

311.312:351.755.3-044.3(100)

**J. J. Brown,**  
Professor,  
School of Mathematical & Physical Sciences,  
University of Technology Sydney,  
E-mail: James.Brown@uts.edu.au

**Дж. Дж. Браун,**  
професор,  
Школа математичних і фізичних наук,  
Технологічний університет Сіднея

**Дж. Дж. Браун,**  
профессор,  
Школа математических и физических наук,  
Технологический университет Сиднея

### **Future Models for Population Census: Can We Have an Administrative Based Census Without a Population Register?**

There is an increasing interest in producing a census-like count of a population's characteristics without the use of a full traditional census. In this paper we review some of the models that were used in the 2010/2011 round of censuses within Europe (and English speaking countries of OECD); and discuss the possibility of administrative-based alternatives in countries that do not have, at this time, a full population register.

**Key words:** *Population Census, Administrative Data, Population Registers, Coverage Errors, Population Estimation.*

**Introduction.** Census-based information is still the bed-rock of population statistics for countries around the world. It provides a unique snapshot of the population down to the lowest levels of geography enabling efficient local planning, efficient allocation of resources, as well as providing the basis for annual population bench-marks for ongoing survey estimation. In June 2015 the United Nations Economic and Social Council adopted a resolution on the 2020 World Population and Housing Census Programme stating:

‘Population and Housing Censuses form the foundation of national statistical systems and area critical source of information for national planning. Their value as an important vehicle for data collection has been reiterated during the course of discussion on the post-2015 development agenda’ [1].

However, there is a clear recognition of the need to modernise the traditional census approach. For example, in 2010 the UK Statistics Authority asked the National Statistician and the Office for National Statistics (ONS) to review the future provision of population statistics in England and Wales, the idea being that the 2011 Census would be the last traditional census. The outcome of the work led to a recommendation that in 2021 a traditional census would still be needed, albeit utilising modern approaches. In the UK Government's response, accepting the recommendation they state:

‘Government recognises the value of the census and its history as a bedrock of statistical

infrastructure. The census provides information on the population that is of fundamental importance to society. At the same time, Government has long said that the census in its current form is outdated and – with modern technology – could be delivered more effectively and more cheaply. Modernising the approach could significantly improve the speed of analysis and outputs. In addition, we are not making the best use of the considerable data that government already collects’ [2].

So while the need to population census type data is not controversial, there is an increasingly diverse approach to census taking around the world as countries respond to changing society and the desire to deliver information in the most cost-efficient way. The 2007 ISI Satellite Conference ‘Innovative Methodologies for Censuses in the New Millennium’ saw presentations showing the range of new approaches that countries were exploring in preparation for the 2010 / 2011 census round (the program of Conference with short papers can be found at [3]). In 2010, the review by P. Valente [4] identified four approaches being used by 40 European countries in the 2010 round of population censuses. It showed that while around half (21 countries including Ukraine) still planned to conduct a traditional census, seven countries were moving from a traditional census in the 2000 round to an alternative; while three countries with no census in the 2000 round were planning an alternative approach in 2010. In this paper, we review some of these alternative approaches and discuss the possible application of an administrative-based

approach in countries without a population register. A more detailed review of some of these approaches as applied in the 2010 round of censuses, with various case-studies, can be found in [5].

**Approach One: Modernising the Traditional Census.** In many countries, census taking has a long history. For England and Wales this goes back to 1801. As early as 1841, the Census was being conducted by a team of local enumerators with individual households completing the forms. More recently, countries have looked to increase the efficiency of data collection by utilising the postal system. The US Census Bureau started using a bespoke address register, the Master Address File (MAF), and post-out for some households back in 1970 [6]. In 2001, the UK Censuses utilised post-back to reduce the cost (and increase the efficiency) of data collection for the first time; but still used enumerators to deliver forms as there was not a suitable address register. However, in 2011 the Office for National Statistics (ONS) conducted a post-out post-back census operation for 100% of households (see [7]) based on a national address register constructed from the postcode address file [8] produced by Royal Mail augmented with other administrative sources. There was also the option to return via the Internet and this was taken by 16.4% of households as reported in the 2011 Census General Report available online [9].

Statistics Canada has pioneered the approach of pushing Internet returns from households. During the 2006 Census they piloted an 'Internet Promotion Response' approach that involved just posting a letter with an internet access code as opposed to the full form with an internet access code printed on the form. The results in [10] showed an internet uptake greater than 60% for the treatment group as opposed to less than 25% for the control group. (The overall internet response rate for the 2006 Census was just over 18%). Consequently, this approach was applied widely in the 2011 Census resulting in 60% of households receiving a letter with an access code and over 50% of households responding via Internet [11]. Other countries are following; in Australia the 2016 Census: '...will be Australia's first Census where more than two thirds of Australia's population (more than 15 million people) are expected to complete the Census online in August 2016. New delivery and collection procedures will make it easier to complete the Census online. In the lead up to 9 August, households will receive a letter from the ABS, addressed 'To the Resident', including a unique login and instructions on how to complete the Census online. Completing the Census online will be fast, easy, secure, environmentally friendly, and help to reduce the cost of the Census to the community' [12].

In the UK, while there will be a 'traditional' Census enumeration it will also most likely build on the use of the address register in 2011 and take the

Statistics Canada approach to deliver a census with predominantly internet returns. So, even in those countries using the 'traditional' census model, there has been tremendous change during the past two decades. There has been a move to using address registers to aid the delivery of forms, use of post as a method of return, and most recently the push of Internet as the preferred method for households to respond.

**Approach Two: Complementing a Traditional Census with Surveys.** For several rounds of census taking, one approach to reducing the census-burden has been the use of a long-form short-form structure. In this case, all households receive a short-form that collects the basic household and individual data, while a sample of households receive a longer more detailed form. Two countries taking this approach in the past were Canada and the US. (Interestingly, the UK has never taken this approach to the collection of data but certain outputs for the 1991 Censuses are based on a 10% sample as not all information collected on the forms was processed.) However, these two countries have moved away from this approach in two different ways. Post 2000, the US Census Bureau introduced the American Community Survey; a large ongoing household survey that: '...helps local officials, community leaders and businesses understand the changes taking place in their communities. It is the premier source for detailed information about the American people and workforce' [13].

The annual survey covers around 3.5 million households, which equates to around two percent of the population. The survey design is based on the US Census Bureau Master Address File (MAF), which also forms the basis of the post-out component in the short-form decennial census, and this allows geographic structures to be incorporated. The aim of this paper is not to give the detail of the ACS design, a full picture of the ACS design and operation can be found in [14] with an update in 2014. However, by using the MAF and a design that rotates around blocks of addresses within specified levels of geography, it produces estimates at differing levels of geography by averaging across successive years with the lowest level of output based on five years of data.

A critical part of the ACS is that, like the short-form Census it complements, response by selected households is compulsory. Under-taking a compulsory collection, in addition to the short-form Census, has been attacked and such a large ongoing survey has been controversial in recent years, see for example 'The Beginning of the End of the Census?' published online by the NYTimes 19th May 2012 [15]. Therefore, in Canada the approach was to replace the compulsory long-form with an optional questionnaire that accompanied the standard census short-form for a sample of households in the 2011 Census; called

the National Household Survey, 2011. This was not without controversy as well; and as a result of it having voluntary response the National Household Survey achieved a response rate of just less than 70%. [16] outlines the comprehensive weighting approach, which utilises the existence of the 2011 Census, to minimise non-response bias; as well as the overall sample design.

**Approach Three: Rolling Census Approach.**

The rolling census approach has been pioneered in France as a survey-based approach that both replaces the traditional population census and also yields more regular data at low levels of aggregation. [17] gives a detailed review of the design of the French Rolling Census as well as outline the approach to estimation. At the heart of the system is the local government unit called a commune. These vary in population size from less than 10,000 to 100,000s; and the rolling census design treats small and large communes differently. Small communes are completely enumerated on a five-year cycle, with 20% of small communes sampled each year. Large communes are split into five rotation groups with one rotation group selected each year. A sample is then selected from the selected rotation group. Each year data collection covers around one seventh of the total population. As with the ACS, the rolling census approach uses averaging across successive years (in the French approach five years) to produce estimates annually at the desired level of aggregation.

The crucial difference with the ACS is that there is no external system to produce population estimates. The ACS, discussed under approach two, is able to bench-mark survey estimates to annual population estimates, which themselves come from rolling forward the latest census. However, with a rolling census system, it is the 'rolling census' that is providing those population bench-marks for other surveys and local planning. However, that does not mean it cannot pull-in external administrative sources to support data collection and estimation. As discussed in [18], correlated aggregated administrative data can help with estimation, and household level administrative data can support non-response adjustment. In the French system, see [17], administrative data is crucial for the maintenance of the address register that under-pins data collection in the large communes. It also plays a role in the annual updating by either providing an updating for the small communes, or as a basis for averaging across years with the large communes.

**Approach Four: Complementing a Register with standard Surveys.** The approach pioneered in the Netherlands is outlined in [19]. It is built on a population register that provides the basis for all social surveys. The full set of social surveys is then integrated into the population register to maximise the information available across a wide range of

topics. This now creates a population level file that has some basic information for all individuals from the register (such as age and sex) and then detailed information for sampled units from a range of surveys. Using a calibration approach as outlined in [20], a full set of census type outputs are created. A key point is that while no single survey has a particularly large sample size, the use of common definitions across core questions ensures the data can be integrated together with the population register. This ensures sample sizes that can support producing tables at the lowest level of aggregation required by EUROSTAT.

By the 2010 / 2011 round of censuses, [4] shows that the concept of a register linked with surveys had been rapidly introduced across central Europe, including Germany. The quality and depth of information from the registers varies across countries and therefore surveys play an important role; but in all cases there is the concept of a population register, not just sets of administrative data.

**Approach Five: Full Register-Based Census.**

The use of registers for statistical purposes has been pioneered in the Northern European countries but has spread to now include Austria. These countries are now turning to their administrative registers to deliver a full range of census outputs. They have a well-established history of population registers and link extensively across administrative data to create the census database in a census year while using a basic population register to give annual updates. In a sense this is just an extension to Approach Four where the registers are so comprehensive additional surveys are not required. However, a country could easily supplement the register(s) with survey information as in Approach Four as this basic register is also the basis for any on-going social surveys.

**The Future: An Administrative-Based Approach?** As already mentioned, the Beyond 2011 Census program run by the Office for National Statistics set about investigating alternatives to the traditional census. These included some of the approaches above (see for example [18] on the rolling census of Approach Three) but at the heart of the program was the desire to introduce something like Approach Five but without the machinery of a formal population register. In reality, it was likely that it would start more like Approach Four and become less reliant on augmenting surveys as the system developed.

The starting point for the thinking was recognising that Government have several administrative systems that contain lists of individuals. These would include tax records, social benefit records such as pension payments, healthcare records, and school enrolment information. The issue is that the administrative systems are not designed to count the population but have been designed to support a particular administrative function. For example, in the UK

context the tax system, which operates in conjunction with the National Insurance system, contains records for all individuals that have ever made payments as those payments qualify the individual for future pension payments regardless of where the individual then resides. Therefore, the raw count of live records from the system far exceeds the current resident population of the UK, although typical most individuals not in the UK can be removed. A framework for dealing with coverage (and other quality) issues is discussed in [21]; looking at both the issues of over-coverage and under-coverage.

One of the first attempts to utilise these general administrative systems was around the 2000 Census in the US. The research work led to the creation of an administrative records census along-side the traditional decennial Census, referred to as the Administrative Records Census Experiment or AREX 2000. The work to create the database is outlined in [22]. While the Census Bureau may not have pursued an administrative based option yet, the work has led to the Statistical Administrative Records System (StARS). This is a research database containing person and address data from different administrative record sources provided to the Census Bureau from other federal agencies; and StARS was utilised in the quality assurance of the 2010 Census.

This approach of individual level linkages was also pursued as part of the work for the Beyond 2011 Census program but at the same-time additional work also considered an aggregate approach. As in the unit record case, the issue is coverage with both under- and over-coverage being a problem for the aggregated administrative data. [23] tackle correcting coverage errors when the administrative data is available in aggregate form for the desired level of dis-aggregation. In their example this is the patient record data for England and Wales dis-aggregated by age-sex groups and local government area. They then combine with accurate population information (for illustrative purposes they use the 2011 Census) that may be from a large-scale national survey and therefore only available at very high (national or regional) levels of aggregation. In a sense, the aggregate administrative data can be thought of as auxiliary information to support small area estimation [24]. As discussed in [18], aggregated administrative data can also provide correlated auxiliary data to strengthen estimation in a rolling census set-up.

An issue with using aggregated administrative data is that while it may provide basic demographics, such as the population by age and sex at low levels of geography, it is not well-suited to providing the richness of variables in a full traditional census. The alternative is to create a population spine by either taking a single administrative source or a linkage of several sources at the level of the individual. If several sources are linked this may add to the richness

of variables although the initial target may still not be much beyond age and sex. [21] identifies a full error framework and then discuss using a linked survey to assess both coverage and quality of the administrative data. From this the administrative data can be corrected and estimates produced. The ability to tackle quality as well as coverage is important. Coverage issues for administrative data often dominate the discussion, particularly over-coverage, but the way administrative data is collated can have quality impacts as well. For example, health data may contain a marker for ethnicity but this may be completed poorly by a health-professional rather than directly reported by the individual. The survey would have a higher quality measurement so could help correct the administrative data. The survey would also be able to increase the richness of variables in the administrative database, and in that sense is similar to Approach Four.

The estimation approach suggested in [21] makes use of the PREG estimator introduced in [25] for measuring coverage of the Australian Census. The difference with this approach over more traditional dual-system estimation as in [26] is its ability to cope with both under- and over-coverage in an integrated way. It also does not require the double survey approach utilised by the US Census Bureau [27]. As it uses the census, or in this case the population spine from administrative data, as bench-marks it can also cope with the quality issue where an individual's characteristics are poorly measured on the administrative data. In the case of the Australian Census this has been developed to deal with poor reporting of indigenous status in the census that the survey then corrects.

One issue with the single survey approach is dealing with over-coverage resulting from emigration. This form of over-coverage is important when dealing with administrative systems rather than a traditional census, and [21] make some suggestions to avoid using the double survey approach but these have never been tested in a real situation. However, when the population spine is created from linking multiple administrative sources, and then combined with a survey, [28] introduces an estimation approach that can cope with this form of over-coverage. The estimation develops the basic dual-system approach within the log-linear framework allowing for a model for under-coverage and a second independent model for over-coverage.

Once you conceptualise the issue as the survey providing the key information with the administrative data providing auxiliary bench-mark information it makes it clear that the administrative data does not need to be correct but can play a powerful supporting role in either a rolling census approach or as the basic population spine as in Approach Four. [29] provides a more general framework in which to fit

this use of auxiliary administrative data with survey data, showing how such an approach can not only be efficient but also reduce non-response bias in any of the survey based estimates.

**Discussion.** There are a range of countries where it is unlikely that a true population register will exist in the (near) future. However, there is considerable effort now in exploring administrative sources. In their preface to the to the Journal of Official Statistics Special Issue on Coverage Problems in Administrative Sources [30], the authors acknowledge the value of census type data while pointing to the considerable financial advantage of replacing a traditional census with an administrative sources. Therefore, it is true that the 2011 Census was not the last traditional census in the UK but the 2021 Census may well be.

These approaches of combining (imperfect but readily available) administrative sources with additional survey data is also an attractive proposition when a full traditional census may be desired but unfeasible due to financial constraints or local political situation. In such situations it is clear we can make a considerable improvement beyond relying on historical census sources. Creating a population spine using either a single or multiple administrative data sources and linking to surveys has the potential to create an excellent alternative to the traditional census, certainly for reasonably high levels of population aggregation.

### References

1. United Nations Economic and Social Council adopted resolution on the 2020 World Population and Housing Census Programme [Electronic resource]. – Access mode : [http://unstats.un.org/unsd/ecosoc\\_2020wp.htm](http://unstats.un.org/unsd/ecosoc_2020wp.htm)
2. UK Government Response to Beyond 2011 Census Recommendations – letter 18th July 2014 from Rt. Hon. Francis Maude MP to Sir Andrew Dilnot [Electronic resource]. – Access mode : <http://www.statisticsauthority.gov.uk/reports---correspondence/correspondence/letter-from-rt-hon-francis-maude-mp-to-sir-andrew-dilnot---180714.pdf>
3. Innovative Methodologies for Censuses in the New Millennium [Electronic resource]. – Access mode : <http://www.s3ri.soton.ac.uk/isi2007/programme.php>
4. Valente, P. (2010). Census taking in Europe: how are populations counted in 2010? *Population & Societies*, 467, 1–4.
5. Baffour, B., King, T. and Valente, P. (2013). The Modern Census: Evolution, Examples and Evaluation. *International Statistical Review*, 81, 407–425.
6. Kaplan, D. L. (1970). Plans for the 1970 Census of Population and Housing. *Demography*, 7, 1–18.
7. Cabinet Office (2008). Helping to shape tomorrow – the 2011 Census of Population and Housing in England and Wales. Cm 7513. London: Stationery Office.
8. The Postcode Address File [Electronic resource]. – Access mode : <http://www.royalmail.com/business/services/marketing/data-optimisation/paf>
9. 2011 Census General Report. Office for National Statistics [Electronic resource]. – Access mode : <http://www.ons.gov.uk/ons/guide-method/census/2011/how-our-census-works/how-did-we-do-in-2011-/2011-census-general-report/index.html>
10. Cote, A. M and Laroche, D. (2009). The Internet: A New Collection Method for the Census. *Statistics Canada's International Symposium Series: Proceedings from the 2008 Symposium 'Data Collection: Challenges, Achievements and New Directions'*.
11. Statistics Canada (2012). Overview of the Census: Census year 2011. Statistics Canada: Catalogue no. 98-302-XWE.
12. Census is going digital. Australian Bureau of Statistics [Electronic resource]. – Access mode : <http://www.abs.gov.au/websitedbs/censushome.nsf/home/2016>
13. American Community Survey (ACS). United States Census Bureau [Electronic resource]. – Access mode : <https://www.census.gov/programs-surveys/acs/>
14. U.S. Census Bureau (2009). Design and Methodology American Community Survey. U.S. Government Printing Office, Washington, DC.
15. Rampell C. The Beginning of the End of the Census? [Electronic resource]. – Access mode : [http://www.nytimes.com/2012/05/20/sunday-review/the-debate-over-the-american-community-survey.html?\\_r=0](http://www.nytimes.com/2012/05/20/sunday-review/the-debate-over-the-american-community-survey.html?_r=0)
16. Statistics Canada (2015). Sampling and Weighting Technical Report. Statistics Canada: Catalogue no. 99-002-X.

17. Brown, J. J. and Beaujolain, E. (2013). Review of the Rolling Census Approach: and other survey-based options. *CPC Working Paper 32, ESRC Centre for Population Change, UK* (<http://www.cpc.ac.uk/publications/home.php>).
  18. Brown J. J. (2013). *Exploring the feasibility of a Rolling Census Design for England*. Office for National Statistics: Beyond 2011 Research Report 2a.
  19. Nordholt, E. S. (2005). The Dutch virtual Census 2001: A new approach by combining different sources. *Statistical Journal of the United Nations Economic Commission for Europe*, 22, 25–37.
  20. Renssen, R. (1998). Use of Statistical Matching Techniques in Calibration Estimation. *Survey Methodology*, 24, 171–183.
  21. Brown J. J. and Honchar O. (2012). Design and Estimation of Surveys to Measure Data Quality Aspects of Administrative Data. *Lithuanian Journal of Statistics*, 51, 5–16.
  22. Judson, D. H. (2002). *Merging Administrative Records Databases in the Absence of A Register: Data Quality Concerns and Outcomes of an Experiment in Administrative Records Use*. Joint UNECE-EUROSTAT Work Session on Registers and Administrative Records for Social and Demographic Statistics (Geneva, 9–11 December 2002), WP5. ([http://www.unece.org/index.php?id=14707#/?](http://www.unece.org/index.php?id=14707#/))
  23. Yildiz, D. and Smith, P. W. F. (2015). Models for Combining Aggregate-Level Administrative Data in the Absence of a Traditional Census. *Journal of Official Statistics*, 31, 431–451.
  24. Rao, J. N. K. (2003). *Small Area Estimation*. New York: Wiley, 344.
  25. Bell, P., Clarke, C. and Whiting, J. (2007). An Estimating Equation Approach to Census Coverage Adjustment. *ABS Research Paper*: 1351.0.55.019.
  26. Brown, J. J., Diamond, I. D., Chambers, R. L., Buckner, L. J. and Teague, A. D. (1999). A methodological strategy for a one-number census in the UK. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162, 247–267.
  27. Hogan, H. (1993). The 1990 Post-Enumeration Survey: operations and results. *Journal of the American Statistical Association*, 88, 1047–1060.
  28. Zhang, Li-Chun. (2015). On Modelling Register Coverage Errors. *Journal of Official Statistics*, 31, 381–396.
  29. Kott, P. S. and Chang, T. (2010). Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse. *Journal of the American Statistical Association*, 105, 1265–1275
  30. Bakker, B. F. M., van der Heijden, P. G. M. and Scholtus, S. (2015). Preface to the Special Issue on Coverage Problems in Administrative Sources. *Journal of Official Statistics*, 31, 349–355.
- All electronic resources are accessed on 31st October 2015.