

SAHSU response to ONS consultation on *The census and future provision of population statistics in England and Wales*

Every ten years, for over 200 years, every household in England and Wales has been required to respond to the Census. The Census provides population statistics that are used by planners, policy makers and researchers, including SAHSU. Improvements in technology and in government data sources offer opportunities to either modernise the existing census process, or to develop an alternative census method that re-uses existing administrative data already held within Government. The Office for National Statistics (ONS) is currently consulting on these options and have narrowed the choices down to two approaches for census-taking in future:

- once a decade, like that conducted in 2011, but primarily online or
- using existing government data and compulsory annual surveys.

No decision has yet been made by ONS and they have undertaken a consultation (<http://www.ons.gov.uk/ons/about-ons/get-involved/consultations/consultations/beyond-2011-consultation/index.html>) from which ONS will publish their findings in 2014. SAHSU is strongly supportive of continuing the once a decade census and has provided a response to the ONS consultation (below).

Q1 What are your views on the different census approaches described in this document (a) An online census once a decade; b) A census using administrative data and surveys)?

The Small Area Health Statistics Unit (SAHSU, see www.sahsu.org) part of the MRC-PHE Centre for Environment and Health based at Imperial College London, makes extensive use of population characteristics and attribute information at the small area level. In most of our studies we use either Census Output Area (COA) data as a proxy for individual characteristics (e.g. socio-economic status as defined by the Carstairs Index which combines four census variables: unemployed male, car ownership, overcrowding and low social class) or Lower Layer Super Output Area data to characterise neighbourhoods.

SAHSU relies on small area data in particular when looking at health effects of environmental exposures that vary greatly over space such as noise or air pollution. Increasing the geographical resolution of these studies (e.g. to Local Authority level) would result in averaging out the exposure contrast (i.e. all Local Authorities would end up with more or less the same level of exposure) and as a consequence the detection of possible health effects or the definition of exposure-response relationships would not be possible.

Another large problem of depending on indirect sources of information (i.e. administrative data as opposed to a Census) is that estimates could be fine overall but very wrong in particular (small) areas, where there may be concerns about environmental pollution, or other factors. If the denominator data are incorrect then the disease risk estimates in those areas will be wrong. The

decennial census provides an extremely valuable “check” with which to correct and re-standardise modelled estimates derived in the inter-censal years.

One of the remits of SAHSU as outlined in SAHSU’s terms of reference is to rapidly assemble information on potential health risks related to environmental exposure. To this end SAHSU has developed the Rapid Inquiry Facility (RIF), software that is now used worldwide in environmental health tracking including the Centre for Disease Control (CDC) in the US. The RIF is based on population information at the small area level, making use of both population characteristics such as age and gender distribution of COAs but also other socio-economic characteristics such as deprivation status and ethnicity. In the UK, the RIF uses census data at the COA level. Without this the small scale analysis necessary to carry out investigations in disease risk around industrial installation would not be possible. Information collected in the decennial Census is therefore of vital importance for SAHSU. The level of detail which could be provided by the use of administrative data and surveys is insufficient for our purposes. For SAHSU, the spatial resolution is of higher importance than the temporal resolution and although the collection of census data every ten years is not ideal, the benefits gained from high resolution population attributes outweigh this limitation.

From SAHSU’s perspective, an optimal solution would be a combination of the two approaches described by ONS: the continuation of a census once a decade to secure a reliable base of small area data and in addition the use of administrative data and/or surveys to update and expand the mid-year population estimates (e.g. by releasing more population characteristics such as ethnicity at the small area level).

Uses and benefits of population and housing statistics

Q2 Please specify any significant uses of population and housing statistics that we have not already identified?

Population statistics, i.e. age and gender distribution, form the denominators for many health statistics, in particular rates of disease and the Standardised Mortality/Incidence Ratios (SMR/SIR). The SMR/SIR are the ratios between the observed and expected numbers of cases in an area, adjusted for the age structure of the population. The disease rates and SMR/SIRs can be adjusted for area-level socio-economic status using a deprivation indicator (e.g. Carstairs index) which is calculated at a small area level. These show the relative mortality or incidence of an area with respect to the general population of England and Wales. They all use the population denominator derived from the census at a small area level and in order to remain comparable over time a consistent method of capturing the population in an area is required.

As we commented previously, the focus of SAHSU research is health analyses at small-area level such as census output areas. It is unlikely that health effects from environmental hazards will be detectable at a larger area such as local authority. SAHSU is involved with Public Health England’s Environmental Public Health Tracking programme which aims to track the health impacts of hazards. Without consistent small area level population data this programme will be unlikely to be able to track any health impacts, because small area level data are crucial to detect potential increases in disease risk at a small area level.

Population data are of great usefulness to help design and monitor representativeness of health studies. SAHSU used small area data to help design the location of the UK Biobank (<http://www.ukbiobank.ac.uk/>) data collection sites, and also have done same for the new LIFE study birth cohort (<http://www.lifestudy.ac.uk/homepage>).

We also support continuation of the ONS Longitudinal Study (a 1% sample of the population with both Census information and linked registry information on cancers and deaths). SAHSU have used this study previously to look at the migration of breast cancer cases and controls near high voltage overhead power lines (Elliott et al *Epidemiology* 2013 Mar;24(2):184-90) and are have analyses in progress looking at very long-term air pollution exposures in several hundred thousand individuals (Ghosh et al, [International Society for Environmental Epidemiology (ISEE) Conference Abstract] *Environ Health Perspect* 2013)

Ward level is sometimes included in definitions of small area level; we use ward level spatial resolution in many of our analyses. For example, we have an Environment and Health Atlas for England and Wales in press with OUP , which is entirely dependent on the census data, both for small-area population estimates to provide the denominator, and for information on powerful disease determinants at small-area scale including ethnicity and indices of deprivation. Another example is our recent study looking at differing trends in cardiovascular disease at ward level (Asaria et al *Int J Epidemiol* 2012 Dec;41(6):1737-49).

We further support the production of small area-level deprivation indices, such as Carstairs, which uses housing statistics along with other Census information. We frequently this as an adjustment factor in our analyses. An example, is our work looking at the health effects of landfill sites (Elliott et al *BMJ* 2001;323:363; Jarup et al *British Journal of Cancer* (2002) **86**, 1732–1736) where we used the Carstairs deprivation index based on small-area census data, as well as the small-area census data themselves to provide the denominator for analyses of cancers. Another example of the potential use of these data is in Public Health England efforts to estimate the “global burden” of disease at local scale, which will also be dependent on small area population and contextual estimates of deprivation derived from census.

Q3 Please specify any significant additional benefits of population and housing statistics that we have not already identified?

The Small Area Health Statistics Unit (SAHSU) was set up specifically to focus on small-area health effects. Without consistent and comparable population statistics over time SAHSU would be unable to carry out its core functions. SAHSU’s terms of reference specifically require it to develop and maintain databases of health data, environmental exposures and social confounding factors at the small-area level. Also in our terms of reference is that SAHSU should carry out substantive research studies on environment and health issues. These are only possible with the ability to monitor health impacts at a small area level over time. SAHSU has been heavily involved in the development of methods for analysing and interpreting statistics relating to small areas and acts as a centre of expertise on these methods. These are all core functions of the unit that would be impossible to continue without population statistics at small area level.

Another part of SAHSU’s core remit is to be able to respond rapidly, with expert advice, to ad hoc queries about unusual clusters of disease, particularly in the neighbourhood of industrial

installations. Cluster analysis is deeply dependent on having accurate denominator data at a very small resolution. Without the small area level denominators SAHSU would be unable to carry out part of its core work on cluster analyses.

Impact of different census approaches on statistical use (i.e. impact of different census approaches on operational or business decisions)

Q4 What would the impact be if the most detailed statistics for very small geographic areas and small population groups were no longer available? High, medium, or no impact?

High

Q4.1 If medium or high, please give further information

Most studies conducted by SAHSU would not be possible without the use of small area level data. A recent example of the type of health research we conduct is the BMJ-published study on Aircraft noise and cardiovascular disease near Heathrow airport. The study had extensive press interest, was the subject of a Prime Minister's Question about considerations of health impacts by the Airports Commission deciding on UK airport capacity (the Davies commission) within 24 hours of publication and has featured in several parliamentary discussions since then. The paper findings were included in the consultation on night flight restrictions at Heathrow, Gatwick and Stansted published November 2013 and will be considered in Howard Davies' (Chairman of the Airports Commission) interim report available at the end of 2013.

Using local authority level data this study would not have been possible because

(i) As with many environmental exposures, the aircraft noise contours don't follow local authority boundaries and small area level data (COA/SOA) gave us the flexibility to provide appropriate denominators matching the contours.

(ii) It was important to take account of ethnic composition of the area- this also applies in many other environmental health investigations. Ethnic minority groups may have higher rates of certain diseases or conditions and the distribution of ethnic groups may vary across the area exposed . There are large South Asian communities living near Heathrow and therefore also having higher noise exposures - but these areas would have higher heart disease rates regardless of any effect from noise exposures as South Asians have higher risk of heart disease. In the paper (as stated) "We used the following cut points: the national average (%) for England and Wales at census output area level (4% for South Asian, 2% for black ethnicity), double the national average (8%, 4%), and 50% South Asian or black ethnicity—areas where these comprised the majority ethnic group. This gave us four categories for each ethnicity, where the reference categories were less than or equal to the national average (%) for that ethnic group ($\leq 4\%$ for South Asian and $\leq 2\%$ for black ethnicity)." We could not have done this additional adjustment if fine resolution ethnicity information for the population denominator were not available at small area levels.

(iii) We needed to adjust for deprivation. Deprivation is strongly associated with health outcomes and is usually also associated with worse physical environmental conditions - not adjusting for it may have biased observed associations between aircraft noise and health outcomes. Deprivation levels vary markedly across areas within districts. We used Carstairs quintiles - again having this at small area level was crucial to be able to match this to the population denominators.

Q5 What would the additional benefit be if more frequent (i.e. annual) statistics about population characteristics were available for areas like Local Authorities and Electoral Wards? High, medium, low or no impact?

High

Q5.1 If medium or high, please give further information

For SAHSU the benefits of producing annual statistics about population characteristics at Local Authority or ward level would be the availability of input data to model annual characteristics at the SOA or COA level. In addition to high resolution mid-year population estimates by age and gender we would be able to also model annual ethnic distributions or characteristics related to socio-economic deprivation. This would improve any studies looking at specific years. The majority of SAHSU studies however average over many years to provide the necessary number of cases of disease (important in terms of statistical power) and a high spatial resolution is in this case more important than high temporal resolution.

Impact of different census approaches on historical research

Q6 Please specify any significant uses of census information for historical research that we have not already identified

SAHSU holds data going back to the 1970s. It is often very useful to examine these data as they can provide information about trends over time and, where there is a recent environmental exposure, what disease patterns were like before the exposure. Census information is used for the denominator and, as collection methods have remained similar over time, a change to the way in which data are collected would make comparisons of health trends at small area level difficult and also preclude any comparison of environmental exposure boundaries (which are unlikely to conform to administrative boundaries) in a before-after situation.

Q7 What advantages or disadvantages for genealogical or historical research can you see from a move to a solution based on archiving administrative data sources?

We see further problems in using administrative boundaries as these frequently change over time and this can create large problems across change years that can be avoided if small area information is available, which allows more flexibility in aggregating to required geographies over time.

Further administrative systems change over time as they are set up for different purposes. A Census reliant on external sources is highly vulnerable to changes in the system for which the administrative source is primarily set up, leading to bias in the data and lack of continuity over time. It also gives

problems in interpretation, such that basic information such as population in an area may become unreliable.

Managing risks

Q8 What are your views of the risks of each census method and how they might be managed?

Our experience in SAHSU of the use of administrative data to produce information for the purpose it was not intended for is that extracting and interpreting the information needed can be surprisingly difficult, which we discuss in following paragraphs.

Firstly, we give the example of using hospital episode statistics (HES) data for health analyses involving birth outcomes. HES data are collected for administrative reasons and we have extensive experience of their use for health analyses. A simple question of defining how many births there are using HES is not easy to answer because of (i) non-random missing data (not all births occur in hospital, not 100% of maternity units provide data) (ii) coding issues relating to how births are defined (iii) interpreting missing and wrongly entered information. These combine to currently give around 94% of all births in England captured using HES (the administrative data system) as compared with near 100% using the purpose-designed births registration information, where all births are all registered by law.

Secondly, we give the example of a study we conducted at the request of a Primary Care Trust to investigate long-term health outcomes of areas exposed to fumes from a chemical fire in 2000. A register of those exposed was compiled using GP records, but when we compared this with Census information we found that the register only captured 93% of residents in the area identified through the Census. While this might seem a small percentage, many of the missing residents were in the more deprived wards that were the highest exposed, for example a permanent caravan site that was closest to the fire. Not taking account of this in the analyses would have provided bias in the results as it would have under-represented the most highly exposed population.

The work carried out by ONS so far has indicated that the method of estimation based on administrative records will only be within in 10% at 88% of COAs and the further 12% will be even less accurate. There will introduce problems with bias as it is likely that it is the deprived, ethnic minorities and the very young that are more likely to be poorly estimated and these are the individuals who are often at the most risk from environmental health effects. Further comments on use of administrative systems are:

- Administrative systems change over time as administrative needs change and this may mean that data suddenly become non-comparable or not available and therefore cannot be used for the census.
- Use of administrative data may have hidden or increasing costs – there will be data provision, processing and interpretation costs.
- Thought will need to be taken as to protecting the census from rapid increases of costs of the administrative data, particularly if the census becomes reliant on a small number of data sources and is essentially a captive market.

- We have concerns around the comparability of a census using administrative data and surveys with previous censuses – this is very useful for our analyses, which often span censuses and need to look at trends over time.

Regarding an online Census: filling in the Census form online is likely to reduce costs, but we do have some concerns about how this will be implemented given that approximately a fifth of households do not have internet access (<http://www.ons.gov.uk/ons/rel/rdit2/internet-access---households-and-individuals/2013/stb-ia-2013.html>) and even where households do, there may be limited ability to use it to return information in certain groups e.g. the elderly. We would expect that other options would be considered to make sure these individuals are included if the decision is made to go primarily online.

Q9 Are there any other issues that you believe we should be taking into account?

We are happy to contribute further to consultation process and to discuss any of the issues identified further with ONS as needed.