



Advisory Group Paper AG (05) 05

2011 Census Testing and Evaluation: Strategy

1. Introduction

Early planning, testing and evaluation is an essential process for the success of the 2011 Census. This process begins with identifying and then testing innovations, allowing time to make changes and then to re-test where necessary. The evaluation of tests also includes the assessment of the procurement for equipment and outsourced services that must be in place for conducting the Census.

Action:

Advisory Group members are asked to note the 2011 Census Testing and Evaluation: Strategy

The 2011 Census Design aims to use: new questionnaires, new delivery and collection methods, early consultation, stronger partnerships, new out sourced technologies and operational targeting methodologies including enumeration, follow-up and publicity. Overall, the testing and evaluation process will enable the Census Division to explore these innovations by confirming, or otherwise, the effectiveness of all aspects in the proposed 2011 Census Design, in particular its feasibility and impact on levels of response. The key areas requiring careful testing are:

- I. The statistical design of the Census, including questionnaire content and design, coverage assessment methods and processing rules;
- II. The operational design of the Census field work, including post enumeration surveys;
- III. The operational design of the Census systems; and
- IV. The ability of external suppliers to provide the required systems, products and services.

Within each of these areas, the key factors that the Office for National Statistics (ONS) would like to test and/or evaluate are:

I. Statistical Design

- the accuracy of all questions, definitions and instructions within the Census questionnaire to provide the required information, particularly the new questions on capturing multiple residence;
- the public's likely response to specific questions, particularly income and the impact on overall and item non-response;
- the impact on quality and coverage of new modes of response, for instance internet and telephone technology, in providing information comparable to that from paper questionnaires;
- our ability to build ways of coverage assessment and adjustment into the census questionnaire and operation, including the design and implementation of a Census Coverage Survey;
- the use of administrative and qualitative data for assessing response behaviour of hard-to-find/enumerate targeted areas and population sub-groups; and
- the effectiveness of our data quality management, including data capture and quality assurance of the results.

II. Operational Design – Field work

- the ability of our pre-enumeration intelligence in supporting the allocation of our field force (such as assigning different enumeration and publicity approaches to different geographic areas and population sub-groups);
- the effectiveness of post-out, both in terms of the impact on response rates and accurately delivering questionnaires;

- the effectiveness of the increased role of local and community liaison, particularly involving partnerships with Local Authorities;
- our ability to manage a flexible field force during the Census operation, mainly during follow-up;
- our ability to gain incomplete/missing data using telephone follow-up via the contact centre;
- the effectiveness of any post enumeration surveys; and
- the accuracy of our cost-quality model.

III. Operational Design – Systems

- our ability to enhance an address list to a sufficient quality that can support post-out and response management,
- the effectiveness of the channel management strategy i.e. the public's likely reaction to the optional self-selected modes of response including the internet and for specific groups the telephone;
- the effectiveness and technical feasibility of response tracking and response management;
- the ability to limit and handle potential duplicate responses arising from the self-selection modes within the channel management strategy; and
- the effectiveness of our down stream processing for outputs of different population bases.

IV. External Suppliers

- the effectiveness and cost of outsourced contract(s) for recruitment, training and pay;
- the effectiveness of the two suppliers of the 2007 Test systems, working within a test environment;
- the ability of the supplied systems interfacing successfully to meet with our output based specifications given the changing technological environment;
- the understanding of the dependencies with external contract requirements on census plans and decision dates; and
- the effectiveness of the postal provider in all aspects of their proposed role.

2. Scope of strategy

There are a number of different stages of testing and evaluation prior to the next Census. These testing stages vary in size and tend to get larger nearer to the Census. Many of the initial stages explore feasibility of ideas and are therefore small in nature. The need for testing and evaluation, and hence expense, increases as the Census Design develops. This strategy sets out:

- the key testing and evaluation stages at a high-level and the key objectives;
- collaboration between the UK Census Offices;
- methods of communication;
- roles and responsibilities; and
- the quality assurance process for both test plans and the results of evaluation.

The scope of this strategy excludes the specific key evaluation questions, methods, data sources, specific responsibilities and timescales which can be found in the 2011 Census Testing and Evaluation Plan.

3. Planned testing and evaluation

The on-going process of testing and evaluation currently falls into six testing stages, assessing an array of items from scoping field procedures, questionnaire design, question testing to e-capture. The various testing activities in this process have an impact on and/or a reliance on the previous and subsequent stages, and as such requirements and financial considerations may change. The objectives within each stage could change and update through the iterative development of the Plan. Thus objectives may be dropped, combined or have their testing factors altered. Alternatively, new testing objectives may be introduced.

The definition of a test within this strategy can refer to either:

- a statistical test – testing feasibility of new concepts using an experimental approach, for example statistically testing the difference in response rates between hand-delivery and post-out of questionnaires;
- an operational test – testing, scoping and proving practicalities of both old and new concepts, for example testing fieldwork using different methods for the proposed address checking task or testing the interfaces between key systems; or

- an evaluation test – collection of information needed for development and evaluation, for example an evaluation survey, a postal survey or a focus group.

Some of the testing and evaluation objectives will involve contact with the public or working with contractors while other objectives will involve only internal liaison. The key objectives drive the design and size of each stage including the type of target area(s) and sub-population(s). The six main testing stages should be viewed as a process of desirable testing and include:

1) Pre-2007 Testing

Key objective – to test preparations and new procedures leading into the 2007 Test. This may include:

2005 Small Scale Testing

- the scoping of the proposed innovation of the pre-delivery address checking task including: method, area type, quality, workload, agency recruitment, training, instructions, supplies, field and office management, health and safety, duration, cost and pay.
- the assessment of the postal system delivering to known and unknown addresses, experience of postal sorting and delivery rounds, and quality of the postal address list.

2005/6 Questions Testing

- the acceptability of questions and why some questions are not acceptable in providing the data users want collected and the ability of the general public to supply this information.
- the acceptability of the format for the 2007 questionnaires including a questionnaire in Welsh but excluding the format of the telephone capture.

2006 Pilot Testing

- the ability to implement census field procedures required for the 2007 Test including a pilot run of the planned pre-delivery address listing task.
- the capacity of postal staff's workload (postman's walk) and the staff's accuracy of sorting and delivery.

2007 Management Information Dataflow Test

- the effectiveness of the dataflow process from field enumeration to systems, to management and back again including connection to the internet in the mock up management information system required for the 2007 Test.

2) 2007 Test Evaluation

Key objective – to test the initial Census Design in a large scale field test. This may include:

2007 Census Test

- the relative effectiveness of the field operation and supporting systems to improve overall response/coverage or reduce differential coverage within the enumeration targeted classification groups.
- the response rates for post-out or hand delivery of census questionnaires and the inclusion or not of an income question.
- the general feasibility of the proposed 2011 Census Design, evaluating whether the key concepts work and work together.
- the impact of interactions between census and the out sourced contracts and services.

2007 Post Enumeration Questionnaire

- the gathering of additional information via a post enumeration questionnaire and/or focus groups of non-responders.

2007 Test Evaluation Survey

- the gathering of additional information from the public to assist in the evaluation of the 2007 Test including quality and the use of internet mode of response via a follow-up interview and/or questionnaire to a sample of responders and non-responders.

2007 Questions Test Survey I

- the effectiveness and acceptability of the 2007 Test questions via follow-up qualitative interviews to responders and non-responders.
- the understanding of why some questions are not acceptable for providing data users want collected and the ability of the general public to supply this information.

3) Pre-2009 Testing

Key objective – to test refinements to the Census Design leading into the 2009 Rehearsal. This may include:

2007/8 Questions Test Survey II

- the effectiveness and acceptability of questions, understanding why some questions are not acceptable for providing data users want collected and the ability of the general public to supply this information, via a post-out survey including follow-up work which may link with the 2008 Field Test.
- the acceptability of the format for the 2009 questionnaires including a questionnaire in Welsh but excluding the format of the telephone and internet capture.

2008 System Integration Test

- the effectiveness of the integration of systems, supporting interfaces with field enumeration and processing including: coverage assessment, edit/imputation and output systems.
- the interfaces between the output systems, particularly coverage assessment, edit and imputation, disclosure control and different population based outputs.

2008 Field Test

- the testing of field operational refinements following the 2007 Test and linking with system integration.

2008 Census Coverage Survey Pilot

- the scoping of operational aspects of a Census Coverage Survey.

4) 2009 Rehearsal Evaluation

Key objective – to test the final Census Design in a large scale field rehearsal. This may include:

2009 Rehearsal

- the effectiveness of all operations (field, questions, processing, coverage and output systems).

2009 Census Coverage Survey

- the effectiveness of the Rehearsal's Census Coverage Survey.

2009 Evaluation Survey Pilot

- the effectiveness of the Rehearsal's Evaluation Survey pilot including Quality.

5) Pre-2011 Testing

Key objective – to test final refinements to the 2011 Census Design. This may include:

2010/11 Final System Test

- the effectiveness of the overall integration of all final systems and volume approximations including output systems.

2010 Adhoc Census Coverage Survey Test

- the testing of refinements to the Census Coverage Survey following the 2009 Rehearsal.

2011 Operational Readiness Test

- the final effectiveness of the required operational readiness of the systems process.

2011 Total Operations and Systems Test

- the effectiveness of the processing systems, by processing actual early returned questionnaires, n-days prior to Census.

6) 2011 Census Evaluation

Key objective – to provide additional information on the 2011 Census and data for planning the next census. This may include:

2011 Quality Survey

- to gather additional information from the public to assist in the evaluation of the 2011 Census.

2011 Evaluation Survey - public perception

- to gather additional information from the public to assist in the evaluation of the 2011 Census.

2011 Census Coverage Survey

- to gather additional information from the public to assist in the evaluation of the 2011 Census.

4. UK collaboration

It is the responsibility of the General Register Office for Scotland (GROS) and Northern Ireland Statistics and Research Agency (NISRA) to develop and conduct their own Census testing and evaluation, as requirements including procurement and subsequent dates differ from those in this strategy. Although this strategy covers England and Wales, many of the evaluation findings and recommendations will also inform GROS and NISRA. Similarly, ONS will learn from the tests conducted by GROS and NISRA, for example the 2006 Census Test in Scotland. The ONS testing and evaluation plans will therefore be developed with high-level collaboration from both GROS and NISRA. This collaboration, including the Welsh Assembly Government, has the benefit that testing and evaluation is not unnecessarily duplicated between the UK Census Offices. The structure for this high-level collaborative work is in place through the UK Census Harmonisation Committee.

5. Communication

This Testing and Evaluation Strategy will be available to all interested parties via the internet on the ONS 2011 Census website. In addition, general progress will be provided via regular updates to the website and shown against each stage of the Testing and Evaluation Plan.

To assist with data storage all related testing and evaluation documentation will be stored in, or linked to, a central database for easy access. This database will be set up and maintained by a testing and evaluation coordinator. The database will contain a file for each of the testing and evaluation stages where allocation of the documentation including any datasets will be clearly assigned to the relevant file.

Procedures will be put into place to ensure that all data is held securely within the database and that confidentiality is maintained particularly when data is collected from the public.

6. Roles and Responsibilities

The planned testing and evaluation process will generate a demanding workload within tight deadlines including planning, preparation, conducting, evaluating and quality assuring. There are a number of distinct roles and responsibilities falling into six categories, which are:

- Coordinator/Advisor – The main responsibility will be to advise on the testing and evaluation requirements thus provide the statistical/research link between Census and Methodology. This will include an overview of the requirements of Census to give recommendations on testing factors and to ensure there are no test duplications or missed factors. There will be the need to provide advice and support during active testing. The Coordinator's responsibility will also be to ensure the tests

can be evaluated by identifying appropriate data to be collected. Overseeing and quality assuring that each stage of the evaluation is being conducted correctly (objectively and within timescales) will also be a key role. Writing the overall 2007 Test and 2009 Rehearsal Evaluation and High-level Evaluation Reports will be a key responsibility. The Coordinator will set up and maintain the central storage of the documentation.

- Workstreams – Workstream Managers will be the commissioning owners of specific testing and evaluation stages providing guidance, resources and quality assurance to the testing. A Workstream Manager may have responsibility to approve/sign-off documentation and will give members of their team lead responsibility to undertake the work.
- Workpackages – Workpackage Managers will have the responsibility to write the Product Descriptions, Design Plans, Evaluation Reports and High-level Evaluation Reports, with the assistance of the Coordinator. The Workpackage Managers may also need to set up working teams and will take the lead in organising and managing the team's work. Many of the testing stages will also involve Workpackage Managers and teams' direct involvement in conducting the active test and evaluation.
- Methodology – Selected representatives from within Methodology Division (MD) will be involved in the development of the testing and/or evaluation approaches. The MD will provide input into the Product Descriptions, Design Plans and Evaluation Reports. The responsibility of MD will be to provide advice on how best to conduct the testing and evaluation. In some cases MD may also be responsible for conducting the testing and/or evaluation.
- Technical Specialists – External and internal technical specialists will be involved in the development of the testing and/or evaluation. For example, all the systems and security arrangements will require technical input. The specialists will be involved in the development of the Product Descriptions, Design Plans and Evaluation Reports. Their responsibility will be to provide advice on how best to conduct the testing and evaluation. In many cases the technical specialist will also be responsible for developing a product and conducting the testing and evaluation.
- Quality Assurance Teams – The quality assurance process itself will involve a range of relevant people from Census, Methodology and a wider field of experts. The membership of the quality assurance teams will depend on the testing and evaluation stage and will be stated within each Product Description.

7. Quality Assurance

The testing and evaluation will follow the standards set out in the 2011 Census Quality Management Plan where all testing and test evaluation will go through an extensive quality assurance process before final approval is gained. Testing will not start until the Design Plans have been signed off by the relevant board (e.g. Census Project or Progress Board). The level of final approval will depend on the testing and evaluation stage and will be stated within each Product Description. The Evaluation Reports will each be quality assured and internally approved by the relevant census staff and quality assurance team before going before an appropriate board or person(s) for approval. The UK Census Design and Methodology Advisory Committee will oversee the quality assurance of the Testing and Evaluation Plan ensuring that the broad range of testing hangs together to underpin the Census Design.

Sarah Walker
October 2005