

2007 Census Test Design and Sample Size

Action requested of Advisory Groups:

Advisory Group members are asked to note the design and sample size for the 2007 Test.

Test objectives

1. The three key aspects of the Test are:
 - an operational Test of the field procedures;
 - a statistical Test of the impact on response of :
 - a) inclusion of an income question;
 - b) postout out compared with hand delivery; and
 - a proof of concept of the key supporting systems.

Test sample size and design

2. The Test will cover approximately 100,000 households and be selected from five Local Authorities:
 - Liverpool (representing Services and Cities)
 - Camden (London Centre)
 - Bath and North East Somerset (Prospering UK)
 - Carmarthenshire (Coastal and Countryside); and
 - Stoke-on-Trent (Mining and Manufacturing).
3. The sample of households will be divided into five equal strata, according to an 'Enumeration Targeting Categorisation' (ETC) which assigns a difficulty level of 1-5 to each census Enumeration District (where level-5 represents the very hardest to enumerate). The ETC has been developed using factors found to be most associated with household census non-response (young adults, ethnicity, income, rent private, rent other and non-standard housing).
4. Within each stratum:
 - half the population will receive a questionnaire by post and half by hand; and
 - half the population will receive questionnaire including an income questions, and half without
5. Thus the sample will be as shown in the following table:

Table 1: Recommended Test Design and Sample Size

ETC Strata	Post-out		Hand-delivery		Total
	Income	No income	Income	No income	
1 (v. easy)	5,000	5,000	5,000	5,000	20,000
2	5,000	5,000	5,000	5,000	20,000
⋮	⋮	⋮	⋮	⋮	⋮
5 (v. difficult)	5,000	5,000	5,000	5,000	20,000
Total	25,000	25,000	25,000	25,000	100,000

6. The five LAs in which the test will be conducted have very different populations. Camden and Liverpool are the only two that have significant numbers of Enumeration Districts in the two hardest strata. Therefore, to achieve a 20,000 households sample in each of these strata, the sample will be much larger in Camden and Liverpool than in the other authorities. The Enumeration Districts within each LA will be selected by the end of May 06 and the likely number of households in each LA is as follows:
 - Liverpool 40,000 households
 - Camden 25,000 households
 - Stoke on Trent 15,000 households
 - Bath and NE Somerset 10,000 households
 - Carmarthenshire 10,000 households
7. For analysis of the delivery method we will compare response rates within each stratum of the ETC, as we would like to understand whether we should choose to use different delivery methods for different area types in 2011.
8. The analysis of the income question will compare the difference in response rates for the whole sample as we will not be considering including an income question only for specific areas in 2011.
9. Using a one-sided Tests at 95% power¹ and 5% significance level, and an expected response rate of 50%, 100,000 households gives a detectable difference in response rate of 6.7% within each ETC stratum and of 3.0% over the whole sample (A discussion of the statistical issues surrounding the test design is in Annex A)
10. However, the fact that the Test is voluntary means that the public will behave very differently from a compulsory census. Whilst the analysis will result in an understanding of whether there are differences in response rates between different options **in the Test**, it will **not** result in an understanding of true response rates **for 2011**, nor the true differences in response rates **for 2011**.

¹ Testing at a power of 95% refers to a 5% chance of concluding there is no significant difference in response rates when in reality there is a difference. A 5% significance level refers to a 5% chance of concluding that there is a significant difference in response rates when in reality there is no difference.

11. Thus any statistical results of the test will only from part of the evidence for final decisions. Final recommendations following the test will depend on a number of (inter-related) factors. For income we will need to consider:

- item non-response to the income question;
- item non-response for all other questions due to inclusion of an income question;
- impact on coverage of people within households;
- the quality of data provided by the income question;
- extent of additional request for individual person questionnaires for privacy reasons;
- the impact on the volume of contacts to the call centre and web self-help;
- public acceptability and understanding; and
- the impact on response and the costs of additional follow-up .

12. Similarly, for delivery method we will need to consider:

- coverage and quality of address lists;
- quality of response for each delivery method;
- public acceptability;
- operational effectiveness;
- the impact on the volume of contacts to the call centre and web self-help;
- postal provider competence; and
- the impact on response and the costs of additional follow-up.

Annex A – Statistical issues surrounding the Test design

Using a one-sided Tests at 95% power and 5% significance level, and an expected response rate of 50%, 100,000 households gives a detectable difference in response rate of 6.7% within each ETC and of 3.0% over the whole sample.

We will use one-tailed tests for analysis as we are only interested in testing whether post-out or the income question decreases response. We are not statistically interested in whether either of the factors increases response. Note that because of this it will **not** be possible to provide statistical conclusions about increases in response if such results are obtained.

We have set the power high, at 95%, (it is usually set to 80%) because if there is a difference in response rates we want to be able to detect it. To conclude that there is no difference when in fact there is one risks a low response, and hence additional cost, in the actual Census for which we would be unprepared.

Operationally a design based on ED clusters was required, because we need to test the task of delivering questionnaires by hand or post to entire areas. Statistically, such a design requires a much larger sample of households to detect significant differences than one where the different methods are assigned to randomly selected households. Table 2, shows the difference in response rates that could be reliably detected with different household sample sizes and varying pre-set probabilities. The rows refer to different powers (P) and significance levels (SL). The table also shows the operational cost of the different sample sizes. The shading indicates the key statistical settings.

Table 2: Detectable differences with different probabilities, sample size and costs

Test Size	100,000	150,000	200,000
Delivery method test per strata			
P=95% / SL=5%	6.7%	5.3%	4.4%
P=90% / SL=5%	5.9%	4.7%	4.0%
P=80% / SL=5%	5.0%	4.0%	3.4%
P=50% / SL=5% Smallest significant difference	3.4%	2.7%	2.2%
P=90% / SL=10%	5.2%	4.1%	3.5%
P=80% / SL=10%	4.3%	3.4%	2.9%
Income question test overall			
P=95% / SL=5%	3.0%	2.4%	2.0%
P=90% / SL=5%	2.6%	2.1%	1.8%
P=80% / SL=5%	2.2%	1.8%	1.5%
P=50% / SL=5% Smallest significant difference	1.5%	1.2%	1.0%
P=90% / SL=10%	2.3%	1.8%	1.5%
P=80% / SL=10%	1.9%	1.5%	1.3%
Operational cost			
	£	£	£
50% post-out, 50% hand-delivery	1,978,935	2,302,861	2,651,764

These calculations assume that we only test the main effects and are not interested in the possible interactions between the delivery methods and the questionnaire types. Our

assumption is that any interactions between the factors would not significantly affect the decisions we make.

Given a 100,000 households total sample, if the response with post-out is more than 3.4% lower than with hand-delivery within an ETC stratum then we can confidently say that post-out reduces response within that stratum. However, if the post-out response difference is lower than 3.4%, then all we can say is that the true difference is unlikely to exceed 6.7%. Because the Test is voluntary, we could not predict the exact scale of the drop in response in 2011.

Given a 100,000 household sample, if the response with an income question is more than 1.5% lower than without then we can confidently predict that including an income question would depress response. However, if the income question response difference is less than 1.5%, then all we can say is that the true difference is unlikely to exceed 3.0%. Because the Test is voluntary, we could not predict the exact scale of the drop in response in 2011.

Other options considered for the Test design

A range of alternative options has been considered, to attempt to circumvent the difficulties of having insufficient sample sizes to draw helpful statistical conclusions. One possibility was to randomly assign parts of the sample to reduce the impact of clustering. This was rejected for assignment of income/no income to households for operational reasons because of the difficulty it engenders in the delivery of the questionnaires (especially for hand-delivery) and what to do with addresses discovered during enumeration. Another possibility was to have part of the post-out sample randomly allocated to households across the whole LA (rather than in ED clusters), but this was rejected because of the additional costs for the address checking and follow-up tasks, and would not replicate the ability of the postal delivery.

Peter Benton

April 2006