An evaluation of computer-assisted occupation coding : results of a field trial

Diane Bushnell, Office for National Statistics, UK

1. Introduction

At the last International Blaise Users' Conference held in Helsinki, I described some work being carried out in the Social Survey Division (SSD) of the Office for National Statistics (ONS) on computer-assisted occupation coding.[i] This paper describes the results of our work.

Over the last two years we have been investigating the use of computer-assisted occupation coding by conducting trials of different coding systems. From the results of the trials, we have concluded that it is feasible to use computer-assisted occupation coding without loss of data quality, and that using a straightforward word-matching system can be better than using a knowledge-based system. However, there are some questions remaining about whether unacceptable levels of bias will be introduced into occupation estimates when using computer-assisted coding.

2. Background

The majority of the surveys in SSD are now carried out using computer-assisted personal or telephone interviewing (CAPI/CATI). Computer-assisted coding is used routinely for most straightforward coding frames, such as country of birth, food items or journey destinations. However, occupation coding is still carried out using a clerical method. As we collect occupation details and code occupation on all our surveys we are very interested in assessing whether it is possible to use computer-assisted coding while maintaining, and ideally improving, data quality.

In the mid-1980s, SSD made a decision to code occupation in the field: the interviewers would collect the occupation details as usual and then code occupation at home, after the interview. Previously, in common with most other survey organisations, we employed specialist coders based in the office. Studies that were carried out at that time[iiiii] comparing office and field coding showed that although interviewers could not achieve the same levels of inter-coder consistency as office based coders, their smaller workloads meant that any systematic bias demonstrated by individual coders 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. Finally, the cost and time savings were large enough to offset any worries about a reduction in data quality.

More recently another trial was conducted comparing office and field coding, since the interviewers had become more experienced[iv]. Although inter-coder consistency was still lower for field coding compared to office coding, the accuracy of coding for the two groups was very similar (Table 1). As in the previous trials, individual coder biases resulted in a greater loss of precision for occupation related estimates when office coded, compared to coding carried out by interviewers.

Table 1 Results of a study comparing office and field coding of occupations

	Reliability*		Accuracy*	
	Office	Field	Office	Field
SOC	0.82	0.74	80.3	76.9
SOC Major group	0.90	0.86	87.8	87.4

* Reliability (inter-coder consistency) was estimated from average agreement between pairs of interviewers (perfect reliability=1). Accuracy was estimated from the percentage agreement between coders and an expert coder (perfect accuracy = 100).

Even though the data quality obtained by field coding has clearly improved since its introduction, we felt that there was scope for further improvement. We hoped that computer-assisted coding would contribute to that improvement.

3. Aims of study

The main aims in carrying out the study were to determine whether computer-assisted coding of occupation by interviewers could

- improve the quality of coded occupation data,
- reduce costs, and
- decrease interviewer burden.

Another, less important, aim was to assess the feasibility of coding occupation during the interview (to further aid the interviewers in assessing which was the most appropriate code).

We designed an experiment to compare two different coding systems to our usual clerical coding method. A discussion of our initial work to decide on suitable systems can be found in my previous paper for the 1995 Blaise conference[1]. We asked the interviewers who took part in the experiment to give us feedback on the two systems, using a semi-structured questionnaire.

4. The study

4.1 Occupation coding

In the United Kingdom, the Standard Occupational Classification (SOC) (OPCS, 1990)[v] is the most widely used occupational coding scheme for social research. The SOC was developed for use on the 1991 Census of Population. Its major use is for assigning individuals to various social class and socio-economic classifications.

The SOC consists of 371 occupational unit groups (SOC codes) which may be aggregated into minor and major groups. The system is hierarchical so, for example, SOC code 270 (Librarians) is in minor group 27 (Librarians and Related Professionals), which is in major group 2 (Professional Occupations).

In SSD the interviewers collect verbatim details on the job title, main duties and responsibilities of the job, qualifications required and industry, as well as asking specific questions about employment status (e.g. whether respondents are self-employed and the number of employees at their workplace). At home, the interviewers go back into the questionnaires, review the occupation details and select the SOC code which they think is most appropriate from a paper coding index, using a set of standard rules and procedures. Finally, they type the SOC code into the Blaise instrument.

4.2 The experiment

We carried out the experiment in three phases. 300 sets of occupation details were extracted from real data collected on the UK Labour Force Survey (LFS). In the first phase, 24 interviewers coded these using the standard clerical method. In the second phase, the same interviewers coded the same set of occupations using a specialised occupation coding system from within a Blaise 2.5 questionnaire. To minimise memory effects, we allowed nine months to elapse between the first two phases and the occupation details were presented to the interviewers in a randomly rearranged order. In the third phase, a different set of 24 interviewers coded the same 300 occupations using Blaise III.

At the first and second phases Blaise 2.5 was our standard interviewing software for the LFS. By the third phase, Blaise III had become standard; all the interviewers had been trained in its use and had been using it for interviewing for a few months. Therefore, at each phase of the experiment, the interviewers were using the same version of Blaise as they used every day for interviewing. This meant that we only needed to provide a minimum of training in the coding procedures.

4.3 The coding systems

- Clerical coding

The system used in the study was designed to resemble the normal coding process as much as possible. A Blaise 2.5 program was written which displayed each set of occupation details (job title, job description, industry, employment status) in the standard way. Using standard

procedures, the interviewer searched through the paper coding index for a suitable match to the job title, using the other occupation information when appropriate. When a match was found the interviewer entered the code into the Blaise instrument. After entering a code, the next set of occupation details were presented until all 300 were completed.

- CASOC/Blaise 2.5

A customised version of the specialised occupation coding system CASOC[vi] (Computer Assisted Standard Occupational Coding) was integrated with Blaise 2.5 for use in the study.

CASOC is a sophisticated knowledge-based occupation coding system, first developed in 1986 and based on the Standard Occupational Classification (SOC).

The matching algorithm takes several factors into account before proposing suitable codes, including a word match comparison and the results of a weighting procedure. Weights are attached according to how commonly the job occurs in the population and how many other possible matches there are for the job title entered. It is possible to use employment status information to further refine the search (e.g. if the respondent is self-employed the job titles which are suggested first are those which are most appropriate for the self-employed).

Our customised version of CASOC was incorporated into Blaise 2.5 (it was not possible to call external programs from Blaise III at the time we developed the system). It was specifically designed to look and behave like Blaise 2.5 so that the interviewers would think they were using the usual Blaise coding facilities.

For the experiment, the job title and employment status information was passed to the matching algorithm; the best matches were displayed in the bottom half of the screen; the other occupation details were displayed in the top half of the screen. The interviewers had the option to search through the list of suggestions, extend the search to include more unlikely codes and to edit the job title to add or change the information used for matching.

- Blaise III

As we knew that Blaise III would become our standard interviewing software we were very interested to see how coding using the facilities provided by Blaise III compared to coding with the knowledge-based system.

Blaise III coding is based on word matching (using trigrams) and/or on stepping through hierarchies of a coding frame. We decided that the Blaise III system would use the job title only for matching but that the other details would be displayed on the screen for information. We had some doubts about how a straightforward word-matching system would cope with the complexities of occupation data but, nevertheless, thought it was worth investigating.

A combination of trigram and hierarchical coding was used in the study. The job title was used as a starting point for the matching process and the other occupation details were displayed in the top half of the screen. The best matches were displayed in the bottom half of the screen; the interviewers could scroll up and down through the suggestions, edit the job title and move up and down the coding frame hierarchy to get more information on the suitability of the codes.

5. Results from the experiment

When we designed the study we had certain expectations and theories about the possible outcomes. We thought that computer-assisted coding would be

- more reliable/consistent,
- possibly more accurate,
- less biased,
- quicker, and
- easier

than clerical coding. In addition we thought that

- a knowledge-based system would be 'better' than a simple word-matching system.

Reliability

Inter-coder consistency, also referred to as reliability, is the extent to which coders will assign the same code when supplied with the same information. If reliability is low then we cannot be sure that codes are not being assigned at random and would have little confidence in any survey estimates derived from the data. Perfect reliability would be obtained if all the interviewers assigned the same code when supplied with the same information. In this study we estimated reliability from the average agreement between all pairs of interviewers: a value of 1 indicating perfect reliability.

We hypothesised that reliability would be higher for computer-assisted coding compared to clerical coding as a computer-assisted system would provide the interviewers with a more consistent set of options to choose from. We thought that a knowledge-based system would present a more suitable shortlist of codes to select from than a word-match system and that therefore the interviewers would pick a code from that list more often, resulting in higher agreement amongst interviewers.

We estimated reliability for occupation codes at the most detailed level of the classification (SOC codes, 371 codes) and at the most aggregated level (Major group, 9 codes).

Table 2 Reliability for clerical and computer-assisted coding methods

	Clerical method	CASOC	Blaise III

SOC	0.73	0.74	0.76
SOC Major group	0.85	0.86	0.88

We found no significant difference in reliability between the CASOC system and clerical coding. However, Blaise III coding was significantly more reliable than coding clerically (Table 2). These results were rather surprising and showed no support for either of our hypotheses. Although the Blaise III system was better than clerical coding, we could not conclude that computer-assisted coding, in general, would be more reliable.

Accuracy

By itself, reliability is not sufficient to assess data quality: reliability may be high even when the codes assigned are incorrect. For example, reliability will be perfect if all coders assign one code to all the occupations, regardless of whether the code is suitable or not. It is, therefore, important to measure the accuracy of coding.

We were not able to go back to respondents to check whether codes assigned to occupations were actually the most appropriate but we were able to look at the agreement between the codes assigned by the interviewers in the experiment and an expert occupation coder. The expert coder had coded thousands of occupations on the 1991 Census of Population and was very experienced in applying the coding rules and resolving queries. We defined coding accuracy as the percentage agreement between the interviewers and the expert.

Accuracy of coding will depend on whether the interviewers are influenced by the coding system used or whether they continue to code as they do clerically. In the first instance, accuracy will depend on how good the system is at suggesting appropriate codes; in the second instance, accuracy will remain the same as for clerical coding. We expected, by definition, that a knowledge-based system must be more accurate than a straight word-matching system.

Table 3 Percentage agreement with expert coder

	Clerical	CASOC	Blaise III
SOC	76.5	75.2	79.5
SOC Major groups	86.9	87.3	90.0

The accuracy of codes assigned using the CASOC system was not significantly different from codes assigned clerically but the Blaise III accuracy levels were significantly higher than both the clerical and CASOC levels (Table 3). Again, this was a surprising result: we had expected the CASOC system to produce the most accurate codes. Moreover, the accuracy of coding using the Blaise III system was as high as that obtained by office coders in the earlier study (Table 1), even at the SOC code level.

Bias

Bias occurs in the data where there is a systematic deviation from the true answer. If accuracy is very high then there is little room for bias in the results. However, where accuracy is not high we must examine whether

the incorrect codes are distributed randomly or form a pattern which may lead to bias in the resulting occupation estimates.

If a coder has a tendency to favour certain codes this will lead to individual coder bias. We normally assume that these biases will cancel out when aggregated over all the coders. This will result in an increase in the variance around the estimates (i.e. a decrease in precision) but no overall bias. When using a computer-assisted coding system, we might assume that the individual coder bias would decrease as interviewers have less scope to favour particular codes. Thus, the impact on the variance would be less than for a clerical process. However, we can no longer assume that there is no overall bias as the system itself may be influencing all the coders in the same way.

Assessing bias is usually very difficult, as the data we collect is usually subject to some sort of error (measurement, recall, sampling, etc.). In this study, we made the assumption that the codes assigned by the expert coder were the correct codes. We looked for evidence of bias by comparing the distributions of the occupation codes found in the experiment with the distribution obtained from the expert coder. Sample sizes from the experiment were not large enough to allow comparisons between distributions at SOC code level but it was possible to examine SOC Major groups.

Table 4 Distribution of codes allocated to Major groups

SOC Major Group		Expert	Clerical	CASOC	Blaise III
		%	%	%	%
1	Managers & Administrators	14.0	15.7	15.9	13.9
2	Professionals	6.3	6.5	6.2	7.1
3	Associate Prof. & Technical	12.0	11.5	11.6	11.7
4	Clerical & Secretarial	14.0	14.0	14.1	14.2
5	Craft and Related	12.0	12.2	12.2	12.1
6	Personal & Protective Service	10.0	9.8	9.5	9.4
7	Sales	9.0	8.0	7.6	8.4
8	Plant & Machine Operatives	9.7	9.7	9.7	9.6
9	Other Occupations	13.0	12.5	13.3	13.6

Each of the methods of coding demonstrates some bias, compared to the distribution of Major Groups derived from the expert coder (Table 4). However, the pattern of bias for each method is different so we cannot draw any conclusions about whether computer-assisted coding is more biased than clerical coding.

Speed

The interviewers were asked to record the time they spent coding the 300 occupations so that we could compare the relative speeds of the methods.

Computer-assisted coding was significantly faster than clerical coding and the Blaise III system was slightly faster than the CASOC/Blaise 2.5 system. Interviewers using the CASOC/Blaise 2.5 system and Blaise III system respectively took 13% and 23% less time to code the 300 occupations, than those using the clerical method.

Although it appears that significant time savings (and therefore cost savings) can be made by using computer-assisted coding, in practice codes will be assigned one at a time, rather than in a batch of 300. Therefore, time savings may not be so substantial in the field.

Ease of use

The interviewers taking part in the trials were asked to fill in questionnaires to provide us with feedback on the use of the systems. Most of the interviewers found using a computer-assisted system easier than coding clerically, although some found ambiguous occupations slightly more difficult to code. Most of the problems which arose were due to deficiencies with the coding frame and index, rather than the coding systems. Despite some reservations, all the interviewers thought that occupation coding should be computerised and looked forward to using such a system in the future.

6. Conclusion

The results of the experiment enabled us to draw some conclusions about the aims of the study.

- Improving quality of coded occupation data

Computer-assisted coding of occupation by interviewers does not necessarily improve the quality of coding, compared to clerical methods. Coding using the CASOC/Blaise 2.5 system developed for our trial was no more reliable or accurate than coding clerically. However, there were significant improvements in reliability and accuracy of occupation coding when using the Blaise III system.

These findings were contrary to our expectations, as we had assumed that the codes suggested by the knowledge-based CASOC system would be correct more often than for the Blaise III system, and that therefore the interviewers would be more likely to select from the suggested codes. We thought this behaviour would lead to higher reliability (through interviewers picking codes from the suggested list, rather than searching around for their own preferences) and to higher accuracy.

Further analysis of the data revealed that the CASOC system was more likely to suggest correct codes than the Blaise III system. When we looked at the first code suggested by the two systems for each occupation, we found that 62% of the CASOC first codes agreed with the expert-assigned codes, compared to only 50% of the first codes suggested by Blaise III. However, the interviewers did not necessarily select these codes, even when the first code suggested was the correct one. Using either system, interviewers selected the first code about 90% of the time when the code was correct. However, interviewers selected the first code, when incorrect, in 23% of cases for the CASOC system and in only 8% of cases for the Blaise III system. A possible explanation for this behaviour is that incorrect first codes suggested by the Blaise III system are more obviously implausible than those suggested by the knowledge-based CASOC

system, leading the interviewers to reject them more often and to search for better solutions.

All the coding systems (including clerical coding) showed some evidence of bias, compared to the codes assigned by the expert coder. There was a tendency for the interviewers using the CASOC/Blaise 2.5 system to agree with the first code suggested, regardless of whether it was correct or not (codes assigned by interviewers corresponded with the first code suggested in 64% of cases, compared to only 50% of cases in the Blaise III system). Although it was clear that interviewers were still using their judgement to assign codes and not just agreeing with suggestions, they may become more dependent on the system suggestions over time and so bias may increase. However, a computer-assisted system will allow us to have more control over the coding process than at present and will allow us to monitor coding quality more easily.

- Reduction in costs

Using either of the computer-assisted systems is likely to result in cost savings due to a reduction in the time interviewers spend coding. Coding with the Blaise III system was slightly faster than the CASOC/Blaise 2.5 system.

There are other cost considerations which must be taken into account besides field coding costs. There will be one-off costs associated with a change to computer-assisted coding, whatever system is used. These costs will include work on development of a system, training of interviewers, changing existing management systems, updating coding instructions and so on. There are also the costs of using and maintaining a system. The cost of introducing either of the systems used in our experiment would be fairly minimal (one of the reasons why we selected them in the first place). In the long term, therefore, it is quite likely that using a computer-assisted coding method will be cheaper than our existing method.

- Decreased interviewer burden

There was a definite decrease in interviewer burden when using the computer-assisted systems. The elements of the systems which the interviewers did not like are relatively easy to change (screen layout, speed and so on). We will be conducting some usability trials to find the best combination of layout, behaviour and speed.

- Ability to code during the interview

All the interviewers thought that coding during the interview, using the systems in the trial, would not be possible, even though they recognised the advantages that it would bring. Their main objection was that the systems were too slow and that spending a lot of time coding during the interview would interrupt the flow of the interview and disrupt the rapport they had with the respondent. They also felt that it was not appropriate to spend a lot of time on occupation when it may appear to the respondent to have little relevance to the survey topic.

In conclusion, we will need to carry out some further work to identify the factors which are relevant to an increase in the data quality, for example whether there was some aspect of the Blaise III coding system, not included in the CASOC system, which could be responsible for the increase in coding quality. We also need to analyse the results on bias in more detail, before we feel confident that we can make a change to our coding procedures.

However, this experiment clearly demonstrates that computer-assisted coding of occupation by interviewers is feasible, without a reduction in data quality, where coding is carried out after the interview. Interviewers can code occupation more consistently and accurately, faster, more cheaply and more easily using computer-assisted coding than using the standard clerical procedures.