

How many people live on this island?

Ludi Simpson

The 2001 Census results have not hit researchers' desks. Instead they are floating just above them as it were, available not just to researchers but publicly on the government's Neighbourhood Statistics website (<http://www.neighbourhood.statistics.gov.uk/>). As I write, only a minority of the data have been released. This includes 250 counts for each of the 100,000+ output areas in the UK (areas of just 100-250 households), and several thousand counts for each local authority District. I have already seen analyses that show housing pressure on inner cities, the inverse care law (more professional care where there is least need for it), racial legends demystified, and growing of inequality during the 1990s.

This note takes a step back from that more interesting work to review the debate on the completeness of the 2001 Census and is loosely based on a contribution to the Radical Statistics conference in February 2003. It looks at the evidence for the official population figure for England and Wales, and the uncertainty about it. I apologise for not including Scotland and Northern Ireland in most of this – one result of devolution is the difficulty in putting together UK-wide statistics.

To jump to the conclusion, a government reaction to the difficulty of measuring population size could be the use of a population register. As part of thinking about balance of this with civil libertarian and privacy concerns, one should question whether a population register would be any less biased than other approaches to estimating population size.

One Number Census – summary

In recognition that census undercount adversely affects its quality, government put into place the largest ever household survey – more than 300,000 households across the UK – to estimate the extent and nature of census undercount. On the basis of these estimates for each part of the UK, records were duplicated to represent the undercount. The 2001 Census output on the website includes these imputed records so that it represents the full population including those who were not enumerated. These One Number Census (ONC) procedures were discussed with census users, who agreed that a less biased database would be a good thing. Most of the methods are well documented on the ONC census website (<http://www.statistics.gov.uk/census2001/>)

IntroOneNumber.asp). An aside: there are many other issues of census data quality that could be considered – the adequacy of imputation both of missing items on returned forms and the ONC imputed whole records, the impact of suppressing all counts of 1 and 2 in England and Wales by rounding them to 0 or 3, the adequacy of the questions, particularly those that were asked newly or differently in 2001. But this note is focused on one issue: can we know whether the adjustments were big enough?

Worry 1: previous estimates didn't match the census

The traditional first check of the census results didn't offer comfort. The One Number Census population was 1.2 million less than the figure estimated before the Census:

England and Wales

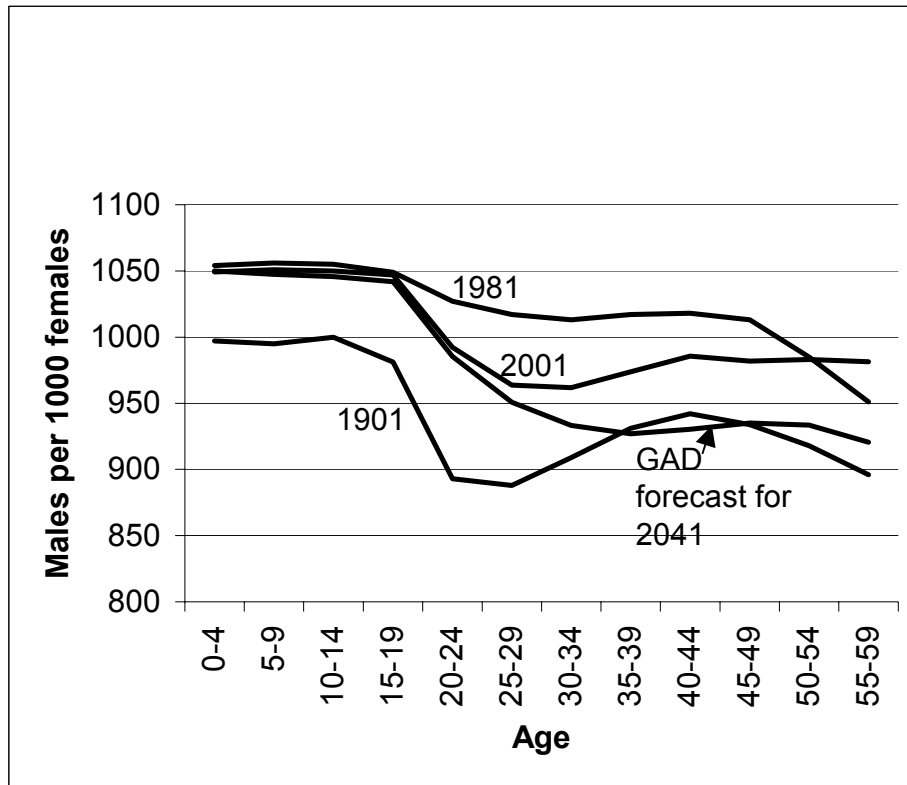
48,843,000	Population enumerated, Census day April 29 2001
+3,199,000	Non-response estimated by One Number Census procedures
=52,042,000	Total One Number Census population, April 29 2001
+46,000	Estimated population change between Census day and mid 2001
=52,084,000	Census-based mid-2001 population
53,174,000	Pre-census population for mid 2001 (GAD projections from 2000)
1,090,000	Excess of pre-census population estimate over census-based population estimate.

Worry 2: Population and undercount sex ratios have radically changed

Figure 1 shows a pronounced dip in the male/female ratio at young adult ages in the 2001 population, not observed in 1981 or other population estimates in the past 50 years. An even bigger dip was evident a century ago; it was due to young men servicing the British empire overseas before returning home. The dip in 2001 is interpreted as young men (more so than women) seeking work outside Britain in the 1980s and 1990s. The latest forecast population assumes that those young men won't return, causing the lower sex ratio to continue into later ages.

This is fine if it's true. The worry is that some or all of the missing young men may not have emigrated but be present but uncaptured by the ONC, despite the hugeness of the operation of census-and-survey. We are left looking for more evidence to check against the census-based estimates.

Figure 1 **Historical, current and prospective sex ratios, England and Wales**



Source: Britton and Edison (1986), GAD (2003)

Table 1 shows that the undercount in the 2001 Census – as measured by the ONC procedures and included in published census figures – was focused where one would expect, in London and cities.

Table 1 Composition of non-response and population

<i>Type of District:</i>	Undercount rate, ONC2001, all persons	Composition of undercount: ratio of undercount rates, ONC2001		Change in population sex ratio
		Male 20-34/ Female 20-34	All 20-34/ All other	20-34 M/F, ONC2001-MYE2000
Inner London	22%	1.11	1.41	-.126
Outer London	10%	1.15	1.95	-.146
Principal cities	9%	1.30	1.64	-.131
Large cities	7%	1.30	1.90	-.146
Small cities	6%	1.36	1.91	-.125
Resort, port and retirement	5%	1.25	2.21	-.073
Other metropolitan Districts	5%	1.25	1.75	-.091
New towns	5%	1.35	1.76	-.063
Industrial areas	4%	1.30	1.85	-.074
Urban and mixed urban-rural	4%	1.35	2.30	-.032
Remoter, mainly rural	4%	1.42	2.13	-.028
England and Wales	6%	1.22	2.06	-0.086

Philip Redfern has pointed out (2003) that the composition of the undercount is less dominated by men in the areas of greatest undercount. The table shows this and that the undercount is also less dominated by young people in the areas of greatest undercount. He suggest that this indicates a bias against young men in the estimation of undercount. That’s not very convincing. This same pattern of undercount could also simply mean that a wider range of people are missed where the census fieldwork has failed most severely; in that case the follow-up survey may have estimated this correctly.

More worrying is the final column in Table 1, also expanding on Philip Redfern’s ideas. It shows that the population sex ratio has tended to drop from the previous population estimates to the new ONC-based population most in the areas of greatest undercount. This would plausibly be due to a bias against young men in estimation of the undercount. But it could also be that men have migrated most from those areas. These would be people who have migrated but not been included by the estimates of migration that went into the pre-census population figures. Surely that extent of missed migration – a net 1.1 million out of England and Wales and mainly young men - will have been noticed in some other way?

Worry 3: international migration – not much evidence, a lot of conjecture

Unfortunately, emigration is the hardest thing to count. When someone leaves the UK, there is not a record in another country of an entry. ONS have revised their international migration estimates in an attempt to meet the implications of the One Number Census. But the revisions are a mass of assumptions, some more plausible than others, some very questionable, but all are assumptions rather than independent evidence. This isn't the place to go into it in detail, but those who are interested should refer to ONS (2003). Official population estimates and projections and methods, and methods for future estimation of international emigration have all been revised to incorporate assumptions that will produce more emigration than is directly measured, with no evidence to support this except an acceptance of the 2001 Census results.

It is possible that some of the 1.1m gap between the old and the new population series is due to inconsistent population definitions. Some people have a residence in more than one country and it is not clear where they should be counted. I personally don't believe that definitional differences could explain much of the gap. Either way, it is more conjecture.

Worry 4: lack of validation through administrative counts.

As part of the ONC quality assurance plans 14,000 diagnostic ranges, for several age-sex groups within each local authority District, were constructed from administrative records. These counts of births, patients, child benefit, pensions, and previous population estimates provided ranges that were intended to be wide enough to highlight possible problems with the ONC estimate should it fall outside them. ONS had already adjusted the census-based population estimates where independent counts of armed forces or students indicated they should do so; and where response rates were based on small numbers from the coverage survey, adjacent age-groups were often combined. But still the final ONC population estimate falls outside those diagnostic ranges in 20% of the 14,000 tests. ONS judged that this meant the administrative counts were not up to the job they were asked to do, not that the ONC had failed in any of these cases. It may be correct that none of these administrative counts are after all good enough singly or together to validate a population estimate; but one more avenue of quality assurance is absent by fiat.

Worry 5: likely dependence of census and coverage survey

The previous worries can be expressed as one: that those missed by the census were more likely to be missed by the follow-up coverage survey: that the success of each is not independent as the ONC methodology assumed. The ONC preparation had demonstrated that a certain level of dependence would not overly affect the results. An undiscussed late addition to the plans did measure some dependence, showing that both the census and the coverage survey missed more households than expected, resulting in a rise of 230,000 in the ONC population total (included in the 52,084,000). However, isn't it clear that there will be dependence also for people missed from counted households? ONS have admitted as much by saying that those who don't want to be counted could not be estimated. Could there be a very considerable number not wanting to be counted? Those working in the informal economy, involved in criminal activity, or resident in households that for tax and benefit purposes do not include them?

It is not unreasonable to think that half a million young men are missing from the population estimate, and perhaps others too. Mortality rates calculated on the revised population show an improvement for every single age since 1991, except 25-34 where mortality has risen. This could be an artefact of too low population denominators. It could be that the 1991 figure was too high. Again: more questions; and no ONS answers except to have faith in the ONC census methodology.

This is not a difficult problem to formulate and discuss, but it is a difficult one to resolve. When census undercount reaches over 3 million, then the impact of those avoiding response or otherwise extra difficult to count may have become very significant.

What are the solutions? Better organisation of fieldwork would help (see RadStats 78). A government that was respected and an economy was strong enough to discourage illegality at the margins would also help. Some would see those things as impossible, but the increase in survey refusal rates suggests that the degree of co-operation is changeable (government surveys that got 80-90% response in the 1980s now get 70-80%). It would help to know a lot more about the non-response ethnographic and other studies of what goes on during the census fieldwork. Aside from these approaches, a purely technical solution is likely to be sought in the further combination of different approaches.

The One Number Census already combines door-to-door enumeration with a survey. The next steps are likely to link administrative records for individuals, or piggyback proposals for a population register and identity cards, which of course will also raise civil liberty issues. Some statisticians are presenting the Netherlands and Scandinavian use of population registers as the way forward, but there is little evidence that

they provide any more complete enumerations of the population than a census combined with post-enumeration survey. With a population register or ID card in would be tempting to use of the 'registered population' as the population estimate for official purposes. This would be a large step away from public services according to need, and a divorce of demography from reality.

The Royal Statistical Society and the Office for National Statistics are organising a two-day conference on 11-12th November which will address many of the technical issues and no doubt raise some of the more political ones: 'Beyond the 2001 Census',
http://www.statistics.gov.uk/events/rss_ons_conf/default.asp

References

- Britton, M. and N. Edison (1986). The changing balance of the sexes in England and Wales, 1851-2001. *Population Trends* 46: 22-25.
- GAD (2003) Population projections, 2001-based. Government Actuaries Department: London.
- ONS (2003) Revised international migration estimates 1992-2001. Press release and methodology note, 12 June 2003. ONS: London. Also on ONS website.

Ludi Simpson

*Centre for Census and Survey Research
Faculty of Economic & Social Studies
University of Manchester
Manchester
M13 9PL
ludi.simpson@man.ac.uk*