Census Transformation and the Future of Population Statistics

James J. Brown[1] (ID) and James Chipperfield[2] (ID)

Keywords

population estimates, census innovation, administrative data

1. Context

Throughout the world, census remains the bedrock of a nation's population statistics. As stated by UN Statistics Division under their World Population and Housing Census Programme:

> The Programme recognizes population and housing censuses as one of the primary sources of data needed for formulating, implementing and monitoring policies and programmes aimed at inclusive socioeconomic development and environmental sustainability. It further recognizes population and housing censuses as an important source for supplying disaggregated data …

> (https://unstats.un.org/unsd/demographic-social/census/)

But the approach to census is changing and has changed over the past two decades. Valente (2019) reported that two-thirds of European countries in the 2020 census round planned a non-traditional approach to their data collection, utilizing either registers only or a combination of register with direct collection. It was less than half in the 2010 round. Within countries that are still undertaking a full direct

[1]School of Mathematical and Physical Sciences, University of Technology Sydney, Broadway, NSW, Australia
[2]Australian Bureau of Statistics, Belconnen, ACT, Australia

Corresponding author:
James J. Brown, School of Mathematical and Physical Sciences, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia.
Email: james.brown@uts.edu.au

enumeration of their population, this process has been updated to make use of new data sources and approaches. Administrative data and satellite data are used to create address lists, rather a fieldwork listing; households are invited to complete their census form online; administrative data are utilized during processing to improve the quality of imputation for item non-response and non-responding households (e.g., using electricity data to help discriminate between dwellings that are occupied, and potentially non-responding, and unoccupied dwellings).

To think about the current context, and future opportunities, it can be useful to recognize that the full census of population and housing has two objectives. First it provides the basic geographic structure of the population in terms of core demographics and household structures available for all (small) geographic areas. In the final outputs this is typically on a usual residency base, where an individual or household is defined to usually reside, even if the actual collection is based on location on census night. Second it provides a rich set of attributes beyond the core demographics covering education, economic activity, basic health measures, as well as other measures on housing conditions and socio-economic status. Some countries, Canada for example, have a history of meeting the first objective via a short-form to all individuals and dwellings, while meeting the second objective via a contemporaneous long-form to a sample of dwellings. The American Community Survey splits the second objective from the decennial census collection by utilizing an ongoing data collection activity that combines across independent annual samples to produce the most granular outputs. The resulting situation is that countries sit in one of three broad categories for their census operations:

1. Full register census. In this set-up, population and housing registers not only provide the basic geographic structure of the population in terms of core demographics and household structures, the first objective; they provide the full richness of census attributes meeting the second objective.
2. Register census augmented with direct collection. This can be thought of as the traditional short-form long-form set-up, where the short-form has been replaced with registers. Population and housing registers provide the basic geographic structure of the population in terms of core demographics and household structures. These data sources are then added to via additional administrative data and direct data collections that may occur at a specific time-point, or built-up from ongoing collections. The current census methodology in Netherlands is an example where there is an increasing use of administrative data but survey data is still integrated for some of the attributes (Stolk 2024).
3. Direct collection census. In this set-up, a country still conducts a point-in-time data collection (or combines across annual point-in-time collections under the rolling census approach) to provide the basic geographic structure of the population in terms of core demographics and household structures, and meet the first objective. In many cases, the full richness of census attributes covering education, economic activity, basic health, as well as other topics, occurs within the same data collection process as well. That can be

to a sample of households/individuals (Canada for example) or to all households/individuals (Australia for example). However, in other cases the data under objective 2 can come from ongoing collections, the American Community Survey being an example. Administrative attributes are being increasingly linked to censuses to either increase the number of attributes or to replace the need for the attributes to be collected by the census (Statistics New Zealand 2019).

For countries in each broad category, how can they continue to innovate? What are the opportunities and barriers to meeting objective 1 for countries moving away from direct collection?

In thinking about innovation in census, it is important to recognize the role census plays goes beyond providing detailed information for small sub-populations and small geographies at a single point in time. In many countries it also plays the role of anchoring ongoing population estimates, which roll-forward from the most recent census using the demographic accounting equation; add on the births, subtract the deaths, adjust for migration. Births and deaths data come from administrative sources; while migration data might be implicit in additions and subtractions from registers, estimated using flows from a variety of administrative sources, or rely on entry and exit surveys.

2. Changing Data Landscape

National Statistics Organisations (NSOs) find themselves at different points on this journey but it is clear there is a revolution happening in our data landscape. No longer are NSOs the only providers of key demographic, social, and economic data. Other organizations are leveraging their own data to provide often more timely data insights than the NSO, but without the same quality guardrails. An example are the indicators of household expenditure patterns, such as CommBank's Household Spending Insights in Australia where the website states:

> … based on real behaviour and transactions, leveraging approximately 7 million retail customers to understand consumer spending patterns and behaviours.

> https://www.commbank.com.au/business/latest/spending-intentions.html

But while 7 million is indeed impressive, it is not necessarily an unbiased picture of all household and individual expenditure that an NSO would wish to represent in an official statistic. However, the data is there and so the role of the NSO is becoming one of a trusted data aggregator of public and private sector data rather than an agency undertaking bespoke collections. This revolution is happening in a social context around public acceptability of the use of their data.

In this context, NSOs looking to innovate and move-up from relying on a direct collection census to utilizing existing data sources, will be looking to pull-in data

from a wide range of providers. In the first instance, as has been the tradition for those NSOs already using registers, that will include government data providers but will need to extend to non-government providers. For those NSOs already utilizing a full register-based approach, additional sources will enable enhancing the attributes it provides. Smart meter data for electricity consumption would allow analysis of usage by demographics at small geographic scale enhancing understanding of household usage patterns. Linking to mobile phone data will allow for a more dynamic picture of population flows throughout the day/week, beyond the static picture of usual residency at a point in time.

3. Methodological Challenges

When there exists an official population register, even if this is not perfect, it provides a basis for defining the usual resident population. In the absence of a formal register, the NSO needs to construct a starting point by combining across administrative sources that capture individual's interactions across government. The PECADO approach laid-out in Dunne and Zhang (2024) outlines this for Ireland but the approach is similar for other countries. The outcome is a list of potential usual residents based on interactions (so called "signs of life") with, or presence on, government administrative systems.

Crucially, in the case of Ireland, there has been an evolution to add a unique identifier to government data sources, and this removes the first challenge; linkage errors. However, in many countries there is no unique identifier available to NSOs to link across health registration data, tax returns, and social benefits data. As a result, record linkage techniques utilize name data (sometimes in an encrypted form) along with other information around date-of-birth, gender, country of birth, and geography to link records belonging to the same person. There has been considerable work in the development and implementation of linking methodologies (Chipperfield et al. 2018). While linkage errors remain a major concern, we are now in a position where we can understand their scale, and potentially adjust for them (see Chipperfield 2020 and Chambers and Diniz da Silva 2020). Post their 2018 Census, Statistics New Zealand used their administrative-based population list to enhance the census with additional responses and the estimation adjusted for linkage errors (Bycroft and Matheson-Dunning 2020).

Creating a population list from linking multiple sources in place of a formal register that is designed to list the population is a crucial first step. As yet no NSO has achieved this with the quality required for a full census-replacement product, although as mentioned above it has been used to enhance census response. Once over the linkage issues, the NSO faces the issues of coverage errors. With direct collection census, under-coverage is often the major focus and for those countries continuing with direct collection that is not changing. Luiten et al. (2020) outline the declining response rates that NSOs are grappling with across their social collections. However, with a move to a list created through linking data sources (as in register-based systems), over-coverage becomes the dominant coverage error. Over-coverage is exacerbated by False Negatives in linkage of the data sources, as

one person may be presented by two or more unlinked records on the combined register.

Over-coverage comes in two guises. First, the erroneous inclusion of individuals that do not belong to the population typically leading to gross over-coverage at a national level. Second, the incorrect placement of individuals into local geographic areas, where over-coverage in the incorrect location nets-out with under-coverage in the correct location at the national level. With lists constructed from administrative or register systems over-coverage occurs due to individuals not updating address information, lag in removal of deaths, and international out migration not being captured when it occurs, but deduced by later interactions (or absence of interactions).

The E-sample approach by US Census Bureau, which goes back to 1990 Census (Hogan 1993), is designed to tackle both guises. However, being sure a record in the E sample is erroneous as opposed to a non-response is difficult. With a starting list created through combining across government administrative sources there may be legal barriers to sampling from the list and attempting to check whether individuals exist in the field. Even without legal barriers, the social license for checking individuals' administrative records is dubious. In place of a field reconciliation, NSOs are approaching this by leveraging the transactional data in administrative systems and using signs-of-life rules to classify records as erroneous (see Chipperfield and Zhang 2025). In essence, removing records that have little evidence of interactions, thereby removing over-coverage, and using estimation approaches such as trimmed dual-system estimation to estimate the under-coverage (Dunne and Zhang 2024).

Tackling over-coverage due to incorrect locations is already integrated into coverage assessment tools, the Australia estimator being one example (Chipperfield et al. 2017). However, what is more complex is that combining across sources does not necessarily lead to a consistent answer for either geography or characteristics. This leads to the idea that in more dynamic populations there may no longer be simple answers. Fractional counting (Bernardini et al. 2022) weights an individual to multiple locations based on their likelihood to be at a particular location, but it implicitly allows for the possibility that individuals contribute to more than one location. Latent class approaches are unlocking the idea that the true concept sits behind the differing measurements seen across systems. This has been implemented by Van der Heijden et al. (2022) in the context of identifying Maori in New Zealand. Building on the dynamic nature, the dynamic population model approach Office for National Statistics (2023) sees the population as a realization of a demographic accounting equation based on several data sources, where the net estimated error associated with the sources is removed from the totals. Systems like this start to move from the concept of a point-in-time census with ongoing population estimation to an integrated system that updates more regularly.

In the current climate, we cannot finish without considering the potential of AI/ML tools as NSOs innovate. In countries pushing forward with direct collection, the push for online collection opens the opportunity for live coding of text responses; and even with traditional paper-based collection there is the opportunity

for increased automated coding. Editing and imputation for data inconsistencies and missing items will be an area of innovation, and provides an alternative for predicting responses when inconsistencies exist across sources. Utilizing the tools to identify priority edits is already an area of work, with recent work by Forteza and García-Uribe (2025) focusing on classifying cases with the most serious errors or omissions. An issue with many administrative systems, and even more mature register-based systems, is that they operate for individuals and not households. While individuals might be attached to a common address, identifying households and their structures is not straightforward. Supervised ML tools have the potential to learn household structures, while unsupervised tools open the possibility of exploring new household structures that emerge from the input data structures of individuals sharing a common location. In other words, the opportunity for new ways to categorize as opposed to always fitting current approaches to new tools and data.

4. The Future

Innovation in population statistics with census at the center is a requirement if NSOs are to continue as the trusted source of data for government and business decision making. Wherever an NSO sits currently it will need to engage in less reliance on direct collection in the medium to long term. For many NSOs the innovation is needed more immediately with downward pressures on response rates to direct collection and on budgets. That move away from direct collection and one-point-in-time measurement opens the door to more dynamic systems of population statistics. Being dynamic enables a better reflection of society in the twenty-first Century, where traditional notions of usual residence are challenged and individuals are more fluid in how they may identify themselves across a range of attributes. There is always a risk when we embrace innovation, but increasingly NSOs will face more risk without innovation.

With an increase reliance on administrative data in censuses, we must manage the risk that people who do not have records on administrative data are less likely to be counted or are implicitly assumed to have the same characteristics of people who do. NSOs must develop methods that are robust against changes to who administrative data captures and what information it collects.

ORCID iDs

James J. Brown (iD) https://orcid.org/0000-0002-7535-2874
James Chipperfield (iD) https://orcid.org/0000-0002-4673-3678

References

Bernardini, A., J. Brown, J. Chipperfield, C. Bycroft, A. Chieppa, N. Cibella, G. Dunnet, M. F. Hawkes, A. Hleihel, E. C. Law, D. Ward, and L.-C. Zhang. 2022. "Evolution of the Person Census and the Estimation of Population Counts in New Zealand, United Kingdom, Italy and Israel." *Statistical Journal of the IAOS* 38: 1221–37. DOI: https://doi.org/10.3233/SJI-220018.

Bycroft, C., and N. Matheson-Dunning. 2020. "Use of administrative records for non-response in the New Zealand 2018 Census." *Statistical Journal of the IAOS* 36: 107–116. DOI: https://dx.doi.org/10.3233/SJI-190591.

Chambers, R., and A. Diniz da Silva. 2020. "Improved Secondary Analysis of Linked Data: A Framework and an Illustration." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 183: 37–59. DOI: https://doi.org/10.1111/rssa.12477.

Chipperfield, J. 2020. "Bootstrap Inference Using Estimating Equations and Data That Are Linked with Complex Probabilistic Algorithms." *Statistics Neerlandica* 74: 96–111. DOI: https://doi.org/10.1111/stan.12189.

Chipperfield, J., J. Brown, and P. Bell. 2017. "Estimating the Count Error in the Australian Census." *Journal of Official Statistics* 33: 43–59. DOI: https://doi.org/10.1515/jos-2017-0003.

Chipperfield, J., N. Hansen, and P. Rossiter. 2018. "Estimating Precision and Recall in Probabilistic and Deterministic Linking." *International Statistical Review* 86: 219–36. DOI: https://doi.org/10.1111/insr.12246.

Chipperfield, J., and L.-C. Zhang. 2025. "Enumeration Errors of Census Imputation." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. Published electronically February 6 2025. DOI: https://doi.org/10.1093/jrsssa/qnaf005.

Dunne, J., and L.-C. Zhang. 2024. "A System of Population Estimates Compiled from Administrative Data Only." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 187: 3–21. DOI: https://doi.org/10.1093/jrsssa/qnad065.

Forteza, N., and S. García-Uribe. 2025. "A Score Function to Prioritize Editing in Household Survey Data: A Machine Learning Approach." *Journal of Official Statistics* 41: 144–71. DOI: https://doi.org/10.1177/0282423X241309971.

Hogan, H. 1993. "The 1990 Post-Enumeration Survey: Operations and Results." *Journal of the American Statistical Association* 88: 1047–60. DOI: https://doi.org/10.1080/01621459.1993.10476374.

Luiten, A., J. Hox, and E. de Leeuw. 2020. "Survey Nonresponse Trends and Fieldwork Effort in the 21st Century: Results of an International Study Across Countries and Surveys." *Journal of Official Statistics* 36: 469–87. DOI: https://doi.org/10.2478/jos-2020-0025.

Office for National Statistics. 2023. "Dynamic Population Model, Improvements to Data Sources and Methodology: Local Authorities in England and Wales, Mid-2021 to Mid-2023." *Office for National Statistics.* https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/articles/dynamicpopulationmodelimprovementstodatasourcesandmethodology/localauthoritiesinenglandandwalesmid2021tomid2023 (accessed March 16, 2025).

Statistics New Zealand. 2019. "Data Sources, Editing, and Imputation in the 2018 Census." *Statistics New Zealand.* https://www.stats.govt.nz/assets/Uploads/Methods/Data-sources-editing-and-imputation-in-the-2018-Census/Data-sources-editing-and-imputation-in-the-2018-census.pdf (accessed March 16, 2025).

Stolk, G. 2024. "2021 Census Provides an Accurate Picture of Our Population." *CBS (Statistics Netherlands).* https://www.cbs.nl/en-gb/corporate/2024/16/2021-census-provides-an-accurate-picture-of-our-population (accessed March 16, 2025).

Valente, P. 2019. "Modernizing the Census in Europe: Traditional and New Methods for the 2020 Round." *N-IUSSP (Weekly IUSSP Magazine).* https://www.niussp.org/environment-and-development/modernizing-the-census-in-europe-traditional-and-new-methods-for-the-2020/ (accessed March 16, 2025).

Van der Heijden, P. G. M., M. Cruyff, P. A. Smith, C. Bycroft, P. Graham, and M. Matheson-Dunning. 2022. "Multiple System Estimation Using Covariates Having Missing Values and Measurement Error: Estimating the Size of the Māori Population in New Zealand." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 185: 156–77. DOI: https://doi.org/10.1111/rssa.12731.