sciendo

Journal of Official Statistics, Vol. 37, No. 3, 2021, pp. 533–541, http://dx.doi.org/10.2478/JOS-2021-0023

# Preface

## 1. Introduction to the Special Issue on Population Statistics for the 21st Century

Reliable population statistics are indispensable for many crucial areas of public policy and planning worldwide. The world these statistics measure is changing rapidly, with population processes gaining pace and acquiring new forms, such as the increasingly fluid mobility and migration. Novel technologies, data sources and analytical methods offer new, better opportunities to deal with the demographic challenges of the future. The role of official population statistics in addressing these challenges – from aiding humanitarian relief for victims of wars and environmental catastrophes, to ensuring sustainability of social security systems for the decades to come – is fundamental. At the same time, the uptake of many of the new methods and approaches for applications in official statistics remains limited.

The aim of this Special Issue of the Journal of Official Statistics on "Population statistics for the 21st century" is to present cutting-edge innovations in the methodology and applications relevant for official population statistics, in order to help the users of statistical products better address the challenges of the 21st century. The articles included in the special issue are clustered in three thematic areas:

1. estimation and forecasting of components of population change (fertility and mortality),
2. census data, their quality, new census forms and methods of estimation, as well as
3. the methods for sampling and small area estimation used in official population statistics.

Collectively, they offer insights into the recent methodological advancements in the population statistics, while bearing substantial promise of applicability in the context of official statistics.

In this Editorial, we first summarise the contents of the Special Issue (Section 2), broken down into the three thematic areas listed above. Subsequently, we offer a general reflection on the current gaps in official population statistics in terms of concepts, definitions, data and methods, and suggest some possible ways and avenues of filling these gaps (Section 3). We conclude by charting a few particularly promising pathways for the development of official population statistics throughout the 21st century.

## 2. Contents of the Special Issue

The Special Issue is comprised of ten articles and a Letter to the Editors, all of which share a strong methodological slant. In general, four articles related to the model-based estimation and forecasting of the components of population change deal with the dynamic aspects of official population statistics, by describing demographic *processes* and *events*, current and future. At the same time, four articles looking both at the census data and

related estimation methods, as well as two articles on sampling methods and small-area estimation procedures, focus on population *stocks – sizes* and *structures*, include a mix of design-based and model-based approaches. We present the three themes in this order, noting that even though their sequence in the Special Issue is to some extent arbitrary, the practical applications in official statistics require methodological advancements in all these areas. In addition, the Letter to the Editors (Lanzieri 2021) offers a cross-theme discussion of some of the conceptual and definitional aspects of measuring populations, highlighting both promising avenues and pitfalls to be avoided in the shaping of official population statistics in the coming decades.

### *2.1. Components of Population Change*

The first four articles in this Special Issue deal with fertility and mortality in the context of their role in the population estimates and projections. To start with, Folkman Gleditsch et al. (2021) offer a comprehensive overview of the way fertility assumptions are being treated and set in (mostly national level) population projections across Europe. The article reveals a diversity of approaches, typically based on a mix of formal modelling and expert input, but persisting in their use of variants to depict the future uncertainty, despite recent developments, for example in the area of stochastic forecasting. The conclusions of the article highlight important gaps in the current approaches, in terms of the methods used, but also communication, transparency, and coordination between different countries, calling for a greater extent of 'peer learning' and knowledge exchange between national and supra-national statistical agencies.

Turning to mortality, but still remaining at the national level, Hilton et al. (2021) offer a methodological extension of the concept of best-practice life expectancy (Oeppen and Vaupel 2002) to a dynamic *mortality frontier* for the whole range of the human lifespan. The authors propose a Bayesian method for estimating and forecasting the frontier, as well as country-specific deviations, based on non-parametric methods utilising general additive models. The approach seems particularly useful for carrying out multi-country mortality forecasts, such as those prepared by international organisations, by allowing long-term trends in the development of the frontier to be distinguished from the less predictable local deviations in individual countries and periods.

Mortality forecasting is also explored by Ševčíková and Raftery (2021), this time in the subnational setting. The authors propose a probabilistic extension to the Bayesian hierarchical model used in the UN World Population Prospects (Raftery et al. 2013, 2014), which is applied to over 400 regions in 29 countries, mainly from Europe. Particularly promising is a method scaling national to subnational-level mortality estimates, and assuming that the subnational trends in life expectancy follow simple autoregressive processes, AR(1). This method proves competitive when compared with standard benchmark approaches, and thanks to its simplicity, has a high potential of being applicable to the practice of official statistics in terms of producing subnational mortality forecasts or projections.

Remaining at the subnational level, Zhang et al. (2021) present a novel method for clustering fertility patterns, aimed at contributing to the production of regional population estimates and, by extension, forecasts. The authors use a Bayesian spatio-temporal mixed-

effects model for fertility, and propose a spatial functional approach for clustering. The application to Portuguese NUTS-3 regions provides strong empirical support for the usefulness of the proposed approach, which is additionally strengthened by comprehensive simulation studies, designed for a range of situations which may occur in official statistical practice.

### *2.2. Census Data and Their Quality*

The second group of articles deals specifically with census data, with particular focus on new census forms, and the assessment of census data quality, especially with regard to census coverage. Among this group of articles, Righi et al. (2021) report on the annual Permanent Population Census implemented in Italy since 2018, which augments a population register with a Population Coverage Survey. The methodological details of this approach are presented, with the sampling design and the extended version of dual system estimator, proposed by Nirel and Glickman (2009), applied to correct for the possible under- and over-coverage of the register-based census. The method allows for a reliable estimation of the sizes of the main socio-demographic strata, while improving the precision of the estimates under some conditions. In this way, the article offers valuable lessons for designing coverage surveys for administrative censuses, which become increasingly popular for gathering population-wide information without some of the efforts and costs associated with traditional censuses.

A similar topic is tackled in the second article in this group (Baffour et al. 2021). This article focuses on methods for dealing with unobserved heterogeneity across individuals present on multiple lists used for applying capture-recapture methods for estimating population sizes. The proposed solution, based on applying latent class analysis coupled with log-linear modelling, additionally allows for dealing with the possibility of dependence between the different lists, where the assumption of independence between probabilities of being included on different lists does not hold. The methods are illustrated using data from the US 1988 Census Dress Rehearsal.

The methodological aspects of capture-recapture methods are also explored by Zult et al. (2021), who deal with the problem of imperfect linkages between the different data sources, which is particularly important if unique keys, such as ID numbers, are not available for linking. The proposed solution extends the earlier special-case methods provided by such authors as Di Consiglio and Tuoto (2015, 2018) for two and three lists, to a general case, utilising the log-linear modelling approach with covariates. The proposed method is tested on four idealised simulation-based scenarios, the results of which allow for a detailed discussion of the possible advantages and limitations of this approach. For statistical agencies, this work is particularly relevant to measuring census coverage, including coverage of administrative data censuses.

The final article under the census theme, by Wazir and Goujon (2021), proposes a demographic method for assessing the quality and validity of census data in the absence of a post-enumeration survey, based on experience with the 2017 population census in Pakistan. The quality assessment is based on the propagation of populations from past censuses, augmented by information from a range of intercensal surveys. The main conclusion is that while the country-level census-based population estimates can be seen

as broadly valid, there are many problems with subnational estimates that would need addressing in order for the related statistics to be reliable. The results offer strong arguments for carrying out a carefully designed post-enumeration survey in Pakistan – a conclusion that is also likely to be valid for many other countries across the world.

### *2.3. Sampling and Small Area Estimation*

The third group of articles in this Special Issue deals with specific issues related to sampling, nonresponse, and small-area estimation. Boonstra et al. (2021) present a simulation-based study of the diagnostic measures for non-ignorable bias in non-probability samples. Even though the extent of this bias is unknown, certain indicators can provide very useful tools for estimating the extent of the resulting selection bias. The authors build on the earlier work by Nishimura et al. (2016) and Little et al. (2020), testing two such diagnostic statistics: the *standardized measure of unadjusted bias* (SMUB) and the *standardized measure of adjusted bias* (SMAB). The usefulness of these measures for assessing the extent of selection bias in non-probability samples for a range of assumptions on the extent of the non-ignorable bias, has potential implications for the practice of official statistics, especially given the trend towards including more non-sample-based sources (administrative records, digital traces, and so on) in statistical production.

Finally, Fúquene-Patiño et al. (2021) look at the problem of small sample estimation, applied to the share of households with emigrant members in the municipalities of Colombia. The proposed method is based on the model originally designed by Fay and Herriot (1979), estimated both under the frequentist and Bayesian paradigm, with the response variable (proportions of households with emigrants) based on Demographic and Health Survey data, augmented with census-based small-area covariates. The results are encouraging, with the estimated emigration levels agreeing with national-level aggregates, and reduced errors in comparison with the direct estimates. The proposed method – and more broadly, an approach to combining survey and census data with the help of small-area estimation techniques – has a potential especially in the context of developing countries, where administrative data at lower levels of geographic disaggregation may be incomplete or entirely absent.

## 3. Towards Population Statistics Fit for the 21st Century

The articles presented in this Special Issue offer very timely insights into some of the most recent advances in the methodology – and indirectly practice – of official population statistics. Still, given the breadth of the topic, the present collection is necessarily limited in scope to a few selected themes, and barely scratches the surface of the contemporary landscape of the methodological developments and outstanding needs in official population statistics. Many other methodological gaps remain, which would need to be filled in order to make the population statistics ready for the challenges of the 21st century, not least with respect to the United Nations' Sustainable Development Goals, the objectives related to the Global Compact for Migration, and other current and future international initiatives aimed at improving the health, welfare and prospects of people in different parts of the world (see UN Statistical Commission 2021 for the current statistical work programme of the United Nations).

The key gap here is the global inequality – as well as, and growing diversity – in the development of statistical systems in different countries. One key aspect is of course related to inequalities in income, level of economic and technological development, and statistical capacity, which varies across the world. However, in the recent decades also other differences started to emerge. A move from standardised censuses and surveys to more country-specific administrative data, commercial data or digital traces, is creating more diversity in statistical systems. As a result, if there is no attempt at coordination of the official statistics efforts in that area nor at standardising the sources, methods and approaches used, the international comparability of data may be affected, limiting the usefulness of measures such as the Sustainable Development Goals, or similar.

Two articles in this collection (Fúquene-Patiño et al. 2021; Wazir and Goujon 2021) explicitly deal with some of the statistical challenges faced by the less economically developed regions and propose innovative solutions. Still, much more work is needed to help streamline the production of official population statistics, especially in some of the world's poorest economies, both in understanding the past trends that have resulted in the current populations, and then in their current measurements. Especially in this context, the fusion of different data sources and the use of 'Big data' sources in addition to the more traditional forms, such as censuses and surveys, bear substantial promise. To that end, various methods that can be developed and thoroughly tested in the countries with higher data availability, can then be transferred to contexts with higher data gaps (and needs), but at the same time, dedicated solutions are needed to help transform the official statistics across the globe. There are important current initiatives, such as the WorldPop project (*https://www.worldpop.org*, see e.g., Tatem et al. 2007 or Qader et al. 2021), on which these developments could build and expand.

At a more general level, to make the official demographic statistics fit for the 21st century, the key concepts related to populations and their dynamics need to be revisited and updated. The underlying conceptual issues, related to defining populations, concepts of residence, migrants and migration events, urban and rural areas, and so on, are ripe for re-assessment: with many new fluid forms of presence of individuals across space, some of the old, well-established concepts, such as permanent residence, do not seem longer fit for purpose. There are examples of promising frameworks looking at the space and time in new, more flexible ways (e.g., Martin et al. 2015), with the aim of offering a more fine-grained picture of populations to the desired spatio-temporal resolution. A matching extension of population definitions – for example moving from static counts of people to place-specific duration 'exposures' – is a promising way forward. At the same time, ensuring international comparability of official population statistics remains essential, as argued in the Letter to the Editor accompanying the current Special Issue (Lanzieri 2021).

In this context, migration statistics are particularly tricky: even in countries with well-established collection mechanisms, such data are still far from being harmonised and flawless, despite regulatory efforts and establishment of broader statistical frameworks (e.g., UN Statistical Commission 2021, §52/109; for a discussion, see also Lanzieri 2021). A promising option is the migration 'mainstreaming' in official statistics, whereby migration indicators are routinely added to existing sources of data, such as surveys (Radermacher and Thorogood 2009; Knauth 2011). The same ideas can apply to other

areas of interest for population statistics, especially where the monitoring of social inequalities is very important, such as in the case of gender identity, ethnicity or race, and so on, which are also changing in terms of how they are being understood and measured (see e.g. ONS 2021).

In order to meet these challenges, there is an urgent need for establishing an international programme of work, as an extension of the current endeavours, which would produce concrete recommendations for countries at the different levels of development of their statistical systems. Such international recommendations would form a base for transforming the official population statistics across the globe. Their implementation would see changes to the statistical infrastructure, and the way the official population statistics are produced and disseminated, by making more efficient use of new and existing data sources, increasing data availability to the users, integration of different data, geo-referencing, or improved communication to the users and general public, for example by developing new methods for the visualisation of population statistics.

From the methodological point of view, some cross-cutting themes in official population statistics become apparent even by looking at the examples of articles in the current Special Issue. There is an increasing use of inferential statistical approaches for different applications in applications in official population statistics, with Bayesian methods gaining increasing prominence, especially since their adoption for global population projections by the UN Population Division (Gerland et al. 2014; for a contemporaneous review of Bayesian Demography, see Bijak and Bryant 2016). Another point is that the traditional purely designed-based view to official statistics is increasingly being complemented by model-based, or hybrid, model-and-design-based approached (see e.g., Little 2004, 2006), which are better suited for dealing with the new and non-standard forms of population data. Communicating the uncertain population estimates or forecasts to the users of official statistics remains a barrier to their wider rollout, but there are already suggestions of how to overcome this (see Bijak et al. 2015).

Despite – or maybe because of – all the challenges, the developments in the areas of new data, methods and practices make this an exciting period in the long history of official population statistics, with many innovations still to come as the century unfolds. To make the most of these opportunities, the key prerequisite is the willingness to embrace change and make headway with implementing novel approaches in official statistical practice. To quote John Pullinger, the former UK National Statistician, "this is the moment when we can make our greatest contribution to society by providing the better statistics that allow for better decisions" (Pullinger 2017).

Jakub Bijak
Johan Bryant
Elżbieta Gołata
Steve Smallwood
*Guest Editors*

## 4. References

Baffour, B., J. Brown, and P.W.F. Smith. 2021. "Latent Class Analysis for Estimating an Unknown Population Size – with application to censuses". *Journal of Official Statistics* (in this issue).

Bijak, J., I. Alberts, J. Alho, J. Bryant, T. Buettner, J. Falkingham, J.J. Forster, P. Gerland, N. Keilman, T. King, A. O'Hagan, L. Onorante, D. Owens, A.E. Raftery, H. Ševčíková, and P.W.F. Smith. 2015. "Probabilistic population forecasts for informed decision making". Letter to the Editor. *Journal of Official Statistics* 31(4): 537–544. DOI: https://doi.org/10.1515/jos-2015-0033.

Bijak, J., and J. Bryant. 2016. "Bayesian demography 250 years after Bayes." *Population Studies*, 70(1): 1–19. DOI: https://doi.org/10.1080/00324728.2015.1122826.

Boonstra, P., R.J.A. Little, B. West, R. Andridge, and F. Alvarado-Leiton. 2021. "A simulation study of diagnostics for selection bias." *Journal of Official Statistics* (in this issue).

Di Consiglio, L., and T. Tuoto. 2015. "Coverage evaluation on probabilistically linked data." *Journal of Official Statistics* 31(3): 415–429. DOI: https://doi.org/10.1515/jos-2015-0025.

Di Consiglio, L. and T. Tuoto. 2018. "Population Size Estimation and Linkage Errors: The Multiple Lists Case." *Journal of Official Statistics* 34(4): 889–908. DOI: https://doi.org/10.2478/jos-2018-0044.

Fay, R.E., and R.A. Herriot. 1979. "Estimates of income for small places: An application of James-Stein procedures to census data." *Journal of the American Statistical Association* 74(366): 269–277. DOI: https://doi.org/10.1080/01621459.1979.10482505.

Folkman Gleditsch, R., A. Syse, and M. Thomas. 2021. "Fertility Projections in a European Context: A Survey of Current Practices and Outcomes." *Journal of Official Statistics* (in this issue).

Fúquene-Patiño, J., C. Cristancho, M. Ospina, and D. Morales González. 2021. "Fay-Herriot model-based prediction alternatives for estimating households with emigrated members." *Journal of Official Statistics* (in this issue).

Gerland, P., A.E. Raftery, H. Ševčíková, N. Li, D. Gu, T. Spoorenberg, L. Alkema, B.K. Fosdick, J. Chunn, N. Lalic, G. Bay, T. Buettner, G.K. Heilig, and J. Wilmoth. 2014. "World population stabilization unlikely this century." *Science* 346(6206): 234–237. DOI: https://doi.org/10.1126/science.1257469.

Hilton, J., E. Dodd, J.J. Forster, and P.W.F. Smith. 2021. "Modelling frontier mortality using Bayesian generalised additive models." *Journal of Official Statistics* (in this issue).

Knauth, B. 2011. "Migration Statistics Mainstreaming." Paper for the 58th Congress of the International Statistical Institute, 21–26 August 2011, Dublin, Ireland. Available at: https://2011.isiproceedings.org/papers/650162.pdf.

Lanzieri, G. 2021. "International comparability of population statistics is essential." Letter to the Editor. *Journal of Official Statistics* (in this issue).

Little, R. 2004. "To Model or Not To Model? Competing Modes of Inference for Finite Population Sampling." *Journal of the American Statistical Association* 99(466): 546–556. DOI: https://doi.org/10.1198/016214504000000467.

Little, R.J. 2006. "Calibrated Bayes." *The American Statistician* 60(3): 213–223. DOI: https://doi.org/10.1198/000313006X117837.

Little, R.J.A., B.T. West, P.S. Boonstra, and J. Hu. 2020. "Measures of the Degree of Departure from Ignorable Sample Selection." *Journal of Survey Statistics and Methodology* 8(5): 932–964. DOI: https:/doi.org/10.1093/jssam/smz023.

Martin, D., S. Cockings, and S. Leung. 2015. "Developing a flexible framework for spatiotemporal population modeling". *Annals of the Association of American Geographers* 105 (4): 754–772. DOI: https:/doi.org/10.1080/00045608.2015.1022089.

Nirel, R. and H. Glickman. 2009. "Sample Surveys and Censuses." In: D. Pfeffermann and C.R. Rao (eds) *Handbook of Statistics, Vol 29A. Sample Surveys: Design, Methods and Applications*. Amsterdam: Elsevier: 539 – 565. DOI: 10.1016/S0169-7161(08)00021-7.

Nishimura, R., J. Wagner, and M. Elliott. 2016. "Alternative indicators for the risk of non-response bias: A simulation study." *International Statistical Review* 84(1): 43–62. DOI: https:/doi.org/10.1111/insr.12100.

Oeppen, J., and J.W. Vaupel. 2002. "Broken Limits to Life Expectancy." *Science* 296(5570): 1029–1031. DOI: https:/doi.org/10.1126/science.1069675.

ONS. 2021. Sex and gender identity question development for Census 2021. London: Office for National Statistics. Online report. Available at: https://www.ons.gov.uk/-census/censustransformationprogramme/questiondevelopment/sexandgenderidentity-questiondevelopmentforcensus2021 (acessed May 2021).

Qader, S., V. Lefebvre, A. Tatem, U. Pape, K. Himelein, A. Ninneman, L. Bengtsson, and T. Bird. 2021. "Semi-automatic mapping of pre-census enumeration areas and population sampling frames." *Humanities and Social Science Communications* 8 (3) DOI: 10.1057/s41599-020-00670-0.

Pullinger, J. 2017. "Statistics are even more important in a 'post-truth' world. Letter to The Guardian". 24 January 2017. *The Guardian*. Available at: https://www.theguardian.-com/politics/2017/jan/24/statistics-are-even-more-important-in-a-post-truth-world (accessed May 2021).

Radermacher W., and D. Thorogood. 2009. "Meeting the growing needs for better statistics on migrants." Paper for the DGINS Conference "Migration–Statistical Mainstreaming", 30 September – 1 October 2009, Malta. Available at: https://ti-nyurl.com/better-statistics-on-migrants.

Raftery, A.E., J.L. Chunn, P. Gerland, and H. Ševčíková. 2013. "Bayesian probabilistic projections of life expectancy for all countries." *Demography* 50(3): 777–801. DOI: https://doi.org/10.1073/pnas.1211452109.

Raftery, A.E., N. Lalic, and P. Gerland. 2014. "Joint probabilistic projection of female and male life expectancy." *Demographic Research* 30(27): 795–822. DOI: https://doi.org/10.4054/DemRes.2014.30.27.

Righi, P., P. Falorsi, S. Daddi, E. Fiorello, P. Massoli, and M. Terribili. 2021. "Optimal sampling for the Population Coverage Survey of the new Italian Register Based Census." *Journal of Official Statistics* (in this issue).

Ševčíková, H., and A. Raftery. 2021. "Probabilistic Projection of Subnational Life Expectancy." *Journal of Official Statistics* (in this issue).

Tatem, A.J., A.M. Noor, C. von Hagen, A. Di Gregorio, and S.I. Hay. 2007. "High Resolution Population Maps for Low Income Nations: Combining Land Cover and Census in East Africa." *PLoS ONE* 2(12), e1298. DOI: https://doi.org/10.1371/journal.pone.0001298.

UN Statistical Commission. 2021. Report on the fifty-second session (1–3 and 5 March 2021). UN Economic and Social Council. E/2021/24-E/CN.3/2021/30. New York: United Nations. Available at: https://undocs.org/E/2021/24.

Wazir, A. and A. Goujon. 2021. "Exploratory Assessment of the Census of Pakistan Using Demographic Analysis." *Journal of Official Statistics* (in this issue).

Zhang, Z, A. Bhattacharjee, J. Marques, and T. Maitis. 2021."Spatio-Temporal Patterns in Portuguese Regional Fertility Rates: A Bayesian Approach for Spatial Clustering of Curves." *Journal of Official Statistics* (in this issue).

Zult, D., P.-P. de Wolf, B. Bakker, and P. van der Heijden. 2021. "A general framework for multiple – recapture estimation that incorporates linkage error correction." *Journal of Official Statistics* (in this issue).