

## **Redesign in a statistical office – sociodemographic statistics in the Office for National Statistics**

Paul A Smith

Office for National Statistics, Newport, UK

Corresponding author: paul.smith@ons.gov.uk

The Office for National Statistics (ONS) has developed a Strategy setting out the areas in which it aims to make progress in the coming years in order to be well-placed to take advantages of the opportunities offered by technology and new data sources, and to meet the challenges of anticipated budget constraints and falling response rates. One key part of this is a project investigating options for future provision of population and small area sociodemographic statistics. I will outline the way the options in this project interact with ONS's existing social survey programme and the system for constructing population estimates as an example of how the strategy is leading the redesign process in the ONS.

Keywords: Administrative Data, Coverage Survey

### 1. Introduction

The Office for National Statistics (ONS) is undertaking a programme of work to investigate alternatives to a traditional population census as a means for providing information on the size and characteristics of the population in England and Wales (with similar work taking place in Scotland and Northern Ireland). More details of this work programme, called *Beyond 2011* are provided in another paper at this conference (Benton 2013). At the same time, the ONS has developed and published the ONS Strategy 2013-2023 (ONS 2013) to set out the directions and priorities for progress in the coming years; this currently a high-level document, with work taking place to define the specific actions which will help in the achievement of these aims. This paper takes the example of the *Beyond 2011* programme, and uses it to show where the opportunities presented by the research can be used to help in the development and redesign of statistical systems throughout the ONS.

### 2. Data Sources

The ONS has a long history of using administrative data as a basis for the provision of some of its statistics, for example the registration of births, marriages and deaths in making population estimates; tax registrations as the basis of the business register; and regulatory information on some economic sectors such as banking as inputs to the national accounts. In the 2000's, information on the origin and use of administrative data sources within ONS was brought together in a catalogue, which highlighted some duplication of data flows and some discrepancies in which data were being used in different parts of the office, and this gave rise to some rationalisation and harmonisation, leading to better consistency between some outputs.

The Statistics and Registration Service Act 2007 (SRSA), which updated the governance of official statistics production in the UK by creating the Statistics Authority, recognised the increasing importance of administrative data in statistics

production, and included two provisions that have underpinned their use. First, a general provision (with some exclusions) that “Any information obtained by the [Authority] in relation to the exercise of any of its functions may be used by it in relation to the exercise of any of its other functions”. Second, a mechanism which enables (but does not require) data sharing between government departments; this process still requires considerable negotiation between departments and the laying of an Order before Parliament, but has meant that it has been possible to gradually bring more data sources together to contribute to estimates. The Code of Practice for Official Statistics (also established under the SRSA) now requires each department to produce a *Statement of administrative sources* (effectively the latest incarnation of the administrative sources catalogue), and the latest statement from the ONS (ONS 2012) contains over 700 separately identified sources.

This increasing availability of administrative data and of a central catalogue has options for bringing together data as a basis for an ‘administrative census’ to be identified, and this is one of the possibilities under investigation as part of the Beyond 2011 programme. There are technical challenges around the matching of data arising from different administrative processes, and from the requirement to preserve the confidentiality of information so that individuals are not identifiable while the data are used for statistical purposes. Providing workable technical solutions for these issues for Beyond 2011 is vital to that programme, but we can also use this as a foundation for a more general approach to combining administrative data and using it as a building block for statistical outputs. The debates over data confidentiality engendered by the Beyond 2011 project should also provide the framework for how other data sources can be stored and used.

Several other projects within the ONS are investigating other uses of administrative data, including work on using tax data in business statistics, and it is also important to ensure that these approaches can be brought together to make a coherent system for using administrative sources.

### 3. Synergies and integration in sociodemographic statistics

The ONS and some other government departments in the UK run a program of household surveys which collect a variety of information about characteristics of the population, their spending patterns and information on wealth and debt, and also covering a number of specialist topics, sometimes regularly and sometimes on an occasional basis. The major, continuous household surveys run by the ONS are:

- Labour Force Survey
- Annual Population Survey (which provides a boost to the sample size to support direct estimates for Local Authorities)
- Living Costs and Food Survey
- Opinions Survey
- Survey of Living Conditions

There were steps towards integrating these continuous surveys, and an Integrated Household Survey (IHS) structure was set up, to combine the data to provide estimates of a range of topics (Smith 2009). The pooled data is weighted to account for the

different designs in the different components, but the full vision of an integrated survey producing all the outputs, with consistency built in to the methods, has not been fully realised. In part this is because developments in some of the components have led to unbundling rather than integration – the Opinions Survey, which includes questions on a repayment basis for other government departments has a quick turnround for data processing, and has been unbundled from the IHS; budgetary pressure has led to rationalisation by the merger of large parts of the former General Lifestyle Survey with the Family Resources Survey (run by the Department for Work and Pensions), with some separate longitudinal follow-up to collect information required for the European Survey of Living Conditions. Remaining questions from the General Lifestyle Survey were added to the Opinions Survey. The net result has been a reduction in the sample size of the IHS, which is now substantially similar to the combined LFS/APS.

In Beyond 2011, in order to produce estimates for some of the variables which are currently provided by the population census, there will be a need for sources additional to the administrative data inputs. These data could be obtained in a number of ways, but it is likely that a survey component will be needed for at least some of the variables which are most difficult to model or derive. Under some scenarios this might be a very substantial survey, comparable to the American Community Survey (ACS). The options for the design of this survey are currently under investigation, but it is likely that there will be benefits from the integration of this survey within the existing program of household surveys. This could once again make the IHS a good model for the production of a range of statistics, and see a closer integration between the census-type demographic information and the more detailed survey information.

The design options for a survey of population characteristics open up some possibilities around data collection, where the use of multiple modes is likely to be needed in order to make collection of some data from large samples practical. This is likely to lead to developments in the way the field force, who collect this information, is managed, and therefore to contribute to systems for managing the collection process.

There is a further set of options in deciding how a coverage survey, designed to assess the extent and characteristics of under- and over-coverage of the administrative sources, should be designed and implemented. The characteristics of such a survey – that it should be as easy to complete as possible and directly administered (both designed to maximise response) – make it more difficult to integrate into an IHS-type framework. But there are still possibilities for efficient use of field resources in administering a range of survey instruments, and in maintaining the management information for all these processes in a harmonised way.

#### 4. Nonresponse

Response rates in social surveys in the UK continue to fall gradually, and this makes it desirable to have mechanisms which help to compensate for non-response bias. These surveys have traditionally used weighting to population estimates (by age, sex and region) to partially compensate for non-response, and the population census has been the fundamental underpinning for the population estimates. In options for alternatives to the census, it is therefore important to maintain a sound basis for population estimates, and this requirement will be part of the assessment.

There is also the possibility that administrative data may be used in place of some data which is currently collected in surveys. There are a number of challenges in setting up a process to do this, including:

- obtaining informed consent to match survey responses to administrative data
- obtaining adequate matching information (quite straightforward for some variables, but some trials of matching employer information with the business register on the Labour Force Survey demonstrated that many people do not know the 'official' designation for where they work);
- coping with differences between the administrative data concept and the one required for survey purposes.

## 5. Small Area Estimation

ONS already produces model-based small area estimates for a number of variables, particularly employment status (ONS 2006) and income (White et al. 2011), and several of these have become established products and been approved as National Statistics (that is, complying with the Code of Practice). The basic approach is indirect estimation ('borrowing strength') using multilevel models based on available predictors, which are often derived from the latest population census or from administrative datasets. In the development of small area estimates there have also been investigations of other variables (for example psychiatric morbidity) where there has proven to be insufficient predictive power among the available predictor variables to allow suitable estimates to be produced.

Indirect estimation was also a component of the estimation process for the 2001 and 2011 population censuses. In these cases the Census enumeration itself provided the predictors, and this was used to distribute the adjustment from dual system estimation to Local Authorities (LAs, which form the main local government level in the UK).

In order to produce small area estimates of characteristics not included in administrative data sources, which is an important component of the existing outputs from the population census, it will be necessary to have some predictors which are sufficiently well correlated with the required variable. In some cases predictors are probably available from existing administrative sources, but we expect that in other cases identifying appropriate variables will be more challenging, although the continuing accretion of administrative data available for statistical purposes does provide additional opportunities. In the medium term the 2011 Census can be used to provide some predictors, but to have a procedure which is robust to changes in the long-term the predictors must come from sources which are updated or renewed regularly.

So a census based on administrative sources with a coverage component may be able to produce reasonable information on population numbers by age and sex within administrative areas (especially LAs). For some other variables in LAs and many variables in smaller areas, there will be a need for some small area estimation if there are to be estimates; and in some cases there may not be sufficient predictive information to construct an appropriate model.

Outside the Beyond 2011 program, the availability of more administrative data, and in some cases linked administrative datasets, provides a richer resource of predictors for the development of small area estimates. There is a link to options for building a data warehouse to make this more generally accessible, though again with regard to confidentiality protection. As further research into small area estimation takes place it also increases the knowledge of approaches to small area estimation and where they work best.

## 6. Delivering the ONS Strategy

The ONS Strategy (ONS 2013, Davies 2013) includes several targets which are supported by work on the development of Beyond 2011, including “Be at the forefront of integrating and exploiting data from multiple sources”. It is important that we use developments in Beyond 2011 to provide tools and processes which fit with the strategy, and which can be reused. There are a number of elements which contribute here, including:

- data storage and integration – a warehousing approach for administrative and other data which makes appropriate variables readily available where they are needed
- data security – ensuring that the processes for accumulating and matching data do not compromise their security and confidentiality
- sound methodology – the various research streams are being used to develop methods whose properties are known, and to provide this evidence as part of decision-making so that the quality and cost of outputs can be balanced

## 7. Conclusions

The development work being undertaken to evaluate a range of options for producing demographic and sociodemographic statistics as an alternative to a population census is creating a range of opportunities for advancing the ONS Strategy. The considerable resources being invested in this work are leading to increased knowledge, experience of applying methods and procedures in different cases, development of infrastructure – initially to support the research and prototype new approaches, but eventually also to enable regular production of estimates, and development of data resources. Although the work being undertaken is strongly associated with the requirements to make decisions at the end of the Beyond 2011 programme and to provide the evidence with which to do that, it also provides prototypes, practices and knowledge which are applicable elsewhere in the ONS, and one of the requirements is to spread that knowledge to improve the whole statistical system.

## Dedication

This paper was due to be given by Frank Nolan, but he died suddenly and unexpectedly in October 2012. Frank was a tremendous thinker about organisational and strategic issues, and I am sure that he would have included many additional insights had he been here to present this paper in person. I have done my best to live up to his legacy, and would like to dedicate this article to his memory.

## References

Benton, P. (2013) "Beyond 2011 - A new paradigm for population statistics?" Invited Paper session 027 at the 59<sup>th</sup> ISI World Statistics Congress.

ONS (2006) "Model-based estimates of ILO unemployment for LAD/UAs in Great Britain, guide for users." Available from <http://www.ons.gov.uk/ons/guide-method/method-quality/specific/labour-market/subnational-labour/model-based-estimates-of-ilo-unemployment-for-lad-uas-in-great-britain---guide-for-users.pdf>

ONS (2012) Statement of administrative sources, Spring 2012. Available from <http://www.ons.gov.uk/ons/guide-method/the-national-statistics-standard/code-of-practice/statement-of-administrative-sources/statistical-usage-of-administrative-or-management-sources/statement-of-administrative-sources.xls>

ONS (2013). ONS Strategy 2013-2023. Available to download from <http://www.ons.gov.uk/ons/about-ons/what-we-do/corporate-documentation/strategies-and-policies/ons-strategy-2013-23/index.html>.

Smith, PA (2009). Survey harmonisation in official household surveys in the United Kingdom. Proceedings of the World Statistics Congress 2009.

UK Statistics Authority (2009). The Code of Practice for Official Statistics, Edition 1.0.

White, N, Dent, A, Clarke, P, Silva, D & Naylor, J (2011). Model-Based Estimates of Income for Middle Super Output Areas, 2007/08, Technical Report. [http://neighbourhood.statistics.gov.uk/HTMLDocs/images/TechnicalReport2007\\_08final\\_tcm97-112199.pdf](http://neighbourhood.statistics.gov.uk/HTMLDocs/images/TechnicalReport2007_08final_tcm97-112199.pdf)