

Those Missing Millions: Implications for Social Statistics of Non-response to the 1991 Census

STEPHEN SIMPSON* AND DANIEL DORLING†

Received 29.11.93

Accepted 28.2.94

ABSTRACT

This article highlights some of the issues raised by incomplete coverage of the 1991 census in Britain, illustrating how far-reaching are its implications across the spectrum of social statistics. The origins and the current state-of-play of the debate over the distribution of the non-response are summarised and the implications then assessed. For many census variables, non-response reduces the observed variation between areas, with consequences for the distribution of resources for 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 which 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 the resulting overall 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 registered is discussed as this may well help to understand census non-response. Such a pervasive degree of uncertainty across almost all figures concerning people has not been experienced in Britain since the last century. There is little reason to expect this uncertainty and the importance of its social policy implications to diminish in the future, without significant changes to census procedures and to the relationship between government and governed.

The census data used in the paper is Crown Copyright. Thanks are due to Tony Champion, Mike Coombes and Sue Heath who commented on a draft of parts of this article. An earlier version appeared in the newsletter of the Radical Statistics Group. Some of the work reported was funded by the British Academy and the Joseph Rowntree Foundation.

* Principal Research Officer, Chief Executive's Office, City of Bradford Metropolitan Council.

† British Academy Fellow, Department of Geography, University of Newcastle-upon-Tyne.

INTRODUCTION

As late as December 1991 it was assumed that the enumeration had gone well (Thompson, 1991):

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. . .

In fact, censuses are never complete counts of the population. Residents can be missed for many reasons. For instance they may be away at the time of the census; they may 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 also be unwilling to respond, perhaps because they feel that their 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 census terminology, the result is a net undercount, or undercoverage, or underenumeration. These terms are exactly equivalent to the term non-response as used more commonly in social survey methodology, which we also use in this article.

In 1991 in Britain the Census Validation Survey, which was designed to find a sample of those missed from the census output, found a level of non-response which would only account for one-fifth of those missing (OPCS/GRO(S), 1994). 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, or over 2 per cent 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 years 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 non-response 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, and this is also discussed below.

But the poll tax is certainly not the only culprit in 1991. In enumerated households in the 1991 census, there were 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 most of these visitors were people who, in 1991, would be seen as homeless single adults or families, at an address on census night which they did not feel was a permanent address (Diamond, 1994).

Equally, institutions – whether medical, care, prison, hostels 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 of these 'visitors' 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 and to expect that a growing proportion of people who were enumerated would not be classified as resident anywhere. Beyond conjecture, what do we know about the response to the 1991 census and the reliability of its results? The rest of this article 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, simple estimations of housing need, and on the allocation of parliamentary seats, if non-response is not taken into account.

As yet the debate is young: government departments currently appear to feel that the non-response is important, but small enough for most users and themselves 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? The answer, for the census at least, is when it is 'imputed'. The census officers 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 their household. If car ownership is missing it is copied from another record which is otherwise similar to the incomplete record for characteristics known to be related to car ownership, including for example 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 per cent (869,000) 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 Dale (1993), 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) and Cole (1993). Across the whole of inner London, 8 per cent 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, 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. The decision to introduce extensive imputation into the 1991 census was imaginative and commendable. However the result is that 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 – those people not included in census output at all.

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

As already mentioned, 80 per cent of those missed by the census were also missed by the Census Validation Survey aimed at finding them. How is this known? Because comparison of the 1991 census adjusted by its validation 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 large surplus of females over males aged in their 20s but not for those in their 40s. These arguments were set down in an article published early in 1993 by the Office of Population Censuses and Surveys, who are responsible for the census, its validation, and the official population statistics for England and Wales (OPCS, 1993d). Since then, a further 145,000 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 has finally been put at 1.209 million, 2.2 per cent of the whole population (OPCS/GRO(S), 1994).

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 and at least until the next census in 2001. While we would accept the arguments that led to the rolled-forward estimates of national population being used in place of the census, the results can clearly not be exact, either overall or for each age group concerned. Adjustments to a count made over 13 years ago are no substitute for a more carefully taken and more accurately validated census.

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

The option of rejecting the 1991 census in favour of 1981-based population statistics is not available for local areas – 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 1991 census estimates of detailed local socio-economic characteristics.

What then can be used for sub-national population estimates, for each of the local authority districts for example? One cannot use the census data alone, because, most simply, there would then 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, 1992), distributing the national non-response evenly throughout the country. In other words they assumed that the missing proportion of 20-year-old men (for example) was the same in Richmond upon Thames as in Liverpool and as in Chester.

This essentially conservative assumption was dropped in the final estimates published by OPCS in the summer of 1993. 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 pointed consistently to much higher non-response in urban areas and particularly in inner London. This evidence included drops in the electoral roll, direct measures for the numbers of infants and young children missed which can be assessed from examining statistics of births in the years immediately before the census (Simpson, 1993a), and the well established experience of social research and other censuses for many years (Marsh, 1993).

The assumptions used in the final mid-1991 estimates might also be labelled conservative because they were based simply on different assumptions of constancy; but they gave a very different picture of the response to the 1991 census. Given that the predominance of young adult men in the national non-response is now accepted, the method identified those areas where there seemed to be a particular shortfall of men and assessed these as most affected by non-response. 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 in 1971 and 1981. The method is described briefly in OPCS (1993e) and precisely in OPCS (1993f). The results are summarised in Figure 1, drawn from OPCS/GRO(S) (1994) which also provides age–sex non-response rates for four categories of Scottish Districts. Census output for highly urban areas is assessed as missing between 20 per cent and 25 per cent of men in their twenties. These areas include the 14 Inner London boroughs, the 9 metropolitan districts, and the 27 city districts in non-metropolitan areas. This is in addition to the missing who were imputed within census output as described above.

This approach to measuring non-response does have considerable merit over the provisional assumptions of an even distribution of underenumeration. But even if it were the best possible approach, it involves making assumptions and thus approximating to an unknown degree; users should be very careful that their interpretations of census and population statistics are not affected. The precise assumptions are dealt with in more detail in Simpson (1993a); in this article we

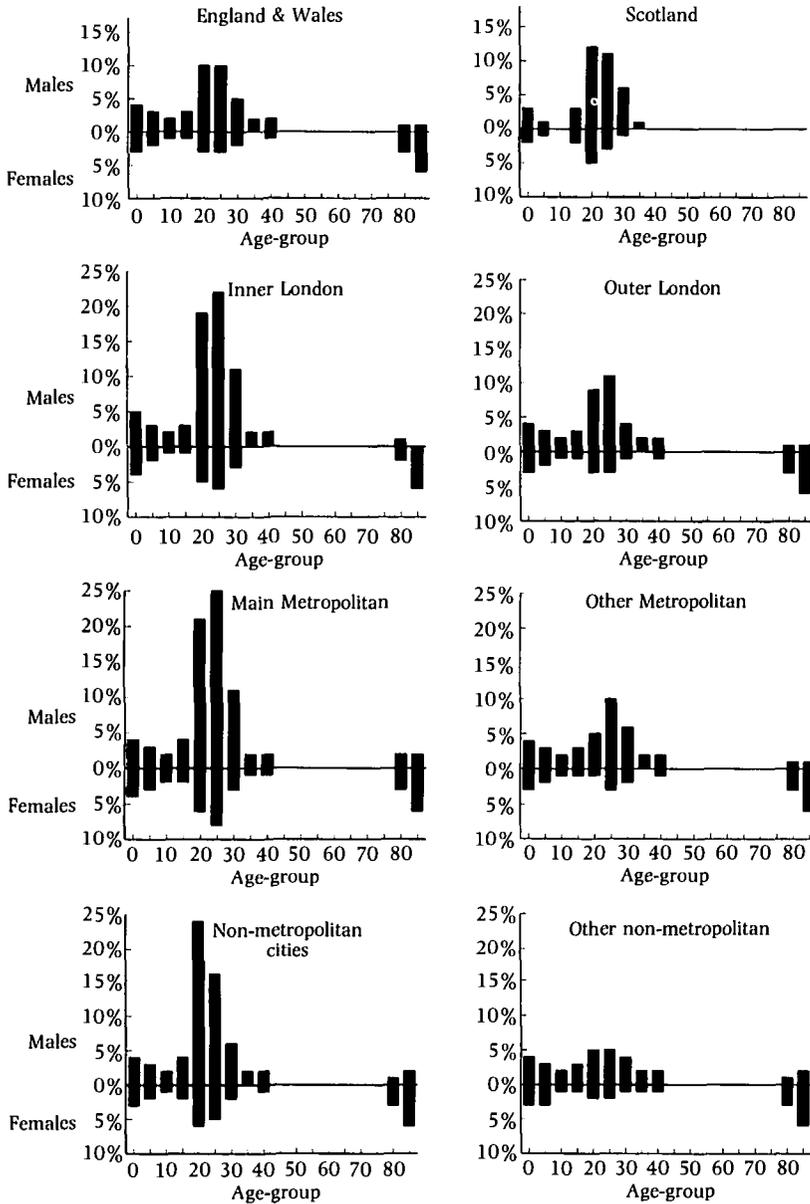


Figure 1. Estimated census non-response by sex and age, England and Wales, Scotland, and six types of district in England and Wales. (Percentage by which the 1991 population estimate (using census definition) exceeds the census usual resident count, for 5-year age-groups beginning with the age shown, and ending with those aged 85 and older.)
 Source: OPCS/GRO(S), 1994.

are concentrating on the implications of a significant degree of unreliability rather than its source.

We may be able to learn a little more of the characteristics of some of the missing 1.2 million from the Census Validation Survey report. At the time of writing, this report is unpublished although it is well over two years since the fieldwork was carried out. But as we have seen, that survey was not a success. It is likely that the most difficult problem in using 1991 census data will remain estimation of the degree of non-response in what is now in effect a biased sample.

Although attention has been paid to non-response of persons as described, and some work is being done on the basic count of households, the overall implications of the resulting estimates of non-response, and the possible other aspects of non-response, have not yet been thoroughly addressed. We have little idea of the bias due to non-response in even simple counts such as employment status or distance migrated. The validation survey which accounted for a fifth of those missing cannot be assumed to have found a representative sample of those people who were not enumerated. 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 need as measured by the census is used to determine variation in resources given, for example to local government in the Standard Spending Assessment. This leads to fining ('capping') those councils which attempt to spend more money than they are assessed as needing to spend.

For several of the census measures used in these calculations, urban areas tend to be at the high end of the scale of need. Since non-response was concentrated in urban areas, their need is likely to be underestimated in the census counts. The best adjustment one can make at present is to apply the age-sex-geography-specific non-response rates provided by the census offices (and displayed in Figure 1), as weights in a census analysis. One can do this whenever the characteristics of need are disaggregated by age or sex or area, as most 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 1991 census) of unemployment, non-white ethnic groups, and chronic limiting long-term illness in households. Because unemployment and non-white ethnic groups are both characteristics of the population that are concentrated in urban areas and to a greater extent among young male adults, this demographic weighting shows them to have been missed by twice the national average non-response of 2.2 per cent. Illness in households, on the other hand, which is concentrated among older age-groups, is missed to a lesser extent. It must be remembered that Table 1 assumes that each group (white and non-white for example) is equally affected by the age-sex-area specific non-response rates of Figure 1, which themselves are estimates with unknown reliability. Nevertheless this kind of adjustment represents the best we can do now.

Table 2 shows the mean impact of not adjusting for census non-response, on the distribution of a grant to local authority districts on the basis of their overall need as measured by (a) the number unemployed, and (b) the number of non-white residents. Main metropolitan districts would lose 3.5 per cent 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 individual districts would in some cases be greater than those illustrated, and less in other cases. The impact would be greater for some variables (non-white youth for example), and less for others.

TABLE 1. *The impact of adjusting for census non-response on national estimates of unemployment, non-white ethnic group, and illness within the household population, England and Wales*

| | Census count (000s) | | Estimated non-response (2)/(1) % |
|--------------------------------------|---------------------|--------------|----------------------------------|
| | Unadjusted (1) | Adjusted (2) | |
| Unemployment | 2.235 | 2.325 | +4.0 |
| Non-white ethnic group | 2.952 | 3.097 | +4.9 |
| <i>Illness, household population</i> | 6.039 | 6.110 | +1.2 |
| All residents | 49,890 | 51,000 | +2.2 |

Sources: Census Local Base Statistics; non-response rates specific to age, sex and district type, from OPCS/GRO(S) 1994.

TABLE 2. *The impact of not adjusting for census non-response 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% |

At present the Department of the Environment uses the unadjusted census counts to determine the *per capita* expenditure of local councils. Percentages are affected by census non-response, to an extent similar to that shown in Table 2.

Tables 1 and 2 show the impact of not adjusting census counts purely on the basis of the officially estimated sex-age-area response rates. The implication is that such adjustments should be made. The unreliability of these estimated response rates, however, implies that 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, even within these age-sex-area sub-groups of the population, then the impact of non-response for the provision of services for these people 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, 1993a). Does this mean that not 20 per cent but 40 per cent of non-white young men were missed? It could be so. Unfortunately, national evidence is hard to come by. In terms of social statistics there is little to fall back on when our national census *and* its validation survey miss a million people.

THE IMPORTANCE OF PRECISION – MORTALITY RATES

A good example of the need for precise population counts are mortality statistics. Standardised mortality rates are the main measure for

charting the changing state of the health of nations internationally. Standardised mortality statistics rely upon an accurate age and sex breakdown of the population upon which to standardise. 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 2). However, for some specific groups the rate has risen slightly when the overall period 1981–5 is compared with 1986–9, in particular for men aged 13–18, 21–24, and for women aged 13–19 and 23. Further analysis shows that a significant component of this rise can be attributed to causes such as 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? Statistics of deaths are thought to be quite accurate (they have been registered along with births 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 3 shows how the official versions of these imply that the population has changed between 1981 and 1991. The cohorts

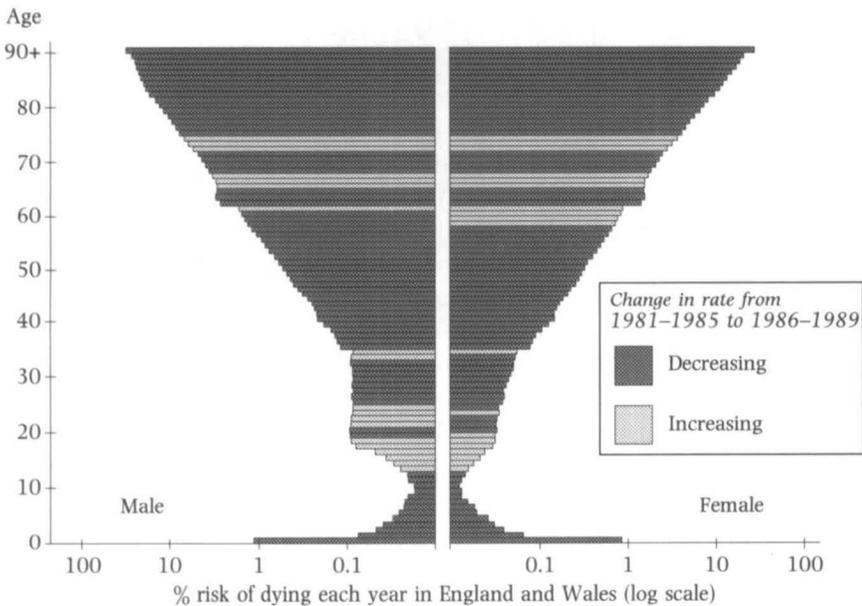


Figure 2. Mortality rate calculated from deaths between 1981 and 1989
 Source: OPCS final mid-year population estimates for 1981 and 1991 and total mortality by age and sex, 1981 to 1985 and 1986 to 1989.

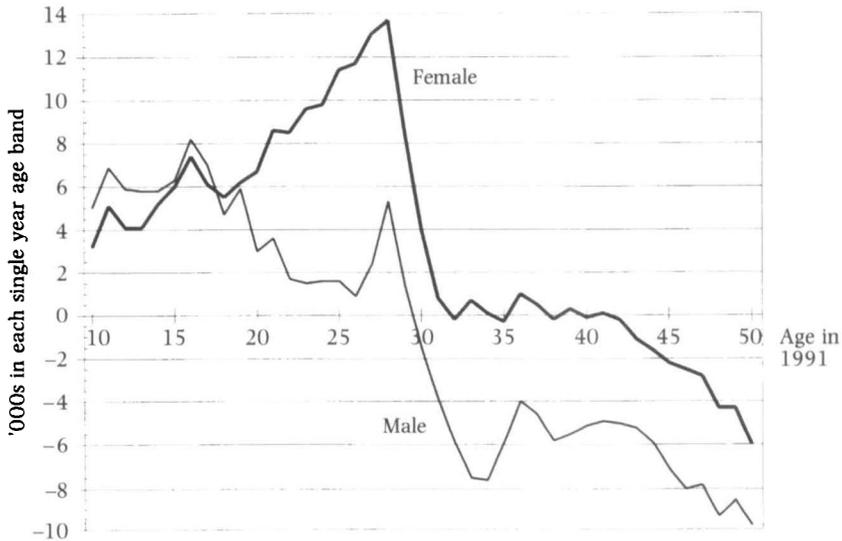


Figure 3. Surplus of net immigration over deaths, England and Wales 1981-1991.

Source: OPCS single age mid-year estimates, 1981, 1989.

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 estimation error in comparing 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 estimate of the number of settled refugees (lack of information about which caused errors amounting to 145,000 people in the final 1991 population estimates before they were revised – OPCS, 1993a).

Our knowledge of whether there has been an increase in teenagers' 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 population estimates of the 1990s are not, in essence, based on the 1991 census. As has been described above, national population estimates still rely on the 1981 census and the

International Passenger Survey since then (plus registration of births and deaths). The International Passenger Survey was never designed for such a role. So, even where an official attempt has been made to estimate the non-response of people, by age and sex, considerable uncertainty remains.

The 1981 Post Enumeration Survey accounted for 90 per cent of the people missed by that census (OPCS, 1985, p. 4) and from it a 95 per cent confidence interval was calculated to contain the non-response in 1981 which ranged from 152,000 to 278,000 people (30 per cent above and below the actual figure – leading to a 95 per cent confidence interval of ± 0.1 per cent in the final mid-year estimate). In the absence of official figures for 1991 confidence intervals we can assume at least this same degree of uncertainty for the estimate of non-response. This is charitable, as the validation survey of 1991 only found 20 per cent of the people presumed to be missing; 30 per cent of 1,209,000 not enumerated in 1991 gives a 95 per cent confidence interval of 363,000 people either way on the revised final 1991 mid-year estimate – an overall rough 95 per cent confidence interval of ± 0.7 per cent on the entire population of England

TABLE 3. *Mortality Rates for young men in 1991, England and Wales*

| Males aged | Estimated mid-1991 population '000s | April 1991 Census Enumerated '000s | Size of** non-response percent | Recorded deaths in 1991 | 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 |

* Confidence limits are presented for the mortality rate assuming that the estimates of non-response may be in error by $\pm 30\%$.

** Non-response is the census subtracted from the mid-1991 population. An approximation is caused by the slight inconsistencies in timing and definition of the two sources, but this does not affect the implications drawn.

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

and Wales. Better estimates may be possible following the publication of the 1991 validation survey. The estimates of uncertainty shown here are likely to be conservative. This uncertainty is higher for certain groups, in particular for young men, as Table 3 above illustrates. The table gives a range of mortality rates making the same assumptions as were outlined above.

Figure 4 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 with which we can use them. Variation estimates are more usually assigned to the numerator, and most statistical tests applied to look for significant change in these rates assume the count of the underlying population to be correct. Until we have some better estimate for the reliability of the mid-year

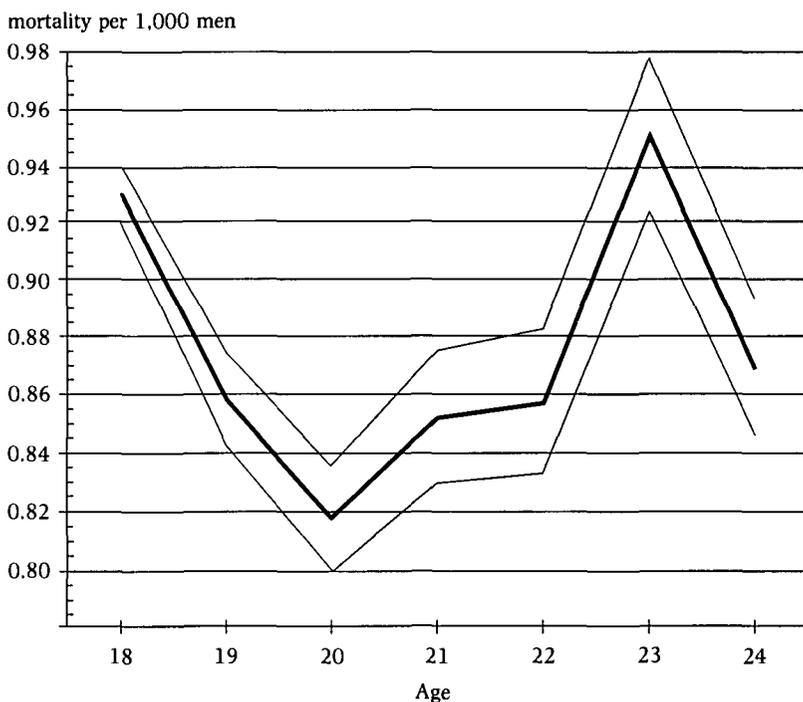


Figure 4. Variance in the mortality rate of men aged 20–24 in England and Wales, 1991 (showing the extent of a 95% confidence level).

Sources: OPCS final mid-year population estimates and total mortality rates for males and the 1981 Census Post-Enumeration Survey Confidence limits rolled-forward, OPCS, 1985.

estimates we will not be able to answer confidently a basic question such as 'are men more likely to die at age 18 or at age 23?' with any degree of certainty. More subtle measures – over time, in geographical sub-divisions of the country, or for particular causes of death will incorporate even larger degrees of uncertainty.

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 underenumeration 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 whole households, rather than individuals, who choose to be missed to evade taxation or for other reasons. It should also be noted that an empty dwelling or a single person household is liable to pay only 50 per cent or 75 per cent 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 0 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 and in each local area. Hence an undercount of just 0.5 per cent in our estimate of household numbers can conceal a possible additional 'need' for 100,000

dwellings. By the late 1980s, official estimates of household numbers became less robust as they had to rely 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 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, one 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 (DoE, 1993); 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 response 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 per cent is not unreasonable when compared to the 2.2 per cent non-response for the population as a whole. Geographically, the county suggested as having the highest rate of missed Local Authority households was London (8.73 per cent), whereas the English region with the lowest rate was East Anglia (2.45 per cent – see Figure 5).

There is corroborating evidence of this non-response 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 200,000 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 6), which

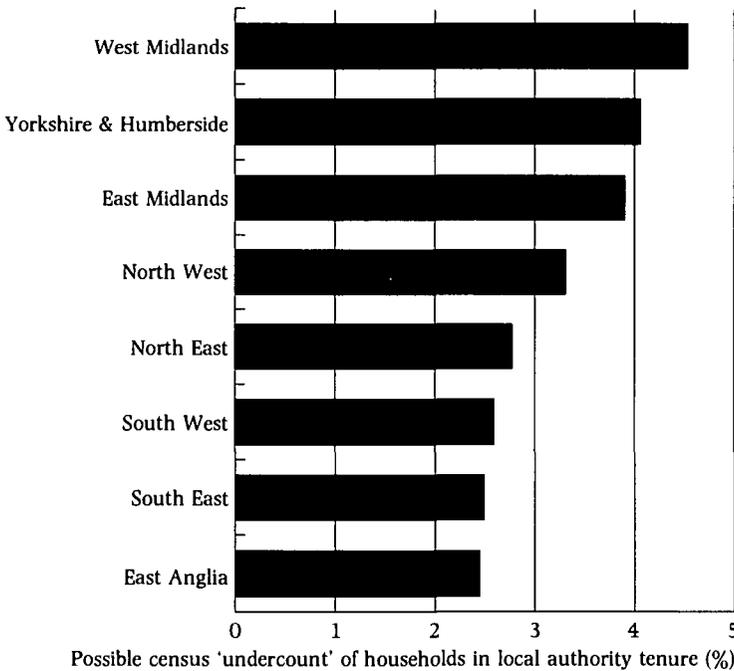


Figure 5. Discrepancy between regional trends and census figures
 Sources: Regional Trends figures for April 1991 by Local Authority (published by OPCS in 1993) and census figures for April 1991 (published by OPCS, 1993)

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 missed households, 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, Table 2.23), 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 7 shows how the estimates

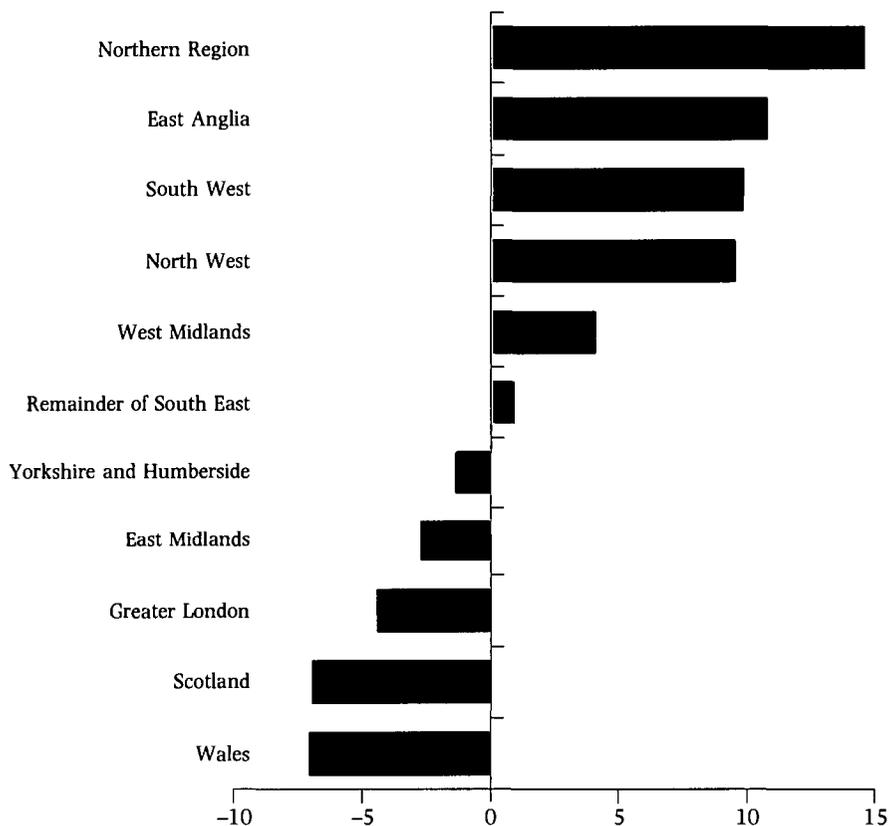


Figure 6. Discrepancy between mortgagee and census figures for home buyers.
Sources: OPCS census figures published 1993 and a parliamentary answer by Stephen Dorell, 1993.

of dwellings were retrospectively 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 4), although no official statement has yet been made.

Table 4 illustrates how even a small estimate (1.8 per cent) of the national non-response of households can translate into a 71 per cent overcount of vacant dwellings and thus have a dramatic effect on national housing debates. The English House Condition Survey figures (DoE, 1993) were not however presented in a succinct form, but have had to be extracted from where they are scattered throughout that

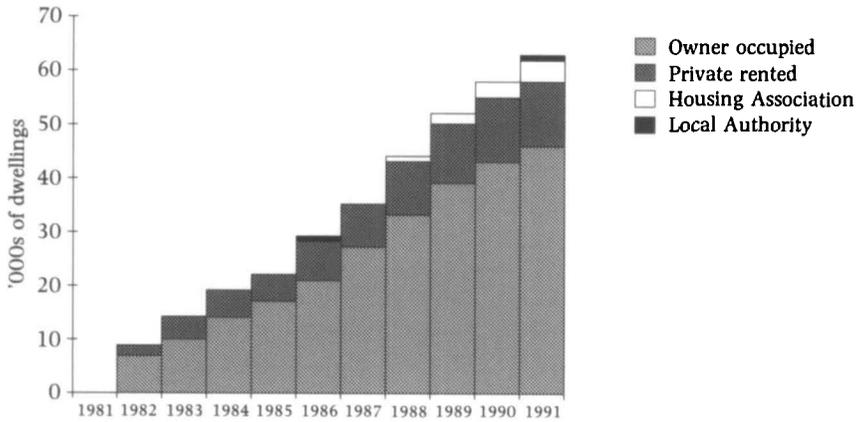


Figure 7. Retrospective additions to the estimate of dwellings in England made in a revision between 1st and 2nd quarters of 1992
 Source: DoE, 1992

TABLE 4. Possible extent of the non-response of whole households using the English House Condition Survey

| England ('000s) | EHCS | Census | Difference from the census per cent |
|------------------|--------|--------|-------------------------------------|
| 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)

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 non-response in the census that falls within the range suggested by the independent comparisons made here.

OPCS, however, do not as yet appear to agree with these estimates. They have previously stated that households were not missed in any

significant numbers. Their latest statement on the subject to date is that:

The CVS has provisionally estimated that in Great Britain some 0.8 per cent of household spaces (187,000), and 0.6 per cent of households (125,000), were not covered by the Census. ... The CVS is the best available source of information on coverage of households. It should be noted that a small number of household spaces and households may have been missed by both the Census and CVS (OPCS/GRO(S), 1994, p. 5).

We feel, in the light of the above discussion, that OPCS still have some way to go in producing an estimate of the household non-response which is justifiable. Some defence is needed of the claim that the Census Validation Survey is the best (and implicitly only) available source of information and that it will have only missed a 'small number of households'. Our analysis agrees independently with the DoE estimates which imply that roughly two times as many missing households were also missed by the Census Validation Survey as were found by it.

To put this argument into context, the number of missed households we think were living in England in 1991, extra even to the revised figure from OPCS, is eleven times greater than the number of Local Authority, New Town, Government Department, and Housing Association dwellings on which construction began in 1991 (Wilcox, 1993, p. 56). This is a very significant discrepancy. OPCS and the DoE must come to an agreement on how many households were missed. That agreement will influence the household need figure projection made well into the next century.

NO REPRESENTATION WITHOUT TAXATION

We have suggested that census non-response has social as well as technical causes. These have also affected the completeness of electoral registration in Britain in a way that will have a marked political effect for years to come. We have already mentioned how the poll tax of 1381 contributed to a drop of one-third in the enumerated population of England. Iain McLean and Jeremy Smith (1994) bring us up to date by 600 years 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 per cent 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 return to the electoral roll, 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 the 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)

DISCUSSION

An incomplete response to the national population census has profound effects across the spectrum of social statistics and for social policy and the social sciences in Britain in general. 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 with confidence 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.

We have given details of four examples above of how uncertainty about the census figures has significant impact on social policy. It is important to identify other applications of census data where users must take particular care because of the non-response. We think these include:

Studies of migration. It is now apparent that the standard methods of monitoring migration within Britain under-estimate migration to

cities, particularly of young people. This bias affects not only the census but patient reregistrations which are used between census years to track the level of migration. A realistic view of migration is essential for many applications, not least in projections of the future growth of Britain's regions.

Survey sampling. Many surveys use the census to establish quotas of interviewees within each social stratum, or to establish the size of the final samples in multi-stage sampling. To do so from raw census figures would perpetuate the exclusion of those missed in the census in future surveys of social conditions. Future opinion polls, for instance, would be biased on this account (although not, ironically, polls of voting intent – where it will be important to exclude people not registered).

Evaluation of change between censuses. The knowledge of how conditions have improved or worsened between censuses is vitally important to the evaluation, and the formulation, of government and other policies. Such change is often finely balanced – small uncertainties induced by poor quality data can affect interpretations. It seems that for the overall count of residents in most census output, the addition of imputed absent households in 1991 had the effect of bringing coverage up to that of the 1981 census (which deliberately excluded people living in wholly absent houses). However, this fortuitous result does not hold for sub-national areas nor for each age-group, making assessment of change very difficult (Rees, 1993; Simpson, 1993a).

Small area statistics. like most social characteristics census non-response exhibits greater heterogeneity between smaller geographical areas than between larger ones. Thus the proportion of census output that is imputed (as discussed earlier in this article) ranges from practically none in most census enumeration districts to over a quarter in many hundreds of enumeration districts, and over a half in some. The true characteristics of the latter areas are clearly rather uncertain. Add to this the unknown local geography of those whom the census missed altogether, and it is rather difficult to have confidence in small differences between the conditions of local communities, and even harder to tell which communities improved and which worsened during the 1980s.

These additional examples demonstrate that the onus should be on the researcher who chooses to ignore census non-response, to explain why s/he is doing that.

How can research cope with this uncertain census data?

In the meantime what can those who use these statistics do? Firstly, advice from the census offices to adjust census figures for non-response (OPCS/GRO(S), 1994) should be taken seriously. Many government and research applications could without difficulty weight census data by the age-sex-area non-response rates already estimated, and by other information that may become available. Weighting for non-response is a common and unquestioned practice in survey analysis.

Beyond this, 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 accurate 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.

Lastly it must be remembered that not only people, but also other social 'units of analysis' are missing from the census tables – households and also families (on which research of this kind has yet to start). Nobody has a firm idea about how many households were missed and so the census, which was supposed to produce figures which could be agreed by everybody as the basis for a debate about housing needs, has become the focus of debate itself with separate arms of the British government producing conflicting accounts. What this all points to is the desperate need for a better census next time round.

Future census-taking

There is no reason why census non-response should suddenly go away and the reliability of our statistics rise magically overnight. The reticence of many people to comply with official requests for personal information is unlikely to be replaced by a new-found sense of common purpose, unless appearing in official lists becomes economically or socially worthwhile. If the state wants people to register it has to be worth people's while to do so. Threats of fines, increased numbers of enumeration officers, or the merging of huge social security computer files will not find people who do not wish to be found.

To some extent the non-response may also have been due to out-moded concepts of household and family structure in contemporary Britain, a 1950s image of each family cosily inhabiting a separate

household, waiting for the enumerator to call. Evidence of such concepts peppers the census tables in which cohabitation has only been fully recognised in 1991, where gay and lesbian couples have been edited out of existence in the technical processing of census forms and where attempts have been made still to identify a single 'head' of each household. Improved questions would recognise the existence of various degrees of homelessness and non-permanent residence at an address. Better preparation would include the more accurate mapping of the areas to be enumerated. Improved fieldwork procedures would send reinforcements to areas where enumerators reported difficulties. Ultimately, improved post-census validation checks are vital to provide not only estimates of those missed but confidence intervals for those estimates, and a good understanding of the reasons why people were missed or chose to be missed.

Census data require accuracy if they are to provide more than blurred images of our society and unfair resource allocations. Those missed, as well as those counted, deserve well informed policy-making.

REFERENCES

- A. G. Champion (1993), 'Population distribution and change in Great Britain, 1981-1991', *Geography Review*, 7:1, 10-15.
- K. Cole (1993), 'The 1991 local base and small area statistics', in A. Dale and C. Marsh (eds.), *The 1991 Census User's Guide*, HMSO, London.
- I. E. Corner (1991), *Modelling Housing Demand*, paper presented to the Statistics Users' Council's Annual Conference at the Royal Society on 26 November 1991.
- A. Dale (1993), 'Fieldwork and data processing', in A. Dale and C. Marsh (eds.), *The 1991 Census User's Guide*, HMSO, London.
- I. Diamond (1994), 'Where and who are the missing million? Measuring Census of Population undercount', in Statistics Users Council, *Regional and Local Statistics*, IMAC Research Ltd, Esher.
- 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, Directorate of Statistics, Government Statistical Service, London.
- DoE (1993), *English House Condition Survey: 1991*, HMSO, London.
- S. Dorell (1993), Written parliamentary answer on MIRAS, Hansard col. 239, Proceedings of the House of Commons, 17 March 1993.
- D. Dorling (1991), *The Demand for Housing in Britain 1971-2001*, NorthEast Regional Research Laboratory Research Report 91/8, University of Newcastle upon Tyne.
- D. Dorling (1993), The 1991 Census and Housing Tenure, BURISA Newsletter, May 1993, 5-7.
- D. King (1990), *Housing Needs in the 1990's, Challenges, Choices and Change*, Institute of Housing Conference, 6-8 April.
- M. Kleinman and C. Whitehead (1989), *Demand for New Housebuilding, 1986-2001*, Development and Planning, Department of Land Economy, University of Cambridge.
- I. McLean, and J. Smith (1994), 'The poll tax, the electoral register, and the 1991 census', in A. Heath, R. Jowell and J. Curtice (eds.), *Can Labour Win?* (forthcoming), Dartmouth Press, Aldershot.

- C. Marsh (1993), 'The validation of census data – general issues', in A. Dale and C. Marsh (eds.), *The 1991 Census User's Guide*, HMSO, London.
- I. Mills and A. Teague (1991), 'Editing and imputing data for the 1991 census', *Population Trends*, 64: 30–37.
- P. Niner (1989), *Housing Needs in the 1990s*, National Housing Forum, National Federation of Housing Associations, London.
- C. Oman (1906), *The Great Revolt of 1381*, Clarendon Press, Oxford, reprinted by Greenwood, 1969.
- OPCS (1985), *1981 Census Post-Enumeration Survey*, HMSO, London.
- OPCS (1992), *Provisional Mid-1991 Population Estimates for England and Wales*, Monitor PP1 92/1, HMSO, London.
- OPCS (1993a), *Revised Final Mid-1991 Population Estimates for England and Wales*, Monitor PP1 93/2, HMSO, London.
- OPCS (1993b), *Mortality Statistics, General*, series DH1 no. 26, HMSO, London.
- OPCS (1993c), *1991 Census Sex, Age and Marital Status Tables*, HMSO, London.
- OPCS (1993d), 'How complete was the 1991 census?' *Population Trends*, 71: 22–5.
- OPCS (1993e), 'Rebasing the annual population estimates' *Population Trends*, 73: 27–31.
- OPCS (1993f), *Rebasing the 1991 Population Estimates – the Use of Sex Ratios*, OPCS Population Estimates Unit, London.
- OPCS/GRO(S) (1994), *Undercoverage in the 1991 Census*, Census User Guide 58, HMSO, London.
- P. Rees (1993), *Using Census Data to Estimate Small Area Demographic Change*, paper at the Institute of British Geographers conference, 6 January.
- A. Sandhu (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*, Scottish Office Library, Edinburgh.
- S. Simpson (1993a), *Measuring and Coping with Local Under-enumeration in the 1991 Census*, paper to conference 'Research On The 1991 Census', Newcastle University.
- S. Simpson (1993b), 'Local analysis of the census undercount', *BURISA*, 108: 11–14.
- A. Thompson (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, Esher.
- C. Whitehead (1993), *Memorandum to the Select Committee on Environment Inquiry into the Housing Corporation*, personal communications.
- C. Whitehead and M. Kleinman (1992), *A Review of Housing Needs Assessment*, LSE Housing, The Housing Corporation, London.
- S. Wilcox (1990), *The Need for Social Rented Housing in England in the 1990s*, Institute of Housing, London.
- S. Wilcox (1993), *Housing Finance Review 1993*, Joseph Rowntree Foundation, York.