

Distr.: General
07 September 2022

English

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Population and Housing Censuses

Twenty-fourth Meeting

Geneva, Switzerland, 21-23 September 2022

Item 6 of the provisional agenda

The long-term future: censuses of the 2030 round and beyond

Census data: a crucial population reference for research and policy

Note by IPUMS at the University of Minnesota*

Summary

The global pandemic proved again that census data are a crucial resource for understanding population level activity and change. Yet scarce resources and public scepticism require that we continue to promote examples of the value of the censuses at every opportunity. This paper reports recent examples of research and policy conducted using census microdata from IPUMS, the world's largest disseminator of census microdata for scholarly research. The paper also suggests that harnessing the potential in new data sources will require continued access to census data. More than 100 countries around the world disseminate microdata sample files from their Population and Housing Censuses through a dissemination partnership with IPUMS at the University of Minnesota. Data are harmonized and integrated across time and space. IPUMS integration and documentation makes it easy to study change, conduct comparative research, merge information across data types, and analyze individuals within family and community contexts. Data and services are available free of charge to registered researchers and policy makers.

*Prepared by Lara L. Cleveland

NOTE: The designations employed in this document do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

I. Introduction

1. The importance of the census ought to be self-evident such that its continued support (both governmental and public) could be taken for granted. Broad and sweeping terms are used to describe the census: it provides a unique and complete snapshot of the population, paints a picture of the country's people, tells us who we are as a nation. The place-based counting of people in their households provides the basis for democratic and representative governance jurisdictions. Censuses are used for allocating substantial governmental funding to subnational regions and localities. They are also important to myriad governmental, non-profit, and private sector planning activities, for new development, or to incentivize investment in areas of need. We often emphasize the value of the census for "evidence-based policy-making" without reference to the specific and explicit benefits derived from the census. The benefits are, indeed, too numerous to count. Yet, it is useful to collect examples and stories of the concrete benefits derived from the census to ensure its continued existence.

2. Facing ever tightening budgets, mounting costs, and public scepticism about privacy and truth in the new information age, national statistical offices must build more compelling cases than ever before to secure funding from their governmental ministries and cooperation from the public for continuing census (or census-like) operations. National statistical offices and their supporters at the United Nations and elsewhere have been compiling concrete examples in compelling storytelling formats to illustrate the ways that censuses are used to address pressing societal challenges and improve the lives of people. IPUMS joins their efforts.

3. National statistical offices, and specifically census organizations, are in the best position to provide examples of planning successes facilitated by census information. Scholarly and policy organizations that make use of derived and downstream versions of census data, like IPUMS at the University of Minnesota, are well positioned to add to the overwhelming list of benefits derived from a country's investment in collecting and disseminating census data. IPUMS International, a census and survey data infrastructure initiative at the University of Minnesota's Institute for Social Research and Data Innovation (ISRDI), disseminates harmonized versions of census microdata samples from more than 100 countries.

4. IPUMS offers a means of disseminating microdata which complements the dissemination activities of National Statistical Offices. Through negotiated partnership agreements with country statistical offices, IPUMS optimizes sample microdata from censuses for research use by standardizing data formats and harmonizing variables across time and place. NSOs disseminate official statistics and statistical products to a large number of publics—citizens, officials, the media, analysts, etc. IPUMS International disseminates microdata on a restricted access basis to researchers who require detailed data on individuals and households to measure and analyse complex relationships, often making comparisons over time and between nations. Per agreements with National Statistical Offices, only vetted applicants with verifiable bona fides and a credible research project are approved to access the international microdata samples.

5. IPUMS maintains a database of scholarly users of the data and compiles bibliographic information about publications that rely on data accessed through IPUMS. IPUMS is, therefore, uniquely positioned to provide information back to statistical offices about the extended usage of census data once files are made publicly available to the research and policy community. This paper describes recent high-impact uses of the data and underscores the importance of continued collection and dissemination of robust, high-quality, government data collections, such as the census. The global pandemic has proved more than ever that census data are still a go-to resource for understanding population level activity and change.

II. IPUMS encourages census data usage by scholars and policy analysts

6. Since 1999, IPUMS International has collaborated with over 100 national statistical offices (NSOs) around the world to facilitate the use of census microdata in scholarly and policy research. Microdata are essential for the study of complex population processes, enabling custom tabulation, multivariate analysis, and multidimensional disaggregation of population-level statistics. IPUMS harmonization and dissemination efforts simplify access and analysis of the world's census microdata.

7. As of 2022, 404 anonymized microdata samples from censuses of 105 countries are available to researchers and students free of charge through the IPUMS online data dissemination system. Truly global in its coverage, the series includes more than 50 samples each from Africa, Asia, Europe, and the Americas. The data series includes contemporary data¹ from 1960 to 2017; the project continues to add data from the 2020 census round as microdata samples become available. Most participating national statistical agencies have entrusted the country's full series of extant census microdata to the project, facilitating intra-national as well as international trend analysis. Ninety-seven (97) census samples from 25 UNECE Member States are available from IPUMS (table 1).

Table 1
UNECE-region contemporary census samples available from IPUMS International

Country	2010s	2000s	1990s	1980s	1970s	1960s
Armenia	2011	2001				
Austria	2011	2001	1991	1981	1971	
Belarus	2009	1999				
Canada	2011	2001	1991	1981	1971	
Finland	2010					
France	2011	2006	1999; 1990	1987; 1982	1975	1968; 1962
Germany				1987; 1981*	1971*; 1970	
Greece	2011	2001	1991	1981	1971	
Hungary	2011	2001	1990	1980	1970	
Ireland	2016; 2011	2006; 2002	1996; 1991	1986; 1981	1979; 1971	
Israel		2008	1995	1983	1972	
Italy	2011	2001				
Kyrgyzstan		2009	1999			
Netherlands	2011	2001			1971	1960
Poland	2011	2002		1988	1978	
Portugal	2011	2001	1991	1981		
Romania	2011	2002	1992		1977	
Russia	2010	2002				
Slovak Republic	2011	2001	1991			
Slovenia		2002				
Spain	2011	2001	1991	1981		
Switzerland		2000	1990	1980	1970	
Ukraine		2001				
United Kingdom		2001	1991			
United States	2015; 2010	2000; 2005	1990	1980	1970	1960

* German Democratic Republic

8. Samples distributed by IPUMS are systematically drawn from the total enumerated population by IPUMS or by the statistical office of the country of origin according to a variety of sample designs. Where possible, IPUMS provides 10 percent samples of census data by

¹ IPUMS International also disseminates historical census data from the 18th, 19th, and early 20th centuries for Canada, United States, United Kingdom, Germany, Norway, Sweden, Iceland, and Denmark.

selecting every 10th household after a random start. Nearly all samples available from IPUMS are cluster samples: they are samples of households rather than individuals.

9. The principal advantage of IPUMS is its reconciliation of census-specific variable codes to produce datasets that integrate records across time and space. The basic goal of variable harmonization is to make data suitable for comparative analysis by applying comparable codes for each variable across all samples in the database. Microdata are integrated so that identical concepts have identical codes. Over 1000 harmonized variables are included in the IPUMS International database.

10. The IPUMS online data access system allows researchers to create customized, pooled data extracts that contain only the census samples, variables, and cases they require. The IPUMS data extract engine generates a pooled dataset containing the requested microdata and the corresponding set of DDI (Document Data Initiative) compatible metadata, including a codebook suitable for constructing a system data file in SPSS, SAS, Stata, or R. The data access system is fully integrated with the variable and sample documentation in a user-friendly online interface, so researchers can make informed decisions as they define their datasets.

11. IPUMS expertise in data and metadata standardization, harmonization, and dissemination is now widely recognized. In addition to harmonizing and disseminating U.S. and International census microdata, IPUMS collaborates with U.S. and international survey organizations as well. IPUMS collections include 115 decennial censuses (including full count census files between 1870-1940) or American Community Surveys and more than 800 surveys. U.S. surveys in the IPUMS collection include all basic monthly surveys and special supplements of the Current Population Survey (CPS) fielded by the Bureau of Labor Statistics; the National Health Interview Survey (NHIS) from the National Center for Health Statistics; the Medical Expenditure Panel Survey (MEPS) collected by the Agency for Healthcare Research and Quality; and the American Time Use Survey funded by the Bureau of Labor Statistics and fielded by the U. S. Census Bureau.

12. Internationally, IPUMS harmonizes and disseminates 600+ health, labor force, and time use surveys of 117 countries in addition to the censuses described above. These surveys include the Demographic and Health Surveys (DHS), where IPUMS' expertise in curating and making easily accessible metadata was sought by the DHS organization collaborators. Other surveys include the Gates Foundation's PMA surveys, multinational time use surveys (in partnership with statistical offices and University College London) and work in process with UNICEF to disseminate harmonization scripts for the Multiple Indicator Cluster Surveys. In all, 154 countries are represented by one or more data sets across the IPUMS collections.

13. The data are ideally suited for use by demographers, economists, public health experts, and a variety of social science researchers. Experts from United Nations organizations, World Bank, and other international policy organizations also use the data extensively. In addition to the data harmonization work, IPUMS provides users with access to extensive curated metadata, including enumeration forms, codebooks, sample design information, and custom documentation detailing IPUMS' harmonization approach and associated comparability notes. IPUMS creates additional tools to aid researchers in leveraging the power of the data such as family interrelationship "pointer" variables, currency conversion options, variance estimation guidance, online tabulation capabilities, and more.

14. More than 200,000 researchers worldwide have accessed one or more of the census samples or surveys across all IPUMS collections for their scientific or policy research projects. Thousands of those researchers are using data from the UNECE region, many of whom are originally from, or currently located in, UNECE countries (outside of the US). Studies using census data from UNECE countries in the IPUMS database address pressing population issues facing the UNECE region including migration, aging, low fertility, labor supply, COVID and climate vulnerability. Even researchers from within the U.S. Census Bureau or Bureau of Labor Statistics use IPUMS versions of their own data files when they need to do quick calculations, cross-temporal, or comparative research due to the easy

accessibility of the files, harmonized variables, and other user-friendly attributes of IPUMS files.

15. Census data are used to study many aspects of population dynamics, population characteristics, and economic development. Among the thirteen broad classifications offered by the IPUMS online bibliography, six account for the majority of citations: labor force and occupational structure; migration and immigration; family and marriage; education; methodology and data collection; and fertility and mortality. Most IPUMS users utilize the data in advanced-degree theses and dissertations or academic journal articles. IPUMS-based research has been published in the leading English-language demography, sociology, and economic journals.

16. Researchers make use of the flexibility of harmonized microdata to study associations between phenomena, change over time, and regional and global patterns. Many of these scholarly articles address the most pressing population issues facing the UNECE region, including aging, migration, low fertility, labour supply, and urban development with direct implications for public policy and planning. Large and nationally representative datasets make it possible to study small subpopulations and subnational regions of countries.

III. Census use cases

17. The census provides the foundation for so many activities in government and the private sector. It also has value far beyond the initial published results. This section highlights only a few of the many ways census data are used to gain new insights about social phenomena, inform policy, and plan for the future.

A. Governmental purpose: representation, planning and funding

18. Population census results help guide the setting of jurisdictional boundaries for elections and allocation of proportional representative seats in governing bodies. For example, in the United States, the data collected by the decennial census is mandated by the U.S. constitution and is used to determine the number of seats each state has in the U.S. House of Representatives. It is also used to redraw small area election boundaries with roughly equivalent numbers of people in each area for local and state elections. (U.S. Census Bureau, 2022). In Switzerland, the population figures are used to allocate National Council seats to the cantons. (Federal Statistics Office of Switzerland, 2022).

19. Nearly every country lists the census as a key source in determining how to allocate central government funding to localities. In the United States, more than one trillion U.S. dollars are allocated to states and local governments annually based on population information from the census. Totals were even higher in 2020 and 2021 when adding in Covid-19 related assistance.

20. Governments, non-profits, and private companies use census information to help them with a wide range of planning activities. Whether areas have adequate schools, health care facilities, transportation routes, and other infrastructure can be assessed with census data. The United Kingdom reports that census counts can help governments understand the workforce and which training policies will help meet the demand for workers in specific occupations (Office of National Statistics, UK, 2022). Australia notes that some public libraries use information from the census to identify language needs in local communities and build bilingual book collections or language support materials (Australian Bureau of Statistics, 2022). The Swiss FSO explains that businesses use population statistics to conduct market studies or choose locations for new branches. (Federal Statistics Office, 2022).

21. In fact, business uses of the census may be more extensive than governments imagine. The extent to which census data helps fuel private sector, economic development and non-profit activity is dependent upon the extent to which census results and data are made publicly accessible. In 2012, when the American Community Survey (an annual survey designed to help keep population estimates current between censuses) came under attack, the loudest opponents of the bill and biggest voices in support of the survey were large organizations

supporting private industry. Organizations like the United States Chamber of Commerce, the National Retail Federation, and the National Association of Home Builders were reportedly “up in arms” (Rampell, 2012). Major retail establishments rely on regular reporting of population information in their ongoing planning processes.

B. Research and policy: downstream uses of census data

22. By working with countries on their census microdata dissemination, IPUMS facilitates a particular type of complex data analysis. Census data, which covers the entire population, is often the only data source, sufficiently large and nationally representative, to enable research about small groups or small areas. Even sample data from the census affords researchers representative information about small groups while preserving a useful level of privacy protection. The sections below illustrate several topics that can be studied using census data. We present examples in three broad categories: understanding vulnerabilities for crisis management; analysis of small groups; and for extending the power of other data collections.

1. Research: responding to the unexpected and understanding vulnerabilities

23. The COVID-19 pandemic reminded us why it is important to have data at the ready to study current events, trends, and plan adaptation to events or new circumstances. When the pandemic hit, IPUMS user registration requests quickly reflected an interest in census data for making population level studies of covid vulnerability. Researchers across the globe were accessing census microdata from IPUMS International for COVID-19-related research. Scholars at universities from the U.S. to Nepal, Columbia to Belgium, Nigeria to China, and elsewhere used IPUMS data to assess population dynamics contributing to COVID-19 vulnerability or spread. Divisions of the United Nations, World Bank, and other policy research institutes similarly accessed IPUMS census data for COVID response and relief efforts.

24. Of particular interest for research on population dynamics of COVID-19, especially at the outset, was information about the age structure of the population, household living arrangements (household size, intergenerational co-residence, etc.), indicators of health vulnerability (age, work status, housing conditions, disability, etc.), healthcare workforce distribution, and migration patterns. IPUMS International census samples include valuable subnational geographic identifiers at the first and second administrative levels, which are especially useful for highlighting particular regions or localities of vulnerability.

25. For example, Esteve et al. (2020) used recent census samples in IPUMS from 81 countries to estimate how age and co-residence patterns shape vulnerability to outbreaks of the disease. They estimated variation in deaths arising due to a simulated random infection of 10 per cent of the population living in private households and subsequent within-household transmission of the virus. The age structures of households in Europe and North America make those populations more vulnerable to COVID generally resulting in higher incidence of direct deaths in the simulation. Coresident patterns of large and multi-generational households in Africa and Asia implied greater within-household transmission of the disease, making older populations in those regions vulnerable. These patterns contributed to higher indirect death rates in those regions (Esteve et al., 2020).

26. IPUMS had been working with UNFPA and Esri to calculate key indicators for UNFPA’s Population Data Platform. When Covid hit, UNFPA quickly pivoted and asked IPUMS to reorganize output and calculate new measures of Covid vulnerability among older adults. Working with ESRI and IPUMS, UNFPA launched a COVID vulnerability dashboard mapping incidence of older adults living without access to electricity, living alone, and living with a disability, and other indicators of vulnerability. The dashboard was published and publicized in an effort to help governments and aid organizations focus their efforts on this potentially vulnerable population. (<https://www.unfpa.org/covid19>)

27. Before the pandemic, the World Health Organization (WHO) was already working extensively on a project to measure and locate the health workforce. In the context of SDG

3c, which aims to "substantially increase health financing and the recruitment, development, training and retention of the health workforce in developing countries, especially in least developed countries and small island developing States," WHO used data from IPUMS International to measure the stock and density of the health workforce. These estimates are publicly available through WHO's Global Health Observatory Data Repository. WHO also uses data from IPUMS International to assess gender equity in the health workforce in its 2019 report *Gender equity in the health workforce: Analysis of 104 countries*.

28. The WHO work on the Health Workforce requires access to detailed occupational coding structures at the ISCO 3-digit, but preferably 4-digit, level and access to relatively detailed geographic information. Only censuses, through high-density samples or full count data, provide sufficient numbers of cases to study the kind of occupational breakdowns at subnational levels for the WHO to carry out this research. The work is critically important as they build policy recommendations for countries where they identify significant mismatches between medical workers and need for care.

2. Research: census for small groups and disaggregations

29. IPUMS International data have been used to track progress towards the SDGs and other measures of economic development. Once again, microdata from censuses are flexible enough and of sufficient size to fulfil the disaggregation requirements of SDG monitoring in the service of leaving no one behind. Census data provide high case counts to enable disaggregation of measurements across age, sex, marital status, geographic regions, migration status, disability status, and so on. Several recent reports published by the Census Bureau, policy organizations, and United Nations agencies highlight the role of census microdata in SDG monitoring and SDG research.

30. Population aging has attracted increased attention in the research community of late. The National Institute on Aging provided one of the most recent grant awards to IPUMS International to support the development of additional IPUMS data enhancements and outreach focused on aging research. Censuses are important in this area due to the fact that studying adults at older ages can only be accomplished in datasets with sufficient cases to examine these groups. The United Nations Population Division (2020) and the U.S. Census Bureau (2020a, 2020b) have each produced recent publications on Aging around the world and features IPUMS data for at least part of the analysis.

31. A 2018 report by the International Organization for Migration (IOM) titled *A pilot study on disaggregating SDG indicators by migratory status* illustrates the potential of using census data to disaggregate national data by migratory status (Jeffers et al., 2018). The authors present a number of SDG and SDG-related indicators disaggregated by migrant status and other individual characteristics where relevant for dozens of countries across all world regions. Data from IPUMS International were also used to calculate indicators and statistics disseminated in the IOM's Migration Data Portal.

3. Research: leveraging census to extend other data sources

32. *Environmental data interoperability*. New data sources, data types, and methods along with greater volume of data mean that data interoperability is an increasingly important consideration in preparing data for broad usage. Geographic referencing is often a means of making datasets "talk to one another." Several recent studies have combined information about place of residence, birthplace, or previous residence (i.e., census location information) along with known timing and location of high pollution concentrations or shock events in order to study the effects of early life exposure to pollutants or environmental circumstances on later life experiences.

33. Clark et al. estimated changes over time (between the 2000 and 2010 US censuses) in disparities by location and race in exposure to outdoor concentrates of a transportation related air pollutant (NO₂). They found that disparities by race-ethnicity decreased from 2000 to 2010, but relative NO₂ exposure disparities persisted, with higher NO₂ concentrations for nonwhites than whites in 2010. Rivas et al. (2019) assess the role exposure to fine particulate

matter (PM2.5) during different prenatal and postnatal windows may play in children's cognitive development at school age in Spain. The study used land use data for exposure information, direct study-related survey of research subjects, combined with an Urban Vulnerability index calculated from the Census.

34. *Small Area Estimation.* Small area estimation has been one approach researchers have taken recently to facilitate disaggregation of national-level findings from surveys. For the past several years, United Nations reporting agencies have been developing Small Area Estimation techniques to disaggregate SDG indicators at subnational area levels. Using census data accessed through IPUMS and from National Statistical offices, researchers have been able to estimate poverty rates and other population characteristics at disaggregated subnational geographic levels by combining data from labour force surveys and censuses. UNSD, UNECLAC, and UNFPA have partnered to create the Small Area Estimation toolkit (<https://unstats.un.org/wiki/display/SAE4SDG/>). The site assembles resources for understanding, conducting, and learning from Small Area Estimation studies (United Nations Statistics Division, 2022).

35. The full set of training materials, including hands on data exercises featuring a census sample from IPUMS, is available on the toolkit website along with recommended software packages, communication guidelines, research examples, and more. The materials are designed to instruct country statistical office analysts and others how to conduct small area estimation to fulfil their reporting obligations for spatial disaggregation across several key SDG indicators. The group will be hosting workshop-style training sessions in the coming months and years.

IV. Conclusion

36. It is impossible to capture all the ways in which the census informs the world around us. However, if we imagine a world without population counts from the census, we can imagine paralyzing impediments to planning and incredible inefficiencies in allocation of resources. Census units and bureaus can make the best case for the governmental planning purposes supported by their work. It is incumbent upon them to do so.

37. IPUMS is also dedicated to contributing to the collective global effort toward documenting the value of the census. Policy recommendations and activities based on research using census data are not always well documented. Policy work shows up in short briefs, meeting presentations, web-based communications, and other reports that are not captured in traditional bibliographic records of scholarly work. IPUMS has begun conducting interviews with key researchers to find out about the policy implications of their work and to encourage them to report such outcomes to IPUMS. We are also working with librarians at the University to identify web-based and other reference sources of policy activity and link it to upstream research using census data. We will make efforts to include more of these examples in our reports to data producers about usage of their data.

38. Many examples of the value of censuses noted and reported by IPUMS are possible only when countries make their data, including microdata samples from the census, available to the public. As countries move increasingly to register-based or register-supplemented censuses, it is imperative that they continue to produce microdata samples and that they provide adequate metadata about methods to properly interpret the data. The advancement of science about human populations is dependent upon it.

39. The era of big data does not eliminate the need for scientific data collection by official governmental authorities. Quite the contrary. We can only understand how to leverage the data deluge when we can calibrate, reference, and quality check new methods and new data sources using sound and scientifically collected reference data. IPUMS will also be engaged in the evaluation of these new data source and methods and will try to suggest ways in which census data might inform the development of methods and approaches to using big data.

References

- Australian Bureau of Statistics. 2022. "The Australian Census." Online, accessed July 31. <https://www.abs.gov.au/census/about-census/australian-census>
- Boniol M, McIsaac M, Xu L, Wuliji T, Diallo K, Campbell J. Gender equity in the health workforce: analysis of 104 countries. Working paper 1. Geneva: World Health Organization; 2019 (WHO/HIS/HWF/Gender/WP1/2019.1). Licence: CC BY-NC-SA 3.0 IGO.
- Clark, L. P., D. B. Millet, and J. D. Marshall, "Changes in Transportation-Related Air Pollution Exposures by Race-Ethnicity and Socioeconomic Status: Outdoor Nitrogen Dioxide in the United States in 2000 and 2010," *Environ. Health Perspect.*, vol. 125, no. 9, p. 097012, 2017, doi: 10.1289/EHP959.
- Esteve, A., Permanyer, I., Boertien, D., Vaupel, J. 2020. "National Age and coresidence patterns shape COVID-19 vulnerability" *PNAS*, 117(28). July 14.
- Federal Statistics Office of Switzerland. 2022. "What is the purpose of a national census?" Online, accessed July 31. <https://www.bfs.admin.ch/bfs/en/home/basics/census/purpose.html>
- Jeffers, K., J. Tjaden and F. Laczko (2018). A pilot study on disaggregating SDG indicators by migratory status. International Organization for Migration, Geneva.
- Office of National Statistics (ONS), UK. 2022. "Why we have a census" Online, accessed July 31. <https://www.ons.gov.uk/census/2011census/whywehaveacensus>
- Rivas, I. et al. 2019. "Association between Early Life Exposure to Air Pollution and Working Memory and Attention," *Environ. Health Perspect.*, vol. 127, no. 5, p. 057002, doi: 10.1289/EHP3169.
- UNFPA. 2022. *Household Structure of Older Persons*. New York: UNFPA.
- United Nations Population Division. 2020. *World Population Aging: Highlights*. New York: United Nations Department of Economic and Social Affairs.
- United Nations Statistics Division. 2022. SAE4SDG, Online, accessed August 2. <https://unstats.un.org/wiki/display/SAE4SDG/>
- United States Census Bureau. 2020a. *Africa Aging: 2020*. Washington, D.C.: US Department of Commerce.
- United States Census Bureau. 2020b. *Asia Aging: Demographic, Economic, and Health Transitions*. Washington, D.C.: US Department of Commerce.
- United States Census Bureau. (2022). Our Censuses. Online, accessed 31 July 2022. <https://www.census.gov/programs-surveys/censuses.html>