

STRATEGIC DIRECTIONS IN HOUSEHOLD SURVEYS DEVELOPMENT: INTEGRATING THE ONS' CONTINUOUS HOUSEHOLD SURVEYS

Nikki Bennett¹

ABSTRACT

The Office for National Statistics (ONS) plans to integrate the Government household surveys on which it leads into a single continuous population survey to achieve better value for money from its survey programme and to deliver a range of new statistical outputs to meet increasing demand for information. A comprehensive integration of the survey process is proposed: an integrated field force of interviewers administering a common modular questionnaire, an unclustered sample and fieldwork design, more integrated processing system, production of outputs from a single common source and a range of new annual outputs on a rolling quarterly basis.

KEY WORDS: Integration, Household surveys, Modular, Unclustered design

RÉSUMÉ

Le Bureau national de la statistique (BNS) prévoit intégrer les enquêtes auprès des ménages du gouvernement à une enquête unique et continue pour assurer la rentabilité de son programme d'enquête et fournir une gamme de nouveaux résultats statistiques qui satisferont à la demande croissante d'information. Une intégration complète du processus d'enquête est proposée et comportera : une équipe d'interviewers sur le terrain qui administre un questionnaire modulaire commun, un plan d'échantillonnage non groupé et du travail sur le terrain, des systèmes de traitement mieux intégrés, la production de données provenant d'une seule source et une gamme de données annuelles produites de façon continue sur une base trimestrielle.

MOTS CLÉS: Enquête de ménage; intégration; modulaire; plan non groupé

1. INTRODUCTION

1.1 Background

In the past decade or so the United Kingdom (UK) has witnessed an increased need for survey information to support policy information needs in government. Sometimes these needs can be met through minor modifications to the design of existing surveys, or new surveys can be set up in order to collect the necessary information. However, both approaches are generally expensive and sometimes inefficient. In trying to minimise these inefficiencies the Office for National Statistics (ONS) in the UK has adopted a more pro-active approach to conducting household surveys in which flexibility and integration are key principles.

This paper describes ONS' work programme to integrate the Government household surveys on which it leads into a single module-based continuous survey to achieve better value for money from its survey programme and to deliver a range of new statistical outputs to meet the ever increasing demand for information. A comprehensive integration of the survey process is proposed: an integrated field force of interviewers administering a common modular questionnaire, an unclustered sample and fieldwork design, more integrated processing system, production of outputs from a single common source and a range of new annual outputs on a rolling quarterly basis.

1.2 Surveys for Integration

The Integrated Household Survey (IHS) will initially bring together the General Household Survey, the Labour Force Survey, the Expenditure and Food Survey, the Omnibus Survey and the English Housing Survey.

¹ Nikki Bennett, Office for National Statistics, 1 Drummond Gate, London SW1V 2QQ, United Kingdom, nikki.bennett@ons.gsi.gov.uk

The illustrative diagram in Appendix A shows how the surveys will be combined into a single module-based approach.

- 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. In 2005 the GHS sample was changed from a cross-sectional to a longitudinal design in order to meet Eurostat requirements.
- 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.
- 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. In recent years the quarterly LFS has been supplemented by a series of annual boost samples in England, Wales and Scotland.
- 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.
- The English Housing Survey is a new survey on the housing situation and conditions of households in England. It will have an annual sample size of approximately 17,000 households, 8000 of which will be followed up by a second survey which will assess the physical condition of the household's accommodation. The survey is a merger of two existing surveys, both of which are widely used in government to support information needs in relation to housing policy.

1.3 Household Surveys in ONS: History and Context

ONS plays a central role in collecting data for its own statistical outputs, and for those of other departments in the Government Statistical Service (GSS). A majority of staff in ONS is involved in the design and collection of a wide range of data sources either for economic or social statistics. This paper focuses on the household survey operations, sometimes also referred to as 'Social Survey'.

ONS and its various predecessors have been carrying out household surveys for over 60 years to provide the sources for social statistics produced by ONS and the GSS. Throughout its existence various survey developments took place in line with the expansion of government-funded survey taking internationally. However, during the 1990s the Social Survey area in ONS underwent greater change than at any time in its history (ONS, 1992). There was an extraordinary growth in the number of surveys launched, and several large surveys were instigated in this period. Other important developments were the use of information technology in interviewing (Computer-assisted Personal Interviewing (CAPI) and Computer-assisted Telephone Interviewing (CATI)) and the move to the competitive tendering of surveys.

The 1990s were a time of change across UK government, with considerable emphasis on the better management and targeting of resources. This affected Social Survey in two ways. On one hand, the drive for greater efficiency and cost-effectiveness changed the way government surveys were commissioned. By 1994 it was usual for government household survey work to be put out to competitive tender. This meant Social Survey competed for surveys along with other survey organisations in the private and voluntary sectors.

On the other hand, new demands for high quality information led to a substantial increase in the need for surveys. While information from administrative records could be used in some cases, in others surveys were needed to provide baseline data for devising policy targets and monitoring progress against them. There was a need for both ad hoc surveys (on a one-off basis, but sometimes repeated after a few years) and continuous surveys (where the fieldwork is on-going and new samples are drawn each year). If UK survey research in the 1980s was characterised by repeat or serial ad hoc surveys, in the 1990s it was marked by the creation or expansion of large continuous surveys which focused on departmental policies. Departments wanted their own continuous surveys, with the flexibility to adjust the content from year to year depending on the policies and targets of most interest.

These rapid and radical changes prompted the need for a re-think of how ONS managed its survey portfolio, and eventually for a comprehensive strategy for how household surveys in ONS were carried out.

The basis for the new vision was laid in the early 1990s with the introduction of survey harmonisation principles (Manners, 1996). Although the large continuous ONS surveys collected information on different topics, all surveys included questions on basic socio-demographic characteristics of respondents and their households. Because these surveys were designed at different times to meet different needs they included slightly different questions and response categories to collect information on the same basic variables. This made it difficult to establish coherence across estimates from different surveys and for data to be pooled to give large sample sizes and more robust estimates. This led to a GSS-wide initiative to produce common survey variables and the implementation of harmonised survey inputs: questions and associated methods such as edit checks. Throughout the 1990s these harmonised survey inputs were developed for a range of basic socio-demographic variables and were implemented on the major ONS surveys and, where feasible, on other surveys carried out on behalf of other government departments in the GSS.

Apart from the lack of coherence in the various data collection instruments for these surveys there were also different sample designs and fieldwork practices, often developed in isolation because the survey design had to meet different requirements. For example, the Labour Force Survey (LFS) operated on an unclustered sample design, whereas the other continuous surveys were based on clustered samples. There were also separate fieldforces for the LFS and the other surveys, each with their own management and working practices.

At the same time there were a number of other drivers which have contributed to the development of a new strategic framework for ONS household surveys (Dunnell, 2005). The drive for cost efficiencies across the public sector means there is an increasing need for efficiencies in data collection and processing. There is also increasing recognition of the need for ONS to produce better information on key social and economic variables between decennial censuses, for a range of policy purposes, and to meet the increasing demand for regional and sub-regional information. Demand for small area statistics has grown rapidly in the last decade.

Increasingly it was recognised that these various new developments could not be accommodated within the existing survey arrangements. Individually, surveys had reached their limit in terms of length and burden and data were not easily pooled across surveys as their designs were different. In addition, the existence and maintenance of separate fieldforces, instruments, and processes represented a duplication of effort and sub-optimal use of limited resources.

These drivers led to the launch of the Integrated Household Survey development project, with design and testing work starting in earnest in 2004/5.

2 IHS DESIGN AND METHODOLOGY

2.1 Design and Fieldwork

The IHS will adopt an unclustered design for all its components, largely based on the sample and fieldwork design for the current Labour Force Survey which already uses an unclustered design. At the same time some LFS fieldwork practices will come in line with the other surveys. Under the IHS, each interviewer will be responsible for delivering all the required interviews in a small geographical area close to where they live, thus reducing travel time and cost and increasing productive contact and interviewing time to help maximise response.

The IHS questionnaire will be designed as a single modular survey instrument comprising:

- a core set of questions administered to the whole sample providing information on key variables for all IHS households and persons;
- topic modules administered to parts of the sample providing information on variables for which sufficient precision to meet policy needs can be obtained from a portion of the IHS sample;

From these two components a small number of viable interview combinations can be formed so that the core can be matched with specific topic modules.

The core sample will be in the region of 220 thousand households and more than 450 thousand adults per annum, making it the largest ever continuous survey to be conducted in the UK.

2.2 Core information

The core set of questions needs to be relatively short and straightforward so that total interview length for core and topic modules remains viable. Indeed, it is essential to the success of the survey that the core questionnaire does not become unduly burdensome. Core questions need to meet all or some of the following criteria:

- a classificatory variable essential for analysis
- an output for which there is a clear requirement for a high level of precision nationally or regionally, and not provided elsewhere
- an output for which there is a clear requirement for reporting at a sub-regional level, for example to local authority or health authority district level, and not provided elsewhere
- question(s) that can be administered by either face-to-face or telephone interviewing, and for which proxy responses are acceptable
- question(s) which would not adversely affect response to the survey as a whole

In early field trials, there was a strict adherence to a common core module of questions at the start of the interview. However, this raised a number of serious concerns about interview flow, the need to return to topics later in the interview and overall interview length. In later trials, the approach was changed so that some topic-specific questions were interwoven with core questions to improve flow, reduce interview length (and also minimises potential for discontinuities in key time series). This represented a trade-off between the ideal and a more pragmatic solution.

2.3 Core sample outputs

The very large sample of core data will provide better quality and more reliable estimates on key social and economic variables between decennial censuses. It will also help to meet the increasing demand for regional and sub-regional information that cannot currently be met by the existing surveys. A new set of rolling annual core outputs will be produced on a quarterly basis in addition to the existing topic-specific outputs. For example, banded household income will be collected on the core module and together with information on topics such as housing, health, employment and education will provide regular information down to local authority district level.

A two-phase weighting process will be introduced for certain IHS outputs. Two-phase weighting has the potential to produce worthwhile precision gains in suitable component survey variables on the IHS, as well as enforcing consistency between component survey variables and the whole IHS sample on selected core variables.

It is proposed that the IHS core microdata should be accompanied by a detailed amount of metadata - particularly in the use of weights, design, and combining core data with component survey data. The IHS component surveys will also require supplementary metadata to set out differences between the 'old' and the 'new' survey and how to combine these data. Documentation will also be necessary on the appropriate dataset to use for particular analyses.

A small working group within ONS has been set up to take forward more detailed work on the IHS outputs and publication strategy over the coming months.

3 DEPENDENCIES

3.1 Infrastructure Requirements

To launch the fully unclustered IHS sample and fieldwork design, ONS is dependent on improvements to its existing Survey Case Management System (SCMS) and developing a new set of survey processes to support the management and delivery of IHS core data. A considerable reorganisation and re-engineering of fieldwork operations to provide an integrated interviewer fieldforce capable of administering the new modular survey in small geographical areas across the whole country is also required.

3.2 Single Integrated Fieldforce

The integrated fieldforce is being phased in over time and this has been successful to date. Activities relating to recruitment, training, management and support of interviewers have already been standardised and integrated, bringing together what was, essentially, two separate fieldforces and functions.

3.3 Improvements to the Survey Case Management System

The launch of the SCMS will enable ONS to allocate work to interviewers in a more flexible and efficient way in the future and better to monitor and manage their work across all surveys. This will be an important first step towards the introduction of the new IHS sample and fieldwork design. Improvements to the SCMS have proved more challenging than anticipated due to a number of issues that have arisen concerning functionality, availability of skilled IT resources and the approach to testing. This has led to a revised timetable for the implementation of both the SCMS and some elements of the IHS design in recent months.

4 IHS IMPLEMENTATION PLANS

4.1 Timetable for Implementation

On current plans, that have changed since the ONS presentation at the SSC 2007 meeting, the IHS will be implemented in stages during 2008 and 2009 as follows:

- January 2008 – introduce the core questions to the Expenditure and Food Survey (EFS), General Household Survey Longitudinal (GSL) and Omnibus surveys) using their existing clustered designs;
- introduce the new IHS core questions to the 'Work' survey module (currently the Labour Force Survey and boost surveys) in October 2008 (was April 2008);
- move to an unclustered sample and fieldwork design for the EFS, GSL and Omnibus modules of the IHS in April 2009 (was October 2008).

Under this plan, the SCMS will be used for case-by-case allocation of interviewers' work from October 2008 fieldwork onwards, although only the LFS and English Housing Survey (EHS) will already have unclustered samples at this point. The unclustering of the remaining modules will not be made until April 2009 following a carefully managed, staged introduction of mixed workloads to minimise the risk of any disruption to response or quality of data. Introducing the IHS core questions on the LFS needs to be done at the same time as the SCMS comes in because of inter-dependent links between the existing questionnaire and the existing case management and processing system.

The EHS module of the IHS will be launched as planned in April 2008 using an unclustered design. It will, however, be based on the existing legacy systems that improvements to the SCMS will replace in October 2008.

5 BUILDING ON SURVEY INTEGRATION

5.1 Household Survey Framework for the Future

Although the current project is working towards the implementation of the existing ONS continuous household surveys, the aim of the project is to build a framework that offers a high degree of flexibility so that new surveys can become part of the modular survey system.

An essential element of the IHS development is to 'future-proof' the survey framework. The sample structure is designed so that wholesale changes will not be required, even if sample size requirements alter substantially. The modular structure of the survey instrument will readily accommodate new topics, while a range of survey types and features can be incorporated within the IHS survey system. For example:

- interview combinations with cross-sectional, quarterly or annual panel designs;
- diary components and other self-completion elements;
- telephone or personal interviewing;
- interviews with all household members or with individuals sub-sampled within a household, or combinations of both.

Hence, the data collection method chosen for each topic module or interview combination can be based on statistical requirements rather than the constraints of a particular survey vehicle.

The significant power of the IHS design is increasingly recognised across the Government Statistical Service. In the last year ONS has successfully proposed to integrate two new surveys in the IHS: the English Housing Survey (from April 2008) and the Longitudinal Disability Survey of Great Britain (from 2009). Both surveys have different requirements in terms of their sample and data collection designs.

As described earlier, the English Housing Survey is a survey on the housing situation and conditions of households in England. It will have an annual sample size of approximately 17,000 households, 8000 of which will be followed up by a second survey which will assess the physical condition of the household's accommodation. The survey is a merger of two existing surveys, both of which are widely used in government to support information needs in relation to housing policy.

The Longitudinal Disability Survey of Great Britain will consist of a baseline survey of approximately 50,000 individuals which will assess their disability status. This baseline survey will then be followed by a number of data collection waves for the group of respondents identified as disabled and for a subsample of the non-disabled population. The survey aims to chart the respondents' experiences of disability over time, collecting information to explore relationships between disability and a range of areas including work, education, income, transport, independent living, social participation and attitudes. Information from the survey will be used to inform work across and beyond Government.

Although these two designs are very different, the flexibility of the IHS framework means the survey designs can be adapted so that optimally they use the benefits associated with the IHS. Both surveys will benefit from using an unclustered sample design, and the associated field work management infrastructure. They will also take on board the IHS core questionnaire, followed by their specific topic module.

On the whole, the benefits from using the IHS framework are either statistical, in cost efficiency, or in both.

5.2 Statistical Benefits

In terms of statistical benefits, this paper has already described the advantages of the IHS core questionnaire in terms of delivering increased coherence across the different survey sources. Given the high profile of the IHS and its ability to provide inter-censal estimates of the UK population it is recognised its outputs will provide key national and regional estimates on UK households and individuals. When new surveys come on board in the IHS they will benefit from a greater coherence with these IHS population-level estimates because they will have collected the information through the same core questionnaire.

A second, related statistical benefit is the ability to use two-stage weighting techniques, in which the estimates from the specific topic modules can be weighted to key estimates based on the full IHS core. This will further increase the precision of certain key estimates.

A third statistical benefit is that the IHS core will be aligned (as far as possible) with the next 2011 Census. This has a number of advantages, one of which is the ability to provide a powerful data source for estimation purposes at small area levels. ONS has carried out work on the development of synthetic estimation at small area levels for various survey estimates after the 2001 Census and this approach has obvious benefits for the production of small area level statistics. Using the Census in this way would provide survey sponsors with detailed data at small area levels which could be used for policy planning purposes.

5.3 Efficiency savings

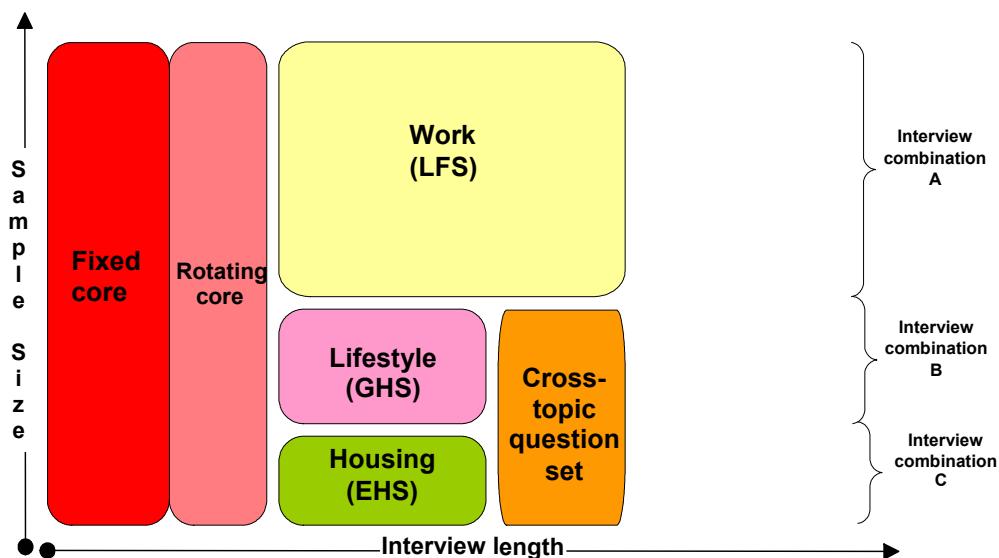
A second series of benefits stem from a number of efficiency savings in survey costs. The new survey framework will deliver savings from integrating the existing two field forces and reorganisation of data collection functions, facilitated by a new Survey Case Management System and communications improvements. Furthermore, because of the larger sample size the IHS will deliver economies of scale in survey management overheads, data processing and quality assurance, all of which means that we should be able to provide high quality surveys at a lower cost.

5.4 The cost / precision balance

Finally, the IHS unclustered sample design (which is unique in the UK) offers clear benefits in terms of the balance between statistical precisions and cost. Because of the unclustered sample new surveys can potentially reduce their sample size (which will reduce costs) to achieve the same level of precision. Or, if survey sponsors choose to, the IHS design can achieve greater precision of some estimates at the same cost. Essentially the IHS framework gives us the opportunity to sample respondents at an optimal level, to produce estimates to the required level of precision within a given cost. The ability to offer this flexibility on the cost/quality dimension of a survey design is a powerful and unique feature of the IHS.

Appendix A

Appendix A: IHS illustrative diagram of modular structure



REFERENCES

- Dunnell, K. (2005) "Towards a single continuous population survey for the UK". *Keynote paper presented at the 2005 Conference of the Association of Survey Computing*. London: ASC.
- Manners, T. (1996) 'Harmonised Questions for Government Social Surveys'. in: *Social Survey Methodology Bulletin*, 38. London: Office for National Statistics.
- ONS (2001) *60 Years of Social Survey 1941 – 2001*. London: Office for National Statistics.