



Economic and Social Council

Distr.: General
25 May 2010

Original: English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Thirteenth Meeting

Geneva, 7-9 July 2010

Item 3 of the provisional agenda

Plans for census quality evaluation

Scotland's Census Quality Survey 2011

Note by the General Register Office for Scotland, United Kingdom

Summary

In addition to the coverage survey, the United Kingdom census offices conduct a smaller post-enumeration survey, the Census Quality Survey (CQS), to evaluate the accuracy of the data by re-asking census questions in a doorstep interview. In 2011 the General Register Office for Scotland (GROS) will conduct the CQS using computer-assisted personal interviewing (CAPI), the first time this method has been used by GROS. This paper describes the aims and design of the survey, including the innovative use of information technology. It also describes our experiences from the pilot survey carried out in 2009, and how we intend to apply the lessons learned in 2011.

I. Introduction

1. The first priority of any census is to obtain an accurate count of the population by key demographic variables, such as age and sex. However, most censuses collect a wealth of other information which is highly useful for a variety of purposes - but only if is accurate.

2. In a self-completion census it is not possible to guarantee the accuracy of every response. Occasionally a systematic error becomes evident during processing, and it may be possible to make a macro-level adjustment to allow for this. Generally, though, the overall accuracy of each variable can only be assessed by carrying out an additional data collection exercise.

3. This paper describes the survey which the General Register Office for Scotland (GROS) will carry out to measure the accuracy of responses to the 2011 Census in Scotland. Section 2 gives a high-level description of the survey and its purpose, after which section 3 goes into more detail of the planned design, including the use of electronic devices as well as the sampling and interviewing methodology. Section 4 describes the pilot survey carried out following the 2009 Census Rehearsal, along with some of the practical lessons learned and how they will be applied in 2011. Finally, section 5 gives a brief summary.

2. Background

4. There are three census offices in the United Kingdom: the Office for National Statistics (ONS) which covers England and Wales, the Northern Ireland Statistics and Research Agency (NISRA) and the GROS in Scotland. In addition to the Census Coverage Survey (CCS), these three offices have in recent times carried out a Census Quality Survey (CQS). Like the CCS it is an interviewer-led post-enumeration survey. Rather than coverage (whether the right *people* were counted by the census) it aims to measure quality (whether they gave the right *answers*).

5. Also unlike the CCS, it is not designed to be statistically independent of the main census. The sample is drawn purely from households that are known to have responded to the census. Individuals within each household are re-asked census questions, and if any differ from their census response, the interviewer will probe for the reason for the discrepancy. In some cases the interviewer may decide that the census response was, in fact, correct. Otherwise, he or she will record a reason for the discrepancy.

6. The achievable sample size for such an intensive interview means that it cannot be used to adjust the census results. Instead, it is normal to publish a report stating the agreement level between the census and CQS for each question, and the main reasons given for any discrepancies. This helps users to decide how much reliance to put on the published census results for each question, and can feed into the design of future censuses.

7. In 2011 GROS will conduct the CQS in Scotland independently of the rest of the UK. It will take place over a period of approximately six weeks beginning in mid-May, following the census on 27 March. Unlike the census itself, it is not compulsory. Computer-assisted personal interviewing (CAPI) will be used, the first time that GROS has conducted a survey in this way. As well as assessing the quality of responses to the 2011 Census in Scotland, it is intended to pilot the possible future use of CAPI for other census-related activities, such as non-response follow-up and the CCS.

3. Design

A. Use of information technology

8. After investigating the options, GROS decided to conduct the survey using mini-laptops running a Microsoft Access-based application that was developed in-house. The interviewer enters data into an Access database via a custom-designed form. Macros written in Visual Basic for Applications (VBA) ensure that the interviewer is routed to the correct part of the form, and presented with the correct options, depending on the response to previous questions.

9. Each interviewer's laptop is preloaded with census data for the households in that interviewer's workload. Since both the census and CQS collect personal information, data security is a high priority. All information on the laptops is protected with a very high level of encryption.

10. The system is designed so that the interviewer does not initially have any access to the respondent's census data. As well as providing an additional level of security, this should ensure that respondents are not prompted, intentionally or inadvertently, to give the same response as they gave to the census. Once the interviewer has entered a response to a particular question, the census response to that question becomes visible.

11. If the two responses are identical, the system records this and the interviewer moves on to the next question. If the system identifies a difference between the census and CQS responses, the interviewer has two options. If the two responses are effectively the same, but a variation in spelling or wording (e.g. "Teacher" on the census and "Primary school teacher" on the CQS) has caused the system to identify a difference, then the interviewer should mark the responses as the same, and move on to the next question.

12. If, however, there appears to be a genuine difference between the two responses, the interviewer will point out the difference to the respondent and probe for the reason. It may become apparent at this stage that the census response was in fact correct as at census day - for instance, the respondent may have changed jobs in the meantime, or he or she may have misunderstood the question in the CQS but given the correct response in the census. In this case, the interviewer will amend the CQS response as appropriate. If the census response was genuinely incorrect, the interviewer will select the most appropriate reason from a list - for instance, the respondent may have misunderstood the census question, or someone else may have completed the census questionnaire on his or her behalf and entered the wrong information.

13. Interviewers are required to regularly upload data directly from their laptops to GROS headquarters, via a secure internet connection. This enables headquarters staff to monitor the progress of the survey, and also to perform preliminary analyses of the data.

B. Sample design

14. GROS does not conduct any surveys other than those connected to the census, and therefore does not employ a permanent team of interviewers. This makes it impractical, in a relatively small one-off survey such as the CQS, to take a random sample of addresses across the whole of Scotland. Instead a clustered design is used - three areas (one urban, one rural and one semi-rural) have been chosen which are considered to be reasonably representative of the country as a whole. In total approximately 1000 households (0.05% of the population of Scotland) will be included in the sample, to be split equally between the three areas.

15. As CQS fieldwork will commence within two months of census day, most census data will not have been captured in time to load it onto interviewers' laptops. The CQS sample will therefore be drawn only from households that respond early to the census, either on paper or on line. Although, as mentioned above, three broad geographical areas have been identified for the CQS, the precise areas within these that will be processed early will not be known until the first returns have been received. This means that GROS faces a challenging timetable for drawing a sample of addresses for the CQS, allocating them to interviewers and loading the relevant data onto each interviewer's laptop.

16. The sample cannot be considered to be statistically representative of Scotland's population as a whole. There are three main reasons for this:

- (a) The sample areas have not been selected at random;
- (b) The sample will consist only of households that have returned their census questionnaire early, in areas with a high proportion of early returns;
- (c) There will be some bias due to survey non-response.

17. This means that no statistically significant results can be obtained from the CQS. Agreement rates between the census and CQS (expressed as a percentage) will be published for each question, but they should be used only as a qualitative guide to the effectiveness, or otherwise, of each question in the census. No confidence limits will be provided.

C. Interview format

18. As noted above, GROS does not employ any interviewers on a permanent basis. CQS interviewers will be selected from among those census enumerators recommended as suitable by their managers on the basis of their work in the census itself. They must, of course, be based in or near the areas covered by the CQS. Each of the three areas is covered by a team consisting of a manager (who carries out some interviews in addition to his or her managerial duties), and six interviewers. Each interviewer is also accompanied to each interview by an assistant, the reason for which is explained below. The workload is approximately 50 addresses per interviewer, with the manager having half this number.

19. Depending on how many people are present, each household interview may take up to an hour. For this reason it is not practical to carry it out on the doorstep, which means that the interviewer needs to enter the house. This means that various precautions need to be taken for the safety of both the interviewer and the household members. The interviewer and assistant interviewer carry photo ID, and can give a prearranged password so that the householder can be confident of their authenticity. There will also be a telephone hotline for householders to report any concerns. Finally, for the interviewer's safety, he or she must not enter any house alone but must always be accompanied by an assistant.

20. At each address, one resident - normally the householder - is asked the household-level questions (e.g. type of accommodation, number of cars). Each resident present (who is willing to be interviewed) is then asked the person-level questions. With the exception of children, household members must be interviewed in person to ensure that the CQS obtains the true response to each question.

4. 2009 rehearsal and lessons learned for 2011

A. Fieldwork overview

21. Following the census rehearsal on 29 March 2009, a CQS rehearsal was carried out during July and August of that year. The fieldwork took place in part of the west Edinburgh area that had been included in the census rehearsal. The field force was equivalent to that planned for each of the three areas in 2011, namely one manager and six interviewers, plus assistants. The interviewers were recruited from among the rehearsal enumerators in a similar way to that planned in 2011.

22. The total number of addresses in the sample was 329. The achieved sample size was 153, or 46.5% of the total. This number would no doubt have been higher if it had not been for a breakdown in communication which led to prospective interviewers being given the wrong dates. As a result, a number of them booked holidays in the middle of the fieldwork period and were not available for a substantial part of the period. Only 7 households are known to have actively refused to take part. The total number of individuals interviewed was 210.

23. In the rehearsal, each interviewer was asked to recruit an “accompanying person” whose role, initially, was purely to be present at each interview for health and safety reasons. In the event, it was found useful to have a second person present to assist with the interview. In 2011 the accompanying person’s role will be expanded, and their title will be “assistant interviewer”.

B. Timing

24. It was initially planned that the CQS should take place in July and August 2011, and the timing of the CQS rehearsal was intended to reflect this. In the event, the processing timetable for the census rehearsal meant that it would not, in any case, have been possible to hold the CQS rehearsal any earlier because no census data would have been available.

25. The result was that the fieldwork took place during a period which probably, for many of the households in the sample, included their summer holiday. This may well have had an adverse effect on the response rate.

26. The survey found evidence that some interviewees, whose circumstances had changed since March, had difficulty remembering their correct details as at census day. For this reason, the CQS in 2011 will take place much earlier. This will not only make it more likely that respondents will remember their circumstances as at census day, but will also make it less likely that the circumstances will have changed in the meantime.

27. A beneficial side effect of an earlier start to the fieldwork is that census enumerators are more likely to be available to work on the CQS, both because they are less likely to take a holiday at that time of year and because there will be very little gap between the end of census fieldwork and the beginning of the CQS, so they are less likely to have moved on to other work.

C. Use of information technology

28. The trial of mini-laptops in the field was very successful - the devices were sufficiently portable, easy to use and had adequate battery life for the purpose. There were some initial teething troubles before fieldwork began, due largely to the fact that there was

insufficient time to test the laptops after all the relevant software had been installed. The intention in 2011 is to allow a longer lead-up time.

29. Due to the security features on the laptops, interviewers had a large number of passwords to remember, which caused some problems. A workaround was found for the rehearsal, but the whole issue will be revisited for 2011.

30. The laptops were fitted with a 3G card so that the interviewers could upload data to GROS headquarters from wherever they happened to be. Unfortunately, despite the fact that the rehearsal took place entirely within a built-up area, a number of interviewers reported poor network coverage. This is likely to be even more of an issue in 2011, when the CQS will include a rural area. For this reason, 3G cards will not be used. Instead, interviewers will use a secure internet connection in their homes to deliver the data uploads.

31. The interview application generally worked well, although there were a few problems with the census data caused by the way in which the data was captured from paper questionnaires and imported into Access. This should not be an issue in 2011, when data will be received directly from the census processing site.

D. Sample representativeness

32. After the completion of the CQS, the achieved sample was analysed by three key variables to see how representative it was of the area as a whole. It was found that the following groups were over-represented:

- (a) People over 60 (especially at the expense of those in their 20s and 30s);
- (b) People of “white Scottish” ethnicity;
- (c) Households who own their home outright (without a mortgage).

33. There are likely to be two main reasons for this:

- (a) These types of people are more likely to return their census questionnaire early, and so be included in the initial sample;
- (b) They are more likely to be available and willing to take part in the CQS interview.

34. The second issue can be partly mitigated by instructing interviewers to vary the time of day and week that they call on each address, and to ensure that they make several efforts to contact each one rather than concentrating on those perceived to be “easiest”. Some weighting of the results may be used in 2011 to rebalance the sample, but it should be emphasised once again that, however representative the sample is of the areas in which the CQS is carried out, it will not be fully representative of Scotland as a whole. The results can only act as a qualitative guide to the success or otherwise of each question in the census.

E. Selected results

35. At household level, the question with the lowest agreement rate between the census and CQS was *How many rooms are available for use only by this household?*, with only 66% of households giving the same response in both cases. In most cases the difference was 1 either way, and the weakness is thought to be primarily in the CQS rather than the census. The census questionnaire lists the types of room that should, and should not, be counted in this question, but this was not always brought to the attention of interviewees in the CQS. In 2011 interviewers will have a ready reference card available containing all the

relevant definitions, and will be trained to check that interviewees have counted the relevant rooms and no others.

36. At individual level, the question *Do you have any of the following conditions which have lasted, or are expected to last, at least 12 months?* had an agreement rate of 76%. This appears to be primarily due to the fact that this is a subjective question - although a respondent may feel on census day that their condition is serious enough to include, and that it is likely to last 12 months, they may not feel the same about it on the day they are interviewed for the CQS. The census question has been redesigned for 2011, which may reduce the number of inconsistencies.

37. The question *Which of these qualifications do you have?* had an agreement rate of 77%. This is a complicated question, with a long list of qualifications to choose from, and has historically been poorly answered in censuses in the UK. Many of the differences were due to people including only their highest level qualifications in the census, presumably because they had misunderstood what was required.

38. At the other end of the scale, there was a 100% agreement rate for the question asking whether a student lives at this address during term-time, although as it is only answered by students this question had a very low sample size. There were four questions where there was only one discrepancy in a sample of over 100, namely *What is your country of birth?* and the questions asking whether the respondent was a full-time student, looking for work or waiting to start a job already obtained (these last three were straightforward yes/no questions).

39. The only significant failure of the CQS rehearsal was in the questions relating to overnight visitors on census day. 15 households contacted by the CQS had said in the census that they had one or more visitors, but only 3 of these said in the CQS that they did so. This is probably because of the time gap - even with the CQS taking place earlier in 2011, it seems unreasonable to expect householders to remember details which may relate to only one night, up to three months after the event. For this reason, it has been decided not to ask questions about visitors in the CQS in 2011.

40. The stated reasons for differences between census and CQS responses were less useful than had been hoped. This was partly because the list that interviewers could choose from was not adequate, and this will be revised before the full CQS in 2011. There were also a number of cases where no reason had been selected. This should not have happened, and in 2011 the interface will be redesigned so that the interviewer cannot move on to the next question without selecting a reason for any difference in the previous question.

5. Conclusion

41. This paper has described the plans for Scotland's Census Quality Survey in 2011. The CQS rehearsal carried out in 2009 demonstrated that the methodology, including an innovative use of information technology, can be made to work, but also highlighted some issues that need to be taken into account when drawing up the final design for 2011. The results obtained from the rehearsal show that the CQS will provide a useful measure of the quality of responses to the 2011 Census in Scotland.
