
Submitted online to the ONS survey

Summary

Data provided by previous censuses of England and Wales have been essential for many areas of socioeconomic, epidemiological and other health research, and also for healthcare planning. For example, they have been used to:

- Provide population estimates for geographical areas that are used to estimate disease burdens.
- Examine environmental exposures in relation to health outcomes in very specific geographical locations.
- Provide characteristic data on the population that are used in health research, such as information on age, ethnicity, and occupation.
- Aid the design of surveys and longitudinal cohort studies
- Study health-related trends in the population over years.
- Provide historical and genealogical information to aid genetic studies of the population.

Many of these uses would not be feasible or would not produce robust results using data provided by the consultation's proposed alternatives to a full decennial census.

Despite some methodological drawbacks of a full census and its higher estimated cost compared to the proposed alternative option (i.e. linked 'administrative' data supplemented with compulsory annual population sample surveys), **the Academy of Medical Sciences recommends that the ONS requests, and be permitted to proceed with, a full decennial census in 2021 (as described in option one)**. The economic benefits provided by the outcomes of health (and other) research that can only currently be derived from census data are likely to offset much of the difference in cost between options one and two.

The proposed alternative option for producing population statistics through data linkage merits further consideration as a means of complementing data produced by full censuses. Data linkage represents a profoundly powerful frontier for health research, if legislation and collection, storage, processing and sharing systems are put in place to support this, and the data are of a high quality. Currently, however, the UK lacks the means to implement linkage on the scale that would be required to produce high-quality population data at the resolution provided by the census, and this shortcoming could severely impede health research and healthcare planning if implemented prematurely.

The Academy therefore strongly encourages future reviews of the process by which the ONS generates population statistics. These reviews should focus on examining options for improving legislation and infrastructure for the linkage of administrative data sources and the provision of annual survey data in years between censuses, and the costs of implementing these options. This will help to improve the value and reliability of census and other publicly held data, and the future cost-effectiveness of producing population statistics.

Introduction

The Academy of Medical Sciences promotes advances in medical science and campaigns to ensure that these are translated into healthcare benefits for society. Our elected Fellowship includes the UK's foremost experts drawn from a broad and diverse range of research areas. The Academy of Medical Sciences welcomes the opportunity to respond to the Office for National Statistics' (ONS) Beyond 2011 consultation on the census and future provision of population statistics in England and Wales.¹ Many of our Fellows have used census data and other ONS statistical products in their research, and have provided examples and guidance to help shape this response. We have focused on the questions from the ONS' online questionnaire that are most relevant to the Academy's remit: namely, the use of census data in health research and what impact the proposed options would have on this research. We would be pleased to provide further evidence for the consultation if required.

Views on the different census approaches described in the consultation document [addressing questions 1, 4, 5 and 8 of the consultation]

Decennial census data (i.e. those provided from a survey of the English and Welsh populations every ten years) are essential for many areas of socioeconomic, epidemiological and other health research, and for healthcare planning and funding. Researchers are able to use census data in a variety of ways (described in detail in the next section), in many cases linking these to their own collected data or other data sources to conduct unique and valuable studies. Healthcare planners, other policymakers, and medical research funders and charities all use census data to derive population level estimates of disease burdens, and to allocate funding for addressing these. Two aspects of the census are worth highlighting as especially advantageous:

- A full survey of the population of England and Wales provides a means of reliably calibrating study statistics against national-level data and for providing reliable denominator data (e.g. using population number estimates to provide the prevalence and/or incidence of diseases per region). It also provides an extremely valuable resource for correcting and reducing errors in modelled estimates of populations and their characteristics that researchers produce for areas of interest during inter-censal years.
- The high resolution of the census provides detailed data on all levels of the population, including individual households that form the smallest area units released by the ONS (i.e. ONS output and super output areas; containing approximately 300 individuals), and this is a unique source of information underpinning much health-related research.

As the ONS reports in its consultation document, there are limitations of full decennial census data.² Generating statistics from this method can be particularly problematic for health research. The census provides a useful 'snapshot' of the population at the point of conduct, but the data take over a year to be released for use. Data are therefore somewhat outdated from their first day of use and become increasingly more inaccurate with time. Societal change means that this may be more pronounced now than in past decades (the population is now much more mobile, and residents more transient). Correcting for inaccuracies in data in the periods between censuses could be improved with a combination of annual surveys and strong data linkage. Therefore the Academy warmly welcomes the ONS' work to consider further implementation of these supplementary data sources.

¹The Office for National Statistics (2013). *The census and future provision of population statistics in England and Wales*. <http://www.ons.gov.uk/ons/about-ons/get-involved/consultations/consultations/beyond-2011-consultation/index.html>

² *Ibid.*

Cost of the census is also an important consideration. The switch to an online format may help to mitigate the rising costs of conducting a full census, but it is difficult to predict the uptake of this option for responding in 2021. Specific demographic groups, such as older individuals and recent immigrants, may much be harder to reach without alternative means of responding to ensure all respondents are engaged. Therefore we support the ONS commitment to providing alternative response options such as posted responses and door-to-door follow-up.

'Administrative' data are forms of information collected on individuals primarily for administrative purposes, such as monitoring in healthcare or welfare systems. These are held by many sources in the UK, including NHS patient registers, HM Revenue & Customs customer information and Higher Education Statistics Agency data. Linkage of these data is an exceptionally promising tool for health research. The 'birth-to-death' model for linkage employed in several Scandinavian countries, where individuals are assigned identification numbers to link many forms of data throughout their lifetimes, is an incredibly powerful source of information for research. Considering this in the context of providing population-level estimates and characteristics data is very relevant for health research in the UK, because health data have not been provided by past censuses.

However, there are important drawbacks to an approach based on administrative data set out in option two (the use of administrative data and compulsory annual surveys), which would severely hinder or prohibit the conduct of some forms of health-related research. Of primary concern would be the loss of accurate and reliable small area statistics, which underpin a wealth of epidemiological and other studies that we describe in the next section. Hard to reach demographic groups would be particularly underrepresented. Even for larger population groups (local authority level or above), there would be areas of concern which the ONS notes in its analysis. These potential sources of bias have important implications for research studies, healthcare planning and spending allocation, that rely on the census for population denominator data. Some sources of information that have been captured well by past censuses for the whole population, such as ethnicity, would be lost if the proposed alternative to a full census were implemented with the current framework. These sources are used extensively by researchers as outcomes of interest or as covariables in statistical analyses. Longitudinal uses and comparisons of data across past and future censuses are another concern of changing to a system based on administrative data. A switch to coding the data collected from a combination of linked administrative data and annual surveys would create major inconsistencies and significant coding differences to those for historically collected data that would need to be addressed.

The Academy is also concerned about the timeframe for implementing option two in the place of a full census in 2021. When the Scandinavian countries transitioned from conducting full censuses to the use of administrative data, they benefitted from a smaller population size, and more sophisticated linkage infrastructure and support than England and Wales currently possess. However, these countries still took several decades to complete the transition effectively. Similarly, New Zealand has a much smaller population than England and Wales and possesses more sophisticated data linkage, but it is not considering replacing a full census with this method before 2030.³ It therefore seems unrealistic to assume that population statistics for England and Wales can be provided to a comparable or better standard than those from a full census by transitioning to the proposed alternative by 2021. The implementation of option two also requires updating legislation to enable administrative data to be used in this way in the UK.

Based on the concerns outlined above, the Academy strongly recommends that the ONS seeks government permission to conduct a full census in England and Wales in 2021. We also recommend that the ONS continues to research the use of data linkage to

³ Statistics New Zealand (2013). *Transforming the New Zealand Census of Population and Dwellings*. http://www.stats.govt.nz/surveys_and_methods/methods/research-papers/topss/transforming-nz-census-pop-dwell.aspx

supplement this, ways to improve estimates of population statistics in years between decennial censuses, and continues to explore available options for producing population statistics in the future. We recognise that the second option proposed by the ONS has substantially lower estimated cost than conducting a full census online in 2021 (i.e. a saving of approximately £200m). However, this alternative does not currently provide viable data for the conduct of much health-related research. The substantial downstream effects of losing census data for use in research without the provision of suitable alternatives have not been quantified, but these will not be trivial. The difference in estimated costs between options is therefore likely to be substantially offset by the highly valuable application of census data in health research, and in other additional uses. The next section describes the application and value of this research in detail.

Uses of population and housing statistics in health-related research [addressing questions 2, 3, and 6 in the consultation]

As described earlier, there are several uses of census data for health-related research and healthcare planning, which are described in turn here.

Denominator data

First, census data have a vital role for a huge amount of research and planning by providing denominator data— estimates of the numbers of individuals in an area of interest, which can be used to provide calculations of disease burden. For example, a 2003 study combined population estimates based on census data from 2001 with data on HIV cases to estimate the total number of adults infected with HIV in the UK in 2003.⁴ Such studies provide epidemiological insights into diseases, and are used to determine public health policy.

Denominator data derived from censuses are also critical for research funders and charities to estimate disease burdens and determine funding decisions, and for focusing the efforts and resources of policymakers on these. For example, Arthritis Research UK is using census data to calculate population estimates of prevalence and incidence for osteoarthritis, rheumatoid arthritis, back pain and high-risk fractures. These calculations will form the basis for the Musculoskeletal (MSK) Calculator: an online tool in development to allow local healthcare planners and providers in England to make informed decisions about the proportionate action required to address the burden of MSK-related conditions in their authorities.⁵

Small area statistics

The uses described above may be possible using proxies for census estimates provided by a combination of administrative data and compulsory annual surveys, because they are generally based on estimates for large areas (e.g. at the level of local authorities). However, estimates from indirect information obtained from currently available administrative data and low proportion representative surveys become increasingly inaccurate for smaller areas, where there are very specific concerns to address about local environmental exposures that could affect health, such as air quality. Thus, statistics from small areas provided by the census are particularly crucial for many epidemiological and socioeconomic studies. As an example of this, recent research linked data on rates of cardiovascular disease hospital admissions to mid-year adjusted output area estimates for London from 2001 to 2005, in order to examine whether living under the flight path

⁴ McGarrigle CA, et al. (2006). *Estimating adult HIV prevalence in the UK in 2003: the direct method of estimation*. *Sexually Transmitted Infections* **82**, iii78-iii86

⁵ Arthritis Research UK (2013). *The Musculoskeletal Calculator*. <http://www.arthritisresearchuk.org/policy-and-public-affairs/policy-priorities-and-projects/musculoskeletal-health-services/the-musculoskeletal-calculator.aspx>

from Heathrow airport is associated with an increased risk of heart disease.⁶ As well as providing illuminating findings about associations between environmental factors and disease, such studies have also helped to shape the formation of planning policies and funding allocation. For example, small area studies have investigated whether there are associations between proximity of residence to mobile phone base stations with risk of developing cancers, and the potential health effects of living near landfill sites.^{7,8,9} Whilst these studies have not always prompted dramatic policy changes, they each add evidence to the base to use for informed decision making.

Currently, the use of census data in studies of this nature is irreplaceable and other means of producing population statistics do not provide fine-grained data required on small geographical areas— the second option for providing population statistics would therefore have an extremely high and detrimental impact on this research.

Population characteristics

Census data also provide characteristics on the whole population that can be used as outcomes of interest (i.e. numerator data), as denominator data broken down into categories (i.e. for analysing subgroups of the population) or covariables for analyses— these include information on individuals' age, ethnicity and occupation. Examining subgroups provides informative results about the highly variable population of the UK: for example, 2001 census ethnicity data were used to help examine how medium-chain acyl-CoA dehydrogenase deficiency (MCADD) disease is distributed in different ethnic groups of newborn infants.¹⁰ A breakdown of mortality risk by occupation data from the 1961 census provided evidence to oppose the suggestion that cooks and butchers are at higher risk of lung cancer because of the environment of their workplaces.¹¹ Postcode information can serve as a useful proxy for a measure of social deprivation, which many epidemiological studies adjust for to ascertain a better estimate of the relationship between exposures and diseases. Using information on characteristics in health research has also had strong impacts on policy: for example, a study used occupational data to examine the risk of pneumonia infection in welders and metal workers.¹² This study, along with ensuing research, led the UK Department of Health to revise guidelines to recommend that workers exposed to metal fumes are immunised against pneumococcal infection.¹³ As described for denominator uses above, alternative data on characteristics derived from administrative data and relatively small annual survey sources become increasingly inaccurate compared to census data for smaller areas of analysis, and currently, only data from a full census will be suitable for conducting some of this research. It is imperative that changes to the collection of population statistics should not curtail monitoring of important characteristics, such as health inequalities and their determinants, in fine geographical detail which allows for the design of better policies for addressing these.¹⁴

Study design

Census data also play an integral part in study design. For example, 'the Lancet' published a series of articles in November 2013 that were based on findings from the Third National Survey of Sexual

⁶ Hansell AL, et al. (2013). *Aircraft noise and cardiovascular disease near Heathrow airport in London: small area study*. British Medical Journal **347**, f5432.

⁷ Elliott P, et al. (2010). *Mobile phone base stations and early childhood cancers: case-control study*. British Medical Journal **340**, c3077.

⁸ Jarup L, et al. (2002). *Cancer risks in populations living near landfill sites in Great Britain*. British Journal of Cancer **86(11)**, 1732-1736.

⁹ Elliott P, et al. (2001). *Risk of adverse birth outcomes in populations living near landfill sites*. British Medical Journal **323**, 363.

¹⁰ Khalid JM, et al. (2008). *Ethnicity of children with homozygous c.985A>G medium-chain acyl-CoA dehydrogenase deficiency: findings from screening approximately 1.1 million newborn infants*. Journal of Medical Screening **15(3)**, 112-117.

¹¹ Coggon D & Wield G (1995). *Mortality of butchers and cooks identified from the 1961 census of England and Wales*. Occupational and Environmental Medicine **52**, 157-159.

¹² Coggon D, et al. (1994). *Lobar pneumonia – an occupational disease in welders*. Lancet **344**, 41-43.

¹³ The Department of Health (2013). *Chapter 25- Pneumococcal*. In Salisbury D, et al. (2013). Immunisation against infectious disease: the green book. The Stationary Office, London.

¹⁴ Marmot M & Goldblatt P. (2013). *Importance of monitoring health inequalities*. British Medical Journal **347**, f6576.

Attitudes and Lifestyles (Natsal-3).¹⁵ This surveyed a random sample of approximately 15,000 adults on various aspects of their sexual behaviour, attitudes, health and wellbeing. The sample was designed to be representative of the British population, and data from the 2011 census were used as a comparator to validate this. Census data also help with the design of longitudinal cohort studies that recruit individuals and repeatedly record information on a range of characteristics, indicators of health and other factors on them over many years or decades. These are powerful resources because measures can be used repeatedly to address many research questions. Cohort studies also rely on representative participant samples so that research findings will be indicative of the wider population. For example, UK Biobank is a major health resource that aims to repeatedly record information from a range of biological samples taken from approximately 500,000 participants that were recruited from 2006 to 2010.¹⁶ Census data from 2001 were used to check that UK Biobank's cohort has a similar age and gender distribution as in the UK population, and to identify eligible participants.¹⁷ This study had initial funding of approximately £62 million (much of it from public sources), thus illustrating how census data can help to shape sensible decisions on investments in large projects. More recent data have been used similarly to aid with the design of the Life Study, a UK-wide birth cohort under development that will recruit up to 90,000 babies as a nationally representative sample to follow over years.¹⁸

Studying longitudinal trends over several censal periods

In addition to research uses of data from single censuses, which only provide 'snapshots' of the countries' populations at single time points, data from several censuses can also be combined to directly provide powerful longitudinal studies. The ONS' Longitudinal Study is a very valuable resource for health research.¹⁹ This has linked census data on 1% of the population of England and Wales from the 1971 census onwards, and the data are also linked with other records to provide additional information on the population, such as birth and death rates. This has been exploited for large studies of health inequalities in relation to individuals' socioeconomic position and social mobility, for example, and many others, such as an examination of trends in breast cancer rates in individuals living near high-voltage overhead power lines over time.^{20,21} Such approaches also indicate the potential of having improved data linkage for complementing population statistics provided by a full census, or an alternative source of equally high-quality data in the future. As discussed in the last section, any transition to a substantially different system for providing population statistics will raise problems with consistency in the coding of data compared to those collected in the past, and there would be a need to ensure valuable research that combines data from past censuses can still be conducted.

Uses of genealogical and historical data in health-related research

Finally, there are some health-related aspects to the studying of historical and genealogical issues which make use of census data. For example, detailed mapping of the genetic structure of the UK population is useful both for measuring historical immigration patterns and because this structure influences patterns of how individuals' genes relate to their risk of disease. Adjustment for the population's genetic structure improves the accuracy of studies which examine how individuals' genes affect their disease risk. Adjustments can be improved by using historical information on the

¹⁵The Lancet (2013). *The Third National Survey of Sexual Attitudes and Lifestyles*. <http://www.thelancet.com/themed/natsal>

¹⁶ UK Biobank (2013). *About UK Biobank*. <http://www.ukbiobank.ac.uk/about-biobank-uk/>

¹⁷ UK Biobank (2007). *UK Biobank: Protocol for a large-scale prospective epidemiological resource*. <http://www.ukbiobank.ac.uk/wp-content/uploads/2011/11/UK-Biobank-Protocol.pdf>

¹⁸ <http://www.lifestudy.ac.uk/homepage>

¹⁹ <http://discover.ukdataservice.ac.uk/catalogue/?sn=5762>

²⁰ Sacker A, et al. (2000). *Comparing health inequality in men and women: prospective study of mortality 1986-96*. *British Medical Journal* **320**, 1303.

²¹ Elliott P, et al. (2013). *Adult cancers near high-voltage overhead power lines*. *Epidemiology* **24**, 184-190.

surnames of individuals and their ancestors from data from past censuses, because surname information acts as a stable proxy to show from where individuals have descended.²²

²² Winny B, et al. (2012). *People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population*. *European Journal of Human Genetics* **20**, 203-210.

This response was prepared by Georgina Apostoli (Policy Intern) with input from the Academy's Fellows. In particular, we'd like to thank Professor David Coggon OBE FMedSci (University of Southampton), Professor Carol Dezateux CBE FMedSci (University College London), Professor Paul Elliott FMedSci (Imperial College London), Dr Anna Hansell (Imperial College London), Professor Matthew Hickman (University of Bristol), Professor Dame Anne Johnson DBE FMedSci (University College London), Professor Neil Pearce FMedSci (London School of Hygiene and Tropical Medicine), Professor Alan Silman FMedSci (Arthritis Research UK) and Professor Anthony Swerdlow FMedSci (Institute of Cancer Research) for their assistance.

For further information, please contact Dr Dylan Williams (Dylan.Williams@acmedsci.ac.uk; +44(0)20 3176 2167).

Academy of Medical Sciences

41 Portland Place
London, W1B 1QH
+44(0)20 3176 2150
info@acmedsci.ac.uk
www.acmedsci.ac.uk

Registered Charity No. 1070618
Registered Company No. 35202