ONS(ONC(SC))01/04

ONE NUMBER CENSUS STEERING COMMITTEE

The Treatment of Movers in the 2001 CCS

1. This paper describes the options for the collection and treatment of information about people who move between the Census and CCS fieldwork periods. These options are evaluated and a recommendation made as to the approach to be implemented.

The Steering Committee are asked to note the paper.

Marie Cruddas
Census Division
Office for National Statistics
Room 4200W
Segensworth Road
Titchfield
Fareham
HANTS
PO15 5RR

January 2001

The Treatment of Movers in the 2001 CCS

1. Introduction

This note addresses the options for the collection and treatment of information about people who move between the Census and CCS fieldwork periods. These options are evaluated and a recommendation made as to the approach to be implemented.[1]

2. Options

The aim of the CCS is to re-enumerate selected postcodes, counting the same population as the Census. Therefore it is important to conduct the CCS as close to Census day as possible. Thus the CCS fieldwork will begin 3 ½ weeks after Census day. Despite this there will be people who move in the interim, those that were resident on Census night and who are no longer resident (out-movers), perhaps replaced by new residents (in-movers).

In the context of the US Census, Griffin (2000) considers three possible ways to deal with movers. In the following these three options are reviewed together with a fourth. The options are:

1) Collect proxy information from the new residents about the out-movers to construct the resident population in the postcode as of census night. (This has been the approach adopted in the trials of the CCS.)

2) Collect information on the current residents at time of the CCS and their locations on Census night and attempt to match to the Census returns from that location.

3) For households with movers, collect basic proxy information on the out-movers and detailed information on the in-movers. Matching information then comes from the proxy return for the out-movers and characteristics for estimation comes from the information on the in-movers.

4) A fourth option- not considered by the US - is to collect no information on movers and assume that movers are just non-responses in the CCS missing at random (ie no different to the non-movers that the survey does enumerate).

The approach taken by Griffin (2000) in the analysis of movers is to assume that movers are a potential source of heterogeneity bias when calculating the DSE. In other words, within a post-stratum (or in the UK methodology within an age-sex group and cluster of postcodes), people who move are '***different***' from those who don't move and may have different 'capture' probabilities. Therefore, this will introduce heterogeneity bias into the estimation defined in Griffin (2000) as

[1] In this note the issue of the legality of collecting proxy information in the CCS is not addressed. We have already discussed this elsewhere and ONS are content that we can collect proxy information on movers. GROS may have a different view in which case their choice of approach is restricted.

$$\text{Bias (DSE)} = - \frac{Nd(1 - c)(1 - m)}{(1 + dcm)(d + 1)} \qquad (1)$$

where N is the population total being estimated

$$d = \frac{\text{number of movers}}{\text{number of non - movers}} \quad d \geq 0$$

$$c = \frac{\text{census coverage for movers}}{\text{census coverage for non - movers}} \quad c \geq 0$$

$$m = \frac{\text{CCS coverage for movers}}{\text{CCS coverage for non - movers}} \quad m \geq 0$$

The exact impact of the bias will depend on the actual values of d, c, and m and their interaction. Realistic values of d will be less than one (less than half the population move) and in fact close to zero but as d increases the potential impact of the bias will increase. Values of c will also, in general, be between zero and one as c measures the coverage of movers in the census relative to the coverage of non-movers. A value of one implies the same coverage in the census for both groups. Finally, m is the same as c but for the survey. Again, realistic values will be between zero and one, with one representing no difference between the coverage of movers and non-movers in the survey. The aim of each option is to get (1) to equal zero by a strategy that makes d zero, or c one, or m one, or closely approximates one of these.

2.1 Option one

If Option one is chosen, the assumption is that by collecting good quality proxy information on the out-movers, the value of m in (1) is close to one. Hence little or no heterogeneity bias due to movers. This is only realistic when d is close to zero (there are very few movers to worry about) and there are potential problems with matching the proxy returns in the survey to actual census returns.

2.2 Option two

In the US, Option two has, in the past, been preferred over Option one. It is argued that, as the survey is counting actual people on the ground, then the value of m in (1) will be closer to one under Option two than it will be under Option one. Under both options the values of c and d will be the same (assuming that out-movers are replaced by similar in-movers).

2.3 Option three

In 1990 the Census Bureau used Option two but there were considerable practical problems matching survey returns with Census forms from all over the country. In addition, the original plans for the 2000 Census would have made Option two very difficult and when the plans were changed it was too late to alter the treatment of movers (Griffin, 2000). This led to the development of Option three that can be considered a composite approach that combines the advantages of Option one and Option two. Proxy information is collected on the out-movers but this only needs to be sufficient to allow matching to the Census returns within the sampled area. Full details are also collected from the in-movers and under the assumption that out-movers are replaced by 'similar' in-movers the in-movers are used in the DSE as the survey returns. This means that the DSE is defined as

$$\text{DSE} = \frac{N_n + N_i}{M_n + \left(\frac{M_o}{N_o}\right) N_i} C \qquad (2)$$

where N_n is the survey return for non-movers

N_i is the survey return for in-movers
N_o is the proxy survey return for out-movers
M_n is the matched count for non-movers
M_o is the matched count using the proxy returns
C is the census return (adjusted for overenumeration etc)

The key assumption underpinning (2) is that the match rate estimated from the proxy information for out-movers is an unbiased estimate of the match rate that would be achieved for the in-movers if the matching could be done. Under this assumption Griffin (2000) demonstrates that (2) will have a bias similar to Option two because the bias depends on the ability to capture the in-movers in the survey relative to the non-movers and by the argument above it will be better than Option one. This approach also has the advantage that the proxy count for out-movers can be quite low compared to the true count of out-movers provided the 'non-response' is at random. Then the ratio, which gives the out-mover match rate, will still be unbiased.

2.4 Option four

Option four relies on the assumption that missing movers in the survey can just be treated as missing at random non-response. In other words, the movers who are missed are no different from the non-movers who are counted and the non-movers who are missed. If this assumption fails then in (1) this is equivalent to setting m equal to zero by design. Therefore, the heterogeneity bias in the DSE becomes

$$\text{Bias (DSE)} = -\frac{Nd(1-c)}{(d+1)} \qquad (3)$$

and the bias in (3) now solely depends on the difference in Census coverage for movers and non-movers and the proportion that move. If the Census coverage is the same for both groups (c equals one) then (3) is zero. However, if the Census also completely misses movers, and this is not at random, the bias simply depends on d and (3) simplifies to

$$\text{Relative Bias (DSE)} = -\frac{d}{(d+1)} \qquad (4)$$

For example, if five per cent move between the Census and the CCS the relative bias given by (4) would be –5 per cent. Intuitively, this is because the five per cent of the population who move are completely missed by both the census and the CCS. However, if the Census capture probability for the 5 per cent of movers is half the capture probability for non-movers (c equals 0.5) the relative bias reduces to 2.5 per cent. Typically the percentage of movers will be small and the differential coverage in the Census will be minimal. Therefore, under the worst case scenario of ignoring movers in the survey the potential bias impact is minimal.

3. UK Context

The arguments made in favour of dealing with movers by using Options two or three are partly dependent on the manner in which the DSEs are used in the US. The approach taken in the US uses weighted survey estimates to construct DSEs. These give directly population estimates at the post-stratum level. In other words, all the quantities in (2) are weighted survey estimates representing reasonably large population sub-groups. However, there are a number of important differences between the UK and US approaches to the One Number Census that inform the approach proposed for the treatment of movers.

- In the UK the DSE estimation takes place in individual postcodes and clusters of postcodes enumerated in the CCS sample (ONS(2000)) rather than at an aggregate. This makes the application of both Option two and Option three problematic. In addition, Option two would require waiting until all the data was processed before matching could be undertaken.

- The time between the census and CCS is very short – 3 ½ weeks – so it is expected that movers will be a 'small' problem (in (1) d will be close to zero). We would expect approximately 1% of households move in the month between Census and CCS. (The 1998/9 survey of housing found that about 12% of households had moved within the last year – DETR website).

- Option one will satisfy the requirement to reconstruct the postcode as of Census day but the quality of the proxy returns may be in doubt.

Thus the choice in the UK is between:

Option one – the current approach

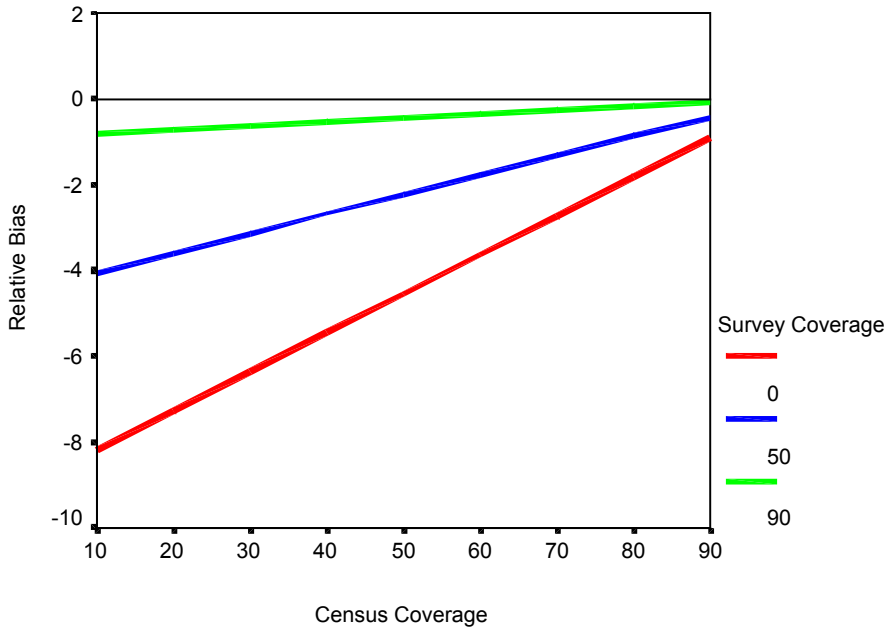
Option four.

These are considered further by looking at some simple examples.

3.1 Example 1

If it is assumed that 1 in 11 people have moved between the census and the CCS Figure 1 shows the bias of the DSE for a situation where the survey collects no proxy information (m equals zero), some proxy information (m equals 0.5 or 50%), and a high percentage of good quality proxy information (m equals 0.9 or 90%) for a range of census response rates for movers relative to non-movers.

Figure 1: The relative bias in the DSE for different levels of relative coverage of movers in both the census and the survey



Both census and survey coverage are for movers relative to non-movers

The ratio of movers to non-movers is 0.1

The results in Figure 1 are based on a situation where the assumption underpinning Option four has failed. Therefore, this simple example illustrates the principle that collecting proxy information is worthwhile. Even when the CCS collects proxy information on 50% of the movers relative to the non-movers the impact of the bias is greatly reduced regardless of the effectiveness of the census at counting potential movers.

3.2 Example 2

For a more realistic situation, where the Census counts potential movers at only a slightly lower rate to non-movers (c equals 0.75) and approximately 1% of households move in the month between Census and CCS, then the relative bias is:

- 0.25% - using no proxy information in the survey
- 0.12% - with a relative coverage for movers of 50% (using proxy information)
- 0.02% - with a relative coverage for movers of 90% (using proxy information)

These results demonstrate that while there is a gain from collecting the proxy information they also demonstrate that there is also little risk from adopting Option four when d is close to zero.

4. Conclusion

It seems clear that the optimal approach for the treatment of movers in the ONC is to collect proxy information from the new residents about the out-movers in order to construct the resident population in the postcode as of census night. This represents Option one and the approach that has been tested in the CCS field trials. However, there is a risk with Option one if the quality of the proxy information is poor, as this will involve a potential bias from matching error. Therefore, in situations where the proxy information cannot be thought of as accurate then Option four, which is a special case of Option one, can be used.

5. References

Griffin, R. (2000). Accuracy and Coverage Evaluation: Dual System Estimation. *DSSD Census 2000 Procedures and Operations Memorandum Series,* **Q-20**, US Census Bureau.

ONS (2000). Estimation strategy for Design Group Estimates by age and sex from the Census Coverage Survey. *One Number Census Steering Committee Paper* **00/03A**