

# Mode effects and maintenance of data quality in a CAPI/CATI survey

*Tim Burrell, Office For National Statistics,*

## 1. Introduction

All ONS computer assisted social surveys use Blaise. As most ONS social surveys are carried out as CAPI, the standard is to design the questionnaire instrument with the ease of the face-to-face (FTF) interviewer in mind. The exception to this standard method of CAI design is the Labour Force Survey (LFS).

The LFS is a rotating panel survey containing five quarterly waves. The first wave is always CAPI, in part because the UK does not have an address or person register with phone numbers. Follow-up waves are usually carried out via CATI unless a telephone number is not given or available at the first interview. This survey is designed for the ease of telephone interviewing. The CAPI design is constrained to suit CATI interviewing. For example, there are no show cards in either mode.

The Annual Population Survey Boost Sample (APSB), like the LFS, has a panel element, in this case four annual waves. The majority of first wave APSB interviews are administered via CAPI. However, where a telephone number can be traced in advance (about 25% of cases) first wave interviews are carried out via CATI. All second and subsequent interviews are carried out via CATI unless the respondent insists on face-to-face interviewing.

The LFS caters for both CAPI and CATI within a single Blaise instrument designed for CATI. Other ONS social surveys have a different (ONS standard) single instrument structure which is more suited to CAPI. For the APSB, a single questionnaire instrument was designed which is flexible enough to use both modes of interviewing and aims to maintain data quality.

The LFS annually-independent sample components comprise, with the APSB, a major new output for the UK called the Annual Population Survey (APS). The APS is designed to measure key census-type variables annually between decennial censuses. It is a precursor for a more ambitious survey, larger in sample and number of variables, called the Continuous Population Survey (CPS) which will start in 2008. The CPS is described in another paper at this conference (Bumpstead, 2004).

In an era where response rates are falling on all household surveys, ONS are looking at ways of increasing or at least maintaining current response rates. One of these methods is to make more effective use of telephone interviewing. Although telephone interviewing is commonly thought to yield lower response rates than face-to-face interviewing for the same survey, it may have more subtle effects when viewed as part of a strategy for dealing with the entire portfolio of surveys. In this perspective, telephone interviewing for appropriate surveys or parts of surveys may provide more face-to-face interviewer time in the field for surveys for which this may make a critical difference to response. In addition, ONS seeks more of the traditional benefits of telephone interviewing: reduced costs.

This paper will briefly describe the structures of the current CATI-oriented LFS questionnaire instrument and the current CAPI-oriented ONS standard instrument structure. It will then examine the design requirements for a new survey instrument suitable for both modes of interview. It will also look at differences arising in how the questionnaire is structured for CAPI and CATI; at how cases can be moved seamlessly between the two modes as real-life situations demand; and at minimising the impact of interview mode on data quality.

## 2. Structures of ONS social surveys

### 2.1. The Labour Force Survey (LFS)

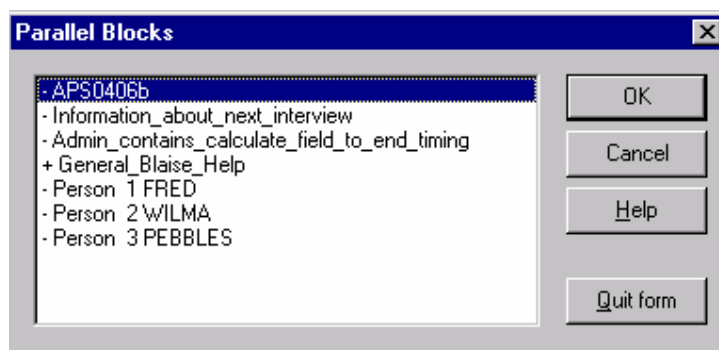
At present, the LFS is the largest single source of survey data available in the UK for monitoring purposes between Censuses. It produces estimates to a level of precision far exceeding any other UK household survey.

The LFS began production in 1990 and was programmed in Blaise I. It was the first ONS survey to adopt CAI. It was, therefore, programmed with simplicity as a strong design driver, given the constraints of 1980s software and hardware. As noted earlier, it was primarily designed to suit a CATI interview as 4 of the 5 waves are usually carried out in that mode. Though CAI tools have advanced and the LFS survey has adopted more up-to-date versions of Blaise (it is now programmed in Blaise 4.6) the essential structure of the instrument has proved highly resilient. Only now, a decade and a half later, do we feel the need to change it.

The design of the LFS questionnaire involves a small household section followed by an individual section for each adult in the household. The LFS type of interviewing structure is called *consecutive* interviewing: where an individual respondent's interview is completed before moving on to the next respondent. Individual interviews are accessible using parallel blocks within the questionnaire. This allows individual interviews to be accessed easily and quickly. At subsequent waves, provision is made to interview the household members in a different order from the first wave interview. Standard ONS social surveys, on the other hand, are designed not to allow access to individual's interviews through parallel blocks. LFS interviews, at a mean of about 30 minutes per household, are shorter in length compared with other ONS social surveys.

Figure 1 shows a screen shot of the box which appears when the option to access the parallel blocks is selected. There are options available to access each person's individual interview. These can be accessed at any stage during the interview, after the household interview section has been completed.

Figure 1.



### 2.2. Other ONS social surveys

Virtually all other ONS social surveys are carried out via CAPI. For most surveys a field interviewer visits a household and interviews all members who are eligible for interview for the particular survey. These types of interview tend to cover a particular subject (such as expenditure or psychiatric morbidity) in great detail, so they are longer in interview length than the LFS. (Several have a mean interview length of around 90 minutes). In most cases a CATI design is not suitable. For many of them, as for the LFS, all adults in the household are eligible for interview (e.g. for the expenditure survey but not for the psychiatric morbidity surveys). The ONS standard instrument structure provides for interviewing all eligible members of the

household. Surveys which are designed to interview only one person in the household use a simplified version of the standard structure.

Where more than one household member is eligible for interview, the CAPI interviewer will attempt to interview as many of the household members as possible on a single occasion. The individual sections are divided into topic areas and the interviewer covers one topic area at a time asking all the individuals in turn before moving on to the next topic. This method (*concurrent* interviewing) maintains the momentum of the interview and retains the attention of all respondents. There is provision for interviewing household members in any desired order and to complete the household's interviews on another occasion for those members not present when the interview is initiated. Concurrent interviewing, with its frequent shifts between respondents, would not be suitable over the phone (unless information is provided by a single proxy respondent on behalf of all household members).

Figure 2 shows how the Blaise interview screen appears during CAPI using the concurrent interviewing method. The sections are divided into tables so that each section is asked in turn to each individual in the household. In this example, the topic block applies to Fred and Wilma but not to Pebbles (person 3 from Figure 1). When one (the order doesn't matter) household member has completed a section, this set of questions are asked of another respondent before moving on to the next topic. At the design stage for new surveys, much work goes into optimising the topic lengths to allow the movements between respondents and between topics to flow naturally.

**Figure 2.**

	AxName	TelProx	Wrking	SchemeETTrn	JbAway	OwnBus
QILO[1]	FRED		2	2	2	2
QILO[2]	WILMA		1			
QILO[3]						

### 2.3. The Annual Population Survey Boost Sample (APSB)

The APSB began in January 2004. This is a new survey comprising data from all annually independent sample components of the LFS (165,000 achieved households) together with a new survey (called the APS boost sample, APSB) of approximately 65,000 achieved household interviews per annum in selected areas of England. The APSB covers a sub-set of the topics in the LFS. By combining the outputs from these samples, we obtain better local area information.

As mentioned previously, the APSB has an annual panel element. Subsequent waves are carried out by CATI and first waves interviews can be carried out either by CAPI or CATI depending on whether a telephone number is available. This mixture of interviewing modes required the development of a tool to suit the needs of CATI and CAPI, and, bring together the elements of concurrent and consecutive interviewing. It is important when doing this to minimise any mode effects which may occur from the different structures of interview and to also ensure that the data has the same output structure irrespective of the type of interview. In other words, the data should appear in the same order in the Blaise data file whether the interviewing mode was CATI or CAPI.

### 3. Development of the APSB questionnaire: design approach

When developing the APSB questionnaire, it was an objective to adopt ONS standards for social surveys. This meant developing a questionnaire based on the CAPI format.

CATI interviewers involved in the pilot had difficulties navigating the CAPI-format questionnaire and accessing individuals' interviews. These problems resulted in pauses in conversation which interviewers felt were awkward and unacceptable. Among the criteria for success for the new APSB structure was for the questionnaire to be easily used in both CATI and CAPI formats.

It was also important to develop the mixed-mode Blaise instrument so that the fields were ordered in the same way whether the interview was CAPI or CATI. This would ensure that the data from the two modes of interview could be extracted in the same way with no matching process after the data collection phase.

The questionnaire has a switch variable at the start of the interview to code whether the case is CAPI or CATI. The design of the questionnaire required different routing depending on whether it was being administered CATI or CAPI. Telephone interviews needed to route each block of questions in turn to one individual before moving on to the next person (consecutive interviewing). The routing of the CAPI interview allowed a new set of rules to be followed causing the questionnaire to be broken up into topics which are looped to each individual in turn (concurrent interviewing).

It was important that there should be no negative effect on the common data structure and that each question should be referenced once within the questionnaire. Our solution was a structure of arrays of arrays which would allow these requirements to be met. We were not deterred by considerations of the arguable inefficiency of such a design, which would result in a high proportion of empty places in the data structure. As there was no performance cost, the design was efficient in ways which mattered in practice, enabling us to work with the standards we have developed for instrument design across surveys in ONS.

The design is best understood by examining first the data structure and then looking at how it was achieved at data collection. In the following example, shown in Figure 3, we are dealing with a block called *QPerson* which is repeated for every person in the household and a block nested within it called *QHealth* which is also an array and which represents here any topic block nested within *QPerson* (i.e. all topic blocks at person level). The example is for a 3-adult household.

If the **CATI** mode is selected the questionnaire uses the person level array to enter data into the first array of the topic (*QPerson*[1].**Qhealth**[1], *QPerson*[2].**Qhealth**[1]). In **CAPI** mode the household members are always routed to the first array of the *QPerson* block (**QPerson**[1]) and responses are entered into the different topic area arrays. To avoid filling the same fields and a 'display order conflict', the answers are coded into the topic area array person number *plus 1* (*QPerson*[1].**Qhealth**[2], *QPerson*[1].**Qhealth**[3]).

**Figure 3: Data structures for CATI and CAPI within a common framework of arrays of arrays as collected**

CATI

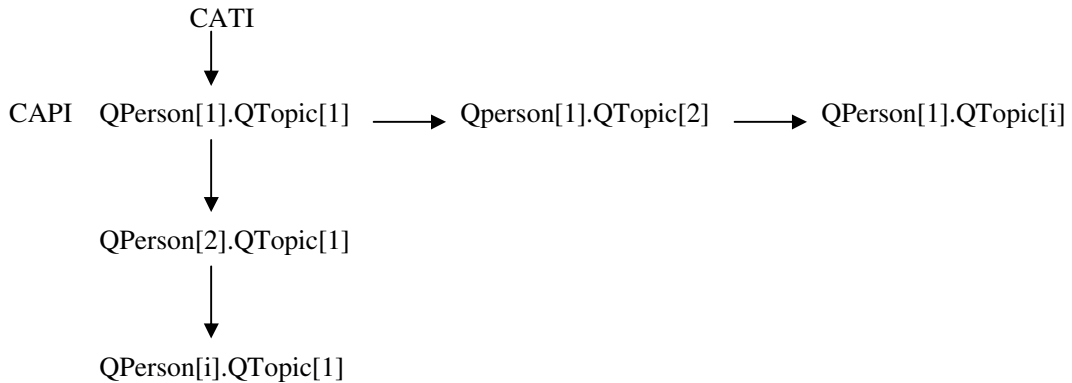
	QPerson[1]				QPerson[2]			QPerson[3]	
	Qhealth[1]	Qhealth[2]	Qhealth[3]	Qhealth[4]	Qhealth[1]	Qhealth[2]	Qhealth[3]	Qhealth[1]	Qhealth[2]
Mickey	,								
Minnie					,				
Donald								,	

CAPI

	QPerson[1]				QPerson[2]			QPerson[3]	
	Qhealth[1]	Qhealth[2]	Qhealth[3]	Qhealth[4]	Qhealth[1]	Qhealth[2]	Qhealth[3]	Qhealth[1]	Qhealth[2]
Mickey		,							
Minnie			,						
Donald				,					

To achieve this structure, the interviewers use different routing for CAPI and CATI, controlled by the switch variable. This is represented schematically in Figure 4

**Figure 4: The CAPI/CATI data collection matrix for each member of the household**



The data collected in the two modes then need to be integrated for output. This is done in Blaise (see below) to the format shown in figure 5.

**Figure 5: Data structures for CATI and CAPI within a common framework of arrays of arrays: for output**

	QPerson[1]				QPerson[2]			QPerson[3]	
	Qhealth[1]	Qhealth[2]	Qhealth[3]	Qhealth[4]	Qhealth[1]	Qhealth[2]	Qhealth[3]	Qhealth[1]	Qhealth[2]
Mickey	<b>CATI</b>	<i>CAPI</i>							
Minnie			<i>CAPI</i>		<b>CATI</b>				
Donald				<i>CAPI</i>				<b>CATI</b>	

## 4. Development of the APSB questionnaire: implementation in Blaise

The block QPerson, referred to in section 3, is shown below in Blaise and is followed by a detailed explanation of each element within the block. The block shown is an extract from the APSB questionnaire and contains only 2 blocks: one block of health questions (QHealth) and a block of questions about caring for family members (QCarer). Figure 6 shows the block as a whole. The following figures comment in detail on sections of it.

**Figure 6: Example of the APSB person block**

```
BLOCK BPerson

  INCLUDE "QHealth.HAR"
  INCLUDE "QCarer.HAR"           {These are the 2 blocks included in
                                  this example}

  LOCALS { BPerson }

      LTLoop1      : INTEGER      {LOCAL used for looping each topic block}

  RULES { BPerson }

      IF QSignIn.IntrType = TU OR QTHComp.NumAdult = 1 THEN {Routing for CATI OR
      QHealth[1]                                           one person household}
      QCarer[1]
      ELSE
      IF QSignIn.IntrType <> TU THEN
          QHealth[1] = EMPTY
          QCarer[1] = EMPTY
      ENDIF
      ENDIF

      IF QSignIn.IntrType = FTF AND QTHComp.NumAdult > 1 THEN {Routing for CAPI}
      FOR LTLoop1 := 2 TO DMHSize + 1 DO
          QHealth[LTLoop1]
      ENDDO
      ENDIF

      IF QSignIn.IntrType = FTF AND QTHComp.NumAdult > 1 THEN
      FOR LTLoop1 := 2 TO DMHSize + 1 DO
          QCarer[LTLoop1]
      ENDDO
      ENDIF

  ENDBLOCK { BPerson }

  FIELDS { DM }

      QPerson : ARRAY [1..16] OF BPerson
```

The routing for this block is determined at a higher level in the questionnaire. This block is an array and is repeated for each individual in the household. This allows parallel fields to be available for individual respondent interviews during CATI.

This block is split into 3 distinct sections – the include files and locals (figure 7); the rules for CATI (figure 8) and the rules for CAPI (figure 8). Each section is explained in detail below.

**Figure 7: Include files and locals**

```
BLOCK BPerson

  INCLUDE "QHealth.HAR"
  INCLUDE "QCarer.HAR"           {These are the 2 blocks included in
                                  this example}

  LOCALS { BPerson }

  LTLoop1      : INTEGER        {LOCAL used for looping each topic
                                  block}
```

The name of the block is *BPerson*. This part of the code contains the person level blocks contain the topic questions which will be asked by the interviewer to each respondent in the household. Each of these blocks is an array. The local is used to loop the topic blocks during a CAPI interview.

The second section contains the routing for CATI.

**Figure 8: Routing for CATI**

```
RULES { BPerson }

  IF QSignIn.IntrType = TU OR QTHComp.NumAdult = 1 THEN {Routing for
                                                         CATI}
    QHealth[1]
    QCarer[1]
  ELSE
    IF QSignIn.IntrType <> TU THEN
      QHealth[1] = EMPTY
      QCarer[1] = EMPTY
    ENDIF
  ENDIF
```

The code to determine that this is CATI is '*QSignIn.IntrType = TU*'. This set of rules also applies if there is only one adult in the household '*QTHComp.NumAdult = 1*'. The rules are the same as the CATI rules if there is only one adult in the household as the sections will be asked in turn with no need to move on to a second household member.

If the interview is CATI or for a single person household, all topic blocks are completed before moving onto the second person's interview. In this cut-down example, Person 1 is asked the health and caring questions before moving onto person 2.

The interview order for a 2 person household is:

Person 1, topic 1  
Person 1, topic 2  
Person 2, topic 1  
Person 2, topic 2

When a CATI interview is completed, data will be stored (as shown in the .bdb file) in:

*Qwhoint.Qperson[i].Qhealth[1].Health*

where *i* refers to the person number of respondent. *Qhealth[1]* refers to the first array of the block of health questions. *Qwhoint* contains the switch variable for CAPI or CATI.

Below is the code used for a CAPI interview of a household containing more than one adult. The code to determine this is: *QSignIn.IntrType = FTF AND QTHComp.NumAdult > 1*

### Figure 9: Routing for CAPI

```
IF QSignIn.IntrType = FTF AND QTHComp.NumAdult > 1 THEN {Routing for
    FOR LTLLoop1 := 2 TO DMHSize + 1 DO                                CAPI}
        QHealth[LTLLoop1]
    ENDDO
ENDIF

IF QSignIn.IntrType = FTF AND QTHComp.NumAdult > 1 THEN
    FOR LTLLoop1 := 2 TO DMHSize + 1 DO
        QCarer[LTLLoop1]
    ENDDO
ENDIF
```

To avoid a ‘*display order conflict*’ error the exact same questions as those used during the CATI interview can not be referenced again. With the first array for the topic questions filled, a CAPI interview will populate the second array of questions

Each block is referenced separately with the same set of conditions. During a CAPI interview, it is required that all eligible members of the household are asked one topic before moving onto the next topic. The interview order for a 2 person household is:

- Person 1, topic 1
- Person 2, topic 1
- Person 1, topic 2
- Person 2, topic 2

The responses will be coded into (as shown in the .bdb file):

*Qwhoint.Qperson[1].Qhealth[i+1].Health*

Responses are always coded into the first persons data (*Qperson[1]*) and into a different array for the topic which reflects the person number of the respondent plus 1 – as the 1<sup>st</sup> array has already been referenced within the questionnaire.

This gives the problem that, at this stage, the data is not coded in the same place in the .bdb file if the interviews are completed in different modes. This is rectified by a piece of code at a higher level than the person block which recodes the responses given during a CAPI interview to the same place as they would have been if the questionnaire had been administered via CATI.



**Figure 10.**

```
FOR LTLoop1:=1 TO DMHSize DO                                {code repeated for each household
                                                            member}
  IF QSignIn.IntrType = FTF AND QTHComp.NumAdult > 1 THEN  {IF CAPI}

    QPerson[LTLoop1].QHealth[1].QHealth1:=QPerson[1].QHealth[LTLoop1+1].QHealth1
    QPerson[LTLoop1].QHealth[1].GenHlth:=QPerson[1].QHealth[LTLoop1+1].GenHlth

    QPerson[LTLoop1].QCarer[1].Carer1q:=QPerson[1].QCarer[LTLoop1+1].Carer1q
    QPerson[LTLoop1].QCarer[1].Carer2q:=QPerson[1].QCarer[LTLoop1+1].Carer2q

  ENDIF
ENDDO
```

This code is repeated depending on the number of people in the household (*FOR LTLoop1:=1 TO DMHSize DO*) and applies if the interview is CAPI (*IF QSignIn.IntrType = FTF AND QTHComp.NumAdult > 1*). This recodes the responses collected during a CAPI interview into the place where they would have been coded had the interview been CATI.

For example, looking in detail at the first line of code within the equation:

**QPerson[LTLoop1].QHealth[1].QHealth1:=QPerson[1].QHealth[LTLoop1+1].QHealth1**

The first half of the equation: *QPerson[LTLoop1].QHealth[1].QHealth1* refers to the place in the data structure where data would be coded during CATI. *LTLoop1* refers to the person number of the respondent in the household. During CATI, data is stored in the person numbers block in the first topic array (*QHealth[1]*).

The second half of the equation: *QPerson[1].QHealth[LTLoop1+1].QHealth1* refers to where data is stored from a CAPI interview. During a CAPI interview data is always stored under person 1 (*QPerson[1]*) and under a topic array corresponding to their person number in the household plus 1, as the first array has potentially been filled by the CATI interview (*QHealth[LTLoop+1]*). This piece of code transfers the data from where it is collected via CAPI to where it would be collected via CATI.

This practice allows cases to be transferred between CAPI and CATI without overwriting any information collected at a previous interview in a different mode.

## **5. Results**

After 7 months of interviewing on the APSB, and over 35,000 completed household interviews, there have been no problems with data collection. Cases have been transferred from one mode to another without data loss. Data from the first 5 months of APSB has been analysed and merged with LFS data, showing that the data structure from the CAPI mode of interviewing is consistent with the CATI mode used on the LFS. Interviewers working on either type of questionnaire have not experienced problems.

Analysing results by mode is problematic since the mode selection process involves confounding factors. The LFS has a random sample covering the whole of England. The APS sample is different as it is used to boost the numbers of interviews in certain areas of England to achieve improved local area estimates. This means that the APSB on its own is not a nationally representative sample like the LFS.

## **6. Conclusions and further work**

This method of designing the questionnaire can be viewed as a success as it has allowed cases to be easily transferred from CAPI to CATI and vice versa without the loss or damage to data quality. Interviewers using the different methods have also had no problems. The complexities described here are transparent to them. From the design viewpoint, the complexities have been handled successfully within ONS's design standards and the requirement that any Blaise code should be understandable and understood by all its researchers.

At present, the APSB is the only ONS social survey which uses this design of Blaise instrument. There are very few social surveys which use telephone interviewing as a substantial part of interviewing. Most only carry out telephone reissues if a CAPI interview was not possible. If telephone interviewing becomes more common at ONS – especially with the prospect of the CPS (mentioned above), the design of Blaise instrument used on the APSB may become an ONS standard. Before reaching this stage, research is required into mode effects which may occur through the different methods. We will also need to assess more fully whether this is an efficient method of questionnaire design for the full range of ONS surveys.