did not have validation data, so to interpret higher observed prevalence levels as a reduction in measurement error requires an assumption that measurement error consists primarily of under-reporting. In this paper, 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 by DI (Section 3). We discuss possible explanations for the small amount of *over*-reporting that is found (Section 4) and 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 DI for panel surveys, or filtered questioning for cross-sectional surveys, could be extended to reduce under-reporting further (Section 6). Section 7 summarizes 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, which was funded by the research methods programme of the UK Economic and Social Research Council. 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 'low income supplemental sample' of the European Community Household Panel (ECHP) survey. This sample was selected in 1997 from respondents in England, Scotland and Wales to the 1994–1996 UK ECHP who exhibited characteristics that are associated with an increased likelihood of low household income (e.g. elderly, single parents or in receipt of income support). A description of the sample design appears in Lynn (2006). Though the sample was not designed to represent all sections of the population equally, it is in important respects similar to the total resident population of England, Scotland and Wales (Jäckle *et al.*, 2004). For experimentation with questions about sources of income, it is an advantage that benefit recipients are over-represented in the sample.

The ECHP interviewed all adult members of sample households eight times at annual intervals. The last wave of interviewing took place between September 2001 and February 2002. The 1163 sample members (in 700 households) who had provided full interviews at wave 8 (2001–2002) of the ECHP were included in the ISMIE study. They were randomly assigned to one of three treatment groups, where the groups are defined by the questioning method that was used. We refer to the groups as the INDI, RDI and PDI groups. The specific questions that were asked of each group regarding unearned income appear in Appendix A. 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 with a median interview length of 5 min. Additionally, an individual interview was carried out with each sample member by using computer-assisted personal interviewing. Individual interviews had a median length of 24 min. A total of 1033 interviews were achieved, representing a response rate of 89%. We refer to these 1033 people as the 'ISMIE respondents'. Fieldwork 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 on data from the previous interview with the same respondent ('wave 8'), but did not

utilize data from interviews with other household members. 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. 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 six types of pension, the second listed 10 state benefits related to disability or injury, the third listed nine other state benefits and the fourth listed eight other miscellaneous sources of income, 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 Appendix A.

A question requesting consent to link administrative data from the Department for Work and Pensions (DWP) to the survey data was asked at the end of the ISMIE individual interview. The DWP is the UK Government department that is responsible for administering state benefits. If respondents answered that they did not 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 among respondents who lived alone, but it did not differ between the three treatment groups that are of interest to this study. For further details see Jenkins *et al.* (2006).

## 2.2. Administrative data

The DWP data were linked to the ISMIE survey data by using non-hierarchical pooled matching based on five criteria. This matching method involved attempting to match independently on each of five criteria and then pooling the results to identify a single match for each survey respondent, as follows. The first match criterion was national insurance number (which ISMIE respondents were asked to supply immediately after the consent question). The other four criteria were combinations of sex with two or three out of date of birth, forename, family name, postcode and first line of address. Among the 14 cases in our data where an ISMIE respondent was matched to more than one person in the DWP data, the modal match was accepted as the correct match, provided that the records matched on at least three of the five criteria. 12 of the 14 cases were matched in this way. The remaining two cases were individually inspected to determine which match appeared to be correct. Among 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.* (2008). All respondents for whom no DWP data were matched are retained in the analysis.

Six of the benefits represented in the DWP data form the focus of the analysis that is presented in this paper: state retirement pension, child benefit, income support, incapacity benefit, working families tax credit (referred to hereafter as tax credit) and housing benefit. State contributory retirement pension is paid to people who have reached state pension age, which at present is 65 years for men and 60 years for women, and have also achieved 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 years and those aged 17 or 18 years in

full-time non-advanced education at a recognized educational establishment. Income support 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. Tax credit 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. Working families tax credit was replaced in April 2003—around the end of the ISMIE fieldwork period—by working tax credit. Tax credit can refer to either. Housing benefit is designed to help people on low income to pay their rent. Three of these six benefits (income support, tax credit and housing benefit) 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 the state retirement pension, at the time of the ISMIE survey in February 2003 (Department for Work and Pensions (2004d), Table C1).

## 3. Effect of interviewing method on measurement error

### 3.1. Estimation of measurement error

Of the social security benefits that are represented in both sources of data, we restrict our analysis to the six described in the previous section as these were the most prevalent among the ISMIE sample. For these six sources of income between 61 respondents (incapacity benefit) and 256 (retirement pension) reported receipt in the survey interview and between 78 (tax 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 1 month during the survey reference period. The survey reference period is from September 2001 until and including the month of the ISMIE interview for the INDI and RDI groups, with mean length 18 months, and from the wave 8 month of interview until and including the month of the ISMIE interview for the PDI group, with mean length 17 months. This is the period about which ISMIE respondents were asked. An equivalent indicator was constructed on the basis of 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 respectively 'false negative' responses, which are cases where receipt is indicated by the DWP measure but not by the survey measure, and 'false positive' responses, where receipt is indicated only by the survey measure. If the DWP measure is taken to be accurate, then false negative responses can be interpreted as cases of survey under-reporting and false positive responses as cases of over-reporting. However, these interpretations should be made with caution, as there may be other explanations for false positive results (see Section 4 below).

We should also take into account that the survey questions ask about receipt 'either yourself or jointly' as three of the six benefits are means tested at the level of the family unit—see Section 2. To minimize the risk of erroneously counting a case as falsely positive, we have counted it as '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 positive results, 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 summarized in Table 1.

**Table 1.**   Definition of the 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 (subscript notation used subsequently)* | *Resultant (assumed) measurement error* |
|---|---|---|---|---|
| No | Yes or no | No | True negative (00) | None |
| Yes | Yes or no | Yes | True positive (11) | None |
| No | Yes | Yes | True positive (11) | None |
| Yes | Yes or no | No | False negative (10) | Under-report |
| No | No | Yes | False positive (01) | Over-report |

**Table 2.**   Notation, table proportions

| | *Survey data* | | |
|---|---|---|---|
| | *No receipt* | *Receipt* | *Total* |
| *Administrative data* | | | |
| No receipt | $p_{00}$ | $p_{01}$ | $p_{0\cdot}$ |
| Receipt | $p_{10}$ | $p_{11}$ | $p_{1\cdot}$ |
| Total | $p_{\cdot 0}$ | $p_{\cdot 1}$ | $p_{\cdot\cdot} = 1$ |

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

Our hypotheses are as follows, where $H_1$ indicates the hypothesis in which we are interested ($H_0$ is the corresponding null hypothesis).

(a)  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:

$$H_0 : p_{10}^c/p_{1\cdot}^c = p_{10}^{\mathrm{INDI}}/p_{1\cdot}^{\mathrm{INDI}}, \qquad c \equiv \mathrm{PDI, RDI};$$

$$H_1 : p_{10}^c/p_{1\cdot}^c = p_{10}^{\mathrm{INDI}}/p_{1\cdot}^{\mathrm{INDI}}, \qquad c \equiv \mathrm{PDI, RDI}.$$

(b) 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:

$$H_0 : p^c_{01}/p^c_{0.} = p^{\text{INDI}}_{01}/p^{\text{INDI}}_{0.}, \qquad c \equiv \text{PDI, RDI};$$

$$H_1 : p^c_{01}/p^c_{0.} = p^{\text{INDI}}_{01}/p^{\text{INDI}}_{0.}, \qquad c \equiv \text{PDI, RDI}.$$

(c) 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:

$$H_0 : p^{\text{INDI}}_{10}/p^{\text{INDI}}_{1.} = p^{\text{INDI}}_{01}/p^{\text{INDI}}_{0.};$$

$$H_1 : p^{\text{INDI}}_{10}/p^{\text{INDI}}_{1.} = p^{\text{INDI}}_{01}/p^{\text{INDI}}_{0.}.$$

(d) 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:

$$H_0 : |p^c_{01} - p^c_{10}| = |p^{\text{INDI}}_{01} - p^{\text{INDI}}_{10}|, \qquad c \equiv \text{PDI, RDI};$$

$$H_1 : |p^c_{01} - p^c_{10}| = |p^{\text{INDI}}_{01} - p^{\text{INDI}}_{10}|, \qquad c \equiv \text{PDI, RDI}.$$

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

In carrying out statistical tests of differences between estimates, we cannot adjust standard errors for the clustering of survey households within postal sectors. This is because the Office for National Statistics, who originally selected the ECHP sample and carried out the initial fieldwork, are unwilling for indicators of primary sampling unit—even if anonymized—to be released. Such indicators are therefore not available to analysts. The effect of this clustering is in any case likely to be small, as the mean number of households per primary sampling unit in our analysis is approximately 3. We do, however, take into account the fact that survey respondents are clustered within households. We do this by using the SVY commands in Stata, specifying households as primary sampling units.

The distribution of our derived indicator, for each benefit and each treatment group, is presented in Table 3, where 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 that are relevant to our hypotheses.

### 3.2. Under-reporting

Among the INDI group, as shown in Table 2, false negative responses depress the survey estimate of the proportion in receipt of the benefit by between 0.0 for retirement pension and 6.0 percentage points for incapacity benefit. This translates to a false negative rate $p_{10}/p_1.$ of between 0% for retirement pension and 50% for incapacity benefit, as illustrated in Fig. 1. Hypothesis (a) was tested by comparing the false negative rates for each form of DI with that for INDI, separately for each benefit (Table 4). DI significantly ($P < 0.05$) reduces the prevalence of false negative results for child benefit for both RDI and PDI and tax credit for 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

**Table 3.**   Income receipt indicators from administrative and survey data (row proportions)†

| Benefit | Treatment group | True negative, $p_{00}$ | True positive, $p_{11}$ | False negative, $p_{10}$ | False positive, $p_{01}$ | Administrative, $p_{1\cdot}$ | Survey, $p_{\cdot 1}$ | Difference, $p_{01} - p_{10}$ |
|---|---|---|---|---|---|---|---|---|
| Retirement pension | INDI | 0.698 | 0.298 | — | 0.004 | 0.298 | 0.302 | 0.004 |
|  |  | (0.028) | (0.029) |  | (0.004) | (0.029) | (0.029) | (0.004) |
|  | PDI | 0.663 | 0.330 | 0.004 | 0.004 | 0.333 | 0.333 | 0.000 |
|  |  | (0.029) | (0.029) | (0.004) | (0.004) | (0.029) | (0.030) | (0.006) |
|  | RDI | 0.664 | 0.321 | 0.011 | 0.004 | 0.332 | 0.325 | −0.007 |
|  |  | (0.030) | (0.028) | (0.006) | (0.004) | (0.029) | (0.029) | (0.007) |
| Incapacity benefit | INDI | 0.882 | 0.057 | 0.057 | 0.004 | 0.115 | 0.061 | −0.054 |
|  |  | (0.020) | (0.015) | (0.014) | (0.004) | (0.021) | (0.015) | (0.015) |
|  | PDI | 0.897 | 0.058 | 0.042 | 0.004 | 0.100 | 0.062 | −0.038 |
|  |  | (0.019) | (0.015) | (0.012) | (0.004) | (0.019) | (0.016) | (0.014) |
|  | RDI | 0.869 | 0.073 | 0.040 | 0.018 | 0.113 | 0.091 | −0.022 |
|  |  | (0.020) | (0.016) | (0.012) | (0.008) | (0.020) | (0.017) | (0.015) |
| Income support | INDI | 0.790 | 0.179 | 0.023 | 0.008 | 0.202 | 0.187 | −0.015 |
|  |  | (0.026) | (0.024) | (0.010) | (0.006) | (0.025) | (0.024) | (0.011) |
|  | PDI | 0.785 | 0.180 | 0.035 | — | 0.215 | 0.180 | −0.034 |
|  |  | (0.025) | (0.025) | (0.012) |  | (0.026) | (0.025) | (0.012) |
|  | RDI | 0.818 | 0.168 | 0.015 | — | 0.183 | 0.168 | −0.015 |
|  |  | (0.023) | (0.023) | (0.007) |  | (0.024) | (0.023) | (0.007) |
| Child benefit | INDI | 0.767 | 0.172 | 0.050 | 0.012 | 0.221 | 0.183 | −0.038 |
|  |  | (0.027) | (0.024) | (0.014) | (0.008) | (0.027) | (0.025) | (0.016) |
|  | PDI | 0.774 | 0.192 | 0.008 | 0.027 | 0.199 | 0.218 | 0.019 |
|  |  | (0.026) | (0.024) | (0.006) | (0.011) | (0.026) | (0.026) | (0.012) |
|  | RDI | 0.770 | 0.208 | 0.018 | 0.004 | 0.226 | 0.212 | −0.014 |
|  |  | (0.025) | (0.025) | (0.009) | (0.004) | (0.026) | (0.026) | (0.009) |
| Tax credit | INDI | 0.901 | 0.057 | 0.023 | 0.019 | 0.080 | 0.076 | −0.004 |
|  |  | (0.019) | (0.015) | (0.009) | (0.009) | (0.017) | (0.017) | (0.013) |
|  | PDI | 0.877 | 0.092 | 0.004 | 0.027 | 0.096 | 0.119 | 0.023 |
|  |  | (0.021) | (0.018) | (0.004) | (0.010) | (0.019) | (0.020) | (0.011) |
|  | RDI | 0.894 | 0.077 | 0.026 | 0.004 | 0.102 | 0.080 | −0.022 |
|  |  | (0.019) | (0.016) | (0.010) | (0.004) | (0.019) | (0.017) | (0.011) |
| Housing benefit | INDI | 0.767 | 0.179 | 0.038 | 0.015 | 0.218 | 0.195 | −0.023 |
|  |  | (0.027) | (0.026) | (0.013) | (0.008) | (0.027) | (0.026) | (0.015) |
|  | PDI | 0.644 | 0.264 | 0.061 | 0.031 | 0.326 | 0.295 | −0.031 |
|  |  | (0.030) | (0.028) | (0.016) | (0.012) | (0.029) | (0.028) | (0.018) |
|  | RDI | 0.668 | 0.274 | 0.029 | 0.029 | 0.303 | 0.303 | 0.000 |
|  |  | (0.028) | (0.027) | (0.010) | (0.011) | (0.029) | (0.029) | (0.015) |

†For the definition of true negative, true positive, false negative and false positive, see Table 1. The columns headed 'Administrative' and 'Survey' show prevalence rates for receipt estimated from the administrative and survey data respectively. The bases are 262 INDI cases, 261 PDI and 274 RDI, comprising all ISMIE respondents who gave consent for DWP matching, with the exception of two cases that were dropped from the analysis owing to missing data on the survey items. Figures in parentheses are estimated standard errors.

is the least prevalent of the six benefits that are included in our analysis and consequently the tests have least power. These findings provide some support for hypothesis (a).

## 3.3. Over-reporting

False positive results appear to inflate the survey estimate of the proportion in receipt of benefit among the INDI group (Table 3) by between 0.4 percentage points for retirement pension and incapacity benefit and 1.9 percentage points for tax credit. This translates to a false positive rate $p_{01}/p_{0\cdot}$ of between 0.4% for incapacity benefit and 2.1% for tax credit. Fig. 2 illustrates
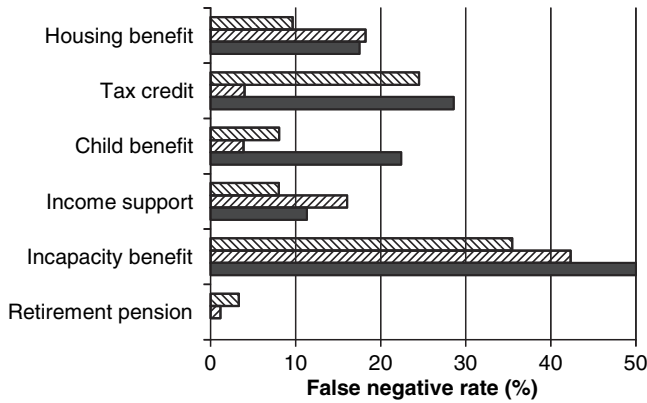
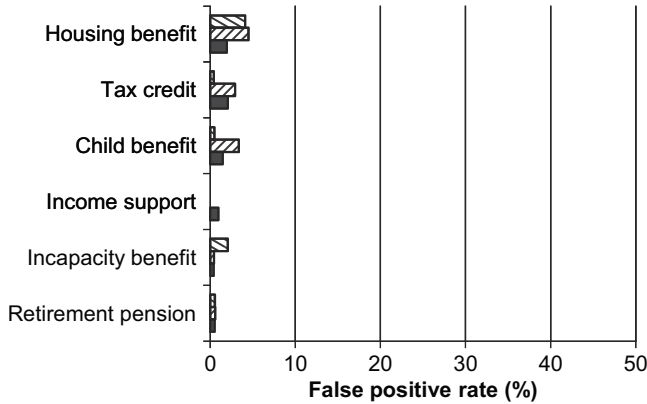**Fig. 1.**    False negative rates for six benefits and three interviewing methods: ▨, RDI; ▧, PDI; ■, INDI



**Fig. 2.**    False positive rates for six benefits and three interviewing methods: ▨, RDI; ▧, PDI; ■, INDI

the false positive rates for the range of benefits. Neither method of DI has a significant effect (at the 0.05 level) on the prevalence of false positive responses for any of the benefits (Table 4). We therefore find no evidence to support hypothesis (b).

### 3.4.   Measurement error

Overall, false positive rates are much lower, for all three interviewing methods, than false negative rates. This is clear from a comparison of the lengths of the bars in Fig. 1 with those in Fig. 2. With INDI, false positive rates are significantly lower for five of the six benefits (Table 4). This supports hypothesis (c) 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 (d) was tested by comparing the estimated error due to measurement in the prevalence estimate—given in the final column of Table 3—between INDI and each form of DI, separately for each benefit. For child benefit, the error was significantly less ($P < 0.05$) with both forms of DI; for both incapacity benefit and housing benefit the error was less with RDI than with INDI.

**Table 4.**   Results of hypothesis tests†

| Benefit | P-values for the following hypotheses: | | | | | | |
|---|---|---|---|---|---|---|---|
| | Hypothesis (a) | | Hypothesis (b) | | Hypothesis (c) | Hypothesis (d) | |
| | PDI | RDI | PDI | RDI | | PDI | RDI |
| Retirement pension | 0.17 | 0.05 | 0.49 | 0.48 | 0.26 | 0.15 | 0.32 |
| Incapacity benefit | 0.32 | 0.12 | 0.29 | 0.05 | 0.000‡ | 0.19 | 0.03§§ |
| Income support | 0.16 | 0.43 | 0.08 | 0.07 | 0.000‡ | 0.08 | 0.49 |
| Child benefit | 0.001§ | 0.01§§ | 0.08 | 0.07 | 0.000‡ | 0.05§§ | 0.04§§ |
| Tax credit | 0.01§§ | 0.39 | 0.27 | 0.13 | 0.000‡ | 1.00 | 1.00 |
| Housing benefit | 0.84 | 0.10 | 0.07 | 0.13 | 0.000‡ | 0.29 | 0.006§ |

†Each hypothesis is tested by using a standard Pearson $\chi_1^2$-test for a difference in proportions. The clustered survey design is taken into account by treating households as primary sampling units.
‡$P \leqslant 0.001$.
§$0.001 < P \leqslant 0.01$.
§§$0.01 < P \leqslant 0.05$.

## 4.   Why does over-reporting appear to occur?

A degree of under-reporting was to be expected, for the reasons set out earlier. Over-reporting is perhaps more surprising, so in this section we discuss how it may arise. There are at least three possible explanations for apparent false positive responses, other than actual over-reporting. In this section, we explore the likely extent of each, to understand better the extent to which observed false positive responses represent genuine over-reports by survey respondents.

### 4.1.   Failure of the matching process
The DWP measure may be incorrect in some cases owing to a failure in the matching process. This could cause a false positive response if a correct record for a particular benefit was present in the DWP data since the respondent is a recipient but was not matched to the 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 that were classified as 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 more than about a third of the apparent over-reporting.

Matching failures could of course also cause an apparent false negative response if a respondent not in receipt of a benefit was incorrectly matched to the DWP record of someone who is in receipt. We believe that such cases are likely to be rare, given the good match on personal details for the majority of survey respondents matched to a DWP record and the general consistency between the two sources of data among matched respondents.

### 4.2. Receipt by other household members

Eligibility for a means-tested benefit is assessed in terms of the income of the 'benefit unit', which is defined as a single adult or a couple living as married plus any dependent children, with payment of a benefit made to one person within the benefit unit: the claimant. Hence, in the case of couples, there will be a record of benefit receipt associated with one member of the couple being the claimant in the DWP administrative data. Thus, recalling that the questions ask about receipt 'either yourself or jointly', a false positive response could occur if the non-claimant member of a couple reports receipt and DWP data are absent from our data set for the claimant. This could happen because the claimant had not responded to the survey or not given permission for the matching or due to linkage failure. To investigate this possibility, we repeated the analysis of Table 3, restricting it to households containing only one adult or one pensioner in the case of retirement pension. Among these subsamples, 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 Department for Work and Pensions 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 in the administrative data are likely to be of low prevalence.

### 4.4. Genuine over-reporting

As the three possible explanations for false positive results 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 recalling dates wrongly (but see Section 5.2 below). Also, some over-reporting could be caused by confusion on the part of respondents between different benefits. Hancock and Barker (2005) reported confusion between attendance allowance, disability living allowance, income support and retirement pension among respondents to the UK Family Resources Survey. We find a few cases in the data of respondents whose responses constitute a false positive response for one benefit but a false negative response 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) that measurement error in benefit income is more likely to occur when receipt 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. Misrecalled 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 shall refer to sample members who according to the administrative data had received the benefit 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 from 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', 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), page 395). Kalton and Miller (1991) provided 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), pages 243–244).

We find that almost half of the false negative responses that were 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 tax 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 among past recipients than among recent recipients (Table 5). It is also apparent that DI was disproportionately successful at reducing the odds of under-reporting among past recipients, as indicated by the lower odds ratios, though these remain high in all cases (Jäckle (2009) reported that 86% of respondents who report receipt of a particular benefit, report receipt for every month in the reference period, and that this proportion does not vary between the three treatment groups).

## 5.2. Misrecalled dates as an explanation for over-reporting

Over-reporting may occur if a respondent had received a benefit during the period immediately before the survey reference period, but not during the survey reference period. To test this

**Table 5.** False negative rates among recent and past recipients by treatment group†

| | *False negative rate for the following groups:* | | |
| --- | --- | --- | --- |
| | *INDI* | *PDI* | *RDI* |
| Past recipients | 0.781 | 0.654 | 0.571 |
| | (0.073) | (0.093) | (0.094) |
| Recent recipients | 0.094 | 0.075 | 0.070 |
| | (0.018) | (0.015) | (0.014) |
| Odds ratio | 34.3 | 23.2 | 17.7 |
| *Base (past recipients)* | 32 | 26 | 28 |
| *Base (recent recipients)* | 265 | 305 | 314 |

†Bases are all cases indicated as recipients by the DWP data; a case is defined as a respondent–benefit combination. Figures in parentheses are estimated standard errors.

**Table 6.** False negative rate by treatment group among wave 8 reporters†

| Benefit | False negaive rate $p_{10}/p_1$. for the following groups: | | | Base n for the following groups: | | |
|---|---|---|---|---|---|---|
| | INDI | PDI | RDI | INDI | PDI | RDI |
| Retirement pension | 0.00 (0.00) | 0.00 (0.00) | 0.00 (0.00) | 73 | 81 | 85 |
| Incapacity benefit | 0.29 (0.12) | 0.00§§ (0.00) | 0.07 (0.07) | 14 | 12 | 14 |
| Income support | 0.12 (0.05) | 0.03§§ (0.03) | 0.03 (0.03) | 41 | 39 | 30 |
| Child benefit | 0.25 (0.06) | 0.04§ (0.03) | 0.07§ (0.03) | 49 | 52 | 59 |
| Tax credit | 0.21 (0.11) | 0.00§§ (0.00) | 0.20 (0.08) | 14 | 17 | 25 |
| Housing benefit | 0.10 (0.05) | 0.05 (0.03) | 0.02§§ (0.02) | 41 | 57 | 58 |

†PDI and RDI are each compared with INDI by using a Pearson $\chi^2$-test. The clustered survey design is taken into account by specifying households as primary sampling units. Figures in parentheses are estimated standard errors.
‡$P \leqslant 0.001$.
§$0.001 < P \leqslant 0.01$.
§§$0.01 < P \leqslant 0.05$.

hypothesis, we constructed two indicators of receipt in the immediate prior period. The first defined the prior period as from March 2001 to August 2001; the second defined it as from September 2000 to August 2001. Among the 56 cases of false positive results 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 positive responses. Recall error in the dates of transitions does not therefore seem to contribute to the observed over-reporting.

## 6. Modifying dependent interviewing designs to reduce under-reporting further

Although DI 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 effect 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 either. Among respondents who *did* report receipt of a particular benefit at wave 8, the effect of DI is clear (Table 6). 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 10 reductions in error rate are significant ($P < 0.05$), but this may largely be due to the small sample sizes within each benefit × treatment group combination.

It therefore seems likely that overall under-reporting rates could be further reduced if the DI questions could be extended to 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 the 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

**Table 7.** Numbers of under-reporters who had reported receipt at past waves†

| Benefit | Reported receipt at wave... | | | | Base |
|---|---|---|---|---|---|
| | 7 | 6 or 7 | 5, 6 or 7 | 4, 5, 6 or 7 | |
| 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 |
| Tax credit | 1 | 2 | 2 | 2 | 7 |
| Housing benefit | 8 | 17 | 18 | 18 | 27 |
| Total | 18 | 31 | 32 | 34 | 79 |

†The base is ISMIE respondents who were deemed false negative for the specified benefit and did *not* report receipt of that benefit at wave 8.

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 what follows we refer to such a design as the $n = i$ design.

In Table 7 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 4–7 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 effect 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 4–7). The $n = 3$ design is almost as effective, capturing 31 (39%) of the cases of under-reporting. Of course, we cannot expect that all 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 among 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 7), so it would be unwise to present empirical estimates of expected error rates for each benefit. But, on average, we would expect that around a 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 8 we report the mean number of DI questions per sample member that would have been asked for each of the $n = i$ designs,

**Table 8.** Mean numbers of PDI questions per respondent under five alternative designs†

| | *Results for the following designs i:* | | | | |
| --- | --- | --- | --- | --- | --- |
| | *1* | *2* | *3* | *4* | *5* |
| Mean PDI questions per respondent | 1.03 | 1.16 | 1.25 | 1.31 | 1.43 |
| Coverage of reporters at wave $t-1$ (%) | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Coverage of non-reporters at wave $t-1$ who under-reported (false negative) at wave $t$ (%) | — | 22.8 | 38.0 | 39.2 | 41.8 |
| Coverage of other non-reporters at wave $t-1$ (%) | — | 2.3 | 3.8 | 5.1 | 7.5 |

†The $n=i$ designs are described in the text. In the ISMIE sample, 17.15% of cases reported receipt at wave $t-1$, 1.28% were non-reporters at wave $t-1$ who were classified as false negative at wave $t$ and 81.57% were other non-reporters at wave $t-1$. A case is defined as a respondent–benefit combination, so there are 6192 cases in this analysis (1032 respondents × 6 benefits).

$i=1,\ldots,5$. For $i=2,\ldots,5$ 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 maximize 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, which is a fairly modest increase. We 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 reduce the mean number of questions asked per respondent, since the follow-up question would only be asked of current non-reporters who have reported receipt any time in the previous $i$ waves.

Aside from previous reports of receipt of the benefit in question, there may be other survey items from previous waves or the current wave that could be used to identify respondents who are eligible for a DI question. For some benefits, there are 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 people of retirement age. In our sample, this would capture the remaining two under-reporters in Table 7, while only increasing the number of other respondents who would be asked the DI question by 14. These 14 respondents were all classified as 'true negative' 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 with dependent children in the household aged 0–16 or 17–18 years and in full-time education. 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 who are asked the question unnecessarily to 207. Again, this would capture the remaining two under-reporters in Table 7, 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. An approach to survey questioning along these lines for retirement pension and child benefit was introduced on the British Household Panel Survey in wave 16, i.e. survey year 2006 (Jäckle *et al.*, 2007).

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 check questions similar to the 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 7). 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 note that some of the question filtering approaches that are 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

Our validation study—the first ever on a study of DI—has revealed that under-reporting is far more prevalent than over-reporting of benefit receipt in survey data, with the net result that rates of benefit receipt tend to be under-estimated. However, the extent of under-reporting varies considerably between the six benefits that were studied, being lowest for the state retirement pension and highest for incapacity benefit.

We 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 of the six benefits examined, under-reporting is by far the dominant component of measurement error under INDI. In consequence, DI—which reduces under-reporting—reduces 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 targeting a DI question at respondents who are 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 on the basis of 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 which is worth pursuing. Of course, in general the optimum value of $n$ may depend on the interval between waves and the temporal stability of the phenomenon under study.

A possible third strategy is to identify other survey variables that predict a tendency to under-report. We have identified misremembering of dates as an important factor contributing to under-reporting that remains even with the DI design that 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 which are related to the tendency to misclassify, 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 or level of education. We could not pursue this strategy further in our study as sample numbers were insufficient to permit modelling of the propensity to under-report among 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 that were 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.

Additionally, we have studied two alternative forms of DI and have found that PDI may be more successful than RDI in reducing under-reporting, though RDI has other advantages such as a reduced length of interview with the approaches explored in Section 6. In general, the relative merits of the two approaches are likely to depend on the nature of the survey questions and the likely nature of measurement error in the absence of DI. These issues are discussed further in Jäckle (2008a, 2009).

Our focus in this paper has been on bias in estimates of receipt prevalence rates. In practice, survey data analysts are also interested in many other types of estimates. The effects of DI on estimation of spell lengths, their determinants and duration dependence were examined in Jäckle (2008b). A particularly important and pervasive manifestation of measurement error for panel analysts is seam bias. Seam bias is the tendency for transitions to be observed at the 'seam' between two reporting periods from successive waves of a panel survey. The effects of DI on seam bias were discussed by Jäckle and Lynn (2007) and Moore *et al.* (2009). Other aspects of benefit receipt are also of interest to analysts, notably the monetary amounts received. The effects of DI on these aspects warrant investigation.

## Appendix A: Source of income questions

### A.1.  *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".

The code is entered for each that applies. The question is 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 or national insurance 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 . . . ?*

## A.2.   Reactive dependent interviewing

For RDI, the independent questions are asked, as in Appendix A.1, followed by the following question, 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>?

## A.3.   *Proactive dependent interviewing*

For PDI, for *each* source of income 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) the following question is asked.

> 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 the following question is asked, starting with 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 takes place for each of cards 2, 3 and 4 in turn (excluding income sources 41 and 59 from the initial proactive question).

## References

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

Bradburn, N. M., Huttenlocher, J. and Hedges, L. (1994) Telescoping and temporal memory. In *Autobiographical Memory and the Validity of Retrospective Reports* (eds N. Schwarz and S. Sudman), ch. 13. New York: Springer.

DeMaio, T. J. (1984) Social desirability and survey measurement: a review. In *Surveying Subjective Phenomena*, vol. 2 (eds C. F. Turner and E. Martin), pp. 257–282. New York: Russell Sage Foundation.

Department for Work and Pensions (2004a) Quarterly bulletin on the population of working age on key benefits. Department for Work and Pensions, Newcastle-upon-Tyne. (Available from `http://research.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. Department for Work and Pensions, Newcastle-upon-Tyne. (Available from `http://research.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. Department for Work and Pensions, Newcastle-upon-Tyne. (Available from `http://research.dwp.gov.uk/asd/asd1/cga_famchild/CGA_FC_May04.pdf`.)

Department for Work and Pensions (2004d) Benefit expenditure tables. Department for Work and Pensions, Newcastle-upon-Tyne. (Available from `http://research.dwp.gov.uk/asd/asd4/medium_term.asp`.)

Department for Work and Pensions (2009) Households below average income: an analysis of the income distribution 1994/95 to 2007/08. Department for Work and Pensions, London. (Available from `http://research.dwp.gov.uk/asd/hbai/hbai2008/pdf_files/full_hbai09.pdf`.)

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. *Research Paper 95-07*. Statistics Canada, Ottawa.

Hancock, R. and Barker, G. (2005) The quality of social security benefit data in the British Family Resources Survey: implications for investigating income support take-up by pensioners. *J. R. Statist. Soc.* A, **168**, 63–82.

Jäckle, A. (2008a) Dependent interviewing: effects on respondent burden and efficiency of data collection. *J. Off. Statist.*, **24**, 1–21.

Jäckle, A. (2008b) Measurement error and data collection methods: effects on estimates from event history data. *Working Paper 2008-13*. Institute for Social and Economic Research, University of Essex, Colchester. (Available from `http://www.iser.essex.ac.uk/publications/working-papers/iser/2008-13`.)

Jäckle, A. (2009) Dependent interviewing: a framework and application to current research. In *Methodology of Longitudinal Surveys* (ed. P. Lynn), pp. 93–111. Chichester: Wiley.

Jäckle, A. and Lynn, P. (2007) Dependent interviewing and seam effects in work history data. *J. Off. Statist.*, **23**, 529–552.

Jäckle, A., Sala, E., Jenkins, S. P. and Lynn, P. (2004) Validation of survey data on income and employment: the ISMIE experience. *Working Paper 2004-14*. Institute for Social and Economic Research, University of Essex, Colchester. (Available from `http://www.iser.essex.ac.uk/publications/working-papers/iser/2004-14`.)

Jäckle, A., Uhrig, S. C. N. and Laurie, H. (2007) The introduction of dependent interviewing on the British Household Panel Survey. *Working Paper 2007-07*. Institute for Social and Economic Research, University of Essex, Colchester. (Available from `http://www.iser.essex.ac.uk/publications/working-papers/iser/2007-07`.)

Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A. and Sala, E. (2006) Patterns of consent: evidence from a general household survey. *J. R. Statist. Soc.* A, **169**, 701–722.

Jenkins, S. P., Lynn, P., Jäckle, A. and Sala, E. (2008) The feasibility of linking household survey and administrative record data: new evidence for Britain. *Int. J. Socl Res. Methodol.*, **11**, 29–43.

Kalton, G. and Miller, M. E. (1991) The seam effect with social security income in the Survey of Income and Programme Participation. *J. Off. Statist.*, **7**, 235–245.

Lynn, P. (ed.) (2006) *Quality Profile: British Household Panel Survey, Version 2.0: Waves 1-13: 1991-2003*. Colchester: University of Essex. (Available from `http://www.iser.essex.ac.uk/survey/bhps/quality-profile`.)

Lynn, P., Jäckle, A., Jenkins, S. P. and Sala, E. (2006) The effects of dependent interviewing on responses to questions on income sources. *J. Off. Statist.*, **22**, 357–384.

Mathiowetz, N. A. and McGonagle, K. A. (2000) An assessment of the current state of dependent interviewing in household surveys. *J. Off. Statist.*, **16**, 401–418.

Moore, J., Bates, N., Pascale, J. and Okon, A. (2009) Tackling seam bias through questionnaire design. In *Methodology of Longitudinal Surveys* (ed. P. Lynn), pp. 73–92. Chichester: Wiley.

Neter, J. and Waksberg, J. (1964) A study of response errors in expenditure data from household interviews. *J. Am. Statist. Ass.*, **59**, 18–55.

Tourangeau, R., Rips, L. J. and Rasinski, K. A. (2000) *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Tourangeau, R. and Smith, T. W. (1996) Asking sensitive questions: the impact of data collection mode, question format and question context. *Publ. Opin. Q.*, **60**, 275–304.

Young, N. (1989) Wave-seam effects in the SIPP. *Proc. Surv. Res. Meth. Sect. Am. Statist. Ass.*, 393–398.