A COMPARISON OF INTERVIEWER AND OFFICE CODING OF OCCUPATIONS

Jean Martin, Diane Bushnell, Office of Population Censuses and Surveys
Pamela Campanelli, Roger Thomas, Survey Methods Centre, Social and Community Planning Research
Jean Martin, Social Survey Division, Office of Population Censuses and Surveys,
St Catherine's House, 10 Kingsway, London WC2B 6JP, UK

Key Words: Occupation coding, Coder variance, Interviewer coding

1. Introduction

Coding information about respondents' occupations in order to derive socio-economic classifications is an expensive and potentially error-prone process. Since such classifications are required on almost all large-scale surveys, there is considerable interest in the most cost effective methods of collecting and coding the relevant information. In Great Britain manual methods are used for coding occupation on all major government surveys. Computer-assisted methods are being evaluated and are likely to offer improvements in the longer term.

In this paper we consider two alternative approaches to manual coding: office based specialist coders; and interviewers carrying out coding at home after the interview. We describe a split sample experiment to compare the two methods with respect to levels of coding reliability and accuracy, and estimate the impact of correlated coder variance on survey estimates. This study of manual methods provides baseline information against which computer-assisted methods can be judged in the future.

2. Deriving socio-economic classifications

In the UK there are two main socio-economic classifications used on government and many other major social surveys: Social Class based on Occupation (SC) and Socio-economic Group (SEG). They were developed by the Office of Population Censuses and Surveys (OPCS) and are based on the Standard Occupational Classification (SOC) (OPCS, 1990).

Both classifications require information about a respondent's occupation and some details about employment status (whether self-employed; whether a manager or a supervisor or other employee). In addition the number of employees in the organisation is required to derive SEG. The employment status questions are normally precoded but the respondent's job title and a description of the work done is recorded verbatim in the interview for later coding to one of 371 three digit codes of the SOC (hereafter called SOC codes). The SOC code together with the other employment status information is used to derive SC and SEG via a look-up table. Not all combinations of SOC codes and employment status information are permissable so procedures have been developed to deal with conflicting information.

The most difficult and time consuming part of the process is coding the occupation details to a SOC code. There has therefore been considerable interest in making the process cheaper and faster, without sacrificing the quality of the coding operation. The traditional approach has been for office-based specialist coders to code the verbatim description to a SOC code using the detailed coding manuals provided by OPCS. Survey organisations using this approach generally have a small number of coders specially trained in occupation coding so all cases requiring occupation coding would pass through their hands.

The advantages of this approach is that coders can be trained and supervised so as to maximise accuracy and consistency of coding. The disadvantage is that this provides an opportunity to introduce bias and correlated coder variance if a group of coders systematically deviates from the 'correct' codes. In addition such coding is expensive and may cause a bottleneck in the survey processing timetable.

In the early 1980s OPCS (the official government survey organisation which carries out most of the largest government surveys in Britain) began experimenting with training interviewers to code occupation at home after completing the interview and then sending the coded information back to headquarters. Early trials (e.g. White, 1983; Dodd, 1985) showed that although interviewers could not achieve the same levels of inter-coder consistency as office based coders, their smaller workloads meant that correlated coder variance had less impact on the precision of the results. Moreover, they learnt what occupation details were important to the coding process and became better at eliciting appropriate information in the interview. The use of this method also resulted in savings of time and money for OPCS and has therefore been in use on all OPCS surveys since the mid 1980s.

Other survey organisations with large numbers of interviewers have been concerned that the additional costs of training interviewers would largely swallow up savings in the costs of coding and that the quality of coding would suffer. Accordingly, Social and Community Planning Research (SCPR), an independent research institute carrying out many social surveys in Britain, including a number of major government surveys, has continued to use office based specialist coders to code occupation for these reasons.

The advent of computer assisted personal interviewing (CAPI) for most major surveys at both OPCS and SCPR has not had much effect on occupation coding procedures thus far: interviewers in both organisations enter details of occupations as verbatim text into their computers for later coding either by interviewers at home or by office coders.

3 The Family Resources Survey

This is a large continuous survey (25000 households per year) carried out since 1990 on behalf of the Department of Social Security covering households' income and financial affairs. It was developed as a CAPI survey from the outset. OPCS and SCPR share the fieldwork equally, assigning interviewer areas to each organisation on a random basis. Occupations are coded for all adults who were in paid work at the time of the interview or who had worked within the preceding 12 months. Thus some 30,000 occupations are coded annually - by interviewers at OPCS and by office coders at SCPR.

There is some evidence that the two organisations are not in fact achieving comparable results in terms of distributions on the two main socio-economic classifications. However, analysis of routinely collected survey data does not allow us to determine the reason for the differences so we carried out an experiment in which coders from both organisations independently coded the same occupation details. It also proved possible to have the sample of cases coded by an expert in occupation coding to provide a standard against which accuracy of coding could be judged.

4 Aims of the experimental study

i) To determine levels of coding reliability of SEG and social class achieved by office or field coders.

ii) To assess accuracy of coding by office and field coders using codes assigned by an expert occupation coder as the 'correct' codes.

iii) To determine levels of reliability for individual SEG and social class codes.

iv) To investigate the impact of correlated coder variance on the precision of the survey estimates.

5 Design of the experiment

A sample of 200 cases with relevant occupation details, 100 originally collected by OPCS and 100 by SCPR, was extracted from the CAPI interview records for the FRS. The number of coders taking part in the experiment was constrained by both cost and availability to five office coders from SCPR and eight field coders (interviewers) from OPCS. The robustness of estimates of intra-coder correlation coefficients (used to assess the impact of correlated coder variance) is affected by the number of coders employed in the study. All coders had been trained in occupation coding and had had experience of coding on a number of surveys. They were given instructions to code as they would normally, and office coders were asked specifically not to discuss problem cases with one another or with a supervisor.

The number of cases extracted was also largely dictated by cost although the sample had to be large enough to study coding at the level of aggregate classifications such as SC and SEG. 200 cases does not allow us to look at the reliability of coding individual SOC codes but the primary interest is in the social classifications which are used in analysis rather than the individual SOC codes which are used to derive them.

The relevant information was presented in the form of a short CAI program (written in Blaise, the CAI software used by both organisations) so that it would be in a form familiar to those doing the coding. Each coder entered the SOC code for each case. The program checked that a valid code number had been entered. Subsequently SC and SEG were derived from an electronic look-up table. Where the employment status was incompatible with the SOC code, the program employed a standard algorithm to determine the most appropriate SC and SEG.

In order to provide a measure of coding accuracy we arranged for the 200 cases to be coded by a member of OPCS Census Division responsible for the maintenance of the SOC who has considerable experience of occupation coding and extensive expertise in the SOC classification. She can therefore be seen as providing the "correct code" on the basis of the information supplied. This is not strictly a test of validity since the information recorded by interviewers does not necessarily represent a valid description of respondents' occupations.

6 Results

6.1 Reliability of field and office coders

In the course of a year OPCS uses around 250 interviewers on the FRS, each of whom is responsible for coding occupation on all cases where this is required, whereas SCPR uses only three coders to code an equivalent number of cases. Considering the wider organisational picture, OPCS has about 800 interviewers coding occupation on a number of surveys over the course of a year compared with around 20 coders employed by SCPR. Since it is more difficult to train and supervise larger numbers of coders to minimise individual variation we would expect greater coding reliability from office than field coders.

Previous studies have not shown much difference in reliability according to the length of the coding frame (eg. Kalton and Stowell, 1979; Collins and O'Brien, 1981; Elliot, 1982; Elliot, 1983). However there is evidence that amalgamating a frame after coding can lead to improvements in reliability (Kalton and Stowell, 1979). Distinctions which lead to different SOC codes do not necessarily matter at the level of SEG or SC so we would expect higher levels of reliability for these classifications than for SOC.

By *reliability* we mean the extent to which different coders assign the same code to the same case. A straightforward index of reliability is the **proportion of agreement** or $\bar{p}$ (see Kalton and Stowell 1979; and also Elliot, 1982). It can be estimated by computing the proportion of all the paired comparisons in which the two codings agree. $\bar{p}$ ranges from 1 if all the coders are in complete agreement on all questionnaires to 0 if none of the coders agree with each other on any of the questionnaires.

An alternative measure which takes chance into account is kappa, $\varkappa$. With a long coding frame like SOC at the three digit level, the probability of two coders assigning the same code by chance will be low, and it is usually sufficient to use $\bar{p}$ (Elliot, 1982). However, $\varkappa$ is required to calculate the inflation factors described in section 6.5.

Although ultimately we are interested in the social classifications, SC and SEG, the results for the detailed SOC codes from which they are derived are also of interest. Office coders were significantly more reliable than field coders ($\bar{p}$ = .82 and .74). There were only slight differences for SEG and SC: the overall levels of $\bar{p}$ were .92 and .91 for office coders and .88 and .87 for field coders, and all were higher than the values for SOC. Despite having a greater number of categories, coding to SEG was no less reliable than coding to SC. This is likely to be because most of the additional discrimination provided by SEG compared with SC depends on the greater use of employment status information in a relatively automatic manner rather than in more detailed use of occupation information.

6.2 Accuracy of field and office coders compared with expert coder

Previous studies suggested that interviewer coders would be less accurate than office coders. Even if interviewers are trained initially to achieve an acceptable level of accuracy with large numbers of interviewers it is difficult to ensure than this level is maintained year in, year out. Office coders have regular access to a supervisor and have more opportunity to discuss problem cases.

However, OPCS has maintained that an advantage of interviewer coding is that interviewers will be sensitive to the information needed to achieve successful coding and will thus adapt their probing of details of respondents' occupations in the light of their knowledge of what will be needed at the coding stage which will contribute to greater accuracy.

We used the codes assigned by the expert occupation coder as indicators of the 'correct' code for each case and for each coder calculated the proportion of cases in which they were in agreement with the expert.

As expected, the levels of agreement for SOC codes showed somewhat lower agreement for the field coders than the office coders (77% compared with 80%).

Despite expecting office coders to be generally more accurate than field coders, to our surprise we found no difference between the two groups with respect to SEG and SC. Both achieved high levels of agreement: 90% for SEG and 89% for SC for both groups.

6.3 Reliability of individual codes

With only 200 cases in the experiment we cannot examine the reliability of the 371 individual SOC codes. Our main interest is in SEG and SC codes, but even with these, some codes had fewer than 20 cases and these were excluded from the analyses reported in this section.

The overall reliability for individual codes can be measured by Q_i which is an index of the amount of agreement between coders on a particular code. (It has a counterpart $\varkappa_i$ which takes account of agreement which may have taken place by chance but which is not used here for the reasons given earlier.) $\bar{p}$ can be decomposed into the reliability of individual codes and is therefore the average of all the Q_i values.

Table 1 shows the reliability of each SEG and SC code separately for the office and field coders. With the exception of three of the SEG codes, reliability for the

Table 1 Reliability (Q_i) of individual codes

	Office coders	Field coders
Socio-economic group		
1.1 Employers:large estabs	*	*
1.2 Managers:large estabs	.94	.91
2.1 Employers:small estabs	.97	.94
2.2 Managers:small estabs	.94	.80
3 Professional: self-empl	*	.92
4 Professional: employee	.80	.74
5.1 Ancillary workers	.82	.87
5.2 Supervisors non-manual	.67	.60
6 Junior non-manual	.94	.91
7 Personal service	.95	.85
8 Supervisors manual	.97	.92
9 Skilled manual	.88	.90
10 Semi-skilled manual	.94	.92
11 Unskilled manual	.94	.83
12 Own-account (not profess)	.98	.99
13 Farmers: (empl'r & manager)	*	*
14 Farmers: own account	*	*
15 Agricultural workers	*	*
Social class		
I Professional	.85	.78
II Managerial & technical	.89	.87
IIIn Skilled non-manual	.91	.86
IIIm Skilled manual	.93	.90
IV Partly skilled	.94	.88
V Unskilled	.94	.83

* fewer than 20 cases

field coders was somewhat lower than for office coders.

6.4 Estimates of correlated component of coder variance

Coders can contribute two types of unreliability to survey estimates: random or haphazard errors in assigning codes; and correlated errors -- the effect of individual coders using the coding frame in a systematically different way from the 'average' coder. In the case of correlated errors, although the individual errors may be thought of as bias, the impact of these systematic differences is to add variance around survey estimates of that particular code category. The population variance around the proportions can consist of (at least) three elements: variance due to sample design, random coder variance and correlated coder variance. Note that the first two of these are typically confounded in estimates of variance around survey means and proportions. What is excluded that therefore leads to underestimates of the true variance is the correlated coder variance. The extent of correlated coder variability can be summarised by ϱ_w, a measure of the average intra-coder correlation -- the amount by which individual coders vary from an 'average' coder (Kalton and Stowell, 1979). The significance of correlated coder variance can be established using Cochran's Q test, and in some cases is estimated as a negative number. As the bounds of ϱ_w are between 0 and 1, these have been set to zero in the results which follow.

Table 2 Intra-coder correlations (ϱ_w x 1000)

	Office coders	Field coders
Socio-economic group		
1.1 Employers:large estabs	*	*
1.2 Managers:large estabs	0.7	3.3
2.1 Employers:small estabs	0.0	0.0
2.2 Managers:small estabs	3.1	2.9
3 Professional: self-empl	*	0.0
4 Professional: employee	-1.3	2.6
5.1 Ancillary workers	*8.3*	3.8
5.2 Supervisors non-manual	3.4	-1.0
6 Junior non-manual	-3.4	2.6
7 Personal service	-5.0	0.3
8 Supervisors manual	0.0	3.9
9 Skilled manual	-1.9	-1.2
10 Semi-skilled manual	-3.5	2.9
11 Unskilled manual	-3.4	-0.1
12 Own-account (not profess)	0.0	0.0
13 Farmers: (empl'r & manager)	*	*
14 Farmers: own account	*	*
15 Agricultural workers	*	*
Social class		
I Professional	-1.3	1.6
II Managerial & technical	6.6	2.1
IIIn Skilled non-manual	3.5	1.3
IIIm Skilled manual	-3.8	3.6
IV Partly skilled	-4.4	-0.9
V Unskilled	-3.4	-0.1

* fewer than 20 cases
Figures in italics indicate significant values

Table 2 gives values for ϱ_w for each of the SEG and SC codes with more than 20 cases, separately for the office and field coders. Out of all the categories tested only one (Ancillary workers and artists, SEG 5.1 coded by office coders) showed a significant difference in usage by coders, although SC II coded by office coders almost reached significance at the 5% level. Several others were also quite large although not statistically significant. Note that with the small number of coders and small sample sizes the standard errors of ϱ_w tend to be large. As the coders are coding to one of 371 SOC codes, and not directly to SEG or SC, a finding of large values for ϱ_w must mean that they are consistently assigning cases to particular SOC codes from which the SEGs or SCs which show correlated coder variance are derived. Rather than discuss these results further we therefore turn to estimates of the effect of correlated coder variance on survey estimates.

6.5 Effect of coder variance on survey estimates

Kalton and Stowell (1979) derive a factor, which we call F_c, which may be multiplied by the estimate of the variance calculated in the standard way to obtain a more accurate estimate of the true variance of the population proportion. This factor is a measure of how much the variance of the estimate will be inflated due to systematic error in coding to SOC.

As can be seen, ϱ_w and $\varkappa_i$ are used in calculating F_c:

$$F_c = 1 + (M - 1) * \varrho_w * (1 - \varkappa_i)$$

where M is the average number of questionnaires coded by each coder and i is the SEG or SC category.

What is also apparent is the importance of workload size in determining the size of the final inflation factor. SCPR normally uses three coders for all occupation coding on the FRS in one year while OPCS uses approximately 250 interviewers. With each organisation coding around 15000 occupations, the average workloads are 5000 and 60 respectively.

Under the assumptions that random coder error and intra-coder correlation are the same for all interviewers and coders, a projected inflation factor PF_c was calculated by substituting the average number of occupations coded on one year of the FRS for the experimental coding workload of 200 occupations. This projected inflation factor gives an estimate of the amount by which the variances around the survey estimates must be multiplied to take account of correlated coder error.

Table 3 shows both inflation factors for SEG and SC. The F_c values are similar for both groups of coders although in some cases the values for office coders are slightly higher.

Although these results appear contrary to the findings on unreliability due to random coder error, where field coders were more unreliable than office coders, a larger sample with more coders is needed to confirm this pattern.

Turning to the impact on results for an average year's worth of occupation coding on the FRS we see that for office coding, where the survey workload is around 25 times higher than the experimental workload, the projected inflation factors (PF_c) are substantially higher than the experimental inflation factors (where there was some evidence of intra-coder correlation). The inflation factors for field coding are lower because the experimental workload per coder was higher than the annual workload.

Table 3 Inflation factors for correlated coder variance

	F_c		PF_c	
	Office coders	**Field coders**	**Office coders**	**Field coders**
Socio-economic group				
1.1	*	*	*	*
1.2	1.01	1.06	1.21	1.02
2.1	1.00	1.00	1.00	1.00
2.2	1.04	1.13	2.07	1.04
3	*	1.00	*	1.00
4	†	1.14	†	1.04
5.1	*1.33*	1.11	*9.22*	1.03
5.2	1.23	†	6.73	†
6	†	1.06	†	1.02
7	†	1.01	†	1.00
8	1.00	1.06	1.00	1.02
9	†	†	†	†
10	†	1.06	†	1.02
11	†	†	†	†
12	1.00	1.00	1.00	1.00
13	*	*	*	*
14	*	*	*	*
15	*	*	*	*
Social class				
I	†	1.07	†	1.02
II	1.19	1.08	5.88	1.02
IIIn	1.09	1.04	3.13	1.01
IIIm	†	1.09	†	1.03
IV	†	†	†	†
V	†	†	†	†

* fewer than 20 cases
† unable to estimate due to negative ϱ_w

The results therefore show that the difference between the office and field coders are even more marked. For three of the SEG categories the office coding method would inflate the variance of the estimates by more than 100% (i.e. more than double the variance) and for two of these categories (SEGs 5.1 and 5.2: Ancillary workers and non-manual foremen) the variance would be increased by factors of 9.2 and 6.7

respectively. Thus, the standard error around the estimate of the proportion of ancillary workers in the sample would be increased by a factor of 3 for office coding.

For office based coding of social class, 4 out of the 6 main groups were not affected by variance due to correlated coder variance but the remaining 2 groups had high values of PF_c. The managerial and technical group (SC II) and the skilled non-manual group (SC IIIn) had factors of 5.9 and 3.1 respectively.

In comparison, the projected inflation factors for field coding were reduced compared to the experimental factors and the largest projected factors were 1.04 for professional employees (SEG 4) and managers of small establishments (SEG 2.2).

7. Conclusions

It has long been recognised that the use of large numbers of coders each with a small workload would potentially reduce correlated coder variance compared with the use of a small number of specialist coders. This study confirms previous findings; it has the added advantage of being based on the latest version of the official government classification of occupations.

It has also been maintained that the disadvantage of using interviewers as coders is that they would not achieve the same levels of accuracy as specialist coders. However, this study shows that interviewer coders can achieve comparable levels of accuracy compared with office based coders, at least at the level of broad social classifications such as SEG and SC although not at the detail level of the 3 digit SOC codes. However, the disagreements among individual SOC codes do not generally affect which SC or SEG the case is assigned to. It is these aggregate social classifications rather than individual SOC codes which are generally used in analysis so the results show than the use of suitably trained interviewers does not necessarily result in loss of quality compared with specialist office coders as has previously been maintained.

The study shows that the normal calculation of standard errors around the survey estimates of proportions in the SEG and social class categories are underestimating the true standard errors, due to the exclusion of correlated coder errors. For field based coding operations the amount of underestimation is negligible but for office based coding procedures the impact of having few coders coding many occupations can result in substantial underestimates of the true standard errors. This also implies a reduction in precision for these estimates for office coding compared to field coding which is not offset by the increased reliability of specialist office coders. The results presented here have assumed that correlated coder variance increases overall variance rather than bias in that individual coder biases cancel out. Of course this is not necessarily the case but we are unable to estimate the extent of bias here.

It is not possible to make direct cost comparisons between the two methods of coding occupation because the cost of a particular method depends on a number of organisation specific features. Part of the reason that the two organisations involved in this study have adopted different procedures is because their cost parameters are not the same. In general, the more interviewers an organisation uses to cover a given volume of fieldwork, the higher the training and supervision costs will be. This makes occupation coding by interviewers, which requires a significant investment in training and supervision, most cost effective for an organisation which is able to use its interviewers intensively, covering its fieldwork with the smallest number of interviewers. Interviewer turnover is also important because of its impact on training costs.

References

Collins, M. and O'Brien, J. (1981), "How Reliable is the Coding Process?" *The Market Research Society Annual Conference Papers*.

Dodd, T. (1985), "An Assessment of the Efficiency of the Coding of Occupation and Industry by Interviewers," OPCS *New Methodology Series*, 14.

Elliot, D. (1982), "Variability in Occupation and Social Class Coding," unpublished report, London: OPCS.

Elliot, D. (1983), "A Study of Variation in Occupation and Social Class Coding - Summary of Results," OPCS *Survey Methodology Bulletin*, 14.

Kalton, G. and Stowell, R. (1979), "A Study of Coder Variability," *The Journal of the Royal Statistical Society, Series C*, 28(3), pp. 276-289.

Office of Population Censuses and Surveys (1990), *Standard Occupational Classification*, Volumes 1 and 2. London: HMSO.

White, P. (1983), "The Continuous Manpower Survey - Feasibility Study," OPCS *Survey Methodology Bulletin*, 15, pp. 33-41.