

## Developing an Integrated Household Survey (IHS) Blaise Questionnaire from Four Major Social Surveys

*Tim Burrell (Office For National Statistics, UK)*

### 1. Introduction

The Integrated Household Survey (IHS – formerly known as the Continuous Population Survey) is a project to integrate four government surveys on which the UK Office for National Statistics (ONS) leads, and replace them with a single modular survey. A comprehensive integration of the entire survey process is proposed: from the creation of a unified field force of interviewers administering a common modular questionnaire, to the processing and production of outputs from a single common source.

This paper first details the IHS project, the paper then focuses on one aspect of the survey's development: developing an IHS Blaise questionnaire from four major social surveys. The IHS is due to begin in January 2008. The development of this has involved two field trials during 2005, a larger scale pilot in January 2006 and there will be a further pilot in 2007. This paper will describe the first attempt at programming the Blaise questionnaire for the field trials, lessons learnt, and the new design for the pilot in 2006. The paper will discuss the issues involved in designing and programming a Blaise questionnaire to handle the different topics and rules of the four surveys with a view to producing coherent and consistent outputs.

### 2. Background

The aim of the IHS is the need for UK National Statistics to produce better information on key social and economic variables between decennial censuses. These demands cannot be met within current survey arrangements.

The central objectives of the CPS are to:

- develop a world class modular survey system, better able to meet the information needs of the 21<sup>st</sup> century;
- provide more coherent, better quality information on which Government, stakeholders and the wider user community can base decisions;
- increase the precision of most existing statistical outputs at national, regional and sub-regional levels;
- create a range of new outputs, including inter-censal estimates of key socio-demographic variables at the sub-regional level;
- develop a survey system with the flexibility to accommodate other surveys at a later stage.

There are four surveys planned for integration. These are:

- the Labour Force Survey (LFS);
- the General Household Survey (GHS);
- the Expenditure and Food Survey (EFS);
- the National Statistics Omnibus Survey (OMN).

All the surveys attempt to conduct interviews with every household member, with the exception of the Omnibus survey where interviews are administered to one

randomly selected household member only. The surveys differ in their use of Computer Assisted Personal and Telephone Interview (CAPI and CATI). More information on these surveys is in Appendix A.

### **3. Design of the IHS questionnaire**

The proposed IHS involves a single modular Blaise questionnaire instrument which meets the information needs of the individual surveys. This questionnaire will comprise of a core set of questions, which will be administered to the whole sample and topic modules. The topic modules will meet the needs of the four surveys and will only be administered to a proportion of the sample. The limited number of different interview combinations will be composed of both the core and selected topic modules. Interviewers in an integrated field force will administer all interview combinations.

Designing a suitable core module is essential to the success of the IHS. The core module is intended to collect information about key Census type variables and other socio-demographic and labour market indicators. Because the core module is administered first, and to all respondents in the survey, it needs to be concise and relatively straightforward. This core module will be used as an aid to weight topic data as it will have a very large base of responding adults.

This core module will be followed by survey specific questions from the original four surveys. In some cases, topics covered in the different surveys are common. For example, detailed questions on employment and education are required on the GHS and LFS type interview combinations but not necessary on the EFS and Omnibus.

There are a number of other procedural differences between the surveys involved in the IHS. LFS interviewers readily accept proxy interviews as the interview is short and the information collected mainly factual. Proxy interviews are discouraged on the GHS and EFS as a lot of the information collected would only be known by the individual. The Omnibus does not accept proxy interviews due to the high volume of opinion questions.

The EFS has a diary questionnaire. Each individual within a sampled household must complete a two week diary of expenditure. Interviewers who have only ever worked on the LFS will not be experienced how to 'sell' a much longer survey involving a diary. Other differences between the LFS and GFF surveys include rules on whether to interview students living away from home and how to deal with multiple households.

One of the main challenges is the development of a questionnaire which can route effectively through the relevant topics and guide the interviewer to use practices that they may not be familiar with.

### **4. Questionnaire design for field trials**

The first IHS field trial took place in late 2005. The purpose of the early field trials was to both test the questionnaire but also to test working patterns and most importantly a new training scheme to educate all interviewers in the different practices on each of the surveys. The introduction of the IHS will involve a number of changes to interviewer working practices as well as a new questionnaire. At present, ONS has a field force dedicated to the LFS. Another field force deals with all other ONS surveys (the exception being the International Passenger Survey

which also has a unique field force based at ports and airports). This field force is named the General Field Force (GFF). The IHS involves combining these two large field forces. All interviewers from these field forces will need to be able to work on all the surveys involved in the IHS.

Field force integration presents more problems than just that of interviewers needing to be taught new surveys with different subject matter. There are many practical differences between the two types of surveys (LFS and others). The main difference is in field pattern and the amount of time each interviewer has to complete their cases. LFS interviewers receive a weekly batch of cases. At the end of each week all cases must be returned. There is some scope to re-issue these cases but the hope is that all cases are dealt with in one week. In the following week, a new batch of cases is issued. On GFF surveys, cases are drawn in monthly batches. Cases are drawn and sent to interviewers on a monthly cycle. Interviewers have one month to work through all the cases. This allows far more flexibility in working patterns. One of the goals of the IHS is to test how these working practices can be integrated.

Although the trial was not too concerned with outputs and consistency with current estimates (due to the small sample size) it was still important to have a functional integrated single questionnaire based on current ONS standards and practices. The next section describes the development of the Blaise questionnaire for these field trials.

Since 1990, there has been an ongoing programme of harmonisation of questions, concepts and classifications. Household surveys at ONS and across the wider UK Government Statistical Service (GSS) have adopted many common standards. At ONS we have a bank of standard Blaise question blocks based on these harmonised topics as well as other ONS standard topics. For example, household composition, ethnicity, country of birth and standard questions on employment, health and qualifications. Where these standard questions are available and if they are appropriate, a survey can take the block of questions and insert them into a questionnaire. Therefore the same questions are used across all surveys. In an integrated survey they can be administered across all module types as a 'core' set of questions, if appropriate. The IHS aims to build on and improve standardisation.

Before any cases are sent to interviewers they are assigned a code depending on what the type of interview the case is. This is essential for the interviewer to know what to expect and also for routing through the questionnaire.

#### **4.1. Core interview**

The first stage in designing the questionnaire was to establish which topics should be included in the core interview. The core includes information about key Census type variables and other socio-demographic and labour market indicators.

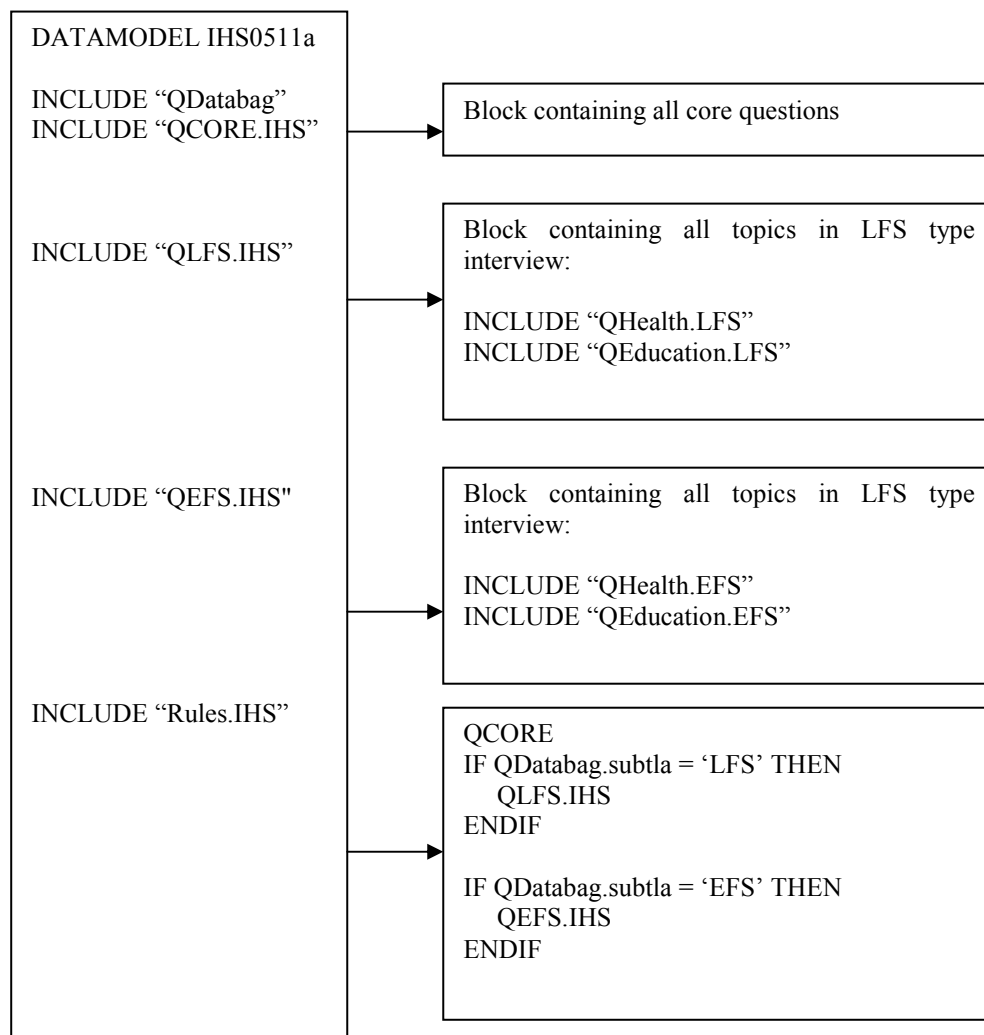
Some surveys collect more detailed information on certain topic areas. For example, an Omnibus type interview only asks a few education related questions to code the respondent's highest qualification, the LFS has a far longer set of questions to establish the same derived variable as well as many other education related variables. The shorter set of questions were included in the core and where necessary, supplementary education questions were asked later in the interview.

#### **4.2. Programming the topic modules**

The four original survey questionnaires are already programmed in Blaise. For simplicity, in terms of programming and time, once the core questionnaire had been devised and programmed, it was possible to remove these core questions from

individual survey questionnaires and add the survey specific questions into the IHS questionnaire. This was fairly straight forward, due to way we programme Blaise questionnaires at ONS. Figure 1 shows the simplified version of the Datamodel.

Figure 1. Datamodel structure for field trial questionnaire

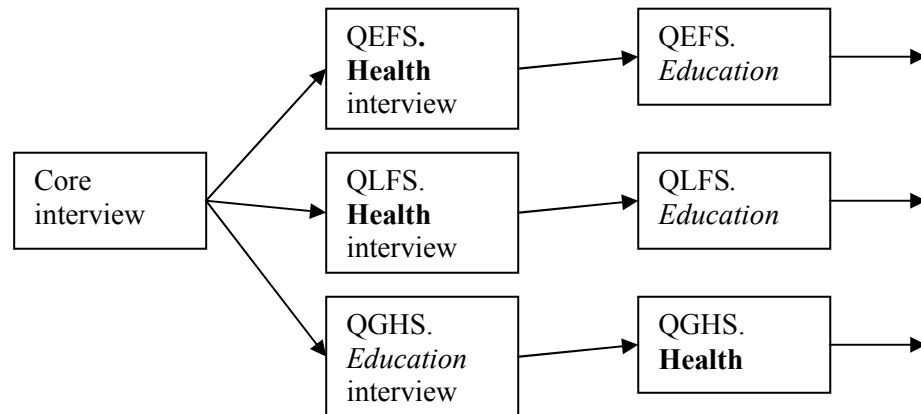


The diagram above shows a simplified version of the Datamodel. The right side of the diagram shows an example of the contents of the INCLUDE files. The first block, 'QDatabag' is where information is stored about the type of interview. All core questions are contained in the block 'QCore.IHS'. The survey specific topic blocks follow and are contained within an include file for each of the different survey types. QLFS.IHS contains the survey specific questions for the LFS which are not already in the core. Likewise, QEFS.IHS contains the survey specific questions for the EFS.

The questionnaire, programmed using this method, worked well in the field. The interviewer could complete the interview and all the correct information was asked depending on the type of interview. There were however some problems. Firstly, repetition of questions, while the core records basic health and education, some modules follow these up in great detail. This interview involved moving away from these topics and having to go back to them at a later stage. For example, key education questions are asked in the core. Then more detailed education questions are required at a later stage in the interview. This resulted in the interview being disjointed and respondents asking 'haven't I already answered that question?' In

this model, there is no consistency in subject order. For one interview combination health was followed by education. For another combination education was followed by health. Interviewers found this confusing as interview did not flow very well. An example is provided in figure 2.

Figure 2. Questionnaire structure for field trials:



One of the goals of the IHS is to have an integrated processing system where data checks can be performed on the data as a whole. It is desired that following the interview a standard set of processes using standard tools can be carried out on the whole dataset before being delivered to researchers for in depth analysis. At present, the same checks take place on the four surveys but are using different tools. An integrated processing system will remove this duplication of effort and cut down the number of analysis tools used.

The structure of the Blaise questionnaire used in the field trial led to problems at the data processing stage. This is because the same variables appear in the datafile more than once but with different names. For example, *QEFS.Health* and *QLFS.Health* contain the same variables but they are listed separately in the datafile. (Figure 3 provides an example)

Figure 3. Data structure during field trials:

Type of interview	Case number	Core info	QEFS. Health	QLFS. Health	QEFS. Education	QLFS. Education	QGHS. Education
EFS	001	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
EFS	002	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
LFS	003	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
LFS	004	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
GHS	005	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
GHS	006	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

The datafile was very large due to the duplication of fields. This meant that it was not possible to process the whole datafile using standard ONS systems. The datafile had to be split into topic type before it was processed.

On the whole the questionnaire worked in the field. However, problems with repetition within the questionnaire and interview flow, and subsequent problems with processing data meant we had to restructure the questionnaire for the pilot.

## 5. The pilot questionnaire

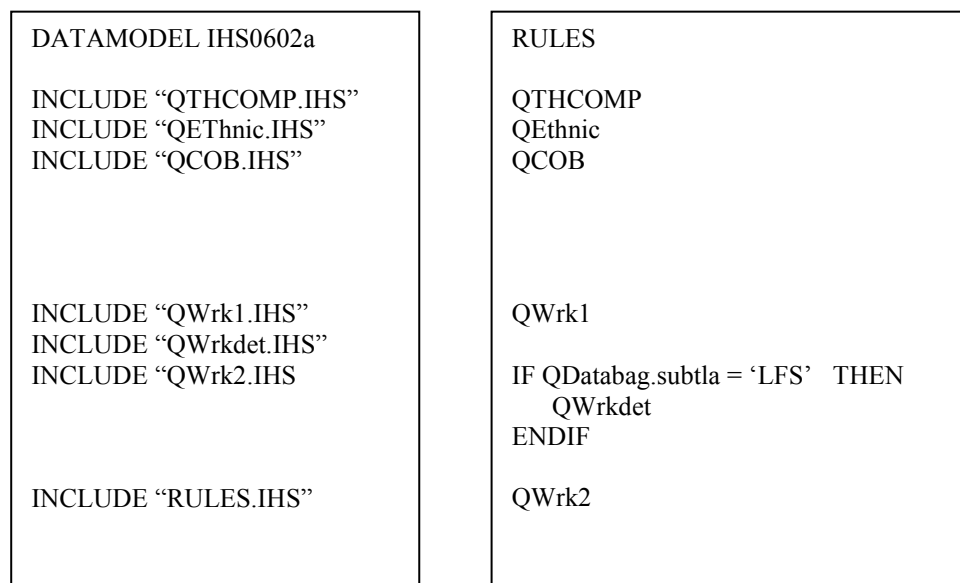
The next stage in the IHS development was a larger pilot in starting in February 2006 Previous field trials had established that interviewers could adapt to working on the IHS. This pilot gave us the first opportunity to compare outputs and response rates with the existing surveys.

### 5.1. Design of the pilot questionnaire

For the pilot, we needed to ensure no variables were duplicated. We hoped this would help with both the flow of the interview. This would also provide a platform from which to develop an integrated approach to processing the data.

The core section contains questions on basic employment details. The full LFS style interview contains far more detailed questions on employment. It was decided that for the pilot questionnaire we would ask whole topics together and route to the appropriate questions based on interview type. The example shown in figure 4 illustrates the interview structure for the employment section. The first three blocks listed (*QTHComp*, *QEthnic* and *QCOB*) are blocks which, amongst others, make up the core interview. *QWrk1* and *QWrk2* make up the core information on employment. *QWrkdet* contains all the detailed work information and only applies to the LFS interview type.

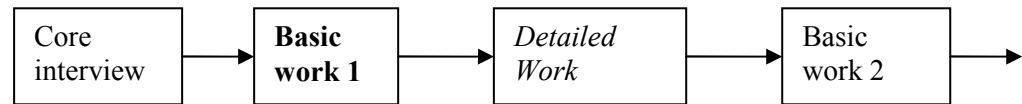
Figure 4. Questionnaire structure for pilot



The diagram above shows, on the left, a simplified example of the Datamodel with each block referenced. On the right the diagram shows the rules section with the detailed work questions only being asked on LFS type interviews. The same practice occurs in other topic areas during the interview, for example, health, education and benefits. The result was that all variables are referenced just once in the questionnaire. This helps with the flow of the interview as questions will always appear to the interviewer in the same order. It also cuts down the size of the datafile produced.

Figure 5 shows the new questionnaire structure. After the first core section the work section is asked. Basic work 1 and basic work 2 (*QWrk1* and *QWrk2* in figure 4) are asked to everyone. Detailed work is asked in LFS type interviews. After the work section, the interview returns to the core, if appropriate.

Figure 5. Questionnaire structure for pilot



With the new structure in place, the datafile is much shorter as all the duplication has been removed. The data can be processed as a whole using ONS standard processing systems. Figure 6 shows an example of the structure of the datafile

Figure 6. Data structure for pilot

Type of interview	Case number	Basic work 1	Detailed work	Basic work 2
EFS	001	<input checked="" type="checkbox"/>	●	<input checked="" type="checkbox"/>
EFS	002	<input checked="" type="checkbox"/>	●	<input checked="" type="checkbox"/>
LFS	003	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
LFS	004	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
GHS	005	<input checked="" type="checkbox"/>	●	<input checked="" type="checkbox"/>
GHS	006	<input checked="" type="checkbox"/>	●	<input checked="" type="checkbox"/>

All basic work information for all interview types is held in the same place and all checks on work information or derived variables can now be run in one go.

## 6. Other developments

The IHS provides an opportunity for ONS to review current practices across the four current surveys. Although we have standards for Blaise programming and screen design and layouts, there are still some differences between surveys. The development of the IHS questionnaire provides an opportunity to remove these differences and standardise the whole questionnaire. A great deal of change is involved for interviewers, one of the goals of the questionnaire is that it should be easy for the interviewers to navigate and use.

### 6.1. Concurrent interviewing

At the time of the IHS questionnaire development, the LFS was the only survey involved which had a panel element. Five interviews are conducted with each selected household on a quarterly basis. All first wave interviews are conducted as face-to-face interviews with follow up interviews by. Due to the high volume of telephone interviews, respondents are interviewed consecutively.

The primary interviewing mode for all the other surveys is face-to-face. These interviews use concurrent interviewing. The different questionnaire topics are broken up and asked in turn to each respondent in turn. This helps keep all respondents interested and involved in the survey.

It was decided that, for the IHS questionnaire, concurrent interviewing would be used for all interview types. This meant a change for LFS telephone interviews. The main criticism from telephone interviewers was: if a case is partially completed, for example, in a four person household, the interview of only two people is completed. On re-entering the case to interview the final two people the responses from the original interview are present in the questionnaire and there is a danger they can be overwritten. This is also a problem of face-to-face interviews but occurs far less.

To avoid this problem, a category was added to an existing field to allow the interviewer to code whether an individual's interview had been completed. If this is the case all the responses from the completed interview were set to 'KEEP' fields and were 'greyed out' in the data entry programme. This avoids any overwriting and was very well received by the interviewers.

## **6.2. Q by Q Help**

For both the field trial and the pilot, we enhanced the online help available to interviewers. The Q by Q online help facility contains procedural and question specific help, this is described in a further paper to be written and presented at IBUC 2006 called 'Training and Learning using Blaise' (Setchfield 2006):.

## **6.3. Additional Blaise features**

We have introduced the following features on the IHS questionnaire which were not standard on ONS surveys before 2006:

- Q by Q help facility, as mentioned above.
- A calculator is now available for interviewers to use at anytime during the interview.
- For certain date fields within the questionnaire, a date picker is available.
- All parallel blocks are now available through using tabs or through using the keyboard.

## **7. Next steps**

The IHS questionnaire is far more advanced now. The process of preparing for the his, which began in 2004, has illustrated the importance of a coherent approach to survey design, each part of the survey process has to work together to achieve the best results. For example, it is important to consider the outputs when writing the Blaise questionnaire..

ONS is currently developing an integrated processing system for the IHS. The next pilot, to be carried out in 2007, will also test this system and attempt to reproduce current outputs.

## **8. References**

Rob Bumpstead, (2004): Integration of UK government household surveys: Development of a modular Blaise instrument for a proposed Continuous Population Survey. International Blaise Users Conference 2004

Colin Setchfield, (2006): Training and Learning using Blaise. International Blaise Users Conference 2006

## **Appendix A**

Survey for integration:

A Labour Force Survey (LFS) has been carried out in the UK since 1973 and in its present form since Spring 1992, providing a wide range of data on labour-market statistics and related topics such as training, qualifications, income and disability.

The General Household Survey (GHS), the first multi-purpose household survey, started in 1971 and covers a wide range of social and socio-economic topics. The main aim of the survey is to collect data on core topics including housing, employment, education, health and family information.

The Expenditure and Food Survey (EFS) started in 2001 bringing together two surveys, the Family Expenditure Survey (FES) and National Food Survey (NFS), that were both well established and important sources of information, charting changes and patterns in Britain's spending and food consumption since the 1950s.

The National Statistics Omnibus Survey (OMN) is a regular, multi-purpose survey that started up in 1990 in order to provide quick answers to questions of immediate interest and information on topics that do not require a full, in-depth survey.



## **Session 2.2: Cati Considerations**

### **Can We Reach You by Telephone?**

*Fannie Cobben*

*(Statistics Netherlands)*

### **Experiences of Usage of Blaise in CATI Surveys**

*Stefania Macchia & Manuela Murgia*

*(ISTAT, Italy)*

### **NASS CASIC Survey Administration System**

*Everett Olbert & Roger Schou*

*(National Agricultural Statistics Service, USA)*

### **A Multi-mode CATI-Web Survey Experience with Blaise**

*Craig Ray*

*(Westat, USA)*

