ONS(ONC(SC))99/09

ONE NUMBER CENSUS STEERING COMMITTEE

Risk in Official Statistics: A Case Study of the 2001 One Number Census Project

1. Attached is a copy of the paper that will be presented by Tim Holt at a plenary session of the RSS Conference on Risk, 12 - 15 July 1999 in Warwick. This paper is circulated for information.

2. 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

June 1999

Risk in Official Statistics:
A Case Study of the 2001 One Number Census Project

Ian Diamond

1. Introduction

A major part of the ONS mission is to deliver reliable statistics to a very broad user community. While many of the data which ONS produce are descriptive all will be subject to measurement error in some way. As a result risk is an inherent part of any ONS activity and the Office invests a lot in the minimisation of risk which can occur in a number of ways. As with all statisticians those at the Office are concerned with minimising the risk that published data are outwith some acceptable level of accuracy. In addition a second risk is that the data, however accurate, will not prove acceptable to the user community. Clearly accuracy is an important component of minimising the risk that the data will prove unacceptable but in addition it is important to ensure the transparency of the methodology which has been used so that users have confidence in all stages of the development of the output.

This paper aims to demonstrate the extent to which identifying and managing risk is an inherent part of the role of the Office for National Statistics through the use of a case study: the work which has carried out since 1996 to develop a strategy to adjust the 2001 Census for underenumeration. The paper will first describe the context in which the research project was set up; it will then describe the research which has led to the proposed strategy, highlighting a number of particular risks; and finally it will describe the consultation with the user community which aims to ensure that the proposed strategy is acceptable. It should be noted that while the research described has largely been undertaken to develop a strategy for the census of England and Wales, colleagues in Scotland and Northern Ireland have also been involved and it is intended that in 2001 a complementary strategy will be carried out throughout the United Kingdom.

2.Context

In 1991 the census of England and Wales was subject to an underenumeration of around two per cent. It should be stressed that such levels are typical in developed countries and most countries have a process for estimating the extent and geographical distribution of this underenumeration. This is done typically via a post enumeration survey where a sample of households are revisited shortly after the census. Those found in the household in the post enumeration survey are then compared with those found in the census so as to produce a population estimate adjusted for underenumeration. In 1991 this post enumeration survey

was called the Census Validation Survey (CVS) and had a dual role: a) to estimate underenumeration; and b) to assess the quality of the census data.

The CVS was a probability sample of households a) which had been enumerated in the census; and b) which had been judged by the enumerator to be vacant or absent. Clearly there was a risk in that the CVS might not find the underenumerated and so to minimise the risk that the final census figure might be unreliable the then OPCS developed a strategy. This used demographic analysis to assess the plausibility of the census figures. The 1981 census figures were rolled forward to 1991 using the standard demographic balancing equation. Vital registration was used to assess natural increase and the International Passenger Survey, Labour Force Surveys and other administrative sources were used to assess net migration. The result of this work was that it was agreed that, although the CVS estimated around one quarter of a million people had been missed by the census unit had underestimated underenumeration by around 1.2 million people in England and Wales.

In addition the results of the CVS indicated that underenumeration was not uniformly distributed geographically. The implication that, for example, a 24 year old man in Inner London had the same probability of being underenumerated as a 24 year old man in Harrogate was simply not plausible. This is contrary to all experience of census underenumeration both in the UK and elsewhere and to have published such data would have meant a high risk that the census would lose credibility with a large part of the user community.

Therefore the then OPCS with input from academic analysis undertook a large amount of further research and developed a deterministic method based on demographic analysis which permitted population estimates adjusted for underenumeration to be made for six groups of local authorities. Within these groups it was assumed that the underenumeration by age and sex was uniform. The resulting estimates were acceptable to the user community and have been used as the base for population estimates since 1991. While these results were acceptable it is important to note that during the process census users received, during 1992 and 1993, at least three different counts for the same area: a raw count, a set of provisional estimates and, finally a set of final estimates. This was of course deeply unsatisfying for the user community.

It should also be noted that the estimates adjusted for underenumeration were produced only at the Local Authority level, meaning that those, for example, for electoral wards were based simply on the raw census counts. This meant that local authority statisticians had to work with a set of outputs which were not internally comparable. As the use of census data for resource allocation at all levels of aggregation increases this presents a problem given that it is accepted that the underenumeration would not be uniform across a local authority. Although Simpson et al (1997) did produce a set of synthetic estimates the lack

of one set of internally consistent estimates does present a risk to the future potential use of the census and to confidence in census outputs.

3. Overcoming Risks: the One Number Census Project

To overcome the risks to the credibility of the Census which would be posed by a similar problem occurring in 2001 the ONS put in process, in 1996, a research project aimed at developing a strategy to adjust the census for underenumeration which would have the following aims:

i) census counts adjusted for underenumeration broken down by age and sex would be available, at a reasonable level of accuracy, for each local authority district – this would enable, for the first time, a proper assessment of the extent of geographical variability in the undercount. It should be noted that local authorities are a key unit of resource allocation. The sum of these estimates would be the national estimate of the population at the time of the census;

ii) estimates of underenumeration would be available down to a very local level – this would mean that all census outputs would be internally consistent and add to 'one number', the national estimate of the population at the time of the census;

ii) the estimates produced by the strategies developed for (i) and (ii) would be plausible and acceptable to the wider user community.

This research project has become known as the One Number Census (ONC) project. My purpose is not to describe the ONC in detail although some description will be necessary. Instead I want to focus on the risk inherent in such a project, and our attempts not just to measure these but to manage them.

The strategy is outlined in Figure 1. Shortly after the census a post enumeration survey (to be known as the Census Coverage Survey (CCS)) will be undertaken. This survey will reenumerate a sample of postcode with a short questionnaire focussing only on the characteristics associated with underenumeration. The survey will be designed for around 100 areas with populations around 500,000. A combination of dual system and regression estimators will provide population estimates, adjusted for underenumeration, by age and sex for these design groups. Regression estimates will then permit estimates for all Local Authority Districts and Unitary Authorities. Finally, individual and household records of those underenumerated will be imputed at the postcode level.

An basic risk for the ONC is that the overall strategy will not command support from the user community, and so as a result the Census outputs, whether statistically sound or not, will be called into question. This requires that from the

first stage there is a process of quality assurance and consultation. This process comprised three main prongs:

i) The research team set up a 'risk register' which highlighted all risks to the successful completion of all aspects of the project. As the strategy develops new risks are identified and old risks are reduced. The dynamic nature of the risk register means that it needs to be regularly monitored by the research team. This is undertaken by a Project Group, chaired by the Head of Methodology at the ONS and including members of all relevant groups at ONS, GRO(S) and NISRA. An example of part of this risk register is found at Table 1. It should be noted that the Project Group also serves as a think tank for the strategy. At different stages of the project over 60 risks have been identified in the register.

ii) To monitor the strategy a Steering Group was appointed. This is seen as the key monitoring group and comprises representatives of the large user communities, most notably academic statisticians and demographers; local authority and government statisticians; and a representative of the Australian Bureau of Statistics. This group meets twice yearly. The purpose of this group is to provide independent and quality assurance, and to promote, by its existence and participation in workshops, user confidence in the project.

iii) A full series of consultation: initially a series of papers was given in 1997 at the RSS (the Cathie Marsh Memorial Lecture) and at a series of seminars organised by the ESRC sponsored Estimating with Confidence project, including a two day workshop in Leeds in May 1998; more recently there have been four 'ONC roadshows' – in London, Manchester, Cardiff and Glasgow; and a set of publications – an initial 1998 consultation paper, a 1999 'Guide to the One Number Census' and a 1999 'What is the One Number Census' which condensed the whole process into two sides!

iv) Quality assurance has also been addressed by insisting that the work was subjected to the highest level of peer review - a series of papers submitted to, and accepted by, leading journals such as the JRSS. In addition papers have been given at leading international official statistics fora, for example the annual Statistics Canada symposia and the annual research conference of the US Bureau of the Census.

Through this process it is intended that the final strategy will be acceptable. However this will only be the case if the strategy itself has overcome many risks. These will now be described.

4. The Calculation of Local Authority Population Estimates

4.1 Census Coverage Survey Design

The basic tool of the ONC will be the Census Coverage Survey (CCS). This will facilitate estimates of underenumeration for about 100 "design groups" - groups of Local Authority Districts with combined population of around 500,000. The use of such a survey is common in many census taking countries. This is undeniably a difficult process that carries inherent risk. The post enumeration survey may not identify people missed in the census as the two groups are inherently similar (Brown et al, 1999). An initial risk would be to accept that this was the only feasible approach thus missing a more effective strategy. The major alternative would be to use administrative data and so a large review was undertaken of all such sources. A successful source (or combination of sources) would need to have national coverage at a small area level, to be comprehensive, and not to be inflated in any way. Some countries, such as those in Scandinavia, have population registers which can be used to monitor (or even take the place of) the census. However it was agreed that at present there was not, for 2001 in this country, the possibility of using administrative sources to assess coverage.

Therefore it was decided to use a post enumeration survey. As previously, in 1991, the CVS addressed both coverage and quality. To minimise the risk that it was inefficient to mix missions in this way it was agreed that there would be a separate quality assessment survey and that underenumeration would be measured by a specialist coverage survey, the census coverage survey (CCS).

The advantage of this is that by concentrating solely on coverage one could use a shorter questionnaire with resulting greater opportunities for high response rate and sample size. However there still remains a large risk. One needs to be sure that the CCS has a very good chance of identifying all those people missed by the census. Such people can either be living in households which were not enumerated by the census or be individuals in households enumerated by the census but omitted from the census form. To identify the former, samples drawn from a list of households identified by the census or indeed from a second household listing may not be optimum. Therefore the strategy proposed was to use an areal sampling strategy. A sample of postcodes would be drawn and interviewers would reenumerate the postcodes but with a smaller questionnaire. This would concentrate only on collecting information on individual and household characteristics associated with underenumeration.

The next question was how to draw the sample. There were a number of considerations:

i) Given that the population of interest – the underenumerated - were likely to be only a relatively small proportion of the population then a large sample size would be necessary. It would simply be infeasible to make very precise estimates of

underenumeration for each of the over 400 local and unitary authorities. However not to have complete geographical coverage would itself pose a risk – that of acceptability. Hence it was decided to divide the country up into a number of 'design areas' which would comprise either large local authorities or groups of local authorities and to make estimates for each design area. It should be noted that it was agreed that there would be a constraint that each local authority would be represented in the sample for a particular design area.

ii) Underenumeration is traditionally biased towards areas with high incidences of multi occupation and poverty (where it tends to be difficult to undertake a census accurately) and to certain demographic groups most notably young males and the very old. Since underenumeration is likely to be concentrated in such areas a good design for national or regional estimates of underenumeration is not the same as a design to estimate underenumeration for each local authority separately. A significant risk to the accuracy of the estimates would be to have a sample which did not include such geographical areas and groups. However it is also important to spread the sample across all the difficult to count groups so as to minimise the risk that underenumeration will also, as society changes, increasingly occur across a broader range of social and demographic groups.

Therefore a model based design was adopted which followed the following steps:

i) divide each design area into a number of strata based on a 'hard-to-count' index (HtC). The HtC index was calculated in a similar way to many of the multiple deprivation indices commonly used in many areas of social and health research but was aimed at factors likely to be associated with underenumeration. Given the size of a postcode and availability of data it was necessary to calculate these for enumeration districts (census areas of, typically, 200 households). The components of the HtC index included multi occupancy, language difficulty and level of private renting. The HtC index was calculated for each enumeration district in England and Wales in 1991 and the distribution divided into quintiles.

ii) the design area was stratified according to the five HtC quintiles and within each stratum a principal components analysis undertaken where the data were the age sex groups. For a number of principal components, a cluster analysis was undertaken to identify clusters within which there would be relative homogeneity. A sample of enumeration districts is drawn from these clusters.

iii) within each enumeration district a sample of postcodes is selected.

Within this basic design there are a number of risks. First data to calculate the HtC index will be available only for 1991 when designing the sample in 2001. This means that the sample will not reflect changes in the composition of areas since that time. While this can be addressed through post stratification it is also intended to consult with local authorities – not I stress on the composition of the sample – but on changes in the distribution of their population since 1991.

Second, one needs to be sure that this design can be used to provide reasonable estimates of underenumeration for an acceptable cost which will of course be reflected by sample size. This risk was addressed through a programme of research which is described more fully in Brown et al (1999). The research used anonymised individual level data from the 1991 census and assumed initially that those data were the 'true' population. Then individuals were removed from the data set according to a set of Bernoulli trials which reflected the probability that an individual would have been underenumerated in the census. CCS samples were simulated and regression estimates of the 'true' population count for each age sex group in each HtC stratum made.

Many censuses were simulated together with a number of CCSs. The results are summarised in Table 2 which gives the median precision of the design area population estimate across five year age sex groups for a number of different sampling strategies including different size design areas, numbers of postcodes sampled per enumeration district and overall number of postcodes. It should be noted that smaller design areas are attractive so as to minimise the number of local authorities per design area and hence to maximise acceptability to the user community. Increasing the number of postcodes per enumeration district is also advantageous as this reduces fieldwork costs.

The results suggest that a sample of around 20,000 postcodes with five postcodes per enumeration district and design areas of around 500,000 would give an acceptable degree of precision. Table 3 shows the implications of these results for a number of different levels of aggregation up to the national level where it is expected that, for England and Wales, the 95% confidence interval has a width of +/- 0.13%. This was supported by the Steering Committee and forms the basis for current planning. However it also presents a number of risks both practical and substantive.

4.2 Methodological Risks

The research on which the above results are based has three important assumptions each of which presents an important risk to the successful prosecution of a ONC. These are:

i) The research described above assumes that the CCS has a 100% response rate. This is clearly unrealistic as there will be people who are missed by both the Census and the CCS. This in itself need not present an overwhelming problem as, by designing the Census to be independent of the CCS, one could simply use a dual system estimator within each postcode to estimate the population of that postcode and then use regression estimators to make estimates of the population of the design areas. It is true that there are likely, at the postcode level, to be a number of age sex subgroups with zero populations and hence

either a) a correction for zero cells in the dual system estimator needs to be made or b) the dual system estimator is used, say, at the enumeration district level. This choice is currently research in progress. However, there remains a risk. This approach is sensitive to the likely dependence between census underenumeration and CCS non-response and also the response rates of the census and of the CCS. Classical dual system estimators assume independence and, despite all best efforts in the fieldwork, it is very likely that there will be a degree of dependence – i.e. the chance of someone being missed by the CCS is greater if they were missed by the Census than if they were not. This can be illustrated by simulating the estimated population for a known population under a number of degrees of dependence. Figure 2 illustrates this and shows that for reasonable degrees of dependence the 95% confidence intervals for the population count overlap the true value. Current research is extending the dual system estimator to include this dependence. It should also be noted that Figure 2 assumes a census coverage of 90% and a CCS response rate of 85%. The dual system estimates are not good where there is either a catastrophic census or a disastrous CCS. This reinforces the need for these to be top class.

ii) a further risk concerns the people covered in the CCS. If only those people observed in the census were observed in the CCS then one would learn nothing about underenumeration. It is essential that a fair proportion of those missed by the census are covered in the CCS if one is to make decent estimates. To quantify this assertion a number of simulations were carried out. These are summarised in Table 4 which shows that, for a reasonable census and CCS coverage, and a reasonable degree of dependence between census and CCS one would need to observe around 66% of those missed by the census. This, effectively, is the practical challenge facing the CCS.

iii) a final risk here concerns the need to match the census records with those of the CCS in order to identify the number of people observed by the CCS but not by the Census. Often when matching records one looks for relatively high levels of matching. However here, given again the rather rare nature of the outcome of interest it is essential that the number of false matches are minimised. This requires that a complex system of both automatic and clerical matching is put into process. ONS has been receiving advice from record linkage experts at the Scottish Home and Health Department and has developed a system which uses probability weights on a number of key variables to assess whether or not an individual from the census is matched with another from the CCS. It should be stressed that below a certain threshold one would not be happy to state whether or not a match had occurred simply on the basis of automatic linkage and so a system of clerical matching has been developed to augment the automatic match. It is intended that the probability of a false match will be under 0.005.

4.3 Practical Risks

As there are, on average, 15 households per postcode the current design represents a sample of around 300,000 households – in itself perhaps the biggest one-off household survey conducted in this country. This has important implications for the fieldwork and results in numerous risks. Many of these risks are similar to those associated with the census fieldwork itself such as: advertising for and recruiting a large temporary fieldforce (the survey is so large that no single company would be in a position to tender to provide the fieldforce); ensuring a top class set of clear maps; or ensuring clear lines of communication from the individual enumerator to the field manager. Others are more particular to the CCS itself. For example, our simulations indicate that for a reasonable census enumeration and a CCS coverage of 85% it will be necessary to gain responses from around 66% of those missed by the census. Initial empirical evidence from the Census Rehearsal is extremely encouraging with very high response rates which would achieve these targets. This requires research on optimum strategies for times of calling and for number of call backs. In addition the respondents may not feel particularly motivated to respond to a survey which asks them about something they have recently already responded on! Therefore they need to be appraised that they are not being checked up on but that it is the census office who are being checked. This requires excellent training of the fieldworkers.

To minimise these risks ONS has set up a CCS team which works in tandem with the ONC research team. Quality assurance and risk minimisation comes through a project board which regularly discusses progress on a dynamic risk register such as that described above for the ONC process. It also reports to the same steering committee thus ensuring that the overall picture is seen by that committee.

The CCS team has also undergone a programme of pilot research with field trials in Brent in 1997, Southampton in 1998 and, currently, as part of the Census Rehearsal taking place in six areas in England and Wales. The results from these trials have been extremely encouraging. For example Table 5 shows response rates from the Southampton test. However there will clearly need to be a full evaluation of the Rehearsal before the strategy for 2001 can be finalised.

5. Quality Assurance of Design Area Estimates

The design area estimates, summed together, will constitute the national estimate of the population by age and sex. However, despite all the controls on the risks described above there remains a further risk that some will not be overcome. This means that in order for the final population estimates to be accepted it will be necessary to provide evidence that they are reasonable given current expectations. This means that a programme of quality assurance will be

necessary between the final production of the estimates and their release. This process will be similar to that undertaken in 1991 which led, at the national level, to the population estimates for 1981 rolled forward to 1991 being deemed to be more plausible than those from the census adjusted by the CVS.

It needs to be stressed that, at this level, population estimation is not an exact science. The strategy is to assemble a range of indicators of the likely population at different levels of aggregation - local authority districts, regions and nations - and to make a judgement, based on a comparison of the census data with these indicators, as to the acceptability of the census figures. It is important to recognise that the bottom line is that one would expect the census data to be the most accurate figure and one would seek evidence that this was not the case. In other words there does not exist a 'gold standard' against which the census estimates will be judged.

The range of indicators will include both demographic and administrative records. Demographic indicators will include population estimates based on the 1991 census data – as these were themselves calibrated from figures rolled forward from 1981 it is to be expected that they will be suspect, by now, to some error. A current research project is estimating the likely level of error in these estimates – the main source of error is, as ever, error in the estimates of net migration. While at a national level estimates of error may not be too extreme, at a subnational level difficulties in estimating net internal migration may make the rolled forward estimates less reliable.

The demographic rolled forward estimates will be compared with the census data using demographic analyses similar to those used in 1991. These include straightforward comparisons of the age sex distributions and analyses of the sex ratios from the observed, underenumeration adjusted and rolled forward estimates.

Other comparisons will be made with administrative sources: these are important and include sources such as birth registrations, pensions registers, school and Higher Education Institution records. Net migration will be estimated using the International Passenger Survey and large surveys such as the Labour Force Survey. Although none of these sources has either a complete demographic, socioeconomic or geographical coverage they can each provide information on important subgroups which are often subject to underenumeration – for example it is well known that very recent births are underenumerated, students are particularly difficult to enumerate and the very old present particular problems. A key source will also be the Family Health Service Agency (FHSA) records. This latter source has been suggested to be a potentially useful source for updating local population estimates. Research as part of the ONC project has indicated that although coverage has improved greatly in recent years the FHSA data do not, at present, provide a base for an individual level comparison at a small area

level. It should be noted of course that many of these ancillary indicators are already used in population estimates.

A final key piece of ancillary information will be the debriefs of the CCS fieldwork staff as these will give some broad, if anecdotal, information on the extent to which the CCS has been a success in different design areas.

The question which remains is, given the comparison with these indicators, how will the judgement be made that the census adjusted estimates are reasonable. It has to be remembered that the indicators are all subject to error at some level and so none can be thought of as a 'gold standard' – indeed the gold standard should be the census.

To address this question one must first recognise that, in general, there appear to be two main risks: a) that the national population estimates are implausible; and b) that there will be implausible results in one or more design areas.

The former risk will occur if the CCS is uniformly unsuccessful in the different underenumeration groups across the nation. This seems highly unlikely but, as described above, evidence for its existence will be assessed by a strategy similar to 1991. If such evidence is found an agreed national age sex distribution will be used to pro rate the population upwards in different subgroups. It is, of course, not possible to write at this stage a formula on which the national age sex count will be based although, clearly, estimates (essentially) rolled forward from 1981, together with reliable national ancillary information on subgroups such as students and pensioners will form the base.

Much more likely is the second risk where fieldwork problems in one or more design areas will result in the estimates being unreliable in those areas. In this case the solution is relatively simple. One will borrow strength from other design areas through a multilevel model which estimates underenumeration in like HtC and demographic groups across the country and, if appropriate, weighted towards those close to the design area(s) with problems. In this case the weighting will be required if there is evidence from the statistical models that, for example, an 18 year old in a difficult HtC group in Manchester has a significantly different chance of being underenumerated to a similar man in Birmingham, Bristol or Bournemouth.

6. Risks in the cascade to Local Authority and very small area estimates

Following the production of design area estimates the full ONC process has two main additional tasks: a) to produce synthetic local authority estimates; and b) to cascade down to postcodes and impute both households and individuals missed within households enumerated in the census. The main risks inherent in these two tasks have been that the research to develop a strategy would be unsuccessful.

It is good then to report that in both cases the research results are extremely encouraging. The estimation of local authority population counts is straightforward and comprises a multilevel model to apportion the design area populations to its constituent small areas. The first risk here is that there is great heterogeneity in similar HtC groups within the same design area. This risk is to be minimised by careful selection of design areas. A second risk is that of model misspecification. In all design areas the data will be subject to extensive diagnostic checks and tests for model misspecification.

The imputation of individuals and households has been a very difficult research project but it is now yielding very encouraging results. The strategy is first to calculate weights of being underenumerated based on inverting the estimated probabilities for a series of multilevel multinomial logistic regressions where the outcome variable categories are: a) in the census and in the CCS; b) not in the census but in the CCS in a household not enumerated; and c) not in the census but in the CCS in a household in the census. These weights are then used to estimate the probability for each enumeration district of there being an individual underenumerated. The subsequent probabilities are then summed across postcodes to observe the number of individuals and households to be imputed in each local authority. As it is necessary for the resulting estimates to be constrained to the already agreed local authority estimates some iteration is necessary. Figures 3 and 4 show, for simulations based on 1991 data, the results of imputing across a local authority for households categorised by tenure. The bold line in both cases is the unadjusted census figures and the broken line is the imputed data. Figure 3 shows the bias – the percentage difference in the estimates across the entire local authority. This is very encouraging as the imputed data reproduces the 'true' distribution extremely well. Figure 4, on the other hand, gives the relative root mean square error of the estimates. These are less encouraging and indicate that there is some error in which postcodes to place individuals. This is clearly a risk but a partial solution may be found in the use of dummy forms – those used by the census enumerator to indicate that s/he feels there is a household at an address but from which a census form has not been returned either by post or to the enumerator.

7. Summary

The census is the largest data collection exercise in this country and is used for the decade following its collection as the base from which £600 billion are allocated to local authorities and health authorities alone. It is used for local planning and the delivery of services and is the evidence on which many commercial and social initiatives are made. It is also the source for enormous numbers of small projects. Indeed it is the only time that we really have a snapshot of our whole population at a very small level. It is therefore key that there is public acceptability of the results which have to be collected at a very

high level of accuracy. This paper has described the risks that were posed to the census following 1991 and has given an overview of the research project in ONS to produce a quality adjustment for underenumeration in 2001 and has highlighted the strategies used in that project to minimise risk.

Table 1 – Part of the One Number Census Risk Register

A. Developmental and Testing Risks

RISK	LIKELIHOOD	IMPACT OF RISK	EFFECTS	SECONDARY EFFECTS	RESPONSIBILITY	CONTINGENCY/ ACTION	STATUS
1. GENERAL MANAGEMENT **1.1Confidence** Failure to communicate the concept of a ONC to users: Internal ONS External	 Low Low	 Medium Large	Customer dissatisfaction		ONC Project team: Ian Diamond, Marie Cruddas	Consultation paper and further consultation to ensure methods fully understood.	Responses received to consultation paper - summary prepared for advisory groups and Census News.
Raising expectations too high	Medium	Medium	Customer and internal dissatisfaction	Ineffective expenditure			ONC Workshops held (April/May 1999) as part of Census Roadshows.
Consultation not adequate	Low	Small	Users criticise method	Undermines Census and mid-year population estimates			
Failure to provide adequate advice to users (as methodology involves degree of complexity).	Low	Small	Undermines user confidence in Census results	Undermines user confidence in P&VS population estimates	Andy Teague Ian Diamond Marie Cruddas	Simple guide to ONC	Guide to ONC (1999) circulated to Users
Damage to ONS reputation if ONC not successful.	Medium	Medium	Undermines user confidence, primarily in Census	Undermines user confidence in other ONS business areas, particularly P&VS	Tim Jones Graham Jones Andy Teague	Develop evaluation strategy for DR and apply findings for 2001 to avoid this.	

Table 2 – Median Precision for 24 Age/Sex groups given fixed national sample size.

National Sample Size	Postcodes Per ED	Median RRMSE of 24 age/sex groups for design group population of		
		1 Million	0.75 Million	0.5 Million
10,000	3	1.78%	2.06%	2.44%
	4	1.77%	2.06%	2.49%
	5	1.72%	2.02%	2.52%
20,000	3	1.36%	1.43%	1.70%
	4	1.37%	1.50%	1.71%
	5	1.35%	1.49%	**1.74%**
30,000	3	1.14%	1.25%	1.43%
	4	1.19%	1.28%	1.47%
	5	1.04%	1.25%	1.44%
40,000	3	1.04%	1.11%	1.31%
	4	1.03%	1.15%	1.28%
	5	1.04%	1.14%	1.25%

Table 3 - Approximate accuracy of total populations for different aggregations

Aggregation	Population	Likely accuracy of ONC estimate
England and Wales	52 Million	±62000
Design Group Level	500,000	±6000
LAD	200,000	±3500
LAD	120,000	±2000
LAD	60,000	±1500
LAD	30,000	±500

Table 4 - Range of dependence for which the DSE gives accurate coverage

Census Coverage = 95% (90%)

CCS Coverage	95%	85%
Expected DSE	99.00-100.28	99.00-100.94
Coverage	99.00-100.59	99.00-101.00
% of those missed		
by Census found	76.77-100.00	**68.69-100.00**
by CCS	86.36-100.00	77.27-92.57
Odds	7.19-0.00	2.77-0.00
Ratio	3.75-0.00	1.79-0.43

Table 5 - Southampton test response rates

100 non residential, vacant and communal establishment properties found

2086 households identified		
1777 interviews achieved	=	**85.2%**
127 no contact	=	6.1%
96 refusals	=	4.6%
86 no information returned	=	4.1 %

Figure 1 – The One Number Census Process

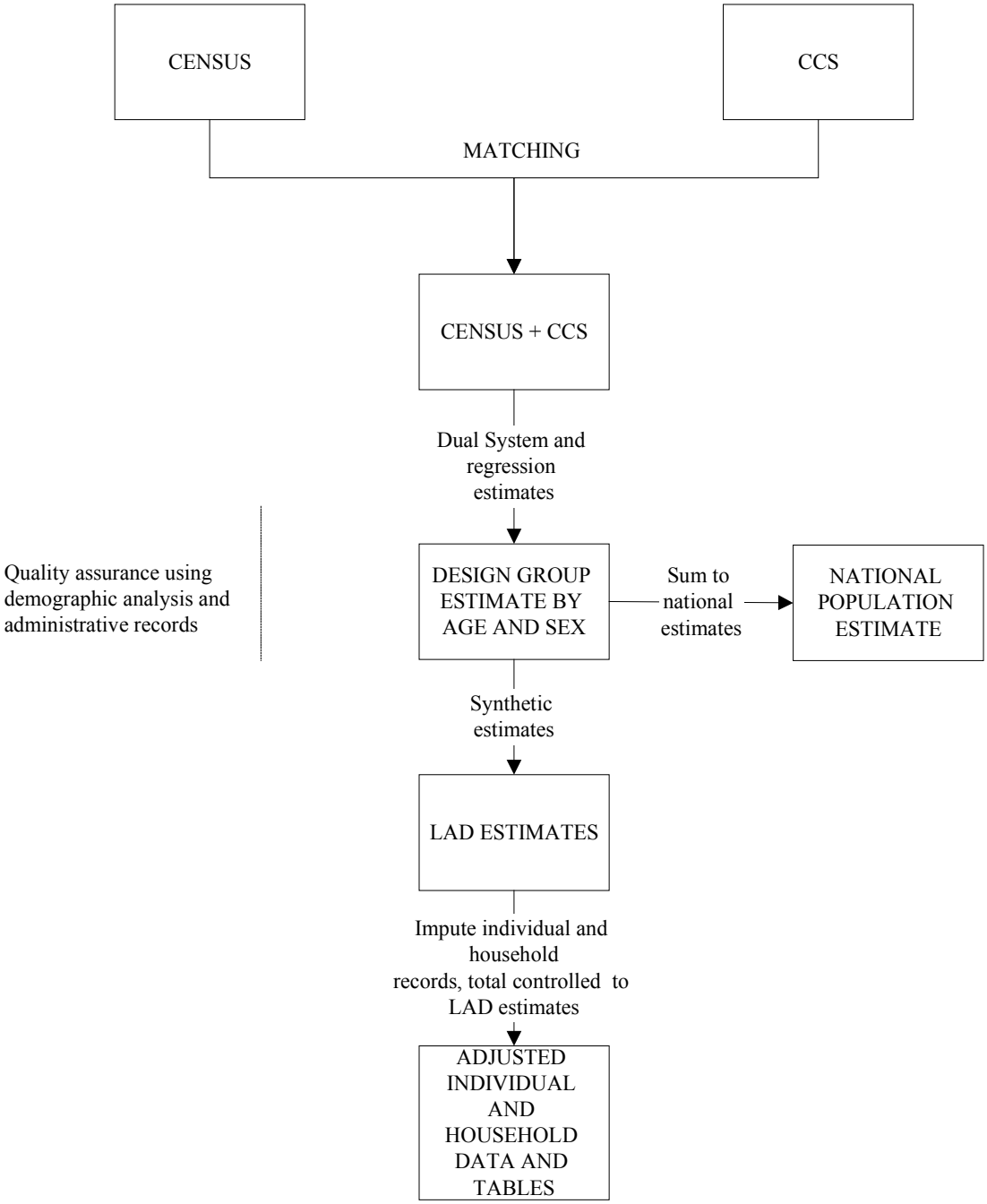

Figure 2 - Estimation of total population with simulated dependence using a Dual System Estimator: Census coverage = 90%

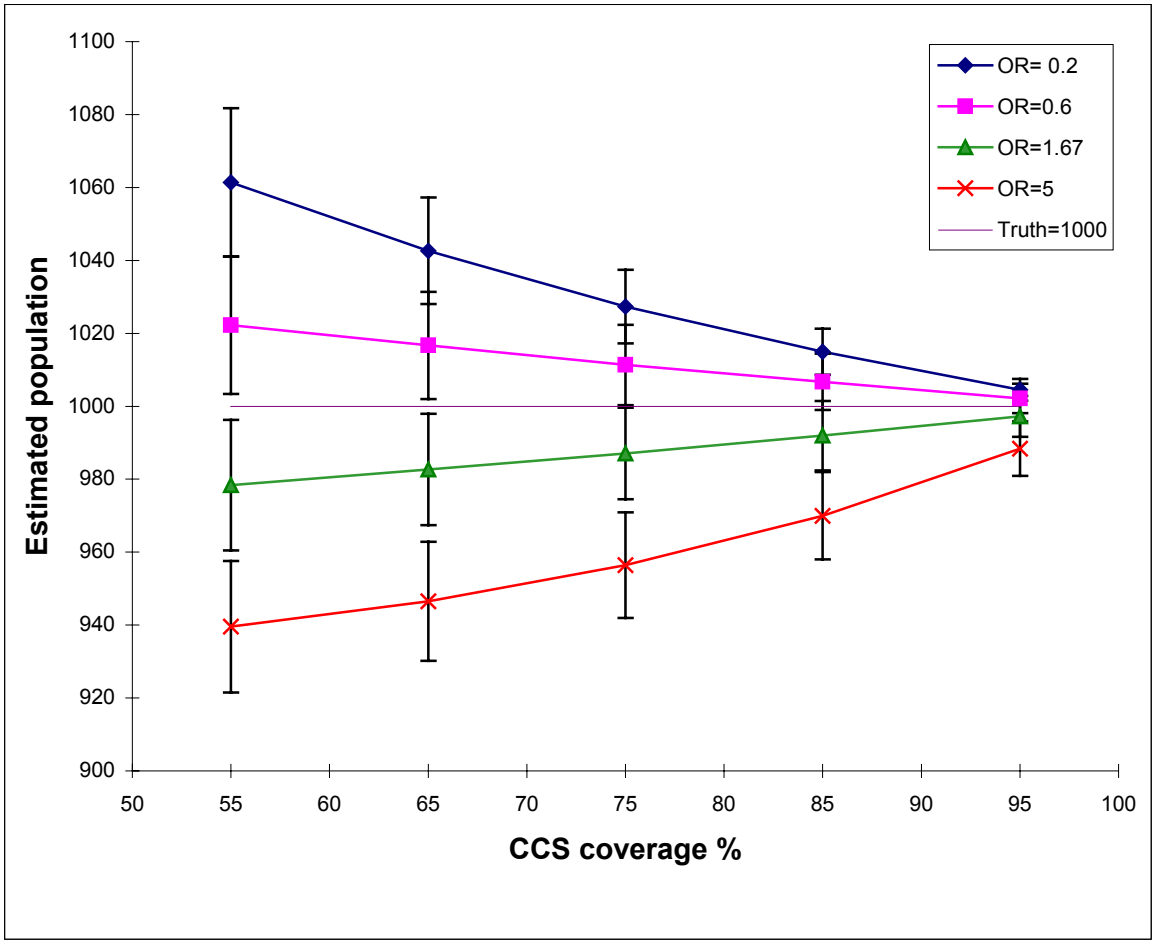

Figure 3 – Imputation of Households: Relative Bias for Tenure

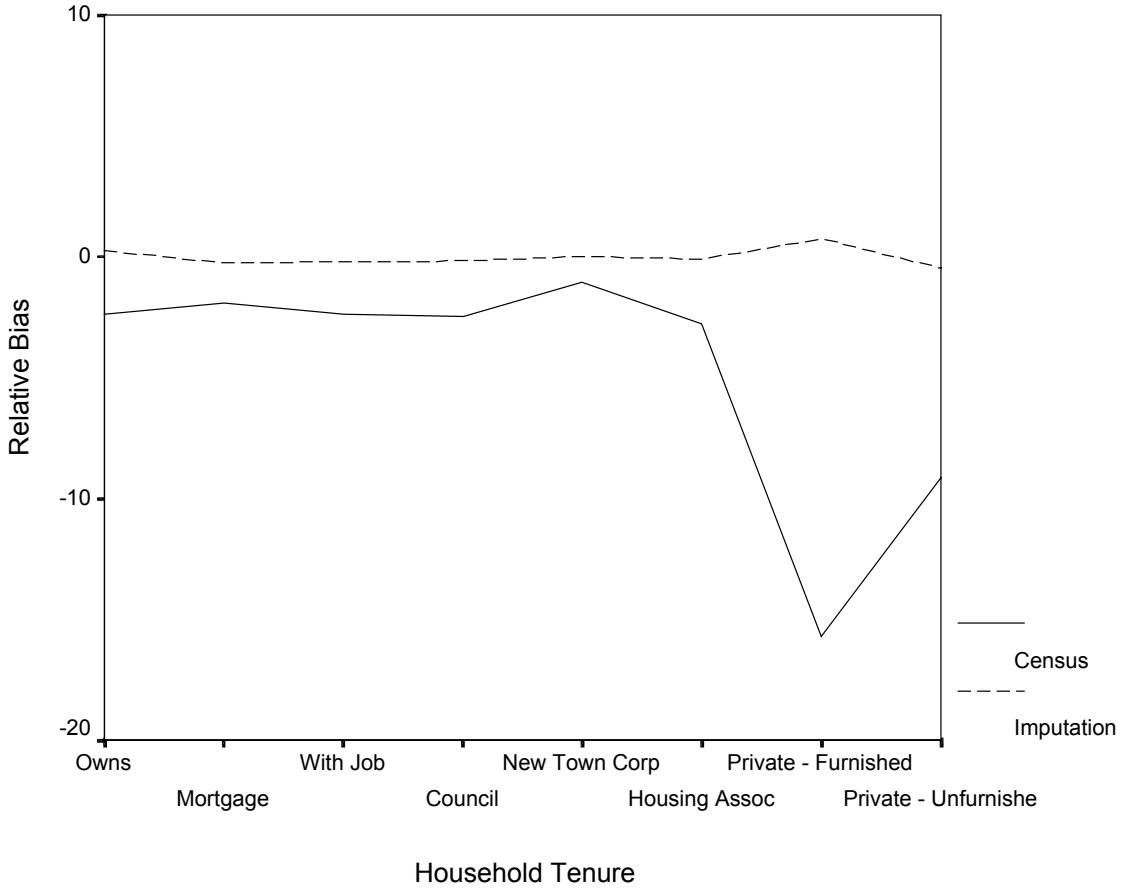

Figure 4 – Imputation of Households: Relative RMSE for Tenure