

# Validating Survey Data: Experiences Using Employer Records and Government Benefit (Transfer) Data in the UK

Annette Jäckle, Emanuela Sala, Stephen P. Jenkins and Peter Lynn  
Institute for Social and Economic Research, University of Essex, UK

## 1 Introduction

Validation studies of survey data are typically limited to a very small number of survey items, to cross-sectional estimates, and to particular sub-populations for which access to records happens to be available.<sup>1</sup> We report here on a validation study carried out in the UK in 2003 with large numbers of validated items, longitudinal data, and based upon a large national sample. However, the representativeness of the validation sample obtained depends on the co-operation of both survey respondents and the providers of validation data and on error in the matching process. In the UK, matching survey data with administrative records is not common practice.

In this paper, we investigate several aspects of the feasibility of validation studies, drawing on experiences from a project on 'Improving survey measurement of income and employment' (ISMIE), funded by the Economic and Social Research Council (ESRC) Research Methods Programme. We focus on the validation of income and employment data. Two validation sources were used: Department for Work and Pensions<sup>2</sup> (DWP) benefit data and employers' records. The former provided histories of benefit receipt<sup>3</sup> and tax credits (e.g. child / disability / housing / unemployment benefits, pensions and income support). The latter provided information on occupation and employment status, gross and net pay, membership of company pension schemes and industry sector.

In the survey interview, respondents were asked for written permission both to obtain their DWP records and to contact their employer. They were also asked to provide information that would facilitate the validation: National Insurance number (equivalent to the US Social Security Number) and employer contact details. Subsequently, DWP records were extracted using a non-hierarchical matching strategy, based on different combinations of identifying variables obtained in the survey (National Insurance number, sex, date of birth, name and postcode), and a survey of employers was carried out (mail, with telephone follow-up).

The design of the validation studies is presented in section 2, starting with the characteristics of the survey sample (2.1), the nature of the DWP data and method used for matching (2.2) and the design of the employer

survey (2.3). In Section 3 we report permission rates and proportions providing matching items for the DWP linkage (3.1) and the employer survey (3.2). Section 4 documents results in terms of match rates for the DWP data (4.1) and response rates to the employer survey (4.2). Section 5 analyses the representativeness of the resulting validation samples, identifying correlates of the measures of success at each stage of the validation process in terms of substantive characteristics of the survey respondents. Section 6 concludes.

## 2 Study design

### 2.1 Household survey

The validation studies use data from the 'low income' subsample of the UK part of the European Community Household Panel Survey (ECHP). This sample was interviewed annually from 1994 to 2001 and since 1997 jointly with the British Household Panel Survey (BHPS) activities. Funding for the ECHP expired in 2001, giving us the opportunity to interview wave eight respondents once more in early 2003 for purely methodological purposes. CAPI interviews were sought with 1,163 individuals in 781 households, of which 1,033 (88.8%) were successfully interviewed.<sup>4</sup> Validation checks were carried out, and the data were fully coded and edited using standard procedures used in the main BHPS survey, so as to provide a realistic test.

The CAPI interviews were based on the BHPS wave 12 (2002 survey) household and individual questionnaires.<sup>5</sup> These collect information about accommodation, tenancy, housing problems, household consumption, demographics and neighbourhoods, health and caring, employment (histories), values and opinions and household finances. Efforts were made to maintain the context of the interview, although both the household and individual questionnaires were shortened by removing sections or questions not needed for the project.<sup>6</sup>

New questions were added to ask respondents for permission to contact their employers and to pass details on to the DWP for matching with administrative records. Employees who had given permission were then asked to provide contact details for their employers. To aid the

<sup>1</sup> See Bound et al. (2001) for an extensive review of validation studies.

<sup>2</sup> The Department for Work and Pensions is the government department in charge of administering benefits in the UK.

<sup>3</sup> The range of cash benefits in the UK largely exceeds those available in the US.

<sup>4</sup> Interviews were sought with all wave eight respondents. Wave eight non-respondents were not interviewed; eligible movers were followed to their new address where possible.

<sup>5</sup> Documentation for the BHPS is accessible at <http://iserwww.essex.ac.uk/ulsc/bhps/doc/index.html>.

<sup>6</sup> For details about changes compared to the main BHPS questionnaire, see Jäckle et al. (2004).

linkage with DWP data, consenting respondents were asked for their National Insurance number (NINO). This was partly to help with data matching, but also to test the feasibility of collecting NINOs and the accuracy of the NINOs provided. Finally, respondents were asked to sign a consent form for each of the validation studies.

It should be noted that, while the ECHP low-income sample is not a representative cross-sectional sample, it represents a wide range of population subgroups and should therefore allow considerable generalisation of methodological findings. It also has the advantage that particular groups of interest, such as those in receipt of benefit income, are over-represented. Additionally, the prior existence of eight waves of data provides very powerful auxiliary data for the analysis of measurement error. In particular, it allows us to investigate the impact of measurement error on estimates of change over time (between waves 8 and 9).

## 2.2 Benefit record linkage

The following section describes the process of matching the survey data with administrative records on benefit receipt held by the DWP. The design was a ‘complete’ record check: validation data was obtained for all respondents regardless of whether or not they had reported benefit receipt.

Respondents were linked to three principal sources of benefit data, held by the Information Centre of the Information and Analysis Directorate (IAD) of DWP: (1) Generalised Matching Service Data (Primary Data, Benefit Details, Benefit History), (2) Housing Benefit Details and (3) Tax Credit Data. These records contain information about current and historic receipt of 17 types of state benefits, including child benefit, housing benefit, working families’ credit, different types of disability allowances, income support, job seekers’ allowance and state pensions. The Primary Data was the key for the linkage, containing only one record per current or historic recipient, with the most up-to-date identifying information known to the DWP.

The variables used for matching were NINO, sex, date of birth, surname, first name, postcode and first line of address. While the NINO was collected in the ISMIE interview, information on the remaining variables stemmed from the sample information held by ISER.<sup>7</sup>

In order to minimise the impact of potential errors in the matching variables (in both the survey data and administrative records) on the success of linking respondents, a non-hierarchical matching strategy was used. The validation sample was matched five times with the Primary Data to obtain an accurate NINO for each of the sample members. Exact matches with the following combinations of variables were used: (1) NINO (without

suffix), (2) Sex, Date of Birth and Postcode, (3) Surname, First Name, Sex and Date of Birth, (4) Surname, First Name, Sex and Postcode, (5) Surname, First Name, Sex and First Line of Address.

The NINOs obtained from the Primary Data were then used as a key for matching with the remainder DWP records. For sample individuals who were matched to more than one person in the Primary Data, (i.e. the different matching criteria produced more than one possible NINO) a separate linking exercise was done for each of the NINOs. We then (as part of the analysis) judged which were ‘true’ matches, by pooling information from variables summarising the different selection criteria and other checks. The inclusion of multiple potential matches enables an assessment of how well each of the criteria operated in practice, in terms of the number and quality of matches (see Jenkins et al., 2004b).

## 2.3 Employer survey

The survey data on employment were validated using records held by employers. Since the sample contained employees from the whole of the UK, obtaining direct access to records was not possible. Instead, employers were contacted and asked to provide information for the period corresponding to the survey interview. To this end, a survey of employers was designed and carried out in several stages. Employers were first sent a questionnaire by post, followed by a reminder letter and eventually a second questionnaire. Employers who had not replied or refused by this stage were then followed-up by telephone.

The mail questionnaire contained a subset of the questions on the respondent’s employment situation used in the household survey. To ensure comparability, the original format of questions was maintained, although the wording was adapted to address the employer rather than the employee. The aspects covered include information on the employer (industry, plant size), job characteristics (occupation, employee/self-employed, managerial duties, usual working hours, working hours arrangements) and income (last gross/net pay, hourly rates of pay, rates for overtime, pension schemes). Each questionnaire was personalised in the sense that the questions referred to the employee by name, and the introduction mentioned the date of the ISMIE interview as a reference period for the information requested.

## 3 Permission and match items

How did respondents react to the validation studies? Of the 1,033 respondents asked for permission to match their data to records held by the DWP, 77.4% consented. These respondents were in turn asked to provide their NINO to facilitate the matching. 88.6% gave their number, 1.5% refused, and 9.9% answered that they did not know it. All employees (434 respondents) were

<sup>7</sup> Since the sample information is verified and updated with every wave of the survey, the quality of these variables is likely to be better than if they had been collected solely during the last wave of the survey.

further asked for permission to contact their employers about their employment situation. 58.5% allowed us to do so and all, except for one, provided contact details for their employers.<sup>8</sup>

Comparing the permission rates for the two studies suggests that respondents are more concerned about flows of information between the survey organisation and their employers, than about third parties accessing their administrative records.<sup>9</sup> Sections 3.1 and 3.2 provide further information about the quality of matching items collected for each of the validation studies.

### 3.1 Collecting National Insurance numbers (NINOs)

NINOs in the UK consist of two letters, followed by a six-digit number and a suffix letter. These three components were entered separately into CAPI by the interviewer. Since NINOs are unique without the suffix, the record linkage was done without this last component. The interviewer also recorded the source of reported NINOs. Most respondents (67.4%) consulted a payslip or other document in order to retrieve their NINO, 30.8% recalled theirs from memory and were positive it was correct, and a mere 1.8% relied on their memory, although they were not sure they could trust it.

In 98.9% of cases, all components of reported NINOs were complete and took plausible values. Only in one case is the middle number missing, while in a further seven cases it consists of six nines. In the BHPS CAPI script, nines are used as a code for 'don't know'. Here this option was not given, however, the use of 999999 looks as though the interviewer had intended to use this code. Looking at the NINO by source shows curious effects: the respondents who recalled their number from memory but were not sure it was correct have all reported complete and plausible components. That is, the missing and questionable values are from respondents who either checked a document, or were sure they remembered correctly. In the cases where NINOs were verified, this points to keying errors on the part of the interviewer.

### 3.2 Employer contact information

Consenting employees were asked to provide the name of a contact person or department which would have information about their employment situation, as well as the contact's complete postal address and telephone number. Collecting good quality contact data turned out to be difficult. As Table 1 shows, all but one employee reported the name of the firm. 61.0%

pinpointed a particular contact person, in most cases (94.2%) specifying both the name and surname. A further 29.1% indicated the Personnel Department, Human Resources or Payroll Office as the point of contact. Finally, 9.8% neither indicated a contact person nor a department. In these cases the questionnaire was addressed to Human Resources.

As far as the address information is concerned, 23.2% of respondents provided both street names and numbers. In most cases (69.3%), however, the information was incomplete: a street name without a number or institutions/industrial estates without street name, or no address information at all (7.5%). Complete postcodes were reported by 43.7%. The remaining employees reported three to five characters of the postcode (17.7%), one to two characters (17.3%), or no postcode at all (21.3%). Finally, complete telephone numbers were given by 81.5%, while 3.5% provided incomplete numbers, and 15.0% did not state a telephone number. Most addresses were therefore incomplete and had to be supplemented.

## 4 Match rates

Having described the co-operation by respondents and their reactions to the request for additional (sensitive) matching information, the following section reports outcomes in terms of match rates for the DWP record linkage and response rates for the employer survey.

### 4.1 Benefit record match rates

Only persons who had received some state benefit since 1999<sup>10</sup> would be represented in the DWP data. Thus, a perfectly successful match would result in a match rate not equal to 100% but equal to the proportion of persons meeting this criterion. This must be borne in mind when considering the results of the match process. Furthermore, this proportion is not known for the study population.<sup>11</sup> The DWP successfully matched 589 (73.7%) of the 799 consenting respondents<sup>12</sup> with the Primary Data, using the five matching criteria outlined in section 2.2. This produced 604 unique NINOs, including 15 cases where two NINOs were obtained for one sample person. 210 cases (26.3%) could not be matched with the Primary Data. The resulting 604 NINOs were then linked

<sup>10</sup> The date since when records were included in the data base varied over the benefits, but for most it was January 1999.

<sup>11</sup> 59% of benefit units in Great Britain received one or more types of benefit in 1999-2000 (a benefit unit is defined as "a single adult or couple living as married and any dependent children") (Ellerd-Elliott et al., 2001). However, it is difficult to translate this figure for benefit units to one for persons, from annual to receipt over 4 years, and from the total population to the "low income" population sampled for ISMIE.

<sup>12</sup> An early version of the data sent to the DWP contained 802 respondents who had given permission. However, for 3 respondents valid survey data could not be derived subsequently.

<sup>8</sup> Respondents' propensity to consent and provide matching information for the two validation studies is analysed in depth by Jenkins et al. (2004a).

<sup>9</sup> See also Singer's (2003) investigation of informed consent, from which she concludes that the respondents' decisions to participate in research are rational, based on perceived risks and benefits.

with the Benefit Details, Benefit History, Housing Benefit and Tax Credit data to obtain details of all DWP benefits claimed by these individuals. The matched data contains 7,615 observations and covers records held for the period 1999 to 2003. The match rates obtained for the five different matching criteria are shown in Table 2. At first glance, matching based on sex, date of birth and postcode (match criterion 2) appears to yield slightly more matches than did a match on NINO (64.1% of consenters compared to 62.0%). A match on sex, date of birth, forename and family name (match criterion 3) linked to only slightly fewer cases than a match on NINO (61.7% compared to 62.0%). Matches on criteria 4 and 5 do noticeably worse, especially the latter.

Among respondents who consented to match and who supplied a NINO, there are clearer differences in linkage rates between criteria: the NINO linkage rate is distinctly higher (69.9%) than the next best (criterion 2: 64.4%). In other words, the reason why linkage rates for criteria 2 and 3 among the sample are much the same as the NINO linkage rate is because they have the potential to link to respondents who did not supply a NINO.

The matched information can be used to get an idea of the accuracy of the NINOs reported in the survey (excluding the suffix, since this was not used for matching). For the 708 respondents who reported a NINO, Table 3 examines whether the probability of achieving a match with the DWP records varied by the source of the reported number. The table distinguishes respondents by the outcome of the linkage: (1) respondents, for whom no match was achieved (25.6%), (2) those for which record linkage was successful and where the NINOs in both the survey report and the DWP records corresponded (69.9%), and (3) successful matches, but with non-corresponding NINOs (4.5%). The probability of achieving a match with a corresponding NINO does not vary with the retrieval strategy used by the respondent. However, since the numbers reported in the survey or recorded by the DWP may be erroneous, some matches with corresponding numbers may not correspond on other identifying variables. The quality of matches achieved by the different combinations of matching variables is analysed in Jenkins et al. (2004b).

#### 4.2 Response to the employer survey

Of the 254 employees who gave permission to contact their employer, one did not provide any contact information. The issued sample size for the employer survey therefore was 253. In total, 72.3% of the questionnaires were completed, of which 51.4% at the postal stage (23.7% after the first mailing, and the remainder 20.9% after the second or third contacts). A further 20.9% were completed after having been followed-up by telephone. Refusals occurred both at the postal stage (13.4%) and the telephone stage (9.1%). No contact was made with a further 5.1%. The main reasons for non-response were concerns about confidentiality and

company policies prohibiting the release of information about employees. Indeed, many employers would not provide any information without the consent form signed by the employee.

The design of the employer survey to include several contacts by mail, followed-up by telephone proved a successful combination. More than half the questionnaires completed at the postal stage were only returned after the second or third contact. The telephone stage was crucial in making contact with employers who had not received the previous mailings. However, many employers were reluctant to provide information about employees over the phone, so postal questionnaires had to be faxed or sent again, once contact had been established by telephone. The mailings were complicated by the limited quality of contact details collected in the survey. On the other hand, the telephone stage proved more time consuming to administer, since it often took many calls and additional faxes or letters until a questionnaire was completed (see Lynn & Sala (2004) for details).

## 5 Bias of validation samples

To complete the discussion of respondent consent, record linkage and co-operation by employers, the following section investigates the implications of these hurdles for the representativeness of the validation samples. Correlates of success at each stage of the validation process are identified, in terms of substantive characteristics of respondents.

### 5.1 Benefit validation sample

Table 4 compares respondents who gave permission for the DWP linkage with those who did not consent. The results indicate that the two groups are comparable in terms of composition by gender, age groups, marital status, qualifications, economic activity, earnings, housing tenure and receipt of benefits recorded in the survey. Using multivariate analysis to investigate the consent patterns Jenkins et al. (2004a) do, however, find evidence of consent bias.

For the record linkage it is not possible to identify bias due to non-matches. We cannot distinguish whether non-matched cases are (1) respondents with benefit records but who could not be matched due to errors (or missing values) in the matching variables, or (2) individuals who genuinely did not receive any benefits during the observation period, and so have no DWP record to be matched to.

Nonetheless, bias added at this stage is likely to be small. The quality of matching information was similar for matched and non-matched respondents. Indeed, 86% of the non-matched cases had supplied a complete and plausible NINO.

## 5.2 Employment validation sample

The characteristics of the validation sample for employment data are presented in Table 5. Comparing employees who gave permission for this study with those who did not, again does not show any significant differences in terms of composition of the groups by gender, age, education, marital status, sector, size of employer's organisation, type of occupation, hours of work and net pay. Jenkins et al. (2004a) take account of selection problems (incidental truncation), caused because only respondents in employment were asked the employer consent question and again find some evidence of consent bias.

At the second stage, the survey of employers, there are no differences in characteristics between employees for whom the questionnaire was returned and those for whom it was not. However, the size of the employing organisation (in terms of numbers of workers) emerges as a significant determinant of success. Larger organisations were much more likely to complete the survey: the questionnaire was returned for 78% of employees working in organisations with 100 or more employees, while only 60% of organisations with less than 25 employees co-operated. Finally, looking at the effective validation sample as a proportion of all employed respondents (regardless of whether or not they gave permission) shows the same results.

## 6 Conclusions

This section provides a summary comparison of the different issues arising in the collection of validation data from employer records and DWP benefit data, focusing on how the representativeness of the effective validation samples depends on the processes of obtaining permission from respondents, gaining access to validation data, and matching survey and validation data. Table 6 reports the samples obtained at the different stages.

The validation sample obtained for the employment data (182 employees) is much smaller than that obtained for the benefit data (589 recipients). This is due, first of all, to the small number of employees in the ECHP low-income sample (434). Secondly, respondents seemed to be more reluctant to give permission to contact their employers (58.5%) than to match their data to records held by the DWP (77.4%), reducing the potential validation samples to 254 employees and 799 benefit recipients. For both studies, it seems that respondents who gave permission to do the validation exercise were happy to provide the necessary matching information. All but one of the employees provided contact details for their employers; 88.6% of those who gave permission to do the DWP matching also provided their NINO, while a further 9.9% said they did not remember their number. Only 1.5% explicitly refused to report their NINO. Respondents' propensity to consent to the validation

studies and provide matching information is analysed in detail by Jenkins et al. (2004a).

Collecting NINOs in the survey therefore proved feasible. Nearly all reported numbers were complete and plausible. 93.9% of matched respondents who supplied a NINO were matched to a DWP record with the same NINO. Only three of these were identified as definite mis-matches. On the other hand, the interviewer check question about the source of the reported NINO does not seem to provide reliable information about the quality of the NINOs collected. Indeed, the main source of errors in the NINOs appears to be typos on the part of the interviewer (see Jenkins et al., 2004b) for an in-depth analysis of the NINOs reported in the survey).

The next stage, gaining access to the validation data, was easy for the benefit records, since they are all held by the DWP. For the employment information, we depended on the co-operation of nearly 253 employers in completing the survey ('nearly' because there are some respondents, particularly couples, who work for the same firm). Non-response by employers reduced the size of the final validation sample to 71.7% of consenting employees – or 41.9% of employees in the respondent sample.

Finally, matching survey and validation data was straightforward for the employment data, but critical for the DWP records, depending largely on the quality of matching variables and the matching strategy employed. 73.7% of the permission sample (57.0% of all respondents) were matched to their benefit records. For the non-matched cases, however, it is impossible to know whether they were 'true non-matches' (respondents who had never received a benefit), or 'false non-matches' (respondents who should have been matched, but could not be due to errors in the matching variables). Jenkins et al. (2004b) analyse the quality of matches and implications of using different combinations of matching variables for the success of matching survey and administrative data.

The first hurdle of obtaining permission did not introduce bias in terms of key substantive (univariate) characteristics of respondents in either of the validation studies. However, gaining co-operation from employers introduced some bias as far as characteristics of the data holder (employer) are concerned.

## References

- Bound, J., Brown, C., Mathiowetz, N. (2001) Measurement Error in Survey Data, in: Heckman, J.J., Leamer, E., *Handbook of Econometrics*, Vol. 5, 3705-3843.
- Ellerd-Elliott, S., Horsfall, El, Butt, N., Dhanecha, N., Pickering, E. (2001) Family Resources Survey Great Britain 1999-2000, DSS.
- Jäckle, A., Sala, E., Jenkins, S.P., Lynn, P. (2004) Validation of Survey Data on Income and

Employment: the ISMIE Experience, ISER Working Paper No. 2004-14. Colchester: University of Essex.  
 Jenkins, S.P., Cappellari, L., Lynn, P., Jäckle, A. Sala, E. (2004a, forthcoming) Patterns of respondent consent: evidence from a general household survey.  
 Jenkins, S.P., Lynn, P., Jäckle, A. Sala, E. (2004b) Linking household survey and administrative record data: what should the matching variables be? ISER Working Paper No. 2004-23. Colchester: University of Essex.

Lynn, P., Sala, E. (2004) The contact and response process in business surveys: lessons from a multimode survey of employers in the UK, ISER Working Paper No. 2004-12. Colchester: University of Essex.  
 Singer, E. (2003) Exploring the Meaning of Consent: Participation in Research and Beliefs about Risks and Benefits, *Journal of Official Statistics*, Vol. 19 (3), 273-285.

**Table 1: Completeness of employer contact information**

Address components reported	Frequency	Percent
Total	254	100.0
Name of firm	253	99.6
Contact		
Person	155	61.0
Department	74	29.1
Missing	25	9.8
Address		
Complete (street and number)	59	23.2
Partial (street without number)	108	42.5
Others *	68	26.8
Missing	19	7.5
Postcodes		
Complete (6 or 7 characters)	111	43.7
Incomplete (3–5 characters)	45	17.7
Incomplete (1–2 characters)	44	17.3
Missing	54	21.3
Telephone number		
Complete	207	81.5
Incomplete	9	3.5
Missing	38	15.0

\* Others refer to institutions such as schools, council hall, industrial estates.

**Table 2: Record linkage rates (%), by criterion**

Criterion and matching variable(s)	ISMIE sample	All who gave consent to match	Those matched on at least one criterion	Supplied NINO	Did not supply NINO
1. NINO	47.9	62.0	84.0	69.9	–
2. Sex, date of birth, postcode	49.6	64.1	86.9	64.4	61.5
3. Sex, date of birth, forename, family name	47.7	61.7	83.7	62.6	55.0
4. Sex, postcode, forename, family name	41.6	53.8	73.0	54.5	48.4
5. Sex, forename, family name, address line 1	33.6	43.4	58.9	44.4	36.3
At least one of the above	57.0	73.7	100	74.4	68.1
Sample N	1033	799	589	708	91
(as % of all who gave consent)		(100)	(73.7)	(88.6)	(11.4)

**Table 3: Outcome of matching by source of National Insurance Number**

Source of NINO reported in survey	Outcome of survey and record linkage (row%)			Total
	No match	Match: same NINO	Match: diff. NINO	
NINO Taken from payslip or other document	123 (25.8)	330 (69.2)	24 (5.0)	477 (100.0)
NINO Remembered: sure correct	56 (25.7)	156 (71.6)	6 (2.8)	218 (100.0)
NINO Remembered: not sure	2 (15.4)	9 (69.2)	2 (15.4)	13 (100.0)
Total	181 (25.6)	495 (69.9)	32 (4.5)	708 (100.0)

**Table 4: Characteristics of respondents by permission for benefit validation**

Characteristics	Total (Frequency)	Permission for DWP linkage * (row %)
All	1,033	77.4
Sex		
Male	429	77.6
Female	604	77.2
Age		
16–35	242	78.9
36–50	292	73.6
51–65	216	77.3
66 +	283	79.9
Marital status		
Married/widowed	600	77.8
Separated/divorced	164	76.8
Never married	268	76.9
Highest academic qualification <sup>1</sup>		
Any qualifications listed	384	75.0
None of these	647	79.0
Economic activity		
(Self-) employed	489	76.7
ILO unemployed	27	63.0
Econ. inactive	517	78.7
Total pay		
£0	544	77.9
£1–299	148	76.4
£300–899	145	81.4
£900 +	128	78.1
Housing tenure		
Owned or mortgage	411	78.4
Rented	575	77.9
Rent free/other	31	64.5
Benefit receipt <sup>2</sup>		
NI retirement pension	317	80.8
Incapacity benefit	77	79.2
Income support	179	81.0
Job seeker's allowance	35	77.1
Child benefit	208	77.9
Family credit	94	80.9
Housing benefit	273	78.4
Council tax benefit	323	77.7

\* Differences in characteristics between consenting and non-consenting respondents are tested using a two-tailed Pearson Chi2 test. At the 5%-level none of the differences are significant.

<sup>1</sup> Listed qualifications include youth training certificates, apprenticeships, clerical and commercial qualifications, nursing qualifications, teaching qualifications, university diploma, degree, higher degree.

<sup>2</sup> For all other benefit types recorded in the survey, differences are not significant either.

**Table 5: Characteristics of employees by consent and completion of employer survey**

Characteristics		Total (Frequency)	Employees given permission (row %)	Employer survey completed (row %)	Effective validation sample (row %)
All		434	58.5	71.3	41.7
Sex	Male	190	56.3	74.8	42.1
	Female	244	60.2	68.7	41.4
Age	16–35	152	53.3	71.6	38.2
	36–50	181	63.0	71.1	44.8
	51 +	101	58.4	71.2	41.6
Education	Any qualification listed	223	59.2	73.5	43.5
	None of these	210	58.1	68.9	40.0
Marital status	Married / widowed	209	60.3	68.3	41.2
	Separated / divorced	68	61.8	81.0	50.0
	Never married	157	54.8	70.9	38.9
Sector	Private company	314	55.7	70.9	39.5
	Civil service.	120	65.8	72.2	47.5
Size of organisation	< 25	159	51.6	59.8 **	30.8 **
	25–99	111	63.1	75.7 **	47.8 **
	100 +	161	62.1	78.0 **	48.5 **
Type of occupation	Manager, administrator, professional	54	57.4	83.9	48.2
	All other	375	58.4	69.9	40.8
Hours of work	< 25	111	54.1	65.0	35.1
	25–36	95	68.4	73.9	50.5
	37–40	171	56.7	75.3	42.7
	41 +	53	58.5	67.7	39.6
Take home pay	< £300	148	53.4	65.8	35.1
	£300–899	144	66.7	70.8	47.2
	£900 +	128	59.4	79.0	46.9

\* Differences in characteristics between groups are tested using a two-tailed Pearson Chi2 test. None of the differences are significant at the 5% -level, except for those marked (\*\*) which are significant at the 1% -level.

<sup>1</sup> Listed qualifications include youth training certificates, apprenticeships, clerical and commercial qualifications, nursing qualifications, teaching qualifications, university diploma, degree, higher degree.

**Table 6: Samples obtained for the employer and benefit record checks**

Samples	Employer survey	DWP record check
Respondent sample	434 (employees)	1,033 (all respondents)
Permission sample	254	799
% of respondent sample	58.5%	77.4%
Matching information given	253 (employer's address)	708 (NINO)
% of permission sample	99.6%	88.6%
% of respondent sample	58.3%	68.5%
Validation sample	182	589
% of permission sample	71.7%	73.7%
% of respondent sample	41.9%	57.0%