EUROPEAN POPULATION CONFERENCE 2022

Annual meeting 2022

Section Data and Methods

Estimating Age Patterns of Net-Migration across European Municipalities

Daniela Ghio, Claudio Bosco, Fabrizio Natale, Jan Loeschner, Anne Goujon *

* Joint Research Centre European Commission

Abstract (250)

Migration plays a key role in the development of local areas challenged by demographic change and ageing. However, the lack of age-specific data and long time series limit the assessment of migration impacts, even in the European Union (EU) with a common data collection system since 2007.

The paper explores patterns in population and net-migration age structures in more than 100,000 EU Local Administrative Units (LAUs) over the period 2011-2019. We harmonize EU population datasets using (2011) population censuses and the latest (indicatively 2019) available statistics provided by National Statistical Institutes. Exploiting these datasets, we estimate age-specific net migration by adapting census survival methods to handle the large EU coverage and variations in data collections coming from different national sources. Specifically, we apply life-table models to assess changes by cohort and derive age-specific net-migration rates applying a residual method. The estimated age-specific net migration rates are validated using pooled official statistics available for small samples of LAUs with limited territorial and temporal coverages.

The resulting net migration rates by age at finer geographical resolution could contribute to enhancing research on the determinants of territorial disparities across the EU during intercensal periods. Covering the whole EU territories, the new datasets at LAU level complement the figures available from the 2011 censuses, anticipating the release of gridded census statistics foreseen for 2023.

INTRODUCTION

European Union (EU) countries currently experience low to negative population growth rates with persistent differences across regions (de Beer et al., 2012). These differences become even more evident when looking at finer territorial units, where ageing has important implications not only for economic development, but also for the social cohesion of populations (Goujon et al., 2021; Kashnitsky et al., 2020). All demographic determinants contribute to explaining differences in the share of population living in territories; yet, migration plays a major role in re-shaping population size and age structures at local level, as conceptualised by several authors (Davis, 1965; De Vries, 2013; Dyson, 2011, Rees et al. 2020). Short and medium term migration effects are age-specific: older age groups may be likely to move to rural areas after retirement, whereas urban areas may be more attractive to younger age groups for studying and working (Goujon et al., 2021).

Nevertheless, most studies focus on national or regional levels (De Beer at al. 2012, 2014; Kashnitsky et al. 2020), accounting for data availability and policy relevance (for instance, De Beer et al., 2012, 2014 define NUTS2 as the prime level of EU regional cohesion policies). Thus, less attention has been paid at local levels (Gutiérrez Posada et al., 2018; Sabater et al., 2017).

Limitations in migration data exist at global level (Willekens et al, 2016). Many countries do not publish migration-related data; according to IOM and McKinsey & Company (2018), only 45 of the 193 United Nation members report statistics on migration flows. Inconsistencies in definitions and measurements across origin and destination sources contribute to affect comparability. For instance, out of 45 countries that report migration-flow data, only 24 use the same definition of migrant (IOM and McKinsey & Company, 2018). In the EU, Member States are obliged to provide harmonised migration flow statistics to Eurostat (Regulation (EC) No 862/2007). However, these datasets are covering the local levels only partially, despite recent efforts to combine annual demographic and migration statistics with more frequent socio-economic information (e.g., micro-data from Labour Force Surveys) As a result, gaps in migration data at local level remain, reflecting the difficulties existing within countries in the harmonisation of their data collections.

This research aims at contributing to filling the gaps by providing estimates of age specific net-migration rates in more than 100,000 EU Local Administrative Units (LAUs), over the intercensal period 2011-2019, which broadly corresponds to the latest round of population censuses and age-specific population data availability from national statistical institutes (NSIs). Results will be key to support research on the role of migration as a determinant of territorial disparities and changes in age structure across the EU Municipalities.

POPULATION DATA BY AGE IN THE EU

Eurostat releases yearly statistics on the population and the demographic components, fertility, mortality and net migration, by the harmonized NUTS3 classification of regions. However, higher geographical resolutions would be needed to explore spatial patterns in age structures in relation to population dynamics at different levels, e.g., in urban and rural areas.

Population datasets at municipal level in the EU are primarily collected by NSIs, but are not always publicly available. It should be noted that the 2021 EU Census round, likely available from 2023, will include data at a spatial resolution of a 1km grid (in accordance to the Inspire Directive). Until these new data become available, 2011 Censuses limit the analyses of recent evolutions in population dynamics at the local level to appraise the role of net migration.

A key contribution of our study consists in collecting and harmonizing data on population by age at the lowest available geographical resolution for most recent years, complementing data at LAU2 level available from the 2011 Censuses.

We collected data from online sources or by contacting NSIs, through the support of Eurostat, for the following 19 countries: Austria, Belgium, Bulgaria, Czechia, Denmark, Finland, France, Germany, Hungary, Italy, Lithuania, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain and Sweden. Population data at LAU2 level were either not available or could not be extracted and processed for Cyprus, Greece, Croatia, Estonia, Ireland, Latvia, Luxembourg and Malta.

The data were downloaded mostly manually from web sites and in two cases by using Application Programming Interfaces. The downloaded data were characterized by different formats, naming conventions. Through a series of data processing and imputation steps using the GNU R software (R Core Team, 2020), the data were transformed into a cohesive dataset with indication of country, year, NUTS3 code, LAU code (according to GISCO 2018 classification), age (in single years) and population size.

Further details on the coverage and validation of the data are provided in the Supplementary Material.

METHOD

We carry out the analysis using the highest administrative granularity of LAUs. We set the temporal framework of the analysis accounting for the latest population Census in 2011 from Eurostat and the more recent official statistics on LAU population structures published by NSIs (the reference year of these statistics varies from country to country, ranging from 2017 to 2019). Because the 8-year interval differs from the conventional 5-year age group interval (or its multiple) commonly applied to aggregate population structures, we decided to stratify the population datasets at the beginning and end of the reference period in single-year age groups. Furthermore, due to heterogeneity in the dates of the 2011 population censuses across MS (for instance, May 9th in Germany, December 31st in Estonia and February 2nd in Bulgaria), we harmonized the interval of the analysis by weighting the first year using a monthly fraction of the year.

The core of our empirical strategy consists in the adoption of the life table, looking at the cohort survival of people living in the same territory and sharing the same year of birth. Simplifying, we assume that the probability of dying during one year of ageing corresponds to the probability of dying within one- calendar year. Thus, individuals become one-year older at the beginning of each calendar period systematically[1]. From the probability of dying, the following functions are derived:

i) the probability of survival, as the complement to one of the probabilities of dying for the selected age-groups/cohorts during the reference-period;
ii) the numbers of years lived collectively by survivors within the age/period interval;
iii) the survival ratio, describing the average mortality conditions during the reference period, as the proportion of people, among those who survive to a given age/period, who live on and attain the next age level (UNDESA, 1970). The life table survival is also known as projective probability of survival, because of its application in the demographic projection methods (Livi Bacci, 2006). Modelling a life table stationary population, this stands for the probability that individuals of same birth *theoretical* cohort will still attain older ages (or be alive one or a number of years later).

Data not being available at LAU level, we rely on 2019 mortality data at NUTS3 level[2] (Eurostat, 2021). Consistently with Coale and Demeny (1966), we compile life tables at LAU level, under the assumption that people living in LAUs, within the same NUTS3 territory, are exposed to similar mortality regimes.

We use the life table to model the *theoretical* cohorts at the end of period under the hypothesis of closed populations or no migration during the whole reference interval. To deal with the open-ended age group, we calculate net migration for age-groups that are below the life expectancy at birth.

We adopt the survival ratio method (UNDESA, 1970) to formalize the computation of *theoretical* cohorts as follows:

$$p\_t\,{}_{t+a}^{x+a,i} = p\_o_t^{x,i} * \frac{L_{t+a}^{x+a,i}}{L_{t+a}^{x,i}}$$

where

i	represents the LAU unit
a	corresponds to the age and period interval
x	corresponds to the age group in 2011
$x+a$	corresponds to the age group in 2019, with (x+a) $\ll e_t^i$ life expectancy
$p\_o_t^{x,i}$	represents the observed cohort p^i living in LAU at age x at the beginning of period
$L_{t+a}^{x+a,i}$	represents the number of person-years lived by the stationary population in the NUTS3 where the LAU is located, at age x+a

[1] For the rest of analysis, we apply a synthetic indication of one-year age/period as whole, rather than notations of exact age and specific point in time.

[2] We chose 2019 mortality data in view of the small differences in age-specific mortality from 2011 to 2019, except for differentials in the seasonal influence of the flu, particularly in 2015. This could not have been the case, if we had decided to extend our analysis to 2020 and later since the COVID-19 pandemic affected mortality substantially and diversely at the local level.

$L_{t+a}^{x,i}$	represents the number of person-years lived by the stationary population in the NUTS3 where the LAU is located, at age x
$p\_t_{t+a}^{x+a,i}$	is the *theoretical cohort* that would be expected in 2019 at age $^{x+a}$living in LAU ^{i}, under the hypothesis it has been closed to migration during the *a* interval
$\frac{L_{t+a}^{x+a,i}}{L_{t+a}^{x,i}}$	is the prospective survival probability of individuals belonging to the cohorts aged *x* of a stationary population closed to migration to attain the age *x + a*

The difference between the *theoretical cohorts* and observed cohorts in t+a is the estimated net migration during the period t, *t+a* :

$$nm_{t+a,t}^{x+a,i} = p\_t_{t+a}^{x+a,i} - p\_ob_{t+a}^{x+a,i}$$

where

$nm_{t+a,t}^{x+a,i}$	represents the net migration by cohort over the reference period obtained as residual value from the comparison between the *theoretical* and observed cohorts at the end of the reference period
$p\_t_{t+a}^{x+a,i}$	represents the *theoretical cohort* in 2019 as derived from the application of prospective survival probability method
$p\_ob_{t+a}^{x+a,i}$	represents the observed cohorts as recorded by official statistics

The model algorithms were implemented in R language within the GNU R computing environment (R core team 2020).

We disentangle the *annual theoretical cohort* net migration dividing the estimated cohort net migration by the number of years of the reference period, which can be interpreted as the estimated average annual net migration behaviours experienced by cohorts of survived individuals who have attained the mean age *x+a*. Therefore, we compute the *annual theoretical cohort* net migration rate as a proportion of the estimated *annual theoretical cohort* net migration (numerator) to the people exposed to the risk of migration (denominator). This allows the calculation of the (*x+a/2*) age *theoretical cohorts*, the arithmetical mean between *x age* cohorts observed at the beginning and *x+a age* cohorts at the end of the reference period, and the crude annual age-period-cohort net migration rate occurred for the (*x+a/2*) age *theoretical cohorts* of survived individuals during the mid-period (t+a/2) respectively.

LIMITATIONS

The estimated net migration suffers from some theoretical and operational limitations. For instance, by definition, we cannot distinguish immigration and emigration flows, nor intra-EU mobility and international migration, therefore not accounting for differentials within the observed population, e.g., between natives and immigrants.

Furthermore, net migration estimates may be affected by inaccuracy in existing population statistics: under-representations of cohorts at the end of the period may drive overestimates of population changes during the reference period; thus, errors in the population counts are reflected in the estimated net migration. To control and when possible partially reduce the bias, we adopted an adjustment strategy, based on the reverse survival method. This consists in the use of the reciprocals of the survival ratios to calculate the expected cohorts that would have been x years old at the beginning of the reference period. The rationale is that cohorts at time t comprise the survivors at the end of the reference period (time $t+a$), plus the migrants and deaths that occurred during the interval (a). The two methods give different estimates of the cohort-period net migration, following different timings of the migration and mortality events (concentrated at the beginning and the end of the period). To compensate errors in the scheduling of cohort events, that should be rather distributed over the interval, we calculate the adjusted age-cohort-period net migration rates, as the average of the two estimates, referring to migration events which occurred in the middle of the interval for the mean age ($x+a/2$) cohorts of survivors.

To assess the appropriateness of our model specifications, we compare the net migration rates obtained from the above-mentioned methodology to the population change. This check gives an indication of the coherence and regularity of the demographic components across age-groups. Theoretically, when appropriate life tables are available and cohort population datasets are accurate, the estimated net migration rates should fairly approximate migration tendencies by cohort along the population lifetime. The method validation is performed applying Eurostat NUTS3 and NUTS1 datasets, as well as official statistics provided by the Austrian National statistical Institute (Statistik Austria) at LAU levels.

METHOD VALIDATION

We further checked the trends in the spatial distribution of populations across time in the EU regions. Using migration data available from NSI, we validate the method at (a) NUTS3 level (through projections from 2019 to 2027) and (b) NUTS1 level (over the period 2011 to 2019); (c) we also validate the results for a selection of LAUs in Austria.

(a) European NUTS3 using Eurostat demographic projection datasets from 2019 to 2027, baseline scenario

We check the consistency of the estimated net migration rates with the rates derived from the official demographic projection datasets provided by Eurostat at NUTS3 level, baseline scenario, for the period from 2019 to 2027 (8-year interval)[3]. We compare the Eurostat annual age-period-cohort net migration rates with: estimates of annual age-period-cohort net migration rates derived from the application of the prospective survival probability method; estimates of annual age-period-cohort net migration rates derived from the application of the reverse prospective survival probability method; estimates of the adjusted annual period-cohort net migration rates, and; the annual exponential population growth rates.

[3] Results for all European NUTS3 are available upon request.

The comparison points at increasing divergences at older ages, e.g. for cohorts aged 40+, mirroring the increase in mortality along age groups. Nevertheless, we conclude that the method generates results in line with the demographic projection trends.

(b) European NUTS1 using Eurostat datasets from 2011 to 2019

We extend the method operability at NUTS1 using official statistics available from Eurostat on immigration and emigration from 2011 to 2019.

The method is consistent with the generalization at NUTS1 for all cohorts, with some few exceptions for young adult and older cohorts. Specifically, higher estimated migration rates derive from under-enumeration of mortality levels for cohorts aged 18-24 years, whereas lower migration rates are a consequence of over-estimation of mortality among cohorts aged 66-72 years.

(c) LAU datasets on populations and internal and international migration made available by the Austrian NSI (2021) from 2011 to 2019.

When comparing estimates of annual migration rates with annual period-cohort rates derived from official statistics, discrepancies are not higher than standard error levels (0.04). Thus, we can conclude that the method seems to be satisfying in profiling age-specific migration rates at LAU levels.

CONCLUSION

The role of migrations in shaping demographic territorial differences has been well recognized and supported by the literature (Davis, 1965, Rees and Kupiszewski ,1999, Rees et al, 2017). However, this role has not been sufficiently understood at local level because of the lack of age-specific population data at high spatial resolution level.

We undertook a major effort in handling the large coverage and variations in the data coming from different national sources to establishing a dataset of population by age at LAU level for the entire EU. It complements the figures available from the 2011 Census, anticipating the release of gridded census statistics foreseen for 2023.

By exploiting this new data set, we estimate age-specific net migration by adapting census survival methods to handle the large coverage and variations in datasets coming from different national sources. Addressing gaps in migration data, we contributed to enhance research on determinants of territorial disparities across European Municipalities during intercensal periods.

REFERENCES

Austrian Institute of Statistics, 2021, https://data.statistik.gv.at/web/meta.jsp?dataset=OGD_bevstandjbab2002_BevStand_2016, login on August 8th 2021

Batista e Silva, F., Dijkstra, L., & Poelman, H. (Forthcoming). The JRC-GEOSTAT 2018 population grid. JRC Technical Report. Forthcoming [JRC Technical Report]. https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat

Bell M, Charles-Edwards E, Kupiszewska D, Kupiszewski M, Stillwell J, Zhu Y. 2015. Internal migration data around the world: Assessing contemporary practice. Population, Space and Place 21(1): 1–17

Chenery, H. B. (1960), "Patterns of Industrial Growth," American Economic Review, vol. 50, pp. 624-54.

Chenery, H. B. (1979), Structural Change and Development Policy, London: Oxford University Press

Coale Ansley J. and Demeny Paul, Regional Model Life Tables and Stable Populations (Princeton, Princeton University Press, 1966).

Castles, S.; Miller, M., 1993, International Population Movements in the Modern World. Vancouver, Canada, Canada: The Macmillan Press.

Chenery, H. B. (1960), "Patterns of Industrial Growth," American Economic Review, vol. 50, pp. 624-54.

Chenery, H. B. (1979), Structural Change and Development Policy, London: Oxford University Press

Champion A (ed.). 1989. Counterurbanization: The Changing Pace and Nature of Population Deconcentration. Edward Arnold: London

Davis, K. (1965). The Urbanization of the Human Population. Scientific American, 213(3), 40–53. JSTOR.

De Vries, J. (2013). European urbanisation: 1500-1800 (1. iss. in paperback).

Dyson, T. (2011). The Role of the Demographic Transition in the Process of Urbanization. Population and Development Review, 37, 34–54. https://doi.org/10.1111/j.1728-4457.2011.00377.x

DEGURBA, Correspondence Table Degree of Urbanisation - Local Administrative Units, see: https://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_DEGURBA, login on August 12th 2021

European Commission. Statistical Office of the European Union. (2021). Applying the degree of urbanisation: A methodological manual to define cities, towns and rural areas for international comparisons: 2021 edition. Publications Office. https://data.europa.eu/doi/10.2785/706535

EC, 2020, European Commission public initiative, Rural development and Long term vision for rural areas, https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12525-Rural-development-long-term-vision-for-rural-areas_en, login on July 29th 2021

Eurostat, 2021 Urban-rural typology, Statistics Explained: https://bit.ly/eurostat2017urbrur, login on July 29th 2021

Eurostat-LAU, https://ec.europa.eu/eurostat/web/nuts/local-administrative-units, login on August 12th 2021

Fielding A. 1989. Migration and urbanisation in Western Europe since 1950. The Geographical Journal 155(1): 60–69.

Gompertz B (1825). On the Nature of the Function Expressive of the Law of Human Mortality; and on a New Mode of Determining the Value of Life Contingencies, Philosophical Transactions of the Royal Society, Vol. 115, No. 27, pp. 513-585.

Goujon, A., Jacobs-Crisioni, C., Natale, F., Carlo, L., Jean-Philippe, A., Filipe, B. e S., Claudio, B., Alessandra, C., Daniela, G., Sona, K., & others. (2021). The Demographic Landscape of EU Territories. In Publications office of the European Union. http://dx.doi.org/10.2760/658945

Gutiérrez Posada, D., Rubiera Morollón, F., & Viñuela, A. (2018). Ageing Places in an Ageing Country: The Local Dynamics of the Elderly Population in Spain: AGEING PLACES IN AN AGEING COUNTRY. Tijdschrift Voor Economische En Sociale Geografie, 109(3), 332–349. https://doi.org/10.1111/tesg.12294

Hastie, T. and Tibshirani, R. (1986) Generalized additive models. Statistical Science, 1, 297-310.

Kabisch, N., & Haase, D. (2011). Diversifying European agglomerations: Evidence of urban population trends for the 21st century: Diversifying European Agglomerations. Population, Space and Place, 17(3), 236–253. https://doi.org/10.1002/psp.600

Kashnitsky, I., de Beer, J., & van Wissen, L. (2020). Unequally ageing regions of Europe: Exploring the role of urbanization. Population Studies, 1–17. https://doi.org/10.1080/00324728.2020.1788130

Keyfitz N (1984). Choice of Function for Mortality Analysis: Effective Forecasting Depends on a Minimum Parameter Representation, in Vallin, J., J,H, Pollard, and L. Heligman (Eds.), Methodologies for the Collection and Analysis of Mortality Data. IUSSP, Ordina Editions. Liege, Belgium. pp. 225-241.

IOM, and McKinsey & Company. 2018. “More than Numbers: How Migration Data Can Deliver Real-Life Benefits for Migrants and Governments.” International Organization for Migration (IOM) and McKinsey & Company. https://publications.iom.int/system/files/pdf/more_than_numbers.pdf.

Islam, N. (2014), Towards a sustainable social model: Implications for the post-2015 agenda. United Nations DESA Working Paper No. 136, ST/ESA/2014/DWP/136

ISO-3166, CODES FOR THE REPRESENTATION OF NAMES OF COUNTRIES AND THEIR SUBDIVISIONS

Lewis, W.A. (1954), “Economic Development with Unlimited Supplies of Labor,” The Manchester School, Vol. 22, No.2, pp. 139-191.

Livi Bacci M., 2006, Introduzione alla demographia, Loescher Editor.

Perpiña Castillo, C., Aurambout, J. P., Proietti, P., Jacobs, C., & Lavalle, C. (2021). A demographic assessment of EU remote areas by 2050 (JRC Technical Report No. JRC124458). https://publications.jrc.ec.europa.eu/repository/handle/JRC124458

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ravenstein, 1889, Laws of Migration, Journal of the Royal Statistical Society Vol. 52, No. 2 (Jun., 1889), pp. 241-305 https://doi.org/10.2307/2979333, https://www.jstor.org/stable/2979333

Rees P, Kupiszewski M. 1999. Internal migration and regional population dynamics in Europe: a synthesis. Population Studies No.32; Council of Europe: Strasbourg

Rees P., Bell M., Kupiszewski M., Kupiszewska D., Ueffing P. , Bernard A. , Charles-Edwards E. and Stillwell J., 2017, Popul. Space Place;23, doi: 10.1002/psp.2036

Rogers, A.; Castro, L., 1983, Regional Migration Differentials. In: IIASA Nations Working Paper. Laxenburg, Austria: International Institute For Applied Systems Analysis.

Rowe, F., Bell, M., Bernard, A., Charles-Edwards, E., & Ueffing, P. (2019). Impact of Internal Migration on Population Redistribution in Europe: Urbanisation, Counterurbanisation or Spatial Equilibrium? Comparative Population Studies - Zeitschrift für Bevölkerungswissenschaft, 44, 201-233. https://doi.org/10.12765/CPoS-2019-18en

Sabater, A., Graham, E., & Finney, N. (2017). The spatialities of ageing: Evidencing increasing spatial polarisation between older and younger adults in England and Wales. Demographic Research, 36, 731–744. https://doi.org/10.4054/DemRes.2017.36.25

Sánchez, A.; Andrews, D., 2011, To Move or Not to Move: What Drives Residential Mobility Rates in the OECD? In: Working Papers 846: 40 [doi: 10.1787/5kghtc7kzx21- en].

UNDESA, United Nations Department of Economic and Social Affairs, 1970, Population Studies, n.47, Manual on methods of estimating population, Methods of Measuring Internal Migration, Manual VI

van den Berg, L., Drewett, R., & Klaassen, L. H. (1982). A Study of Growth and Decline. Elsevier. https://doi.org/10.1016/C2013-0-03056-3

Zelinsky W. 1971. The hypothesis of the mobility transition. Geographical Review 61: 219–249

SUPPLEMENTARY MATERIAL

DATA COVERAGE AND VALIDATION

To match geographical units in the national data to the GISCO 2018 classification, we employed in cascade fuzzy string matching of national names and codes against the GISCO data and geocoding using OpenStreetMap Nominatim. To harmonize age classes, the population was redistributed to single ages in proportion to the national age structure reported in Eurostat statistics. For some cases, the data were afterwards re-aggregated to standard 5-age year groups. For others, when the data were only available for higher geographical units, the downscaled figures at LAU level were estimated by applying the age structure from the NUTS3 level from Eurostat to the population totals derived from a 1km population grids produced in JRC (Batista e Silva et al., forthcoming).

To verify the correctness of the processing of the data collected from the NSIs we conducted two types of validations. First, the data by age were re-aggregated and compared to totals by age at national level for the same reference year obtained from Eurostat statistics. This validation was intended to verify the absence of anomalies in the processing of the age groups into standard intervals. The results of this validation indicate that the age structure at LAU level in all MS reflect almost exactly the national totals.

A second step of validation consisted in looking at the differences between the LAU total populations, against the total population reported at the higher geographical level of NUTS3 in Eurostat statistics. This step of validation was aimed at verifying the correctness of the reclassification and localization of the original geographical units reported in the national data into the standard GISCO classification of LAUs.

As synthetic indication, we calculate the median values of the percentage differences between the population re-aggregated from LAUs and the population in the corresponding NUTS3 regions. In most countries the differences are contained between +2% and -2%. The highest differences were recorded in small countries and in countries for which data at LAU level needed to be imputed using total population from high-resolution grids and age structure from national data.