

THOSE MISSING MILLIONS Implications for Social Statistics of Undercount in the 1991 Census

Daniel Dorling and Steve Simpson

Abstract

This paper highlights some of the issues raised by the census undercount, illustrating how far-reaching its implications are across the spectrum of social statistics. The origins and the current state-of-play of the debate over the distribution of the under-enumeration are summarised, and the implications then assessed. For many census variables, the undercount reduces variation between areas, with consequences for the formulae used to resource local services. The implications for the reliability of the most basic of health statistics are raised through mortality rates. A rough estimate is made, and supporting evidence presented, of the possible numbers of households who were missed by the census. Some implications for housing needs assessments are illustrated along with the effect of the sudden upwards revision of official estimates of the dwelling stock and great uncertainty over the number of vacant dwellings in the country. Finally, recent research on the electoral implications of people's choice not to be enumerated is discussed as this may well be linked to census under-enumeration. Such a pervasive degree of uncertainty across almost all figures concerning people is new and there is little reason to expect it to diminish until the benefits of being known to the State outweigh the advantages of not being a statistic.

Introduction

As late as December 1991 it was assumed that the enumeration had gone well:

"The latest Census of Population in Great Britain was held in April 1991. One of the aims of the Census is to count all people in Great Britain on a particular night. This is relatively easy to do for people who live in houses or flats, or who are staying in a communal establishment such as a hotel, hostel or hospital. Buildings are easily identified and the people living within them enumerated...."
(Thompson, A., GRO Scotland, December 16th 1991)

In fact, censuses are never complete counts of the population. Residents can be missed for many reasons. They are away at the time of the census. They respond incompletely - for example a significant number of new-born babies are omitted from most censuses as if they are not yet considered to have an independent existence. Residents may be unwilling to respond, perhaps because they feel the census records

will be used against their interests, Residents may not be asked to respond because of a failure of the census field procedures. And finally but very importantly, residents may respond but not all be counted as residents because the census procedures have not coped with transient homeless and some other categories.

In 1991 in Britain the census validation survey which is designed to find a sample of those missed from the census output, in fact failed to find even half of them. How we come to know this is discussed below. The result is that we know rather little about the characteristics and the whereabouts of the census missing, now thought to amount to 1.2 million, over 2% of the total number of residents. Thus estimates based on the census are just what the word 'estimate' implies: figures with a degree of unreliability, not the hard baseline that social statistics in census year are usually taken to be.

So little do we know about the types of people missed, that we cannot really say with any confidence which of the reasons for under-enumeration were very significant in 1991. Perhaps they all were. The recent precedent of large-scale attempts at non-registration for an official register - the poll tax which was current at the time of the census - has been suggested as a major reason for the missing million. Parallels have been drawn with the 1381 introduction of a shilling poll tax and the consequent evasion and peasants revolt. At that time the poll tax payment lists were the only measure of the size of the population in England, which thereby apparently decreased from 1,355,201 in 1377 to 896,481 in 1381. In many villages, dependants were omitted from the lists, particularly young women (Oman 1906: Appendices 2 and 3). Interestingly, unbalanced sex ratios in the census were the clue used in 1991 to assess where the missing population were concentrated.

But the poll tax is certainly not the only culprit in 1991. In enumerated households in the 1991 census, there are an estimated 200,000 'visitors' who had no usual residence at another address (OPCS 1993a). If these had been real residents who feared that census residential status would count in the collection of the poll tax, would they not have omitted themselves altogether? So it has been suggested that it is likely that most of these visitors were homeless single adults or families, temporarily at an address on census night which they did not feel was a permanent address (Diamond 1993).

Equally, institutions - whether medical, care, prison, or student halls - were instructed to include as visitors those who had been living there for less than six months, presumably on the grounds that they were temporarily away from a household. But there was no equivalent instruction to householders to include persons recently moved to an institution, and many would have moved from a household in which they were the only person.

So there are clear reasons to expect a sizeable non-response in the census. Beyond conjecture, what do we know about the undercount and the reliability of the 1991 census results? The rest of the paper describes the official estimates of census non-

response, and goes on to illustrate the importance of recognising non-response when using and interpreting census and related statistics. We give examples of misleading effects on resource distribution, trends in mortality rates, estimation of housing need, and allocation of parliamentary seats, if non-response is not taken into account.

As yet the debate is young: the current (November 1993) government line appears to be that the undercount is important but small enough for most users to ignore. We argue otherwise.

Official Estimates of 1991 Census Non-response

The missing included within census output

How can a missing item be included? - when it is *imputed*. The census offices impute items missed off census forms; for example if a person has no marital status indicated on the census form, it is inserted based on the age of the person and the characteristics of the other members of that household. If car ownership is missing it is copied from a non-missing record which is otherwise the same as the incomplete record for characteristics known to be related to car ownership, including housing tenure.

In 1991 for the first time and because greater numbers of non-responding households were expected, whole households that did not respond were imputed. Based solely on the number of people and rooms noted or guessed by the enumerator, detailed records for the household and the persons within it were created by copying them from other forms that had been returned. 1.6% (869 thousand) of residents in the Great Britain census output have been imputed in this way. The imputation method is described by Mills and Teague (1991) and the extent of imputation of absent households can be seen from census reports; its geographical variation and the characteristics of the imputed records are summarised by Sandhu (1993). Across the whole of inner London, 8% of all residents in census output are created in this way, fictitious but for the enumerator's evidence that someone existed at the non-responding address.

Overall, the inclusion of these 'imputed residents' is probably better than ignoring the enumerators' evidence altogether, and it is hard to think of a better way of imputing the details than that which was used. Thus there is already within the census output a very considerable degree of unreliability for many areas. Now we consider the other, larger, part of census non-response - that not included in census output.

The missing not included in census output: creating national population estimates by age and sex

As already mentioned, most of those missed by the census were also missed by the census validation survey aimed at finding them. How do we know? Because analysis of the census adjusted by that survey shows a gap of nearly a million, half of them in their twenties and early thirties, when compared to the numbers expected by taking

the 1981 census and adding births and deaths and international migration for the intervening ten years. In particular, the sex-ratios for young adults were unbelievable in the census-based estimates, showing a surplus of females over males aged in their 20s but not for those in their 40s. These arguments were set down in an OPCS article (1993d) early in 1993. Since then, a further 145 thousand residents have been added to the numbers missing, also mainly young adults and predominantly male - due to a late communication from the Home Office to OPCS about the increase in settled refugees in the late 1980s. The total number of residents in Britain missed from census output therefore is now 1.202 million, 2.2% of the whole population (from OPCS 1992a and OPCS 1993a).

The 1981 census, rolled forward in this way, has been generally accepted in place of the 1991 census results as far as national population estimates go - and this will continue to be the case throughout the 1990s until the next census in 2001 (assuming better procedures then). While we too would accept the arguments that led to the rolled-forward estimates being used, the results are clearly not exact, either overall or for each age.

The missing not included in census output: creating sub-national population estimates by age and sex

So why have a census at all?! Well, the option of substituting 1981-based statistics is not available at local level - because the measures of sub-national migration required to roll forward local 1981 populations by age and sex are just too unreliable. And of course, there is nothing to replace the census estimates of most other characteristics at local level.

So what do we use for sub-national population estimates, for each of the local authority districts for example? One cannot stick to the census alone, because then there would be a discrepancy between the sum of the sub-national figures and the national figures which as we have seen are not based on the census. There has been a continuing debate on this issue since October 1992, when OPCS prepared their provisional census-based population estimates for 1991 (OPCS 1992b), distributing the national undercount evenly throughout the country. In other words they assumed that the proportion of 20-year old men missing was the same in Twickenham as in Liverpool as in rural Berkshire.

This essentially conservative assumption (dropped in the final estimates published by OPCS in the summer of 1993) was defended by drawing attention to the lack of evidence for other assumptions, and by making caricatured conjectures about the number of rural new-age travellers. We call it a conservative assumption (not 'neutral' as some would have had it) because it relied on an assumption of constancy, ignoring evidence which did in fact exist and which consistently pointed to an undercount much higher in urban areas, and particularly in inner London. Some of this evidence is summarised in Simpson (1993), including drops in the electoral roll, direct measures

for the numbers of infants missed which can be assessed from examining statistics of births in the year immediately before the census, and the well-established experience of social research and other censuses for many years.

The assumptions used in the final mid-1991 estimates might also be labelled conservative because they were simply based on different assumptions of constancy; but they gave a very different picture of the undercount in the 1991 census. Given the predominance of young adult men in the national undercount now taken for granted, the method looked for those areas where there seemed to be a particular shortfall of men and assessed these as most affected by under-coverage. More precisely, the assumption was that for ages 20-34, in each of ten types of district in England and Wales, the ratio of men to women relative to the national average should be the same in 1991 as the average of 1971 and 1981. The results are given in detail in OPCS/GRO(S) (1993). Highly urban areas are assessed as missing up to 25% of men in their twenties. These areas include the 14 Inner London boroughs, 9 metropolitan districts, and 27 city districts in non-metropolitan areas. This is in addition to the missing who were imputed within census output as described above.

While this approach to measuring the undercount does have considerable merit over the provisional assumptions of an even distribution of undercount, we are arguing that even if it is the best possible approach, it involves a high degree of arguable assumption and therefore approximation to an unknown degree; users should be very careful that their interpretation of census and population statistics is not affected. The precise assumptions are dealt with in more detail in Simpson (1993); in this article we are concentrating on the implications of a significant degree of unreliability rather than its source.

Perhaps we will know more of the characteristics of some of the missing 1.2 million when the Census Validation Survey reports are published (it will then be well over two years since the fieldwork was carried out, a very serious lapse given the importance the issue has subsequently taken). But it is likely that the most difficult problem in using the 1991 census data will remain estimation of the degree of under-enumeration in what is now in effect a biased sample.

Although attention has been paid to the undercount of persons, and some work is being done on the basic households count, the overall implications of this undercount and the possibilities of other undercounts have not yet been thoroughly addressed. Policy makers are now forced to make decisions based on information known to be defective while underfunded government statisticians face a decade of writing detailed caveats and putting wide confidence limits on much that they produce. A number of examples follow.

The Importance of Applying What We Do Know - Resource Distribution

Census statistics are used in several contexts to distribute resources from central government to local services. Variation in census measures of need are used to determine variations in resources given, for example to local government in the Standard Spending Assessment. This leads to fining ('capping') those councils that attempt to spend more money than they are assessed as needing to spend.

For several such census measures used in these calculations, urban areas tend to be at the high end of the scale of need. Since under-enumeration was concentrated in urban areas, their need is under-estimated unless the census counts are adjusted. The best one can do at present is to use the age-sex-geography-specific rates of under-enumeration provided by OPCS as described above, as weights in a census analysis. One can do this whenever the characteristics of need are disaggregated by age and sex and area, as many are. In survey terminology, this is no more than post-stratification using estimated response rates for different strata.

Table 1 shows how this weighting affects the national estimates from the census of unemployment, non-white ethnic group, and illness in households.

Table 1: The impact of adjusting for census non-response on national estimates of unemployment, non-white ethnic group, and illness in households

	England and Wales Census count (000s)		Estimated under-enumeration
	Actual (unadjusted) (1)	Adjusted (2)	(2)/(1)
Unemployment	2,235	2,325	+4.0%
Non-white ethnic group	2,952	3,097	+4.9%
Illness in households	6,039	6,110	+1.2%
All residents	49,890	51,000	+2.2%

Sources: Census LBS output; rates of under-enumeration specific to age, sex and district type, from OPCS/GRO(S) 1993.

Because unemployment and non-white ethnic group are both characteristics that are concentrated in urban areas and to some extent among young male adults, this demographic weighting shows them to have been under-enumerated by twice the national average non-response of 2.2%. Illness in households on the other hand, which

is concentrated among older age-groups, is under-enumerated to a lesser extent.

Table 2 shows the mean impact of *not* adjusting for census under-enumeration, on the distribution of a grant to districts on the basis of their need as measured by (a) the number unemployed, and (b) the number of non-white residents. Main metropolitan districts would lose 3.5% of such a grant if it is calculated on the basis of raw census counts, rather than on the basis of census counts weighted by the officially estimated response rates. Rural areas would gain. The impact on district comparisons would in some cases be greater than those illustrated (non-white youth for example), and less in other cases.

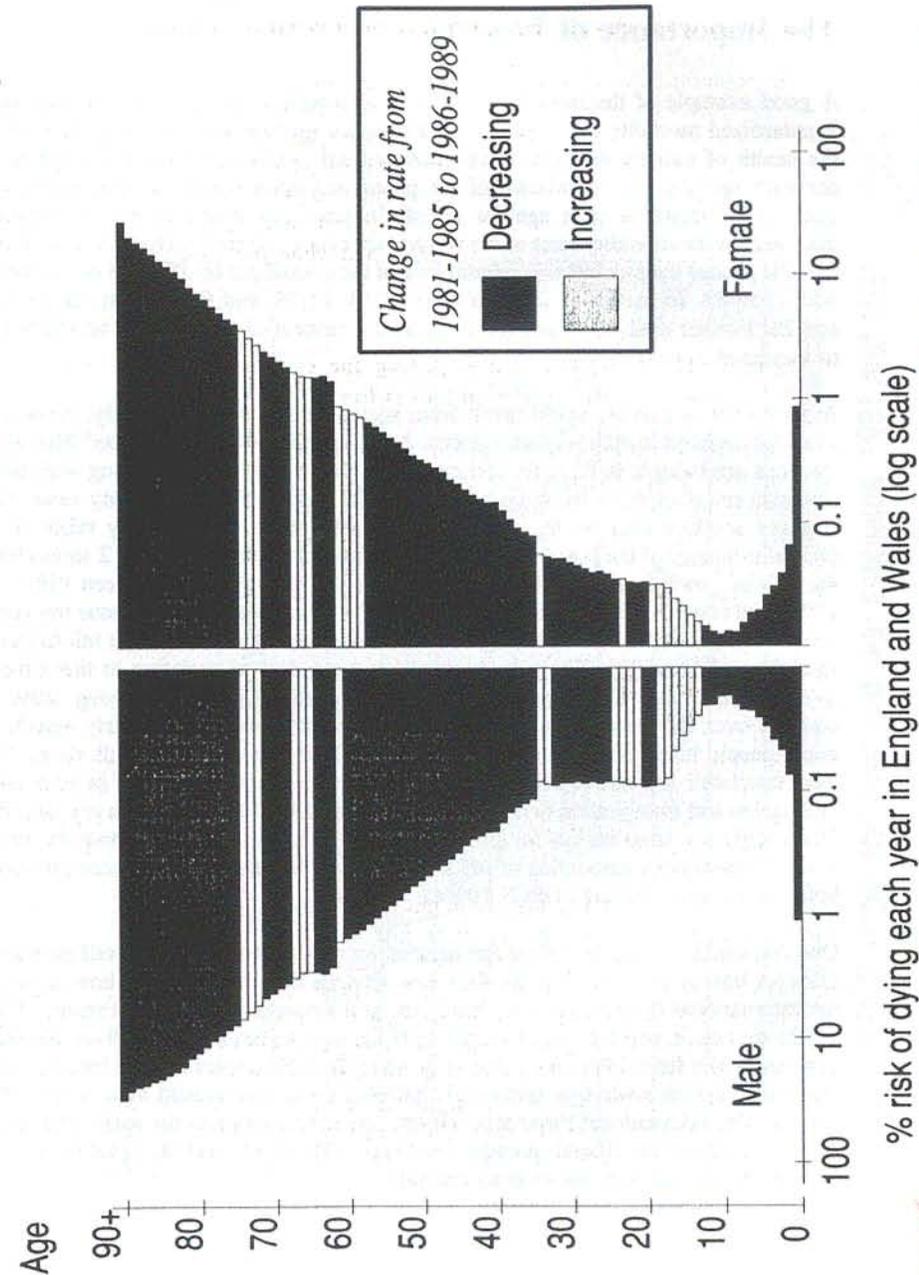
Table 2: The impact of not adjusting for census under-enumeration on distribution of a grant within England and Wales

(a) Grant distributed according to census unemployment	
Main metropolitan districts	-3.5%
Non-city, non-metropolitan districts	+1.7%
(b) Grant distributed according to census non-white ethnic group	
Main metropolitan districts	-2.1%
Non-city, non-metropolitan districts	+2.1%

Tables 1 and 2 show the impact of not adjusting census counts purely on the basis of what is now officially estimated to be sex-age-area response rates. The implication is that such adjustments should be made. The unreliability of the estimates, however, implies that even adjusted census figures should not be treated as providing precise measures (even of these simple statistical interpretations of need) as is the case when capping local authorities. Much less should raw census figures be so used.

Of course, age, sex and broad geography may not account for all the variation in census coverage. If it turns out from the census validation survey when it is published, or from other sources, that unemployed and non-white residents were more often missed than those employed or whites *within* these demographic sub-groups of the population, then the impact of non-response will be rather greater. This seems quite likely. In Bradford for example, non-white young children were twice as likely to be missed by the census as white young children (from comparison with recent birth records in Simpson 1993). Does this mean that not 20% but 40% of non-white young men were missed? It may well be so. Unfortunately, the evidence is hard to come by.

Figure 1: Mortality Rate Calculated From Deaths Between 1981 And 1989



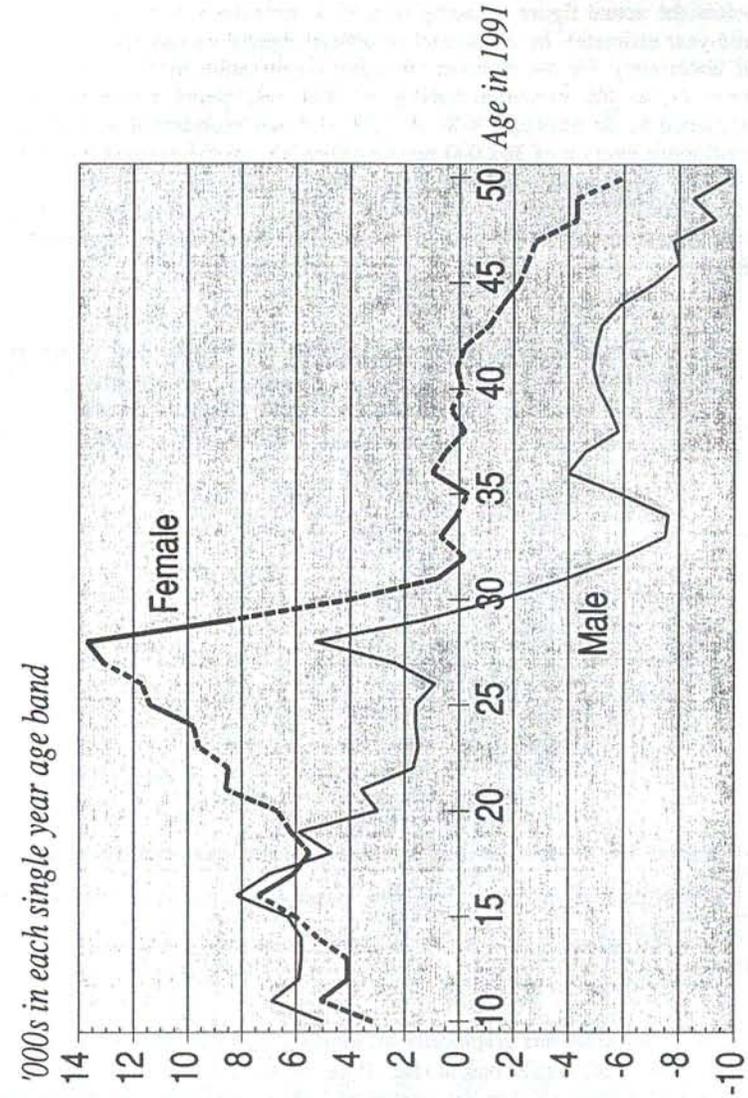
The Importance of Precision - Mortality Rates

A good example of the need for precise population counts are mortality statistics. Standardized mortality rates are the main measure for charting the changing state of the health of nations internationally. Standardized mortality statistics rely upon an accurate age and sex breakdown of the population upon which to standardize. An analysis of deaths at each age for all of England and Wales shows, in general, decreases in the mortality rates as we would expect (see Figure 1). However, for some specific groups the rate has risen slightly when the overall period 1981-85 is compared with 1986-89. In particular for men aged 13-18, 21-24, and for women aged 13-19 and 23. Further analysis shows a significant component of this rise can be attributed to increased rates of suicide.

A great deal of concern could result from such a finding and, eventually, resources could be diverted to address this concern. A key question then is, is it true? Mortality statistics are thought to be quite accurate (they have been registered along with birth statistics since 1836 - Champion 1993). But the calculation of mortality rates also requires accurate data on the populations at risk, which are normally taken from official estimates of the population for each sex in each age band. Figure 2 shows how the official versions of these imply the population to have changed between 1981 and 1991. The cohorts which were aged 10 to 30 in 1991 grew over the previous ten years due to a small surplus of net immigration over deaths, while there was a fall in those aged 30 to 50 in 1991. There is, however, a great deal of variation to this simple pattern as the graph shows, and it is this variation which is the underlying cause of concern over the variation in mortality rates over time. We can be fairly sure how many people have died, but much less certain about how many are still alive. The most unreliable element of this variation is thought to be the estimate of international emigration and immigration derived from the International Passenger Survey, and the Home Office's contribution for people seeking asylum (lack of information about which caused errors amounting to 145,000 people in the final 1991 mid year estimates before they were revised - OPCS 1993a).

Our knowledge of whether there has been an increase in teenager's overall mortality rates (or how large this has been) thus now depends to some degree on how accurate the International Passenger Survey has been, as the national mid-year estimates of the 1990s are not, in essence, based on the 1991 census. As has been described, national level estimates from 1991 onwards owe more to the 1981 census and the International Passenger Survey since then (plus registration of births and deaths) than to the 1991 census. The International Passenger Survey was never designed for such a reliance. So, even where an official attempt has been made to estimate the undercount of people, by age and sex, uncertainty remains.

Figure 2: Surplus of net immigration over deaths to England and Wales 1981-1991



Source, OPCS single age band mid year estimates, 1981 and 1991

The 1991 Census Validation Survey has yet to be released. The 1981 Post Enumeration Survey (OPCS 1985, p.4) calculated that a 95% confidence interval to contain the undercount in 1981 would range from 152,000 to 278,000 (30% above and below the actual figure - leading to a 95% confidence interval of 0.1% in the final mid-year estimate). In the absence of official figures we can assume this same degree of uncertainty for the estimate of under-enumeration in 1991. This is charitable, however, as the validation survey of 1991 only found a minority of the people presumed to be missing. 30% of 1,209,000 not enumerated in 1991 gives a 95% confidence interval of 363,000 people either way on the revised final 1991 mid year estimate (an overall rough 95% confidence interval of 0.7% on the entire population of England and Wales). This uncertainty is higher for certain groups, in particular for young men, as the table below illustrates. The table gives a range of mortality rates making the same assumptions as were outlined above.

Table 3: Mortality Rates for young men in England and Wales in 1991

Males Aged	Mid Year Estimated '000s	Census Enumerated '000s	Size of** Undercount percent	Recorded Deaths actual	Mortality* Rate per million
13	296	287	3%	79	264-269
14	292	289	1%	90	308-309
15	305	301	1%	133	435-438
16	320	317	1%	178	556-558
17	330	324	2%	245	737-746
18	354	340	4%	329	920-941
19	377	354	6%	323	842-874
20	394	365	8%	322	799-835
21	393	358	10%	335	829-875
22	406	367	11%	348	832-882
23	409	370	11%	389	923-978
24	422	383	10%	366	845-892

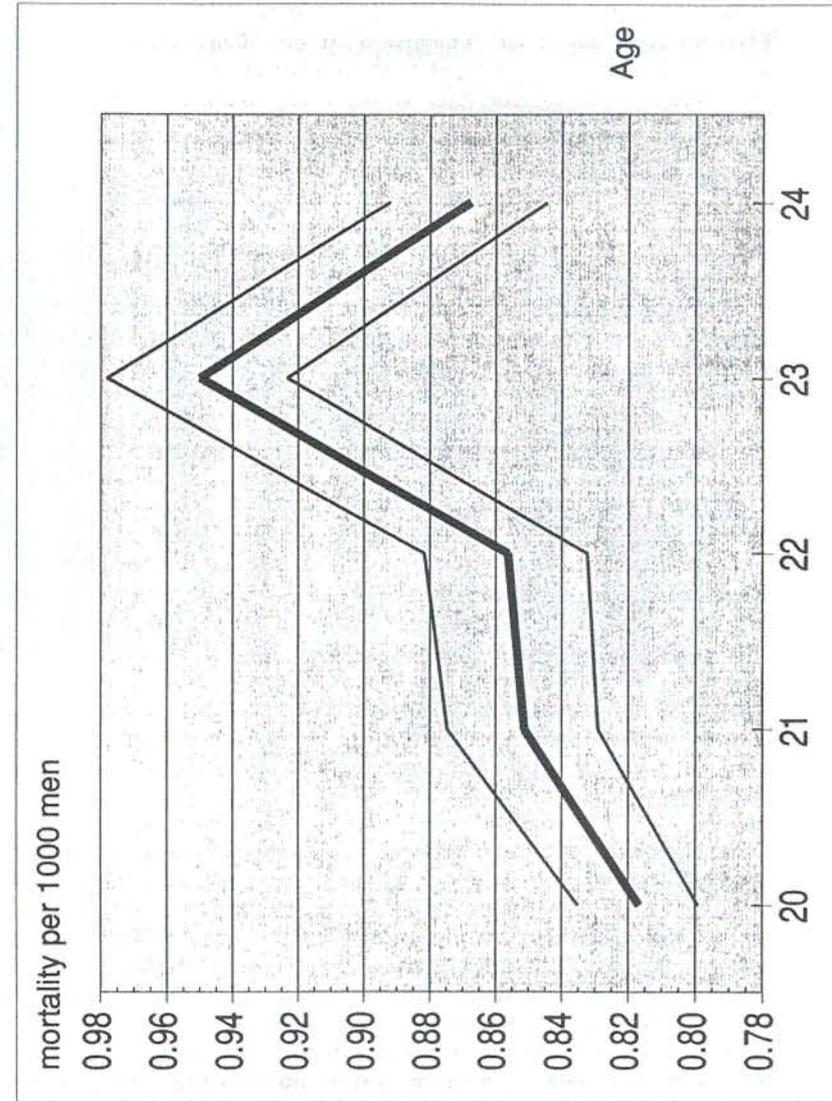
Source: 1991 mid year estimates, mortality statistics, and census sex age & marital status tables (OPCS 1993a,b,c)

* Confidence limits are presented for the mortality rate assuming the undercount estimates may be out by +/- 30%.

**Not an exact estimate of the undercount as slightly different population definitions are used for the two sources.

Figure 3 illustrates this graphically by showing the mortality rate calculated from the above table with confidence limits. OPCS do not publish confidence limits for the mid-year estimates despite the increasing lack of confidence which we can place on them. Variation estimates are more usually assigned to the numerator, and most statistical tests applied to look for significant change in these rates assumed the count

Figure 3: Variance in the mortality rate of men aged 20-24 in England and Wales, 1991 showing the extent of a 95% confidence interval (see text for sources)



of the underlying population to be correct. Until we have some estimate for the reliability of the mid-year estimates we cannot confidently answer a basic question such as "are men more likely to die at age 18 or at age 23?" with any degree of certainty.

The Importance of Assumptions - Household Numbers

"There is considerable scope for inaccuracy arising from the data and estimates used. Even in the base year the estimate of the number of households is likely to be based on extrapolated Census results ..."

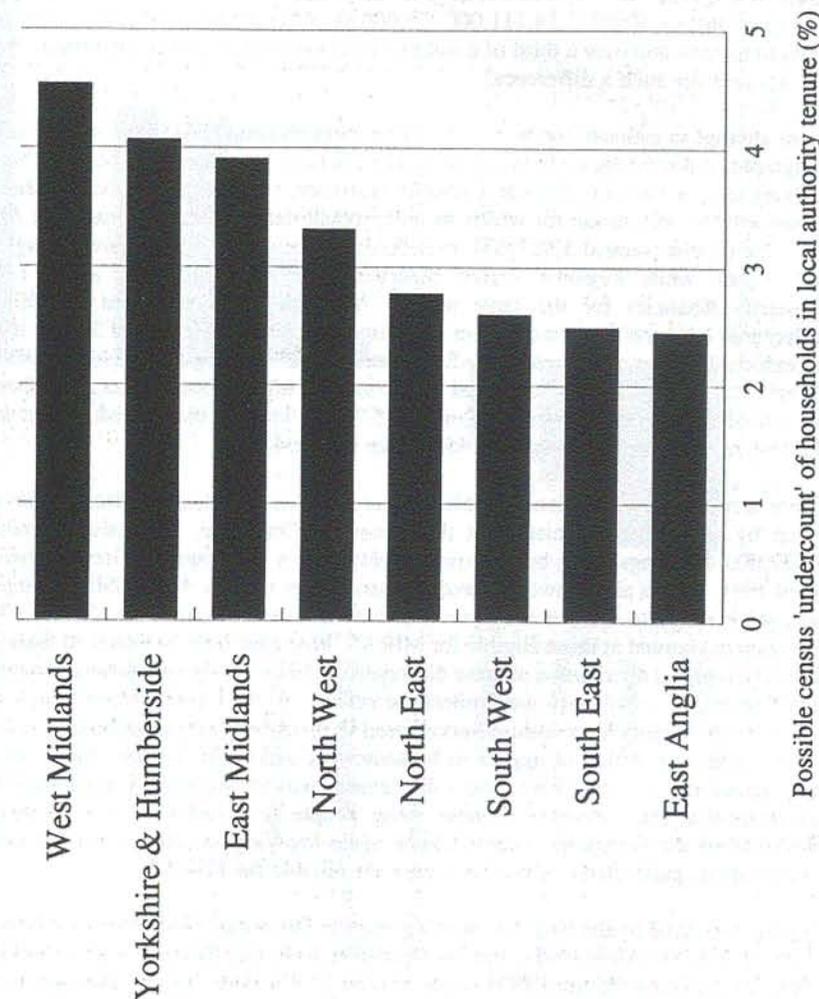
(Corner, 1991, p.4)

Although it has been agreed that the 1991 census missed over a million people, there is as yet no consensus as to how many households the enumerators and the imputing procedures may have missed. But does such under-enumeration matter? Households cannot be as important as people, can they? Standard spending assessments, mortality rate estimates, votes and schools' rolls may not depend on them, but important local decisions and national lobbying is made on the need for and the provision of housing. As the community charge is withdrawn and the council tax emerges, it may well be more and more whole households, rather than individuals, who choose to be missed. An empty dwelling or single person household is liable to pay only 50% or 75% of the full council tax rate respectively.

In a memorandum to the Select Committee on Environment's inquiry into the Housing Corporation, Christine Whitehead produced figures showing how the estimated extra need for social housing in Britain per year ranged from none to 128,000 additional units, depending on whose estimates you believed (Whitehead 1993, Kleinman & Whitehead 1989, 1992, Niner 1989, Wilcox 1990). All these estimates, however, are based on the Department of the Environment's estimate of how many households there are to be housed. Their latest estimate for England in 1991 was 19,036,000 - still based on population estimates from 1989 (DoE 1991; for a critique of DoE methodology see King 1990).

If the estimate of household numbers is only slightly incorrect the effect on the estimated need for new housing would be dramatic as, very simply, this 'need' is based on the difference between the number of dwellings and households in the country. Hence an undercount of just 0.5% in our estimate of household numbers can conceal a possible demand for one hundred thousand dwellings. By the late 1980s official estimates of household numbers became less and less robust as they had to rely more and more on non-census figures. The DoE projection for the year 2001, for instance, has had to be revised upwards in both 1988 and 1991 by about 600,000 households each time for England and Wales (Dorling 1991). Because of this uncertainty great hopes were placed on the 1991 census to clarify the confusion

Figure 4: Discrepancy between Regional Trends and Census Figures



surrounding such basic housing statistics and, in particular, to provide a national count of dwellings which had not been done, systematically, since 1931.

The 1991 census, however, only enumerated 18,766,000 households with residents in England, over quarter of a million fewer than were thought to exist. The latest figure from the Department of the Environment is that which they use in the English Housing Condition survey of 1991 - 19,111,000; 75,000 more households than were previously thought to exist and over a third of a million more than the census enumeration. What can account for such a difference?

In an attempt to estimate the household undercount Dorling (1993) has looked at the geographical discrepancies between the census count of households in Local Authority housing and the number of Local Authority tenancies. Households in Local Authority tenure are the only group for which an independent national count is made. In April 1991 the census counted 3,927,000 households in Local Authority tenure in England and Wales, while Regional Trends reported there were 4,050,000 active Local Authority tenancies for the same month. Although there are slight definitional differences between these two figures, the discrepancy for this group of 3.13% is not unreasonable when compared to the 2.2% undercount for the population as a whole. Geographically the county suggested as having the highest rate of Local Authority household under-enumeration was London (8.73%), whereas the English region with the lowest rate was East Anglia (2.45% - see Figure 4).

There is corroborating evidence of this undercount from other tenures through answers given by government ministers in the House of Commons. The census counted 9,279,000 dwellings being bought (presumably with a mortgage) in Great Britain in April 1991. This is almost two hundred thousand fewer than the 9,460,000 mortgagees who were stated to exist at the time in a written answer in Hansard (Dorell 1993; mortgagees counted as those eligible for MIRAS, Mortgage Interest Relief At Source). The geographical distribution of these discrepancies has a very varied pattern, however (see Figure 5), which may well reflect the reliance in their construction which was placed in the Family Expenditure Survey used to produce the regional breakdown. In short - both sets of figures appear to be somewhat unreliable and the objects which they define are in neither case clear - the census count of households buying and the government's best estimates of how many people or couples have a mortgage. Nevertheless the mortgagee count is higher, again implying possible household under-enumeration, particularly as not all buyers are eligible for MIRAS.

Figures produced by the DoE for dwelling stock in December 1991 (for Great Britain) were 22,984,000 (DoE 1992), and for December 1990 this figure was 22,815,000, a crude interpolation for April 1991 would thus be 22,871,000. The census enumerators estimated that there were 23,000,468 dwellings in April 1991, some 130,000 more purpose built homes than were previously thought to exist (under the census definition). Figure 6 shows how the estimates of dwellings were retrospectively

Figure 5: Discrepancy between Mortgage and Census Figures for home buyers

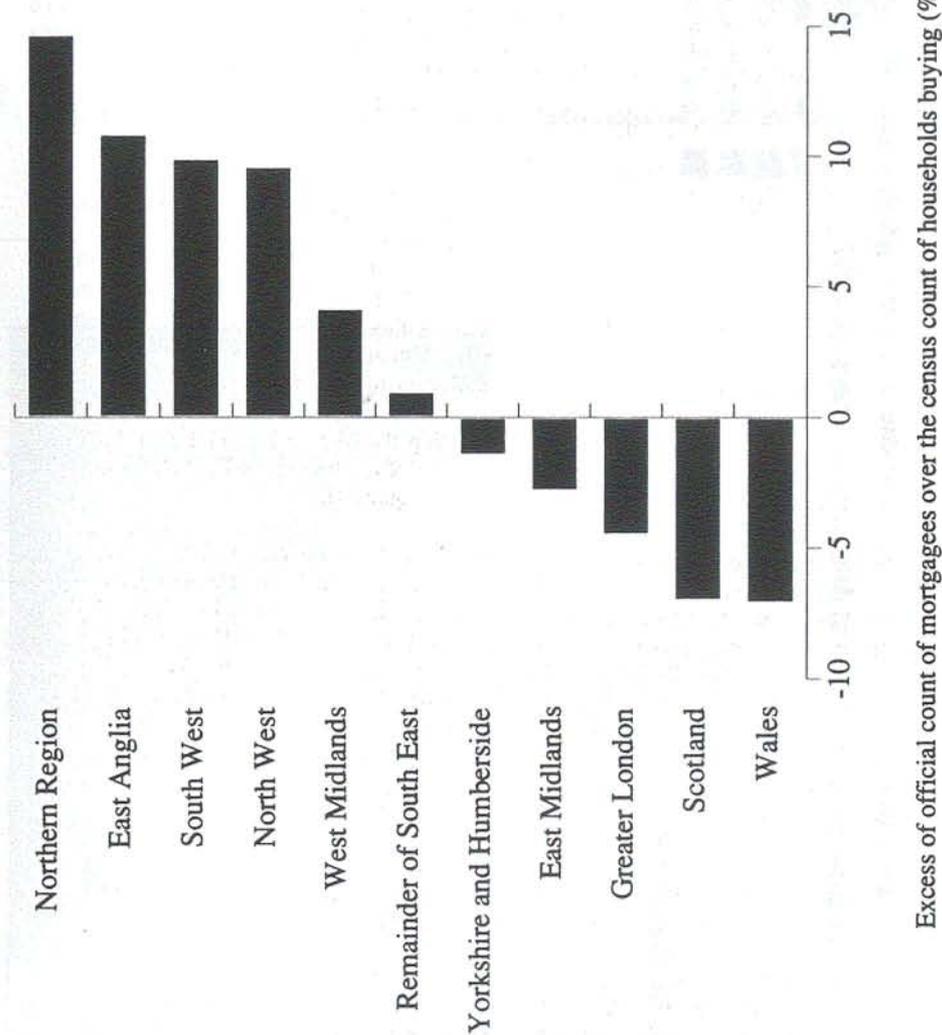
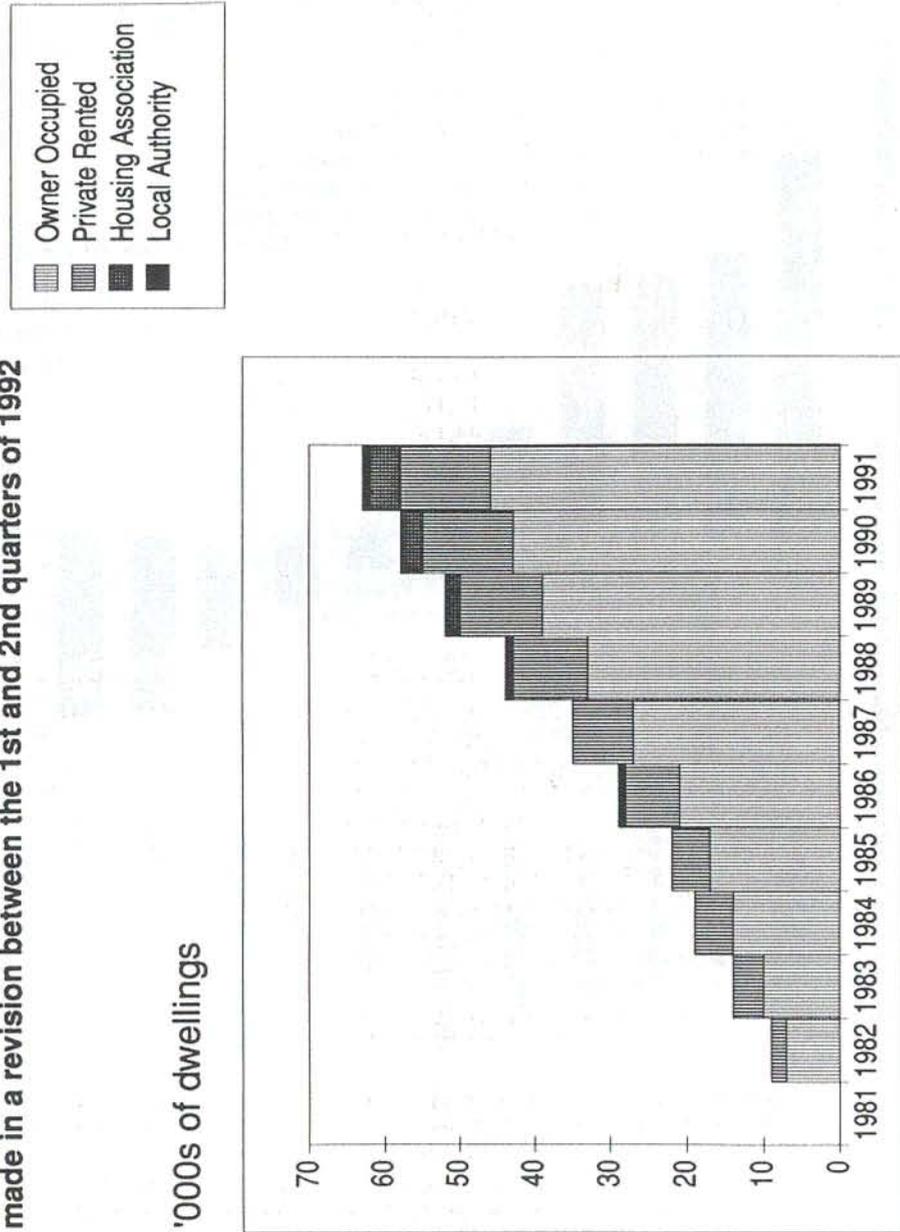


Figure 6: Retrospective Additions to the estimate of Dwellings in England made in a revision between the 1st and 2nd quarters of 1992



revised upwards during 1992 (although no reference was made to the impending release of the new census statistics then).

The DoE also appear, however, to have revised downwards the estimates of vacant dwellings, as compared with the census (see table below), although no official statement has yet been made:

Table 4: Possible Extent of the Household Undercount using Official Sources

England ('000s)	EHCS	Census	Difference from the census
Dwellings	19,700	19,671	-0.1 %
Households	19,111	18,766	-1.8 %
Population	48,100	46,337	-3.7 %
Vacant dwellings	639	1,092	70.9 %
Vacancy rate	3.2 %	5.6 %	71.1 %

Source: 1991 census (OPCS) and 1991 English House Condition Survey (DoE, 1993)

The above table illustrates how even a small estimate (1.8%) of the national under-enumeration of households can translate into a 71% overenumeration of vacant dwellings and thus have a dramatic effect on national housing debates. The English House Condition Survey figures (DoE 1993) were not presented in a succinct form however, but had to be extracted from where they are scattered throughout that report. Intuitively, and with the evidence presented above and from evidence that many properties thought vacant in Scotland were occupied (Scottish Office 1993, unfortunately no figures have been published for England and Wales) we suspect that more faith should be placed in the DoE figures than in the raw census counts. The DoE have assumed a household undercount that falls within the range suggested by what independent comparisons can be made.

No Representation Without Taxation

Undercounts are about far more than official spending, international health league tables or lobbying for house building. The reason for the undercount can have as marked an effect as its statistical implications. We have already mentioned how the poll tax of 1381 contributed to a drop of a third in the enumerated population of England. Iain McLean and Jeremy Smith (1994) go on to estimate that 520,000 electors in England and Wales may not have registered due to the poll tax of 1990 and

that, if they had registered and 40% had voted for the opposition parties in 1992, then the Conservative party would not have been able to form a government at the last election.

The abolition of the poll tax will not end this effect, even if non-registered voters returned, because the fourth general review of the Parliamentary Boundaries of England and Wales will determine the pattern of seats and hence representation until well into the next century. The 1991 electoral register is being used to construct these seats, even though it is thought to be one of least well completed of modern times, and to be heavily biased in its omissions:

"the 1991 electoral register probably under-reports the number of young, poor, mobile, and ethnic minority citizens across the whole country. This gives the opposition parties a disadvantage which will endure until 2006. By contrast, the advantage they gained in 1992 among those still on the register from the unpopularity of the poll tax was evanescent. Resentment against the tax had been defused by the Conservative's intention to abandon it. By contrast, its structural effects remain an obstacle to Labour's prospects of winning again. [There is] no tendency, up to 1993, for individuals to re-enrol on the electoral register now that the poll tax is dead. If they never re-enrol, the advantage gained by the Conservatives in 1992 will remain as long as their generation remains in the population."

(McLean and Smith, 1994, in draft)

Conclusion

The undercount has profound effects across the spectrum of social statistics. From people to health to housing to politics it permeates all our supposed facts with a previously unknown degree of uncertainty: How many people are there? Where are they? How old are they? How healthy is the population? How much health care does it need? How much housing does it consume? How much housing is there to consume? How many households are there to consume it? What government would the people choose if they were all registered to vote? Why do they choose to be hidden? To whose advantage is it not to know? . . . The more finely poised the situation, the more accurate the underlying data needs to be. If we want to know how things have changed or how they are now balanced we require a degree of accuracy above that which current official statistics can give us.

There is also no reason why the undercount should suddenly go away and the reliability of our statistics magically rise overnight. People are unlikely to get tired of not filling in forms, to find a new sense of "common-purpose" - seeing appearing in official lists as economically or socially worthwhile. If the state wants people to

register it has to be worth their while to do so. No amount of threats of fines, more enumeration officers, or the merging of huge social security computer files will find people who do not care to be found. Give people a reason to be found if you actually want to find them.

In the meantime what can those who use these statistics do? It is now imperative that error rates and confidence limits be estimated and included in publications where appropriate. This should be done not only to provide a more true reflection of the actual state of knowledge, but also to persuade those who decide how to fund and organise the collection of these statistics of the importance of being more thorough. If government ministers had to answer parliamentary questions with wide-ranging probabilities provided by their officials then perhaps they would be more concerned about why the figures are so unreliable and then wonder why so many people do not want to be counted.

Acknowledgements

Thanks are due to Tony Champion and Mike Coombes who commented extensively on an early draft of parts of this paper.

Authors' addresses: Daniel Dorling, Department of Geography, Daysh Building, University of Newcastle-Upon-Tyne, NE1 7RU. Steve Simpson, Bradford Metropolitan Council, City Hall, Bradford BD1 1HY.

References

- Champion, A.G. (1993), Population distribution and change in Great Britain, 1981-1991, *Geography Review*, Vol. 7, No. 1, pp. 10-15.
- Corner, I.E. (1991) Modelling housing demand, paper presented to the Statistics Users' Council's Annual Conference at the Royal Society on 26 November 1991.
- Diamond, Ian (1993) Where and who are the 'missing million'? Measuring Census of Population undercount. Conference on regional and local statistics, London: Statistics Users Council.
- DoE (1988) 1985-based household projections 1981-2001, Directorate of Statistics, Government Statistical Service, London.
- DoE (1991) 1989-based household projections 1989-2011, Directorate of Statistics, Government Statistical Service, London.

DoE (1992) Housing and Construction statistics, March 1992, Table 2.23, Directorate of Statistics, Government Statistical Service, London.

DoE (1993) English House Condition Survey: 1991, London: HMSO.

Dorling, D. (1991) The demand for housing in Britain 1971-2001, NorthEast Regional Research Laboratory Research Report 91/8, University of Newcastle upon Tyne.

Dorling D. (1993) The implications of the 1991 census for housing provision studies, University working paper - forthcoming.

Dorell, S. (1993) Written Parliamentary answer on MIRAS, Hansard Col. 239, Proceedings of the House of Commons, 17th March 1993.

King, D. (1990) Housing needs in the 1990's, Challenges, Choices and Change, Institute of Housing Conference, 6-8 April.

Kleinman, M. & Whitehead, C. (1989) Demand for new housebuilding, 1986-2001, Development and Planning, Department of Land Economy, University of Cambridge.

McLean, I. & Smith, J. (1994) The Poll tax, the electoral register, and the 1991 census. In Heath A, Jowell R, and Curtice J (eds) Can Labour Win? (forthcoming) Aldershot: Dartmouth Press.

Mills, Ian and Teague, Andy (1991) Editing and imputing data for the 1991 census. Population Trends 64:30-37.

Niner, P. (1989) Housing needs in the 1990s, National Housing Forum, National Federation of Housing Associations, London.

Oman, Charles (1906) The great revolt of 1381, Oxford: Clarendon Press (reprinted by Greenwood, 1969).

OPCS (1985) 1981 Census Post-Enumeration Survey, London: HMSO.

OPCS (1992a) Great Britain Census National Monitor CEN 91 CM56, London: OPCS.

OPCS (1992b) Provisional mid-1991 population estimates for England and Wales, Monitor PP1 92/1, London: HMSO.

OPCS (1993a) Revised final mid-1991 population estimates for England and Wales, Monitor PP1 93/2, London: HMSO.

OPCS (1993b) Mortality statistics, general, series DH1 no.26, London: HMSO.

OPCS (1993c) 1991 Census sex, age and marital status tables, London: HMSO.

OPCS (1993d) How complete was the 1991 census? Population Trends 71:22-25.

OPCS/GRO(S) (1993) Undercoverage in the 1991 Census. Census User Guide 58 (forthcoming). London: HMSO.

Sandhu, Amarjit (1993) Issues of imputation in the 1991 Census. Census Microdata Unit Working Paper 1. University of Manchester.

Scottish Office (1993) The 1991 Post Census Survey of Vacant Property, Edinburgh: Scottish Office Library.

Simpson, Stephen (1993) Measuring and coping with local under-enumeration in the 1991 census. Paper to conference 'Research On The 1991 Census', Newcastle. Available from the author at Bradford City Council.

Thompson, A. (1991) People sleeping rough at the time of the 1991 census of population, The Statistics of Homelessness, Statistics User's Council Seminar, 16 December, IMAC Research, Kent.

Wilcox, S. (1990) The need for social rented housing in England in the 1990's, Institute of Housing, London.

Whitehead, C. 1993, Memorandum to the Select Committee on Environment Inquiry into the Housing Corporation, personal communications.

Whitehead, C. & Kleinman, M. (1992) A review of housing needs assessment, LSE Housing, The Housing Corporation, London.