DE GRUYTER
OPEN

Journal of Official Statistics, Vol. 31, No. 3, 2015, pp. 515–525, http://dx.doi.org/10.1515/JOS-2015-0031

Discussion

Anders Holmberg[1]

1. Introduction

I would like to thank the editors for the opportunity to comment on the coverage issues affecting administrative data (AD) in this special issue of *The Journal of Official Statistics*. I will follow the definition provided in UNECE (2011) and refer to AD as data collected external to statistical offices, while administrative sources are data holdings that contain information not primarily for statistical purposes, either private or public. My definition of the noun 'survey' includes research that is designed and based on statistics from such sources. Hence, an AD survey or integration survey lacks purpose-built questionnaires, and its original data-acquisition instruments are outside the full control of statistical offices and researchers.

Methodology research for statistics mainly using AD has picked up pace and this special issue demonstrates this fact. One reason is increased worldwide interest in using AD in population censuses. In the last European census, some countries moved away from a traditional census. Others, such as the United Kingdom, New Zealand, the United States and Canada have ongoing census modernisation programs containing significant efforts to investigate the use of AD. However, this interest is not completely new. Scheuren (1999) and the references therein illustrate that it was on the agenda in the US as far back as in the 1980s. Another reason may be that the geographical spread and collaboration between National Statistical Organizations (NSOs) and academia have created a critical mass. Not too long ago, methodological work on AD were restricted to fragments inside NSOs, and in the field of social statistics it was practiced mainly by the Nordic countries, the Netherlands and Slovenia (e.g., see Nordbotten 1966; UNECE 2007; Schulte-Nordholt et al. 2004; Zaletel and Krizman 2008). It is therefore pleasant to see the mix of countries represented in this issue.

If I ignore AD used as auxiliary information in the design and estimation of sample surveys, my personal experience with AD goes back approximately twelve years. During this time I worked with AD methods in business statistics, social statistics, and a register-based census, as well as trying to facilitate an organisational view to improve the use of registers and AD in a national statistics production system. I will reflect on this period and provide some ideas about AD and statistics that have become 'food for thought' after reading these articles, and which (in my opinion) need attention.

My discussion will not focus on the articles' details, but instead make a note of their fit with NSO activities, bearing in mind that NSOs today not only make statistics, some also provide microdata for researchers as a part of their countries' data infrastructure. These infrastructures, which consist largely of AD, are significant and contribute to unlocking the

[1] Statistics New Zealand, PO Box 2922, Wellington 6140, New Zealand. Email: anders.holmberg@stats.govt.nz

value of data – in a safe and trusted way. NSOs have a great opportunity to combine infrastructures for microdata with modernised statistics systems. As public and private AD sources grow, it is vital to align the production systems of official statistics with these infrastructures, with new statistics applications, and with the development of statistical methods. This JOS issue deals with some methodological challenges that follow, namely coverage, linking methods, and subsequent estimation. The estimation techniques proposed will mean that statistical modelling and computer-intensive methods must increase in use. I intend to discuss some points about the opportunity (and challenge) facing NSOs based upon my experience of AD from statistical offices in Sweden, Bolivia, Cambodia, and New Zealand.

2. Data Integration and AD Surveys

As 'flagships' such as population censuses radically change design, it is becoming clearer that the field of survey design is gradually shifting. It is moving from (albeit complex) sample surveys to surveys based on integrated data with AD as a backbone. NSOs that realise and adapt to this change face both an opportunity and a challenge. The opportunity is to use their responsibility and participation to build national data infrastructures and create production environments enhancing integrated survey statistics. As foreseen by Nordbotten and Scheuren, this is cost efficient from a societal perspective and complements sample survey programs by delivering broader, more detailed, and more responsive subject-matter contents. A production environment for integrated data and multiple source statistics would also enable NSOs to play an active role shaping new and alternative data sources and collection methods. Their main challenge is to align the workforce and the production processes.

For NSOs, this means that an end-to-end statistics production process will rely more on data streams with different origins. Production environments must also be able to effectively and efficiently exploit the possibilities of data integration. When doing this, it is necessary to have secure and well-designed IT systems for storage, processing and access, but sound methods are even more essential. NSOs that try to modernise their end-to-end processes with little or no thought to survey designs for data integration risk making bad investments in inadequate IT structures.

2.1. Statistical Modelling and Validation Efforts Will Increase

My first encounters with statistics that relied solely on AD were in business statistics through projects on improving timeliness and accuracy. These projects had only one main data source, which had only one specific use. The tasks therefore resembled those of improving a single-purpose sample survey. Despite the main goal of improving timeliness, the projects spent little effort on data acquisition processes. Instead, the focus was on developing estimation techniques that could provide *rapid (preliminary) estimates* that were robust against bias caused by measurement errors and missing units. Just as in this special issue, statistical modelling played a crucial part.

Six of the articles in this special issue present estimation techniques based on statistical modelling. Five of them (Zhang, Gerritse et al., Chipperfield and Chambers (C&C), Yildiz and Smith (Y&S) and Di Consiglio and Tuoto (D&T)) discuss log-linear models, and one

article, Bryant and Graham (B&G), discusses Bayesian techniques. This is not surprising and it is safe to predict that if more AD is used, all forms of statistical models will play a greater role in official statistics. For NSOs, the challenge would be to explain to users the necessity of models and their impact on statistics quality, particularly when there are many model variants to choose from and different statisticians to trust. It is important to pursue ways of validating model assumptions and estimating errors caused by their violation. I refer to Gerritse et al. as a valuable contribution in this respect. Because of higher recurrence, my experience is that it is easier to validate models in economic statistics than in social applications. One of the abovementioned projects was carried out on monthly statistics, and before introducing a new method we monitored patterns of incoming data over several rounds. We were able to repeatedly compare preliminary estimates based on incomplete data with corresponding final estimates and thereby empirically check competing estimation models (Jäder and Holmberg 2005). This method is not practical with less frequent data collections and definitely not for models proposed for a census. In this case, other validation methods are necessary that may incorporate extra data collections and/or experiments and add to cost. Y&S and B&G give two very different census estimation methods using models. NSOs considering these should look for ways to compare them, which is not straightforward.

2.2. *Linking and Microdata Access*

It is typical for modern statistics using AD to reuse data through integrating and combining different sources. I first came across multiple uses and integration when I worked with Statistics Sweden's Microdata ONline Access system for researchers (MONA). This system contains primarily personal data and has a design that is far more *ad hoc* than the data archives solution advocated by Nordbotten (1966). In MONA, personal identification numbers are available and they provide unique unit record identifiers, which make data integration and high-quality record linkage easy.

Internationally this is unusual – in many environments record linking is a major undertaking that requires significant methodological effort. C&C, D&T and Blackwell et al. illustrate this with different linking aspects. The first two authors present estimation methods in the presence of imperfect linking. Blackwell et al. illustrate the complexities and practical barriers that exist in a big project, such as linking census data with AD. Because of varying circumstances, it is probably unwise to copy Blackwell et al.'s approach exactly. However, the article shows a range of necessary steps and available possibilities by mixing exact/deterministic matching with probabilistic and clerical routines. All this is done to maximise linking rates with as few errors as possible.

Describing the size of the linkage error and compensating for it is indeed a methodological task. Estimates of the true positive rates (the sensitivity) and true negative rates (the specificity) should routinely accompany any linked data. Still, the set of negative links rarely gets the attention it deserves. It is worth looking closely at the records that do not link. This should give good insights into AD patterns, as the false negatives (whenever detected in reviews) are similar to studying the attributes of nonrespondents in a sample survey. The true negatives may reveal other deficiencies in the AD sources – coverage is one of them.

C&C, the references therein, and to some extent D&T, present methods for handling the effect of a certain type of linkage error. Demand for using these methods will increase as a result of NSOs creating research analysis infrastructures with linked microdata. Statistics New Zealand's Integrated Data Infrastructure (IDI) is one of these interesting environments under development. It allows for statistical outputs and research on the transitions and outcomes of *people* through various areas. With a conscious approach to confidentiality and security, the IDI provides analysts with microdata that sometimes are the result of linking multiple datasets. On top of the abovementioned quality traits for linking, transitivity is then introduced as another concern. Blackwell et al. have only one AD source, but NSOs that might, for coverage reasons, want to combine multiple AD sources before linking should study transitivity effects (Sadinle and Fienberg 2013).

2.3. *Coverage and Statistical Units in Production Environments for Integrated Data*

The raw records of many AD sources in MONA and IDI are based on registered events, or (if there is no terminating event) a relation between entities, for example employer/employee, hospital/patient, school/pupil. The records are usually transformed into units of interest such as persons, but sometimes, depending on the purpose, they are kept in their original form as records of employment, treatment, course enrolment and so on. Zhang (2012) uses base units and composite units as a way of understanding the quality properties of integrated AD. This is a useful distinction in studying the interplay between coverage issues and linking, since coverage is defined by the target unit and that unit is not necessarily the linking unit. Linkage errors have a direct effect on coverage, whether the linking unit is the target unit or not.

At the integration/linking stage, reasons other than linkage errors can influence coverage. Zhang's model introduces an alignment stage to sort the relations between base units and composite units in integrated data. It also introduces identification errors and unit errors that are conceptually different but where the effects are similar to those of coverage errors. Burger et al. treat this when they study the effects of setting a single industry code for a composite unit, such as an enterprise unit, when its LKAUs (Local Kind of Activity Unit) have different industry codes. The Swedish register-based census is another recent example where coverage problems arise because of unit errors in integrated data. The post-census evaluation survey indicates that the register-based sources for the census underestimate the number of one- and two-person households and overestimate the number of households with six or more members. Since person coverage is good, the overall effect is an underestimation of the total number of households by 4-5 percent depending on domain (see Andersson et al. 2013). Hence, good coverage of the base unit (person) does not mean good coverage of the composite (household) unit. With access to a greater variety of data sources containing different unit types, NSOs need good functions to handle coverage errors and other problems arising from the integration stage.

I think a flexible and cohesive system for data integration is easier to achieve if the statistical business architecture is built around appropriate base and composite units. Most statistics about society have units related to land, people, or business. In these three spheres, AD is usually available from the public sector. Hence, with legal access or even custodianship of such core AD, the NSOs have better opportunities than others to sort

appropriate statistical units, to standardise the units and to build good infrastructures for multiple-source statistics with such units as a backbone.

Figure 1 shows a simple unit-centric structure with relations between important statistical units in the subject spheres of land, people and business. Complemented with methods for treating the units' time and geographical dimensions, it is a foundation for defining and accessing target units and for applying data-integration methods from a statistical system's perspective. The keystone units in the illustration are base as well as composite units, and in the case of dwellings are both, depending on the statistics question at hand. Dwelling unit is included here to show a unit that establishes a connection between people and land through household/housing statistics. Otherwise each sphere can be expanded and has a set of units not shown here for simplicity. (For example, in a detailed picture the business sphere would have Kind of Activity Units (KAU), local KAUs and legal units – and, if it helps, enterprise groups. The land sphere would have building and entrance units and the people sphere would have household and family units.) In a system structure for integrated data, the geographical attributes in the middle are very important. They are central to the integration apparatus (especially without well-established identifiers) and should not be used only for statistical collection and dissemination processes. Also, by expanding the unit-centric structure below it is also fairly straightforward to put context to and interpret event/activity records as relations between units. A lot of useful AD statistics are based on such data.

Storage, access, and maintenance of the unit data can be done in statistical registers, as described by Wallgren and Wallgren (2007). This can also be done in other ways, for example a system of unit frames which are tied together by a linking methodology and effective data processing capabilities. The unit-centric approach facilitates the development of an environment that can integrate data quickly in a standard, transparent, and interpretable way. A huge benefit is that it enables assessments of various target and accessible survey populations. It also simplifies the interlinking of different subject-matter areas and makes it easier to assess which data are best in a multisource choice situation. The populations in turn can be national benchmarks with well-known coverage properties

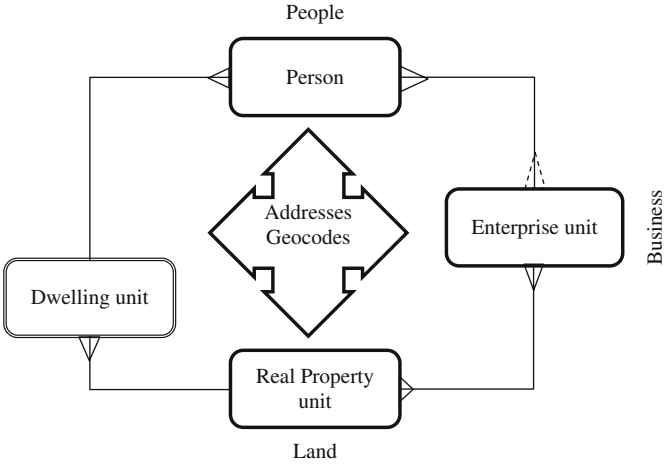

Fig. 1. A unit-centric statistical structure for integrated data surveys

to be used by many in comparative studies as well as official statistics. They can also give meaning to pointless statements against the sampling paradigm such as "...gathering as much as possible, and if feasible, getting everything: N = all" found in the big data literature (Mayer-Schönberger and Cukier 2013, 29). Ultimately, without unit understanding and a sought population it is hard to evaluate what is meant by "all"; surely there are cases when you get more than all. With methodological know-how, NSOs can make sense of integrated data by putting them in context, explaining coverage after linking, and perhaps also improving the quality of AD systems.

3. Development Areas for Integrated Surveys and AD Systems

In this section I highlight some other development areas I considered while reading this special issue and while thinking about how NSOs work with AD.

3.1. Expand the Methods Toolbox Using Geographical AD

Developing a structure such as Figure 1 means we must pay more attention to geographical AD and the location concept. While many NSOs are good at conforming to the geodata evolution when they disseminate statistics, it is still more or less uncharted territory for methodologists designing surveys or working generally with AD. There has been progress in the traditional use of AD, such as standardised solutions to communicate with GIS systems and map the hierarchies of areas relating to national and local administrative geographies, but NSOs seem slow to take up new statistical methods with geospatial data.

I suspect that soon we will see more integration surveys based on geographical linking. These surveys will be based not only on addresses (which require substantial cleaning efforts) but also on geocodes, clusters of geocodes using geohashes and 'snap-to-grid' methods (Heath and Goodwin 2011). Naturally, geographical linking requires good geocoding practices when the AD are created. This already exists partly, both in public and private data, but NSOs should be ready to take advantage of this and regularly add geocodes to their own data collections. This enables easier and more reliable linking between units of different types. It also allows the creation of new types of geodependent composite units.

To give an example, in Sweden practically all electricity meters are geocoded for reasons of repair and reading. The meters are connected to dwellings rather than buildings. By using a geocode link (as one example among others), the chance to infer dwelling occupancy based on electricity consumption is good. This is an interesting option in population censuses and housing statistics. With slight adaption, the ideas in this special issue should be applicable for errors using geographical linkage; the linking articles C&C and D&T are particularly interesting, as is Burger et al.'s contribution. As far as I understand, the classification/coverage problem they treat can also be adapted to composite 'proxy' units linked together by geography. Sometimes you want to classify aggregated composite units (e.g., geographically linked groups of buildings, dwellings or households) that have diverging information on the base-unit level. In another setting, Burger et al.'s approach may also clarify the sensitivity of classification errors on association measures applied on geodependent composite units. Linking composite units is also possible when a base unit option is hard to get or not allowed because of legal constraints.

3.2. *Examine Time Dimension in AD Systems and Analyse Events and Delayed Data*

Time dimension is a critical factor for the coverage and linking of AD. In some sources it can be tricky to distinguish between reference dates and registration dates. There also must be operational solutions for how to relate the data to time, for example the usual residence at a single point in time. Since many AD sources have records that are events or relations and since storage and processing systems often are poorly designed regarding the *statistical* units, studies on units' status-change frequency are rare. A lot can be learned about an AD source by consciously monitoring and analysing unit changes. Changes are not only signals of underlying societal and population changes; they can also be signs of alterations of administrative routines in the source. Moreover, provided that historic or change data are kept, some of the AD retained by NSOs have longitudinal information waiting to be unearthed by computer-intensive pattern recognition methods.

Event or delayed data are also potential sources that can help us to understand how coverage evolves over time. It is not unusual for delayed data phenomena to appear in recurrent business surveys as a survey feedback issue. Often the recommendation is to ignore the information since it introduces estimation bias. However, this practice also neglects coverage errors and the trade-offs are not always straightforward. Delays can sometimes also prevent accurate linking.

Other NSO activities can also benefit from event data in AD. Every day the Swedish population register gets updates on events such as address changes, changes in marriage status, births and deaths. If changes (e.g., moving house, divorce) make people harder to reach, it makes sense to transfer or at least compare this information with that from surveys doing collection and estimation. With survey designs using direct element sampling and a mixed-mode mail and web or CATI collection, this may reduce a significant part of the nonresponse due to no established contact, or reduce bias in calibration estimation.

3.3. *Measuring Coverage, Coverage Targets and Estimation*

This special issue and the work in the Beyond 2011 program (run by the Office of National Statistics to investigate alternative census possibilities in England and Wales, see ONS 2013; Skinner et al. 2013) reveal a focus shift in viewing census coverage. Undercoverage is the most serious issue in a traditional census, and post-census surveys are designed to deal with this, usually through area sample designs that are independent of census collection. However, in a census based on AD, both overcoverage and undercoverage seem likely. These are not expected to be evenly distributed. On the contrary, just cross-examining AD over geography, sex and age is likely to produce complicated patterns of included and excluded units. Therefore it might prove difficult to estimate the extent of both types of coverage errors efficiently using one single survey. The underlying AD mechanisms of the coverage problems can be very different, which is well illustrated by the data in Gerritse et al. In that context it makes sense to view post-census activities as a package of actions with maybe more than one data collection. The practices around post-census data collections and their implications on estimation methods need to be updated, and the solutions are connected to the choice of a dual-system estimation method or perhaps even triple-system estimation as discussed by Griffin (2014). The independence

assumption between sources is highlighted by Zhang and Gerritse et al., and is also considered by Y&S and B&G.

The coverage issue in AD arises when sources are used for statistical purposes. It is the obvious cause of error to study when considering AD because its effect is easily visible when simple estimation/calculation techniques are used. All articles in this special issue address how to minimise or adjust for coverage error. For NSOs, this raises resource use as another related question. Is the coverage issue the biggest one when considering AD? Should one accept no less than close to 100 percent coverage before even considering AD, or can one settle for less and combine AD with sampling techniques and modelling? Although essentially an estimation problem, censuses seem to have a 100 percent coverage target. While this is hard to achieve, it seems reasonable for legislative reasons and because of the census's importance for other social surveys. The trend seems to be that a combination of a traditional area-based frame field collection and AD sources is the choice for achieving this target. The AD source can compensate for undercoverage in field collection if the same people who are hard to reach are present in AD (e.g., through welfare-seeking systems). Alternatively, a field collection can be used in areas where it is believed the AD is poor.

In the business sphere there is a trade-off and often a good reason not to aim for 100 percent unit coverage. This is certainly the case in developing countries, but also applies elsewhere; it would be costly to keep the coverage of small home-based 'household' businesses up to date. To compensate for the undercoverage, other methods involving modelling and household surveys are needed.

3.4. Administrative Data in Developing Countries

In developing countries, the AD systems' maintenance and the contents coordination of planned and made investments are big barriers to using AD for statistics. Coverage issues are a result of these problems, not just in the sources themselves, but also in area frame-based sample surveys and in census practices. The ties between undertakings in population, agricultural and economic censuses and national AD systems are often weak or just occasional. The optimistic view of this is that developing countries may take advantage of 'leapfrogging' and develop AD system structures that facilitate standardisation and multiple uses (including statistics) from the start. Another good point is that a lot can be improved with relatively small means. To give two examples (among many): the first would be to add and enable geocodes in AD, censuses, and surveys. Adding coordinates to units (such as villages) in official databases would greatly improve the quality and simplify the updating of sampling frames and linking possibilities. Lack of harmonisation adds unnecessary burden in studies that combine several sources (e.g., Haslett et al. 2013). Simple actions like creating a standard geocode option for linking would free up analytic resources tied up in data cleaning. The second example would be to establish a statistical business register. By separating the concept of a statistical business register from that of an administrative business register, one can apply methods that achieve better alignment with the needs of national accounts and economic statistics (Wallgren and Wallgren 2007). The actions needed are country specific, but there are good and generic principles to follow. The African Development Bank's report

(ADB 2014) provides relatively exhaustive guidelines for a statistical business register. The guidelines are applicable outside Africa.

Although this special issue does not explicitly refer to developing country problems, the articles are still relevant, as coverage error is the of most concern statistical problem with AD. Some of the articles might be too advanced, but local NSO experts together with external consultants can benefit. In particular, the articles by Blackwell et al. and Burger et al. are good examples for countries such as those in Communidad Andina in South America. These countries have a number of AD sources for land, people, and business already in place and they are working on structures to use them for statistics.

4. A Final Note

Finally, I would like to congratulate the authors, guest editors, and the editors of JOS. Although there is plenty of literature about statistics and AD, a lot of it lacks the rigour that follows from a journal review process. A themed issue on administrative data is timely. With census transformation projects as a major driver, and as the area progresses further with theory meeting practice and vice versa, the future is likely to see a higher proportion of articles about AD methodology. It is an elusive thought (sometimes nursed at NSOs) that statistics based on AD is less complex. Because of society's growing appetite for data, methodologists are looking more closely at previously overlooked areas, and as I stress again the need to integrate data, many questions still need to be answered.

I predict a big increase of papers about AD and statistics, especially studies on the magnitude of error. Are the coverage errors large compared with other error types? What is the difference between having one controlled survey and having multiple data sources, and how is it addressed in terms of total survey design? Which configuration of data sources is the best? Are we sometimes making things worse when data sources are combined?

In particular, I believe the measurement properties of AD compared with those of survey data will be scrutinised. Several studies may conclude that direct collection and sample surveys are needed to adjust and/or guarantee statistical quality. Hence, even if there is a change in paradigm with the death of the sample survey as the *first* option to acquire data, this does not mean that a sample survey sometimes may not be the best option. Regardless of source configuration between AD and self-collected data, the choice depends on trade-offs between error types and cost. From a methodology perspective, it makes sense to further bridge the gap between the survey sampling tradition and the use of AD methods. In doing this, I do not believe that it helps to claim that the methodology applied to AD is a completely distinct and new area. Instead, a holistic view of the survey process is better, identifying where the methodological focus should lie, and when old methods are applicable or new ones need to be developed. NSOs that align their work with methods and their production environment with survey designs backing integrated data are better insured for the future.

5. References

African Development Bank. 2014. *Guidelines for Building Statistical Business Registers in Africa – Laying the Foundations for Harmonization of Economic Statistics Program*. Tunisia: Statistical capacity building division, Statistics Department

African Development Bank. Available at: http://www.afdb.org/en/knowledge/publications/guidelines-for-building-statistical-business-registers-in-africa/ (accessed 14 July 2014).

Andersson, C., A. Holmberg, I. Jansson, K. Lindgren, and P. Werner. 2013. "Methodological Experiences from a Register-Based Census." 2013 Joint Statistical Meetings, Montreal, 3–8 August 2013. http://www.amstat.org/sections/SRMS/Proceedings/y2013/Files/309549_82867.pdf (accessed July, 2015).

Griffin, R. 2014. "Potential Uses of Administrative Records for Triple System Modeling for Estimation of Census Coverage Error in 2020." *Journal of Official Statistics* 30: 177–189. Doi: http://dx.doi.org/10.2478/jos-2014-0012.

Haslett, S., G. Jones, and A. Sefton. 2013. *Small-Area Estimation of Poverty and Malnutrition in Cambodia*. National Institute of Statistics, Ministry of Planning, Royal Government of Cambodia and the United Nations World Food Program. Available at: http://www.wfp.org/content/cambodia-small-area-estimation-poverty-and-malnutrition-april-2013 (accessed 14 July 2014).

Heath, T. and J. Goodwin. 2011. "Linking Geographical Data for Government and Consumer Applications." In *Linking Government Data*, edited by D. Wood, 73–92. New York: Springer.

Jäder, A. and A. Holmberg. 2005. *The Growth Rate Method for Production of Rapid Estimates of the Swedish Foreign Trade. Background facts on Economic Statistics* 2005:9. Statistics Sweden. Available at: http://www.scb.se/statistik/_publikationer/OV9999_2005A01_BR_X100ST0509.pdf (accessed 28 July 2014).

Mayer-Schönberger, V. and K. Cukier. 2013. *Big Data – A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.

Nordbotten, S. 1966. "A Statistical File System." *Statistisk Tidskrift* 4: 99–103. Available at: http://www.nordbotten.com/frame.html (accessed 14 July 2014).

Office of National Statistics (ONS). 2013. *Beyond 2011: Producing Population Estimates Using Administrative Data*. Available at: http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/reports-and-publications/index.html (accessed 14 July 2014).

Sadinle, M. and S.E. Fienberg. 2013. "Generalized Fellegi–Sunter Framework for Multiple Record Linkage With Application to Homicide Record Systems." *Journal of the American Statistical Association* 108: 385–397. Doi: http://dx.doi.org/10.1080/01621459.2012.757231.

Scheuren, F. 1999. "Administrative Records and Census Taking." *Survey Methodology* 25: 151–160.

Schulte-Nordholt, E., M. Hartgers, and R. Gircour. 2004. *The Dutch Virtual Census of 2001. Analysis and methodology*. Voorburg/Heerlen. Available at: http://www.cbs.nl/en-GB/menu/themas/dossiers/volkstellingen/publicaties/publicaties/archief/2005/2001-b57-e-pub.htm (accessed 14 July 2014).

Skinner, C., J. Hollis, and M. Murphy. 2013. Beyond 2011*: Independent Review of Methodology*. Office of National Statistics. Available at: http://www.ons.gov.uk/ons/about-ons/who-ons-are/programmes-and-projects/beyond-2011/beyond-2011–independent-review-of-methodolgy/index.html (accessed 21 July 2014).

United Nations Economic Commission for Europe (UNECE). 2007. *Register-Based Statistics in the Nordic Countries. Review of Best Practices With Focus on Population and Social Statistics*. United Nations Publication. Available at: http://www.unece.org/stats/publications/Register_based_statistics_in_Nordic_countries.pdf (accessed 14 July 2014).

United Nations Economic Commission for Europe (UNECE). 2011. *Using Administrative and Secondary Sources for Official Statistics: A Handbook of Principles and Practices*. United Nations Publication. Available at: http://www.unece.org/fileadmin/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf (accessed 14 July 2014).

Zaletel, M. and I. Krizman. 2008. "The Hidden Side of a Successful Story – Implication of Wide Use of Administrative Data Sources at National Statistical Institutes." Paper presented at the IAOS conference, Shanghai 14–16 October 2008.

Zhang, L.C. 2012. "Topics of Statistical Theory for Register-Based Statistics and Data Integration." *Statistica Neerlandica* 66: 41–63. Doi: http://dx.doi.org/10.1111/j.1467-9574.2011.00508.x.

Wallgren, A. and B. Wallgren. 2007. *Register-Based Statistics Administrative Data for Statistical Purposes*. Chichester: John Wiley & Sons.