ISSN 1831-9424

European Commission

MICSIM

A MICROSIMULATION MODEL FOR POPULATION DYNAMICS

EUR 31979 EN

Joint Research Centre

Contact information
Name: Claudio Bosco
Email: claudio.bosco@ec.europa.eu

EU Science Hub
https://joint-research-centre.ec.europa.eu

JRC138163

EUR 31979 EN

PDF ISBN 978-92-68-18308-3 ISSN 1831-9424 doi:10.2760/343039 KJ-NA-31-979-EN-N

Luxembourg: Publications Office of the European Union, 2024

How to cite this report: Bosco C., Zinn S., Ghio D., Teobaldelli M., Iacus S.M., *MicSim, a microsimulation model for population dynamics.* Publications Office of the European Union, Luxembourg, 2024, doi:10.2760/34303, JRC138163.

MICSIM

A MICROSIMULATION MODEL FOR POPULATION DYNAMICS

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The authors are grateful to Eurostat for providing the high-quality data that have been used in this study. We are also very grateful to Dr. Fabrizio Natale for his valuable support and suggestions throughout the development of the project.

AUTHORS

Claudio Bosco[1]
Sabine Zinn[2]
Daniela Ghio[3]
Maurizio Teobaldelli[4]
Stefano Maria Iacus[5]

[1] European Commission, Joint Research Centre
[2] Humboldt-Universität zu Berlin
[3] University of Toronto
[4] Arcadia SIT S.r.l.
[5] Harvard University

ABSTRACT

Demographers use microsimulation for studying individual life courses and their attributes; events are specified to be the result of stochastic processes based on predetermined probabilistic rules.

In this study, we developed and validated a microsimulation model to reconstruct individual's life courses and their interactions in different regions of the European Union. One of the main objectives of this study was to track migrants' pathways in the context of three population systems, namely Sweden, the Netherlands and Spain, from 2014 to 2018.

We used official datasets as data inputs to analyze and project future population dynamics. A revised version of the MicSim package which is part of the statistical software R has been used to model mobility and migration patterns at large scale. For this purpose, among other things, new functions were added to make the software more efficient concerning runtimes and data handling.

By analyzing the modelling results we conclude that MicSim has the potential to be applied for modelling migration and also more general population movements at a large scale.

The application of the MicSim package would provide policymakers with a valid instrument for the governance of migration accounting for the demographic and social patterns of migrants and their origin-destination contextual environments.

1. INTRODUCTION

Modern societies are characterized by multiculturalism, and this is evident throughout Europe. While this has always been the case and changes in population size and composition are largely foreseeable (European Commission, 2022), the complexity has heightened due to increased mobility and migration (Byram, 2016). In Europe, the current coexistence of varied populations and cultures (European Commission, 2023a), coupled with their dynamic societal development, presents a multitude of political challenges for decision-making (European Commission, 2022), challenges which span from demographic shifts to economic disparities and environmental sustainability (European Commission, 2023b).

For instance, migration played a pivotal role in countering the decline of the working-age population across EU territories from 2015 to 2019 with urban areas experiencing growth due to positive net migration, contrasting with rural and intermediate areas facing declines (Ghio et al., 2022). Coordinated policy measures aimed at economic viability and infrastructure have the potential to counteract this decline and thus attract younger population groups (Ghio et al., 2022).

The COVID-19 pandemic amplified the disadvantageous situation of rural and intermediate areas concerning population density and composition, prompting the European Commission to launch an initiative aimed at pinpointing specific actions and remedies to bolster adaptability and resilience (European Commission, 2023b). As reported by Goujon et al. (2021a) the COVID-19 infection rates exhibited age and gender disparities, with women under 60 more affected due to frontline healthcare roles, while men over 60 faced higher fatality rates, notably in Italy.

In general, population ageing is one of the most serious problems in the EU. It is often accompanied by great economic inequality, which differs greatly from one area of Europe to another. Clearly, changes in the distribution and size of a population have a significant impact on social issues and institutions (Verdugo, 2012). Understanding their potential implications at different levels of management (European, national or regional) is essential for devising and implementing targeted and effective future European policies (Ruiz, 2016; Alvarez Alvarez et al., 2021). However, to implement informed and reactive policy measures, a tailored and, above all, evidence-based approach is needed (Goujon et al., 2021b; European Commission, 2022; European Commission, 2023b).

In this context, having robust tools to help predict, analyse, and evaluate policy impacts is of fundamental importance (Novotný, 2023). In particular, statistical analysis of the dimensions, and evolution of the population over time and space can provide valuable insights into the challenges faced by societies. For example, a sound analysis of migration processes and a sustainable projection of the labor market may lead to a change in migration laws, which in turn may lead to higher labor migration.

Through demographic modelling (especially, population monitoring and forecasting), it is possible to investigate the dynamics and structure of a human population, focusing on its size, composition, and changes it undergoes over time. The changes result from inflows to and outflows from this population and reflect the impact of births, deaths, and migration. When this in- and outflows occur disproportionately concerning fertility, mortality and migration, the age structure of a population or subpopulation changes (National Research Council, 1997). The information necessary to conduct demographic modelling comes from many different sources such as census data, micro or macro surveys, and commercial or administrative data. Due to this large amount of data, it is necessary to synthesize the information to estimate populations and their characteristics (Lomax & Smith, 2017).

For example, to anticipate future population trends and guide decision making, various demographic forecasts have been performed in Europe. Eurostat regularly publishes demographic projections essential for understanding the challenges related to population ageing and labour market dynamics (see for instance: https://ec.europa.eu/eurostat/web/population-demography/population-projections).

Furthermore, the analysis of the numerous socioeconomic and financial variables, of the development strategies implemented and of any existing disparities at supranational, national, regional and local levels allow research institutes such as the Joint Research Center (DG-JRC) of the European Commission, academic organizations and think tanks to contribute with specialized studies on specific aspects of demographic change, such as, for instance, the fiscal impact of immigration (Belanger et al., 2020) and ageing (Belanger et al., 2022) in EU, the implications and needs of long-term care at national (Baldini et al., 2007) or European level (Grubanov-Boskovic et al., 2021; Belmonte et al., 2023) or the transmission of COVID-19 within British communities (Spooner et al., 2021). These projections provide a valuable basis for policymakers, helping them to understand and address the demographic challenges and opportunities facing Europe.Among the tools of demographic modelling, dynamic microsimulation — a computerized technique for simulating trajectories following predetermined probabilistic rules — allows mapping the behaviour of individual entities within a larger population over time, thus providing policymakers with valuable insights into the potential consequences of various policy scenarios. Microsimulation stands out as a powerful and versatile method for leveraging current demographic databases (mostly derived from sample surveys) to trace potential impact of policy-induced changes on individuals, families, and entire populations.

The concepts of microsimulation have been introduced for the first time by Orcutt in 1957 (Orcutt, 1957; Orcutt et al., 1961). The term 'micro' is related to the possibility of estimating individual characteristics and modelling changes individual by individual. The events that individuals experience during their life course are the result of stochastic processes based on predetermined probabilistic rules. Monte Carlo methods are used to simulate events, which involve generating random numbers to determine values from the distribution of the processes being studied (hence the word 'simulation') (Zagheni, 2015). Demographers utilize microsimulation for estimating and projecting populations and their attributes.

The main difference between micro and macro-simulation consists in microsimulation using a sample rather than the total population, individual rather than aggregated data and in not averaging numbers but in repeating random experiments (Van Imhoff, 1998). As stated by Zagheni (2015) "Microsimulation allows population scientists to go beyond some of the simplifying assumptions made in standard macro demographic models".

A large majority of dynamic microsimulation models developed since the 60s have a link with the model of Orcutt (1961). In the last sixty years, many different types of microsimulation models have been developed and applied in different fields to investigate various research questions.

These models have been used for many purposes that include the projection of lifetime behaviours (e.g. labour force participation), policy analysis and the investigation of complex demographic dynamics such as population ageing (Marois, 2021) and population movements (Lutz, 2019). Microsimulation models have been also largely used to evaluate the results of long-term programs; education, long-term care and pensions (Morrison, 2017; Carrière et al., 2008) are among the many different things investigated using this type of models.

An overview of the different applications of this technique in social sciences and in other areas is available in Li & O'Donoghue (2013), Schofield et al. (2018), Kim (2021), and O'Donoghue (2021).

When population interactions and heterogeneity have a central role or the number of variables is large, microsimulation is the most appropriate technique to be applied (Van Imhoff, 1998; Spielauer, 2011). Compared to macrosimulation, microsimulation can accommodate for interactions between individuals and variables much better than 'macro' models (Zagheni, 2015); microsimulation models are the preferred choice when individual pathways are relevant.

Individual-level models are the preferred choice when processes are complex at the macro level but easier to understand at the micro level or when the recording of the history of individuals is relevant. Despite its attractiveness, both to gain insights into population dynamics and for policy-relevant analyses, microsimulation stores important limitations that need to be considered.

Microsimulation is computationally intensive. This was a main problem for early implementations of microsimulation models but the recent developments in computing technology means that this is less of a limitation now than it was in the past. To run a microsimulation model is also necessary to collect detailed demographic data that are often not available or that require a particular effort to be collected. When this type of data is available, the calibration process could be difficult (Zagheni, 2015). Also, the assumption of the importance of selected parameters and the choice of a specific model structure can be considered as major limitations in microsimulation modelling (Çağlayan 2018).

The objective of this study was to track migrant pathways in the context of three population systems (Sweden, the Netherlands and Spain), using official datasets, to analyze and project population dynamics by applying a new version of a microsimulation model. We used a revised and extended version of the R package MicSim. Our analyses on mobility and migration patterns have also been used to validate the new functions that were added to make MicSim more efficient concerning runtimes and data handling, enabling large-scale applications with the package.

The revised MicSim package has shown to be a valid instrument to support policymakers in the governance of population dynamics.

2. METHODS

2.1 THE MICSIM PACKAGE

The MicSim R package (https://cran.r-project.org/web/packages/MicSim/index.html) implements a generic continuous-time microsimulation model suitable for a wide range of applications in demography. MicSim simulates individual life courses that are defined as a sequence of states that individuals pass over time and the time interval between the states (Zinn, 2014).

Continuous-time multi-state models are commonly applied to characterize individual life courses. The stochastic process used for describing individual life courses usually comes from the family of Markovian processes (Kijima, 1997).

MicSim is a discrete-event simulation model where each demographic event brings changes in the state of an individual. The model considers an open population whose life courses are specified to evolve along three continuous time scales: individual age, calendar time, and the time spent by an individual in the current state. The model is parameterized with an initial population (called base population), transition rates, and when applicable, a population of migrants (Zinn, 2014).

MicSim can manage a wide range of situations relevant for population projections and forecasts. It implements two classes of stochastic models: non-homogeneous semi-Markov processes and non-homogeneous continuous-time Markov processes.

Non-homogeneous continuous-time Markov processes are used to specify life-course events over age and calendar time; with semi-Markov processes duration dependence is also considered. This second type of models assumes that the propensity of individuals to change their state may also depend on the amount of time passed in that state.

Before life courses can be constructed, the transition pattern applied during simulation has to be detailed. Information on the events that should be considered have to be provided to the model by the user. Two types of information are required by the MicSim package: first, for all events and time scales along which dependency is assumed, transition rates and, second, a transition matrix indicating which transition rate applies to which transition or event, respectively.

MicSim has been designed to be very flexible. In its first version (1.1.0) the model was mainly applied to small datasets. One of its main limitations was the computational time required to simulate the life courses of large numbers of individuals, especially when a large number of attributes is associated to each individual.

Considering the need to carry out large-scale simulations (e.g., European level), the MicSim package was revised concerning its runtimes requirements. We removed from the source code of MicSim any dependency from other packages (e.g., 'chron' R package) that negatively influenced its computational time. MicSim has also been extended with several new functions to handle time and date data. A strong work for optimizing the code has been done and checks for fertility transitions leading to newborns more efficiently were implemented.

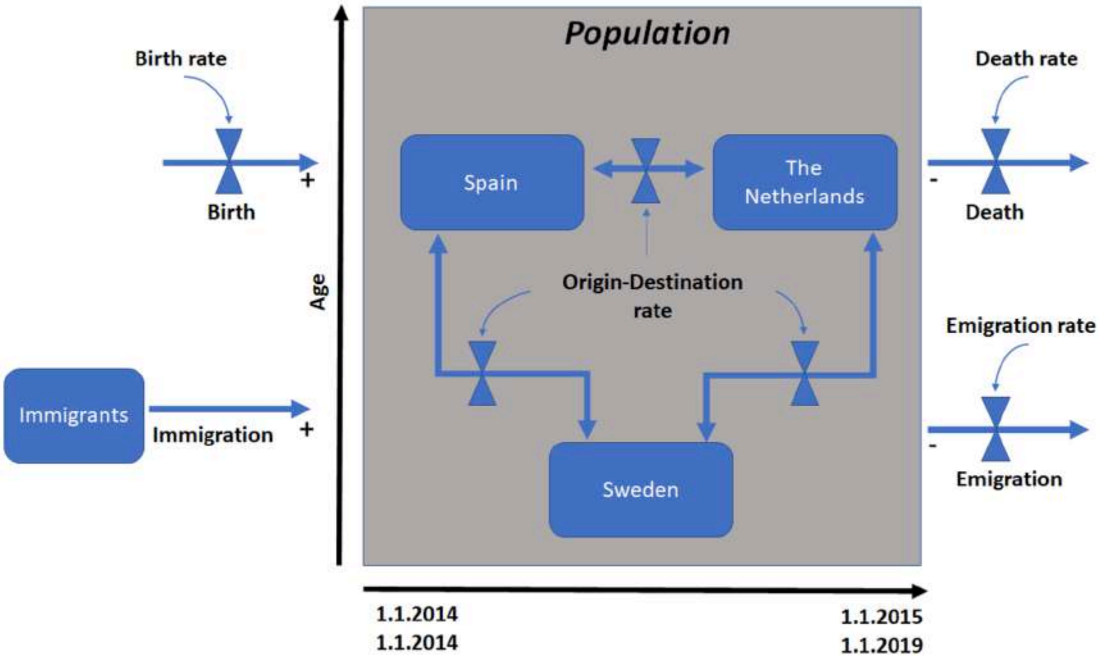
Besides these major changes, further innovations were added to the model. A link with the mother ID was added to newborns together with the possibility to inherit attributes from the mother such as, for example, nationality. New semantic checks on input data were introduced as well. Semantic checks mitigate inconsistencies between input data, parameters and outputs; the use of terse array-based constraints emphasizes the focus on the coherent flow of the information and data among modules (Bosco, 2019; Bosco et al., 2013; De Rigo et al., 2013). These include checks on the time and age ranges in the initial population, in the population of immigrants and the transition rates provided to the model.

All these changes led to version 2.0.0 of the software. In this version, large sets of data can be managed, allowing the application of the model at the European level and for a broader set of applications. Since MicSim is part of the statistical software R, it is open source and available under the GPL-2 license (https://cran.r-project.org/web/packages/MicSim/index.html).

2.2 THE MICROSIMULATION MODEL

The microsimulation model of this study considers mortality, internal movement between Spain, the Netherlands, and Sweden, emigration out of these countries, and parity specific fertility as events. The simulation runs along two simulation horizons: one running from the 1st of January 2014 to the 1st of January 2015, and one running from the 1st of January 2014 to the 1st of January 2019. We selected two different time intervals for testing the model in a short and in a longer period. The decision to run the model over 5 years and not on a longer time is due to the availability of the data necessary to run and validate the model.

FIGURE 1. Diagram of the microsimulation model.



In total, ages from 0 to 100 years are simulated. Figure 1 illustrates the model. MicSim requires transition rates for all events, along all possible ages and the whole range of the simulation horizon considered. Since MicSim works in continuous time, these rates have to be delivered for continuous age and calendar time.

2.3 DATA

Microsimulation uses data specified at individual or household level. MicSim works with a virtual population consisting of all individuals that are considered during simulation. It can be regarded as a longitudinal sample from a synthetically constructed population that resembles a real one. As input MicSim requires an initial population, transition rates for all events considered during simulation, and the distribution of migrants entering the virtual population along simulation time.

2.3.1 THE INITIAL POPULATION AND THE POPULATION OF MIGRANTS

To equip our model, we utilized Eurostat data (https://ec.europa.eu/eurostat/web/main/data) related to three European countries: Spain, the Netherlands and Sweden. Our data include information on a wide variety of demographic characteristics such as age, sex, place of origin and citizenship.

The input microdata related to male and female population aged between 0 and 99 years were obtained using the Eurostat Datasets of population on the 1st of January by age and sex (https://ec.europa.eu/eurostat/cache/metadata/en/demo_pop_esms.htm) with the reference year 2014. The initial population fed into MicSim covers 10% of the total population of Spain, Sweden and the Netherlands (about 7.5 million individuals). Considering that MicSim works on a continuous time base, individuals need a date of birth to be modelled. The day and month of birth of each individual in the microdata were randomly calculated taking into account the yearly age group and leap years.

Data related to international immigration were obtained using information on the total immigration in each of the selected countries after removing the internal movements. The modelling input related to immigration microdata was created using the Eurostat dataset of immigration by five-year age group, sex and citizenship with an age definition equal to 'reach' (https://ec.europa.eu/eurostat/cache/metadata/en/migr_immi_esms.htm).

Similarly, to the microdata for the permanent resident population, the dates of birth and immigration of each individual were randomly calculated taking into account the five-year age group and leap years.

2.3.2 TRANSITION RATES

From Eurostat datasets, we also derived all the transition rates by country, age and sex related to mortality, emigration and internal movement among the analyzed countries. The rates were calculated for each age class considered in the simulation (i.e., from 0 to 99 years of age). In both of our simulation settings (i.e., for the simulation 2014-2015 and the simulation 2014-2019), we assume transition rates to be constant along calendar time.

2.3.2.1 MORTALITY RATES

The Eurostat datasets of mortality at national level by age and sex (https://ec.europa.eu/eurostat/cache/metadata/en/demo_mor_esms.htm) and of population on the 1st of January by age and sex (https://ec.europa.eu/eurostat/cache/metadata/en/demo_pop_esms.htm) were used to calculate the annual probability of dying by age and sex.

For the simulation running from the 1st of January 2014 to the 1st of January 2015, the annual probability of dying was specified as the ratio between the annual flow and the mid-year population, calculated as the arithmetic mean of the initial (1st of January 2014) and final (1st of January 2015) stocks of the population.

For the simulation performed over the time- interval 1st of January 2014 – 1st of January 2019, the probability of dying was calculated as the ratio between the average annual value of the flow (e.g., average annual mortality calculated in the five years) and the mid-year population, calculated as the arithmetic mean of the initial (1st of January 2014) and final (1st of January 2019) stocks of the population in the reference period.

The probability of dying was converted to transition rates by using equation 1 (Fleurence & Hollenbeak, 2007):

(1)

$$r = -\frac{1}{t}\ln(1-p)$$

where *p* is the probability of dying by country, age and sex, *r* is the transition rate and *t* is the unit of time.

2.3.2.2 EMIGRATION AND INTERNAL MOVEMENT RATES

There is not a uniform definition of 'an emigrant', and, very often, methodologies for computing emigration rates are conditioned by data availability, see also (Willekens et al., 2017). Our way of calculating emigration rates is not the only one either, but it is a pragmatic one that yields feasible emigration rates.

We computed the rates related to internal movements between Spain, the Netherlands and Sweden, and emigration out of these countries by using the emigration and immigration by five-year age group, sex and citizenship(https://ec.europa.eu/eurostat/cache/metadata/en/migr_immi_esms.htm) dataset from Eurostat.

Since Eurostat data for immigration and emigration refer to five-year age groups, we derived emigration and internal movements rates for each yearly age, required by the model, from the five-year age group flows, considering they are constant for the five-year age class period to which they refer.

For the simulation running from the 1st of January 2014 to the 1st of January 2015, the rates were computed as the ratio between the annual flow and the mid-year population, calculated as the arithmetic mean of the initial (1st of January 2014) and final (1st of January 2015) stocks of the population. For the simulation performed over the time interval: 1st of January 2014 -1st of January 2019, the rates were calculated as the ratio between the mean annual value of the flow in five years and the mid-year population, calculated as the arithmetic mean of the initial and final stocks of the population in the reference period.

2.3.2.3 FERTILITY RATES

Data relating to cohort-specific (i.e., by age and parity) fertility rates and parity progression ratio (i.e., the proportion of women who progress from one parity to the next with parity ranging from 0 to 4+), for women aged 15 to 49 years old, were calculated using data collected from the National Institutes of Statistics of Spain, Sweden and the Netherlands. Specifically, we use the number of women by age group and parity to calculate the age-specific fertility rates by the number of births. So, we derive the age-order specific rates. We checked the consistency of the 'typical pattern of the average parities', which increases with age, by computing the cumulative order-specific fertility rates to be compared with the total fertility rate.

In our simulation the number of births is estimated by applying fertility rates to women only (i.e. "female-dominant" model) and, therefore, the number of males is assumed to be irrelevant to rates of childbearing (Preston et al., 2001). Consequently, births belong by definition to the same segment (i.e. nationality at birth) of the population as the mother (Preston et al., 2001).

3. RESULTS AND DISCUSSION

To validate whether the implemented model and the MicSim software deliver feasible output, we compared the results of the two simulation settings (2014-2015 and 2014-2019) with Eurostat data related to mortality, fertility, migration and living population in the same time intervals.

3.1 MORTALITY

The model's predictions related to mortality are very close to the official data from Eurostat. Figure 2 shows very similar results for the number of female and male deaths aggregated at five-year age groups predicted by the model (for the time interval 1/1/2014 – 1/1/2015) and the data from Eurostat in the Netherlands. The good performance of the microsimulation model (and the related software) is also confirmed by the analysis over multiple runs for all the countries, where, again, the median of the model's prediction is in line with the official data.

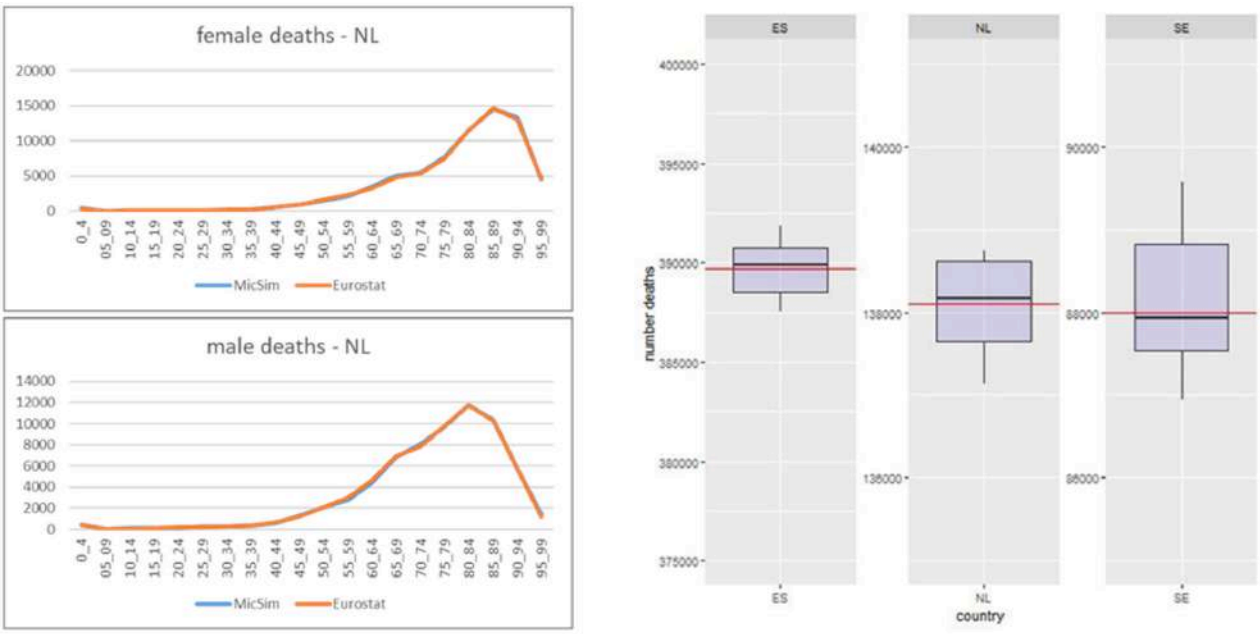
A similar result in predicting the number of deaths was obtained for all the countries also in the time interval 1/1/2014 – 1/1/2019 (see Annex I).

3.2 MIGRATION

Also for migration, the model well predicts the number of emigrants for all the considered age classes.

Looking at the time interval from January 2014 to January 2015, the left panel of Figure 1 shows the comparison between the number of emigrants predicted by MicSim's microsimulation model and the corresponding official data from Eurostat. The right panel of Figure 3 depicts the distribution of internal movements between Spain, the Netherlands, and Sweden over ten runs of MicSim using boxplots. The black horizontal lines mark the median of the simulated numbers of internal movements and the red line is the related number projected by Eurostat. We see a high concordance between simulated data and the Eurostat projections. One of the main benefits of dynamic microsimulation is its capability to mimic heterogeneity in populations by stochastic fluctuations. This fluctuation is also visible in the boxplots of Figure 3. All in all, the stochastic pattern is feasible, the interquartile ranges (depicted by the grey areas) are not as far and over- or underestimation is not an issue.

FIGURE 2. On the left, the trend by five-year age groups related to male and female deaths in the Netherlands for the time interval 1/1/2014 – 1/1/2015. In orange are the Eurostat data and in blue the values predicted by the MicSim model. On the right, boxplots related to ten modelling runs reporting the difference between the simulated projections and Eurostat data for the total number of deaths in Spain (ES), the Netherlands (NL) and Sweden (SE). The red line represents Eurostat data, and the black line represents the median of the different simulations.



Looking at the results obtained by applying the model in the time interval January 2014 – January 2019, we find similar results concerning our MicSim's model capability to reproduce the Eurostat predictions. However, we also detected deviations we are going to discuss.

We have found differences in the number of female emigrants in the Netherlands for the age groups 10-14 and 60-64 years. This difference is not evident in the 2014-2015 simulation setting. For Spain, no significant deviations between the emigration figures can be found either.
Figure 4 reports the corresponding figures by 5-year age groups related to female emigration in Spain and the Netherlands. Also for males, the data show a similar difference between Spain and the Netherlands and some variability affect also the modelling predictions related to Sweden (see Annex I).

Within the model, mortality such as emigration can be mapped by a non-homogeneous continuous-time Markov process, it means the model, from a statistical point of view, calculates the number of deaths and emigrants applying the same approach. Because mortality, as previously mentioned, is always well predicted by the

FIGURE 3. On the left, the trend by five-year age groups (x-axis) related to the total number of male and female emigrants in Spain for the time interval 1/1/2014 – 1/1/2015. In orange are the Eurostat data and in blue the values simulated by the MicSim model. On the right, boxplots related to ten modelling runs reporting the difference between the simulated projections and Eurostat data for the internal mobility among Spain (ES), the Netherlands (NL) and Sweden (SE). The red line represents Eurostat data, and the black line represents the median of the different simulations.

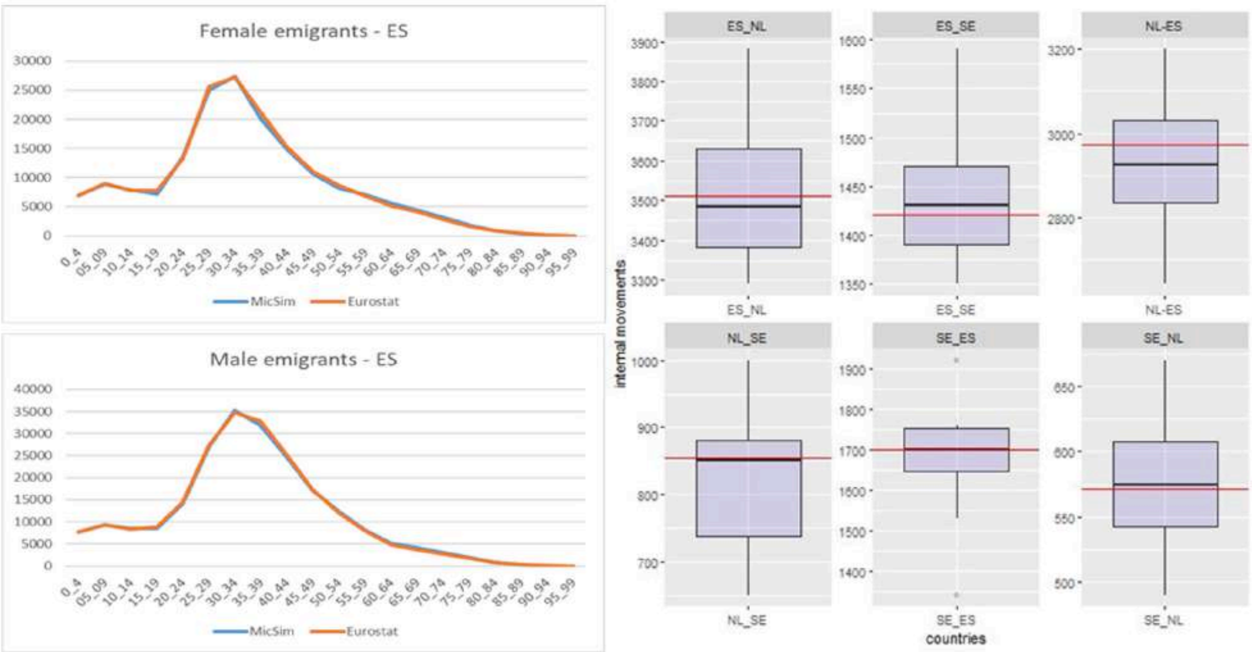


FIGURE 4. The trend by 5-year age groups (x-axis) related to the total number of female emigrants in Spain (ES) and the Netherlands (NL) for the time interval 1/1/2014 – 1/1/2019. In orange are the Eurostat data and in blue the values simulated by the MicSim model.

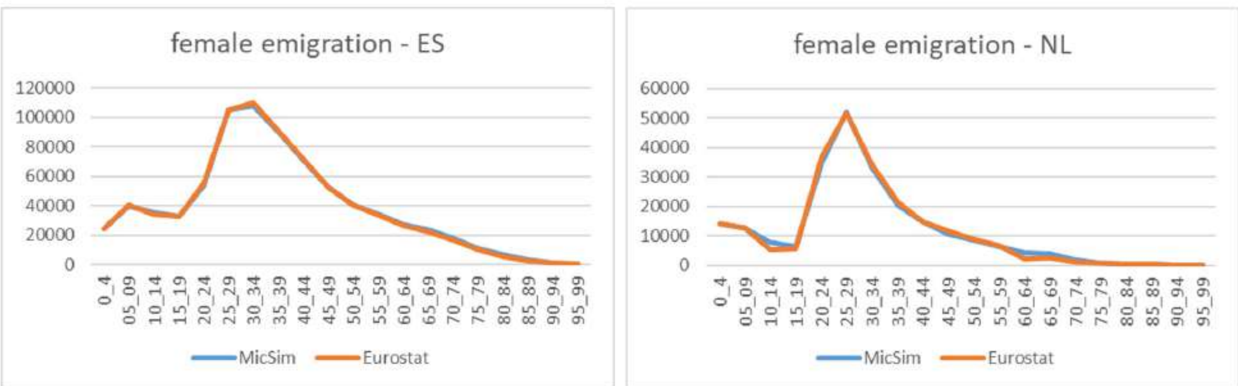
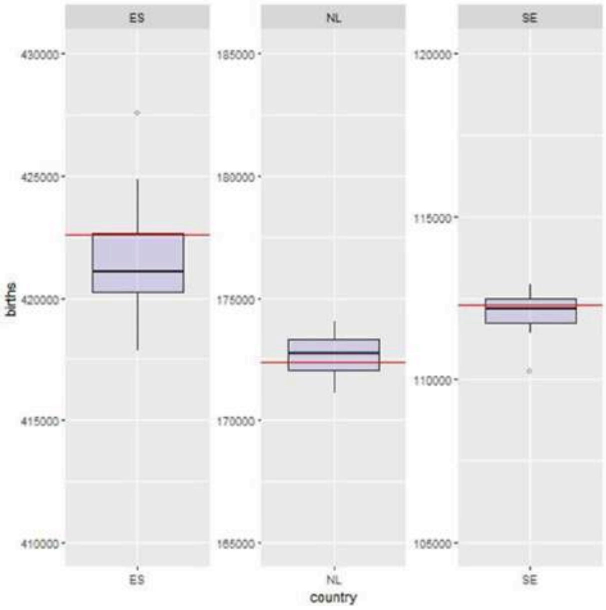


FIGURE 5. Boxplot related to ten modelling runs showing the difference between the simulated projections and Eurostat data for the number of births calculated over a year, from January 2014 to January 2015. The red line represents Eurostat data, and the black line represents the median values over ten simulation runs.



model in all the countries and for all age classes and sex, the modelling architecture is not at the basis of the observed 'fluctuations' affecting migration estimates. The only difference is in the transition rates used in the model.

Within the Eurostat database, the data related to emigration are only available aggregated in classes of 5 years. Because the R package MicSim performs continuous-time microsimulations for population projection, as already mentioned in the data section, we assigned a rate to each age considering they are constant for the five-year age class period to which they refer. In addition, because one of the aims of this study was also to test the performance of the MicSim package in poor data conditions, we decided to calculate a single mean rate, for sex and age, for all the 5 years of the simulation. The use of a 'mean' migration rate during our simulation is one of the main causes of the difference with Eurostat data.

3.3 FERTILITY

Also regarding the number of newborns predicted by the model, our predictions are very close to official data from Eurostat. Considering the time interval 1/1/2014 – 1/1/2015, just for Spain the model's uncertainty is remarkably higher when compared to Sweden and the Netherlands and the median of the simulated data differs from the Eurostat number (see Figure 5).

By analyzing the model's prediction from January 2014 to January 2019 we noticed a slight underestimation of the model when compared to official Eurostat data (in the order of 1-2 % - see Annex I). We suspect that this difference is due to the different databases used to estimate the fertility rates. The MicSim model uses data from the National Statistical Offices to calculate the parity progression ratio and from this the parity-specific fertility rates that are fed into the MicSim model; and not the official Eurostat data. The reason for this approach is that the Eurostat data do not contain sufficient parity information to calculate parity-specific rates.

3.4 LIVING POPULATION

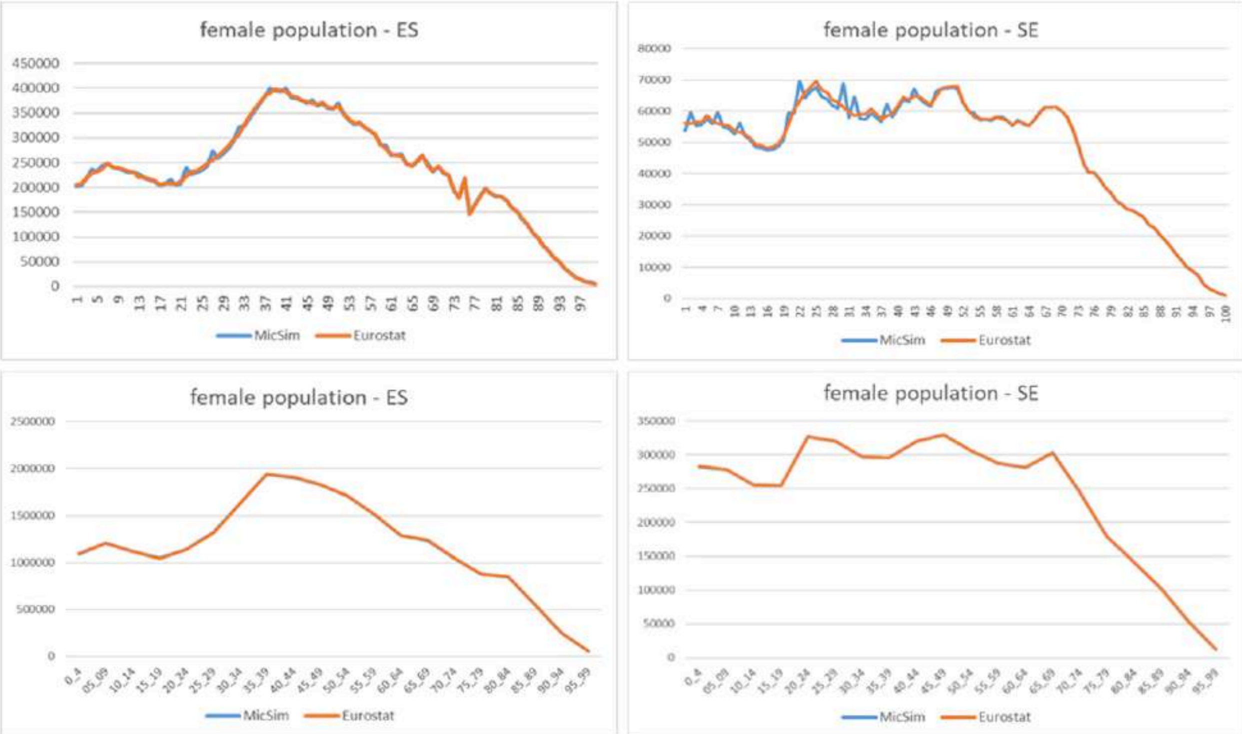
MicSim is also able to predict total population numbers with high precision and accuracy as compared to the Eurostat benchmark.

Figure 6 shows an example of the living female population in Spain and Sweden for the simulation horizon January 2014 - January 2015 (individuals residing in the country on the 1st of January 2015). The upper panel depicts the population numbers for one-year age groups and the lower panel shows them aggregated at five-year age groups.

Although overall there is a high degree of agreement between the simulated figures and the Eurostat values, there are notable fluctuations in the simulated values around the Eurostat benchmark for the one-year age groups, particularly strong in Sweden.

This fluctuation is most likely caused by the use of data for emigration and internal movements aggregated in classes of 5 years. Because we are using within the model a mix of rates from 1 to 5 years, the model's results should be analyzed using a 5-year interval. The aggregation of the simulation results over five age years leads to a levelling of these fluctuations.

FIGURE 6. Comparison between predicted female and male total population by age classes (x-axis) in Spain and official Eurostat data. On the bottom the data aggregated using 5-year age groups and on the top the values for each age class.



By looking at the time interval 1/1/2014 – 1/1/2019 for estimates related to the total living population by sex and age class we noticed a similar trend but also some differences. Figure 7 shows the female population by 5-year age classes; if for Spain the model's predictions almost perfectly overlap with Eurostat data, in the Netherlands (but a similar behaviour is visible also in the data related to Sweden – see Annex I) for the classes ranging from 15 to 35 years the model under- and over-estimate when compared with Eurostat data but the mean of these under- and over-estimates looks still in line with data from Eurostat.

As mentioned before, the use of a 'mean' migration rate during our simulation is probably one of the main causes of this difference with Eurostat data but not the only one.

As reported in the section on data, to create the initial population and the population of immigrants required by the model, it was necessary to randomly attribute to each individual a day and month of birth and immigration. By creating different sets of data for the initial population and the immigrants and running the model over 5 years applying the same rates to both initial populations, we noticed some differences in the final results. Figure 8 shows an example of these differences. A different date of birth is responsible for moving people from one transition rate to another within the same simulation interval. This could bring to some differences in the final population, especially for age intervals with the major differences between transition rates.

FIGURE 7. Comparison between predicted female total living population by 5-year age groups (x-axis) in Spain and the Netherlands versus official Eurostat data.

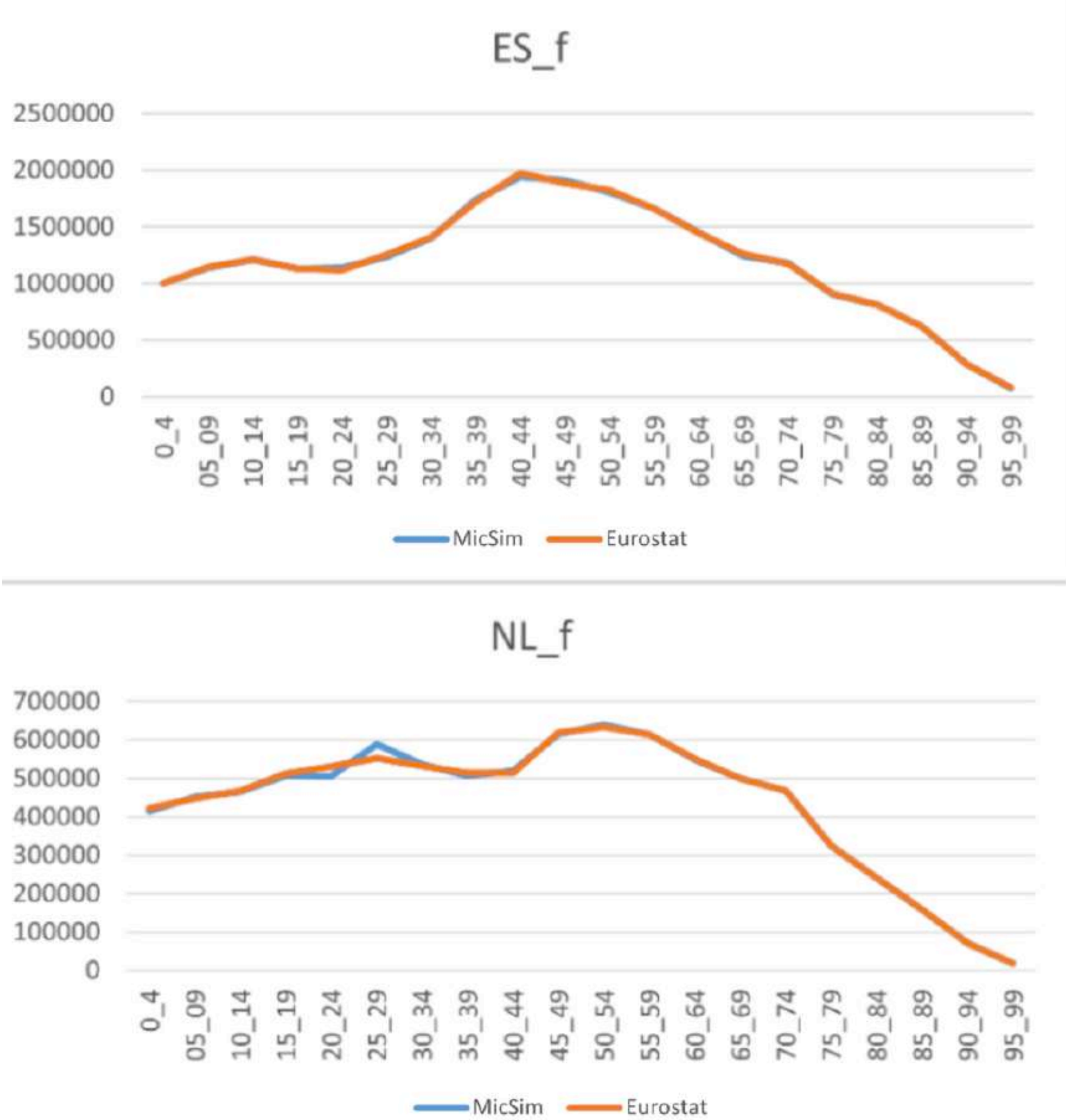


FIGURE 8. Comparison between two different total populations of only males in the Netherlands obtained by applying two different sets of data for the initial population fed into the model.

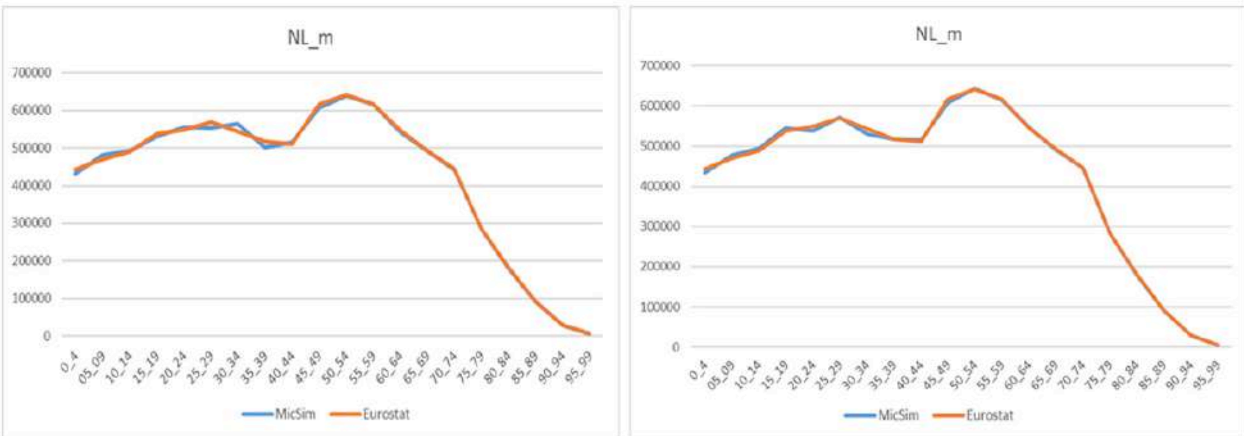
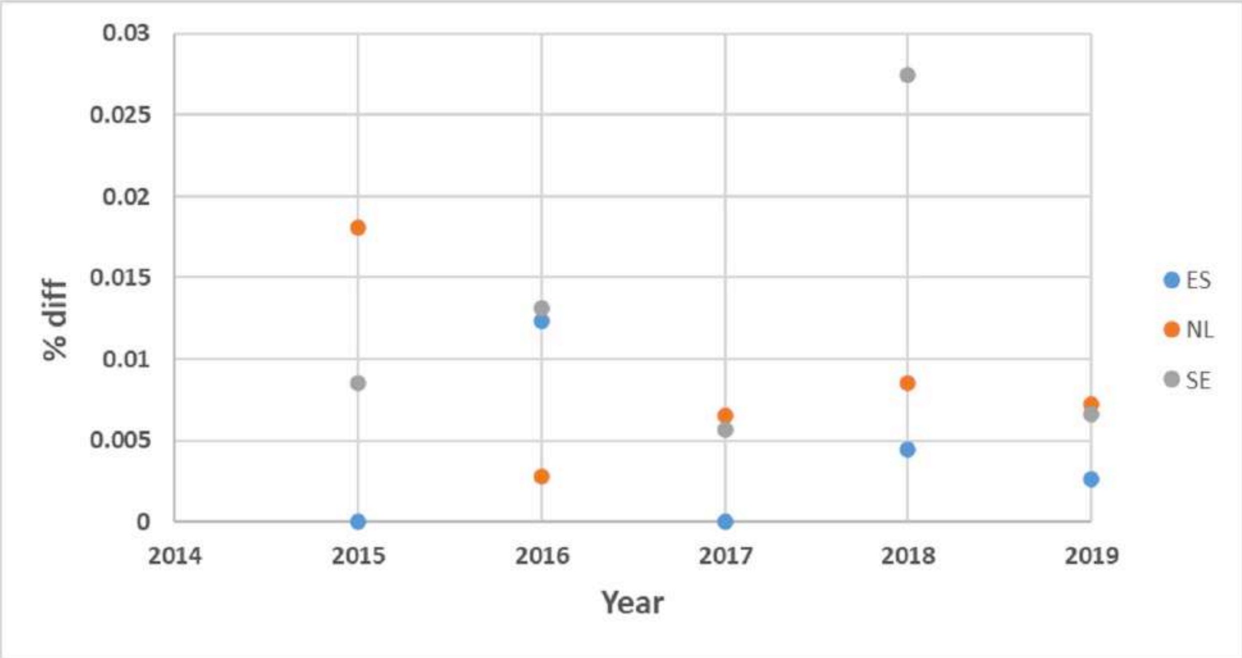


FIGURE 9. Percentage distance between the total population reported by Eurostat (https://ec.europa.eu/eurostat/cache/metadata/en/demo_pop_esms.htm) for all the years from 2015 to 2019 and the total population calculated using information reported by Eurostat for fertility, mortality, emigration and immigration.



Another possible reason for these discrepancies with official Eurostat data could be linked with the dataset we used to extract the initial population, the population of immigrants and all the rates applied in the model. Figure 9 shows the difference existing in some countries and for some years between the total population reported by Eurostat and an expected population calculated by applying the Balancing Equation of Population Change (equation 2):

(2)

$$N(t) = N(0) + B[0,t] - D[0,t] + I[0,t] - O[0,t]$$

where :
N(t) = number of people alive at time t
N(0) = number of people alive at time 0
B[0,t] = number of births between time 0 and t
D[0,t] = number of deaths between time 0 and t
I[0,t] = number of immigrants between time 0 and t
O[0,t] = number of emigrants between time 0 and t

An imbalance in equation 2, sometimes referred to as an “error of closure” (Preston et al., 2001), could contribute to an increase in the distance between predicted and official data, especially for emigration where statistical offices find it difficult to have precise estimates.

4. CONCLUSIONS

Because of the results we discussed in the previous paragraphs, we conclude that MicSim has the potential to be applied for modelling migration and also for more general population movements at a large scale, even if data are scarce or not available for each year of the simulation period. Although the architecture of the model is overall efficient and validated by the present work, fluctuations in the estimates may be due to the use of rates calculated with data having a different periodicity from that used in the simulation (e.g. five-year data used for annual simulations) or to the 'inclusion in microdata of information generally not available in aggregated data but at the same time having a high degree of detail, such as the exact date of birth and/or immigration of each individual.

Nevertheless, MicSim's model can be used for forecasting and projections and could serve as a 'new' tool for measuring the impacts of immigrants on native populations in different countries. The use of MicSim for modelling migration could be an added value for policymakers to take decisions for managing these processes.

Ongoing developments of the package include the inclusion of linkages between individuals. In this way, for example, families and the inheritance of behaviours such as the transmission of fertility through the maternal line can be modelled and simulated. Another extension of the MicSim package is the inclusion of a household structure to model, for example, the intergenerational transmission of home ownership.

REFERENCES

Alvarez Alvarez, M., Boccardo, S., Bosco, C., Choumelova, D., Conte, A., Ghio, D., Goujon, A., Kalan-taryan, S., Loeschner, J., Mcmahon, S., Natale, F., Schumacher, G., Scipioni, M., Sermi, F., Tarchi, D., Tintori, G. and Kompil, M., *Atlas of Demography*, Publications Office of the European Union, Lux-embourg, 2021, ISBN 978-92-76-32463-8, doi:10.2760/57296, JRC124598, [accessed Nov 17 2021, at URL: https://migration-demography-tools.jrc.ec.europa.eu/atlas-demography/]

Baldini, M., Mazzaferro, C., Morciano, M., 'Assessing the implications of long-term care policies in Italy: a microsimulation approach', *Politica economica*, 1, 2008, pp. 47-72, doi: 10.1429/26966

Belanger, A., Christl, M., Conte, A., Mazza, J. and Narazani, E., *Projecting the net fiscal impact of immigration in the EU*, EUR 30407 EN, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-23424-1, doi:10.2760/68847, JRC121937.

Bélanger, A., Mazza, J., Sabourin, P., *Demographic Microsimulation for the Estimation of the Fis-cal Impacts of Immigration and Ageing in Europe*, EUR 30947 EN, Publications Office of the Eu-ropean Union, Luxembourg, 2022, ISBN 978-92-76-46207-1, doi:10.2760/829011

Belmonte, M., Grubanov-Boskovic, S., Natale, F., Conte, A., Belanger, A. and Sabourin, P., *Demo-graphic microsimulation of long-term care needs in the European Union*, Publications Office of the European Union, Luxembourg, 2023, doi:10.2760/941182, JRC135303.

Byram, M., Aase, L., Beacco, J., Cavalli, M., Coste, D., Crisan, A., Fleming, M.F., & Maradan, O., *Multi-cultural Societies, Pluricultural People and the Project of Intercultural Education,* 2016.

Bosco, C., *Semantic array programming in data-poor environments: assessing the interactions of shallow landslides and soil erosion, Doctoral dissertation*, Loughborough University, 2019.

Bosco, C., Rigo, D. D., Dijkstra, T., Sander, G. & Wasowski, J., *Multi-Scale robust modelling of landslide susceptibility: Regional rapid assessment and catchment robust fuzzy ensemble*, in International Symposium on Environmental Software Systems, Springer, Berlin, Heidelberg, 2013, pp. 321-335.

Çağlayan, Ç., Terawaki, H., Chen, Q., Rai, A., Ayer, T. & Flowers, C. R., 'Microsimulation modeling in oncology', *JCO clinical cancer informatics*, 2, 2018, pp. 1-11.

Carrière, Y., Keefe, J., Légaré, J., Lin, X. et al., *Projecting the Future Availability of the Informal Support Network of the Elderly Population and Assessing Its Impact on Home Care Services*, Statistics Canada, 24, 2008.

De Rigo, D., Rodriguez-Aseretto, D., Bosco, C., Di Leo, M., & San-Miguel-Ayanz, J., *An architecture for adaptive robust modelling of wildfire behaviour under deep uncertainty.* In Environmental Software Systems. Fostering Information Sharing, 10th IFIP WG 5.11 International Symposium, ISESS 2013, Neusiedl am See, Austria, October 9-11, 2013, Proceedings 10, Springer Berlin Heidelberg, 2013, pp. 367-380.

European Commission, *The EU in the world of 8 billion people*, In Atlas od Demography, Knowledge Centre on Migration and Demography (KCMD) Data Portal, November 15 2022, [ac-cessed Feb 9 2024, at URL: https://migration-demography-tools.jrc.ec.europa.eu/atlas-demography/stories/AoD/2/S3.9].

European Commission, Bongiardo, D., Blasco, A., Icardi, R. et al., *Atlas of migration 2023*, Edited by D., Bongiardo, I., Crespi, and I., Sofos, Publications Office of the European Union, 2023a, [accessed Feb 9 2024, at URL: https://data.europa.eu/doi/10.2760/17123].

European Commission, *The impact of demographic change – in a changing environment*. Staff Working Document, SWD, 2023b, 21 final, 17 January. [accessed May 6 2024, at URL: URL: https://commission.europa.eu/system/files/2023-01/the_impact_of_demographic_change_in_a_changing_environment_2023.PDF].

Fleurence, R. L., Hollenbeak, C. S., 'Rates and probabilities in economic modelling: transformation, translation and appropriate application', *Pharmacoeconomics*, 25, 2007, pp. 3–6.

Ghio, D., Goujon, A., Natale, F., 'Assessing the Demographic Impact of Migration on the Working Age Population in European Regions', *Demographic Research*, 46, 2022, pp. 261-272, doi: 10.4054/DemRES.2022.46.9

Goujon, A., Natale, F., Ghio, D., Conte, A., Dukstra, L., *'Age, gender and territory of COVID-19 infections and fatalities*', Journal of Population Research, 2021a, doi.org/1https://commission.europa.eu/index_en0.1007/s12546-021-09263-3.

Goujon, A., Jacobs-Crisioni, C., Natale, F., Lavalle, C., (Eds), *The demographic landscape of EU territories: challenges and opportunities in diversely ageing regions,* EUR 30498 EN, Publications Office of the European Union, Luxembourg, 2021b, ISBN 978-92-76-27239-7, doi:10.2760/658945, JRC123046.

Grubanov-Boskovic, S., Ghio, D., Goujon, A., Kalantaryan, S., Belmonte, M., Scipioni, M., Conte, A., Gómez-González, E., Gomez Gutierrez, E., Tolan, S., Martinez Plumed, F., Pesole, A., Fernandez Macias, E. and Hernandez-Orallo, J., Healthcare and long-term care workforce: demographic challenges and potential contribution of migration and digital technology, EUR 30593 EN, Pub-lications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-30232-2, doi:10.2760/234530, JRC121698.

Kijima, M., *Continuous-time Markov chains, in Markov processes for stochastic modeling*, Chapman & Hall, 1997, pp. 167-241.

Kim, D., *Applications of Microsimulation Models to the Social Determinants of Health and Public Health.* In New Horizons in Modeling and Simulation for Social Epidemiology and Public Health, edited by D. Kim, 2021, doi:10.1002/9781118589397.ch9.

Li, J. & O'Donoghue, C.., 'A survey of dynamic microsimulation models: uses, model structure and methodology', Int. J. *Microsimul*. 6, 2013, pp. 3–55.

Lomax, N. M. & Smith, A. P., '*Microsimulation for demography', Australian Population Studies*, 1(1), 2017, pp. 73-85.

Lutz, W., Amran, G., Bélanger, A., Conte, A., Gailey, N., Ghio, D., et al., *Demographic scenarios for the EU: migration, population and education*, EUR 29739 EN, Publications Office, Luxembourg, ISBN 978-92-76- 03216-8, JRC116398, 2019, doi:10.2760/590301.

Marois, G. & Aktas, A., 'Projecting health-ageing trajectories in Europe using a dynamic microsimulation model', *Scientific reports*, 11(1), 2021, pp. 1-15.

Morrison, R. J., *Rates of Return in the Canada Pension Plan: Sub-populations of Special Policy Interest and Preliminary After-tax Results*, in New Frontiers in Microsimulation Modelling, Routledge, Abingdon, 2017.

National Research Council, *The Use of Multi-State Life Tables in Estimating Places for Biomedical and Behavioral Scientists*: A Technical Paper, Washington (DC), National Academies Press (US), 1997, PMID: 25121278.

Novotný, V. (2023). *Europe's demography and what to do about it.* In European View, 22(2), 2023, pp. 205-212, doi:10.1177/17816858231205403

O'Donoghue, C., *Introduction, in Practical Microsimulation Modelling*, Online edn, Oxford Academic, 2021, doi:10.1093/oso/9780198852872.003.0001.

Orcutt, G.H., 'A new type of socio-economic system', Rev. *Econ.* Stat, 39, 1957, pp. 116–123.

Orcutt, G., Greenberger, M., Korbel, J. & Rivlin, A., *Microanalysis of Socioeconomic Systems: A Simulation Study*, Harper and Row, Manhattan, 1961.

Preston, S., Heuveline, P., and Guillot, M., *Demography: Measuring and modeling population processes*, Wiley-Blackwell, 2001.

Schofield, D. J., Zeppel, M. JB., Tan, O., Lymer, S., Cunich, M. M., Shrestha, R. N., 'A brief, global history of microsimulation models in health: Past applications, lessons learned and future directions', *International Journal of Microsimulation,* 11(1), 2018, pp. 97-142, doi:10.34196/IJM.00175.

Spielauer, M., 'What is Social Science Microsimulation?', *Social Science Computer Review*, 29(1), 2011, pp. 9-20.

Spooner, F., Abrams, J.F., Morrissey, K., Shaddick, G., Batty, M., Milton, R., et al., 'A dynamic mi-crosimulation model for epidemics', *Soc Sci Med*, 291, 2021, 114461, doi: 10.1016/j.socscimed.2021.114461

Van Imhoff, E. & Post, W., *Microsimulation methods for population projection, in Population: An English Selection, special issue New Methodological Approaches in the Social Sciences,* 10(1), 1998, pp. 97-138.

Verdugo, R.R., *English Language Learners: An Exercise in Applied Demography.* In Opportunities and Challenges for Applied Demography in the 21st Century, Chapter 6, Edited by N., Hoque, and D. Swanson, Springer B.V. Press, Dordrecht, Heidelberg, London, and New York, 2012, pp.83-98.

Willekens, F., Zinn, S., Leuchter, M., 'Emigration Rates From Sample Surveys: An Application to Senegal', *Demography*, 54(6), 2017, pp. 2159-2179, PMID: 29039116, doi:10.1007/s13524-017-0622-y.

Zagheni, E., *Microsimulation in demographic research,* in International Encyclopedia of the Social and Behavioral Sciences, edited by J. D. Wright, 2nd edition, Vol 15, Oxford, Elsevier, 2015, pp. 343-346.

Zinn, S., 'The MicSim package of R: An entry-level toolkit for continuous-time microsimulation', I*nternational Journal of Microsimulation*, 7(3), 2014, pp. 3-32.

LIST OF FIGURES

ANNEXES

ANNEX I.

Mortality

FIGURE A-1. The trend by five-year age groups related to female deaths in Spain for the time interval 1/1/2014 – 1/1/2019. In orange the Eurostat data and in blue the values predicted by the MicSim model.

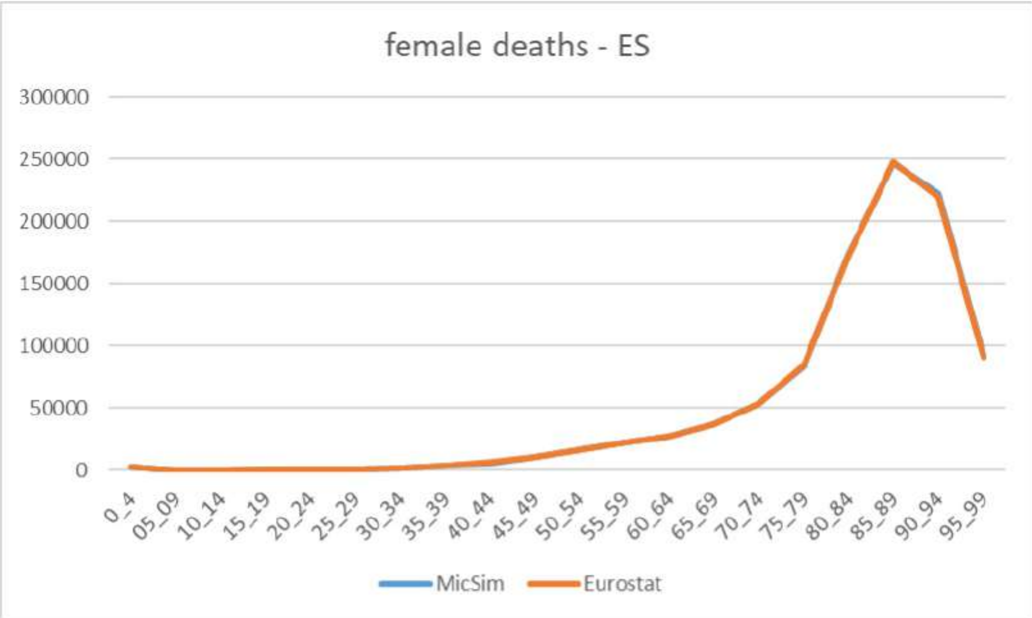


FIGURE A-2. The trend by five-year age groups related to male deaths in Spain for the time interval 1/1/2014 – 1/1/2019. In orange the Eurostat data and in blue the values predicted by the MicSim model.

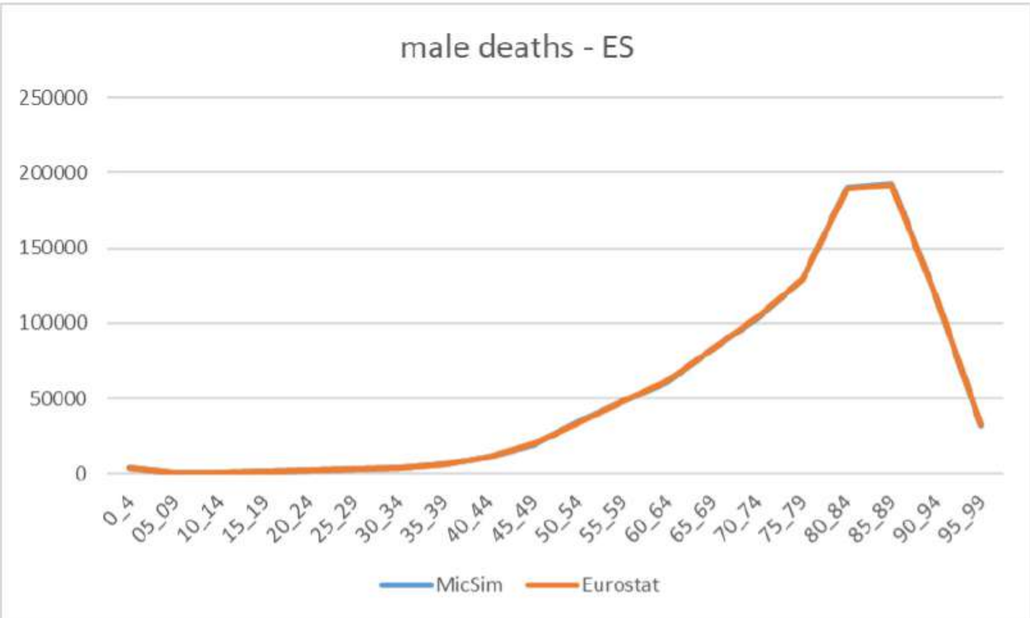


FIGURE A-3. The trend by five-year age groups related to female deaths in the Netherlands for the time interval 1/1/2014 – 1/1/2019. In orange the Eurostat data and in blue the values predicted by the MicSim model.

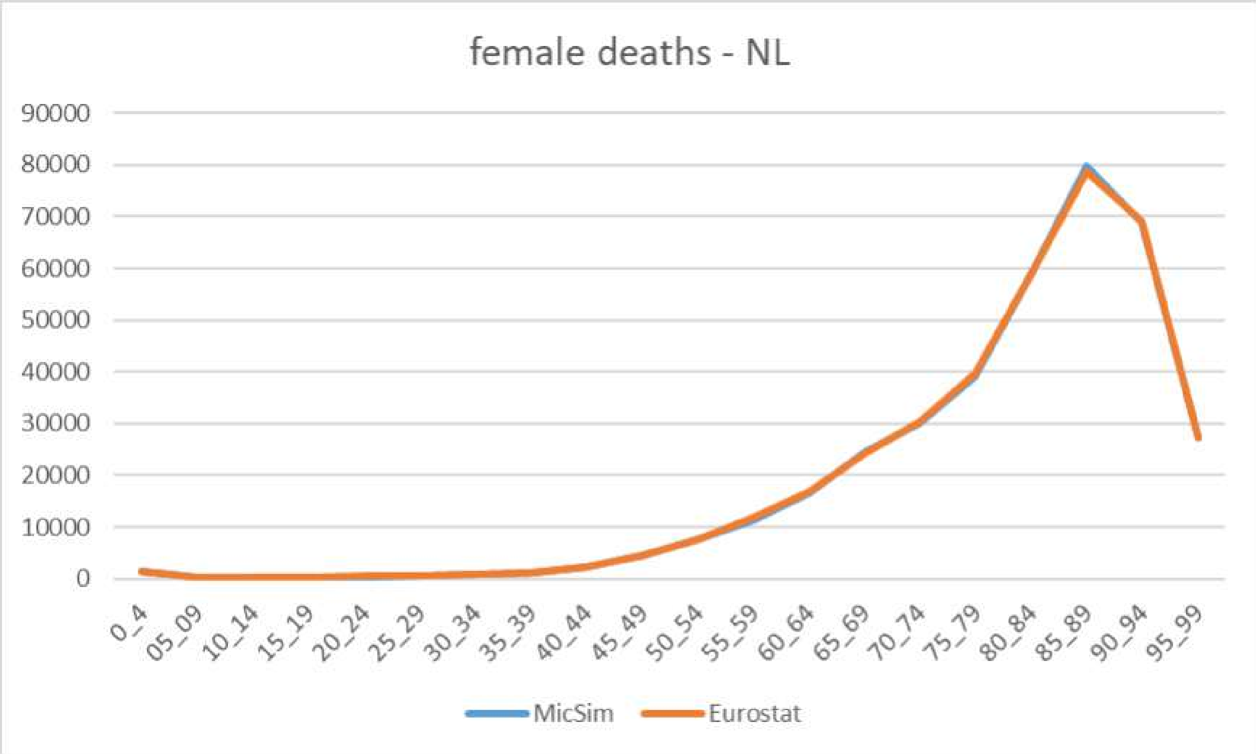


FIGURE A-4. The trend by five-year age groups related to male deaths in the Netherlands for the time interval 1/1/2014 – 1/1/2019. In orange the Eurostat data and in blue the values predicted by the MicSim model.

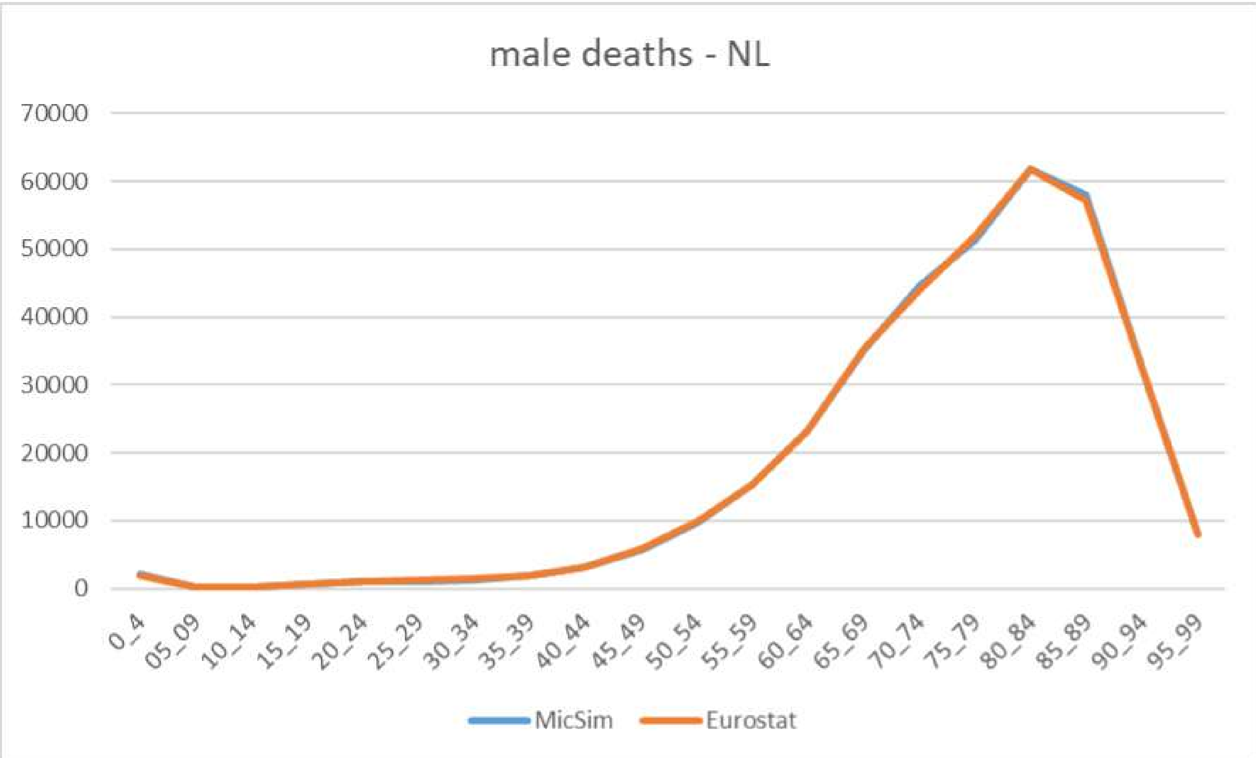


FIGURE A-5. The trend by five-year age groups related to female deaths in Sweden for the time interval 1/1/2014 – 1/1/2019. In orange the Eurostat data and in blue the values predicted by the MicSim model.

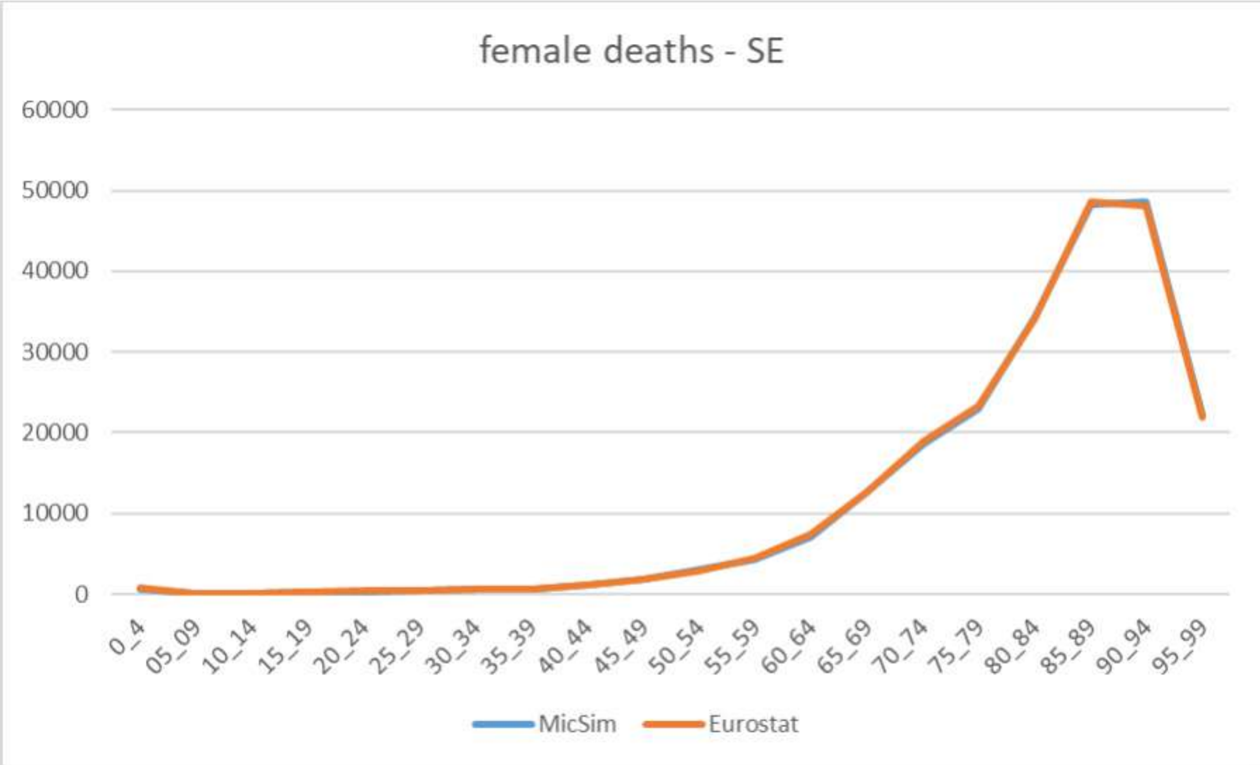
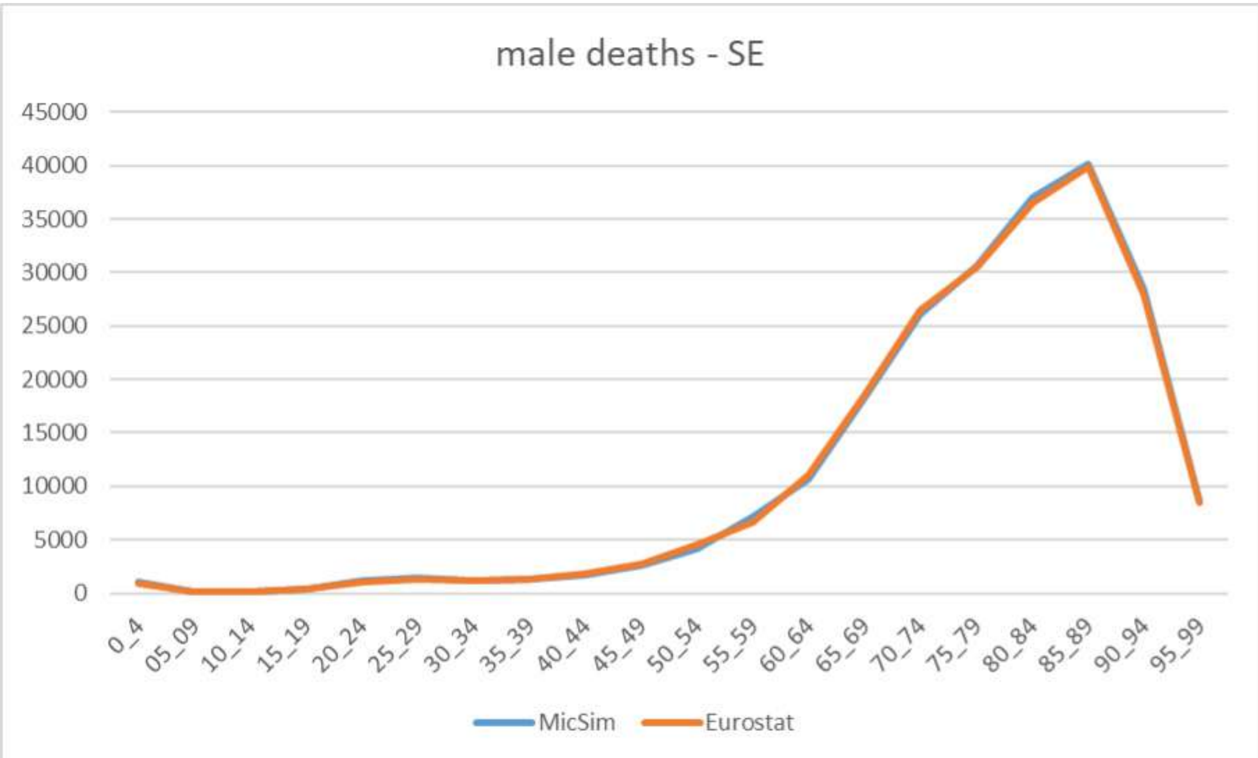


FIGURE A-6. The trend by five-year age groups related to male deaths in Sweden for the time interval 1/1/2014 – 1/1/2019. In orange the Eurostat data and in blue the values predicted by the MicSim model.



Emigration

FIGURE A-7. The trend by 5-year age groups (x-axis) related to the total number of female emigrants in Sweden for the time interval 1/1/2014 – 1/1/2019. In orange the official Eurostat data and in blue the values simulated by the MicSim model.

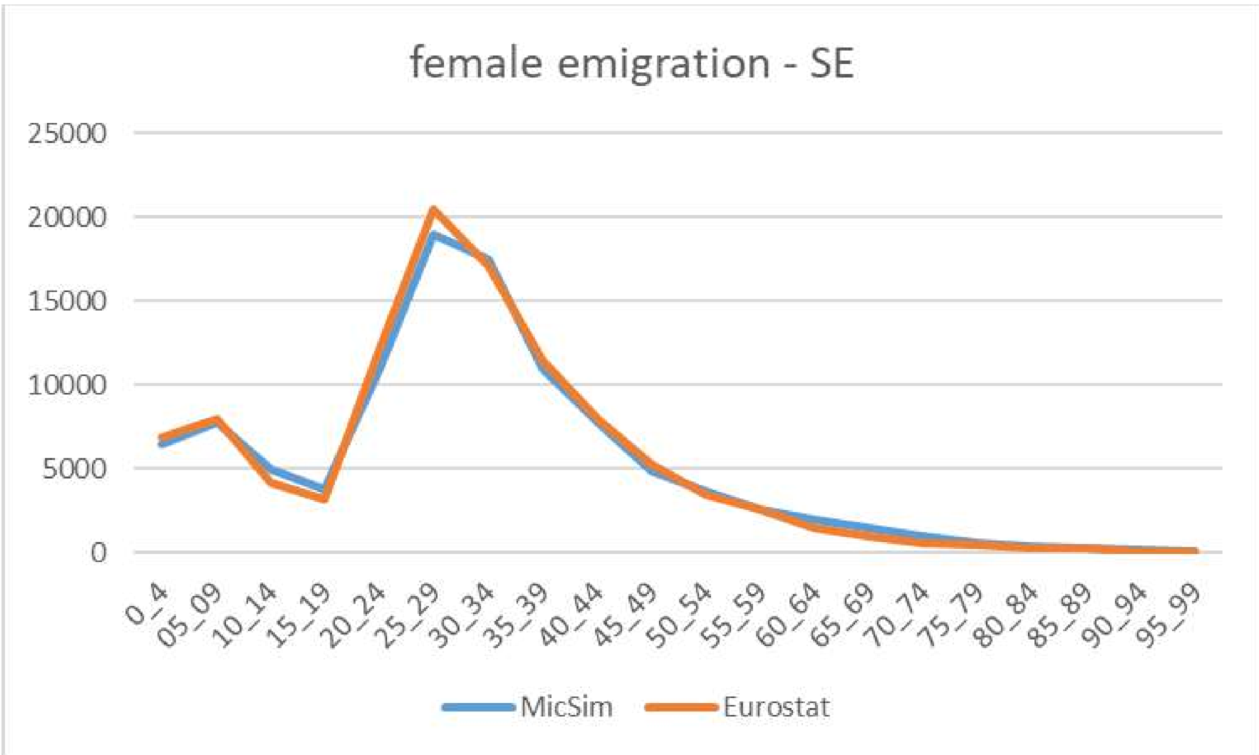
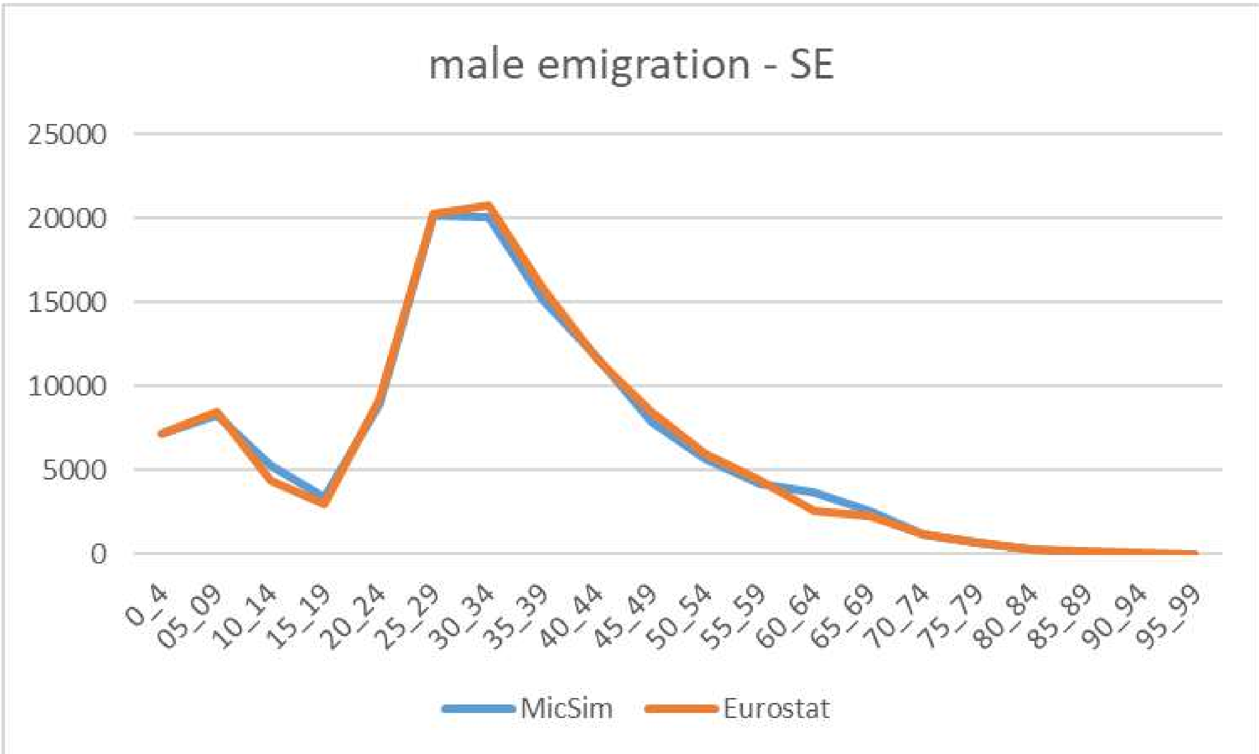
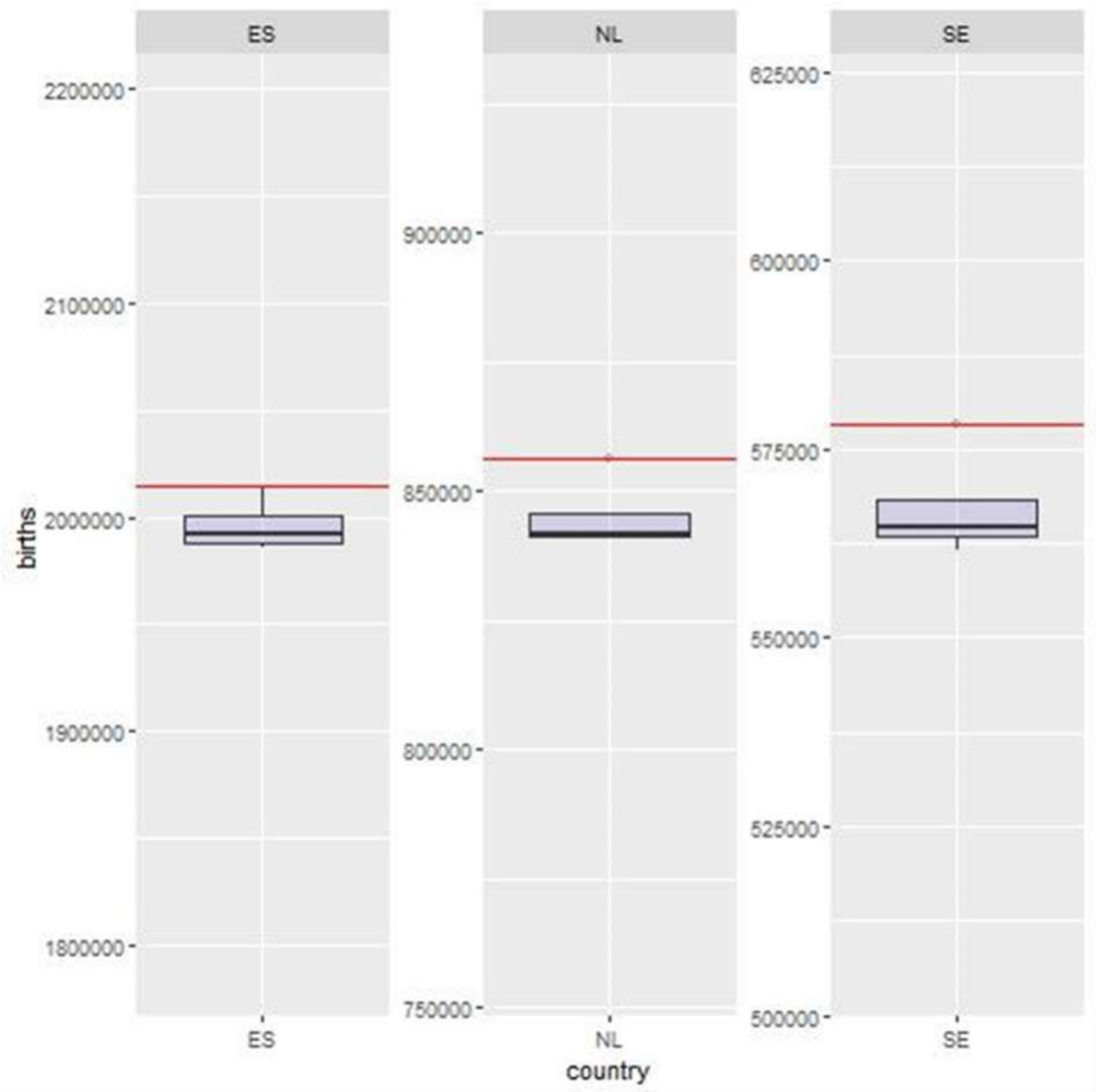


FIGURE A-8. The trend by 5-year age groups (x-axis) related to the total number of male emigrants in Sweden for the time interval 1/1/2014 – 1/1/2019. In orange the official Eurostat data and in blue the values simulated by the MicSim model.



Fertility

FIGURE A-9. Boxplot related to five modelling runs showing the difference between the simulated projections and Eurostat data for the number of births calculated over five years, from January 2014 to January 2019. The red line represents Eurostat data, and the black line represents the median values over five simulation runs.



Living population

FIGURE A-10. Comparison between predicted male total living population by 5-year age groups (x-axis) in Spain and official Eurostat data for the time interval 1/1/2014 – 1/1/2019.

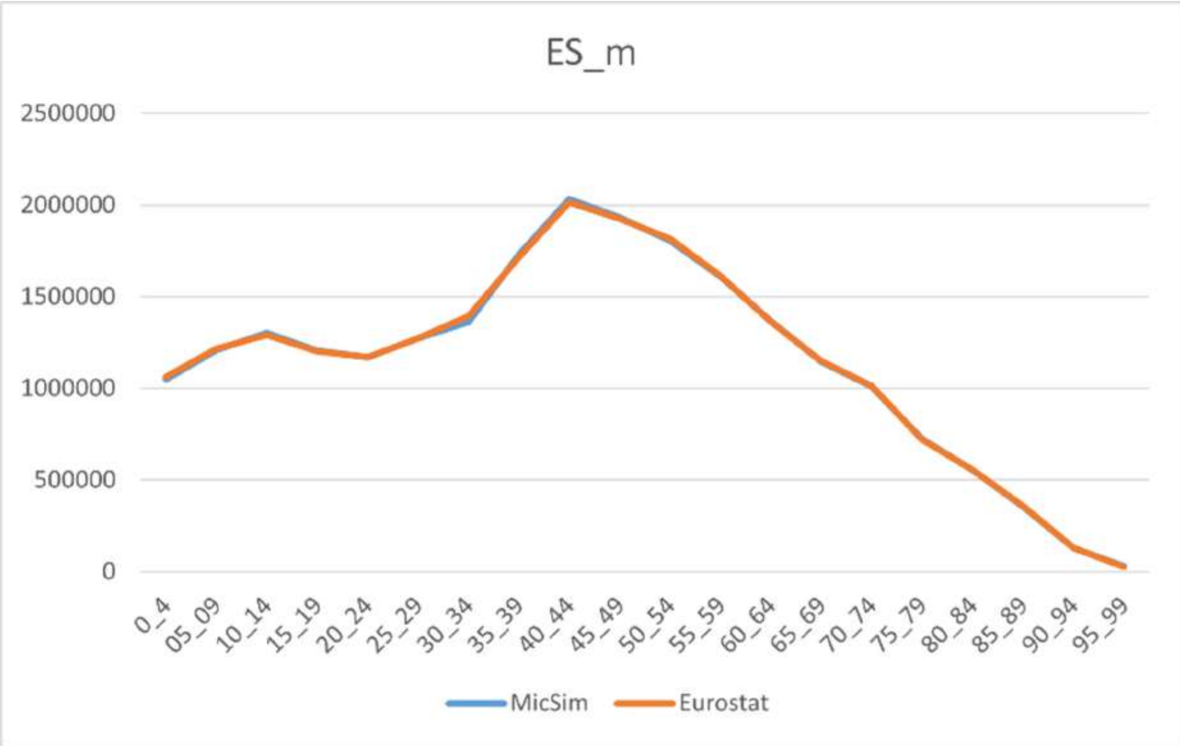


FIGURE A-11. Comparison between predicted male total living population by 5-year age groups (x-axis) in the Netherlands and official Eurostat data for the time interval 1/1/2014 – 1/1/2019.

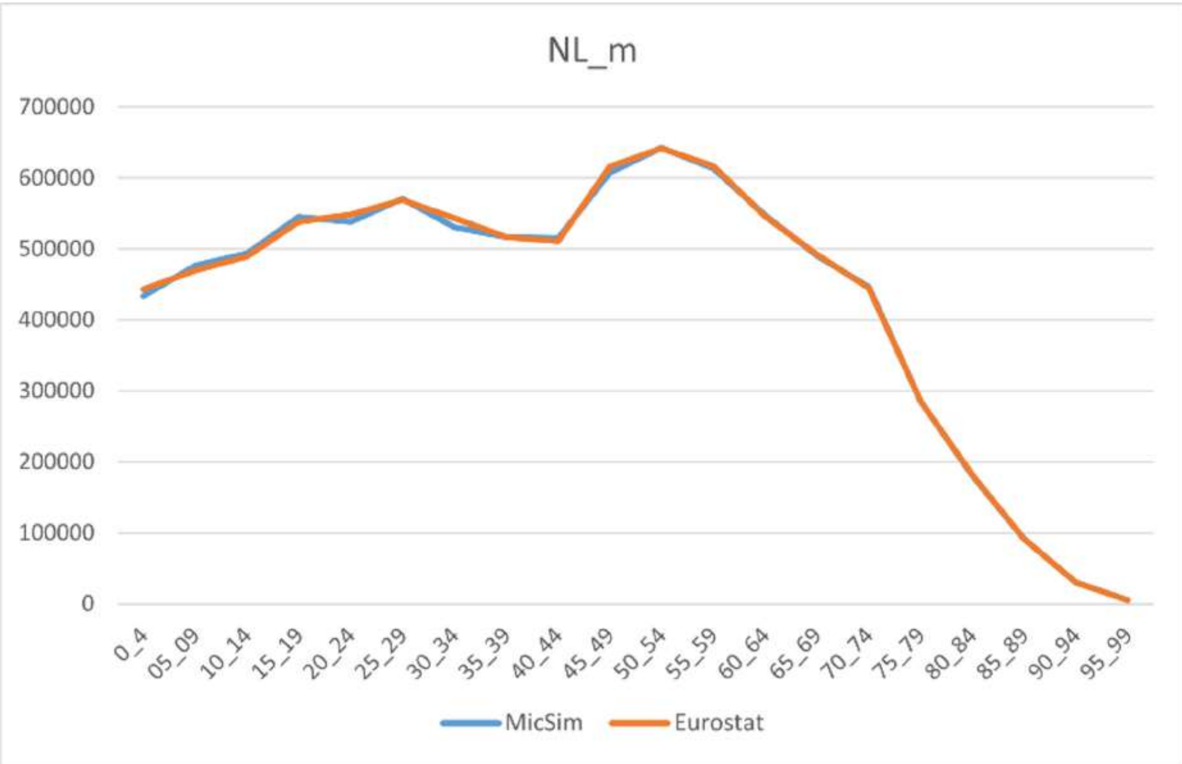


FIGURE A-12. Comparison between predicted female total living population by 5-year age groups (x-axis) in Sweden and official Eurostat data for the time interval 1/1/2014 – 1/1/2019.

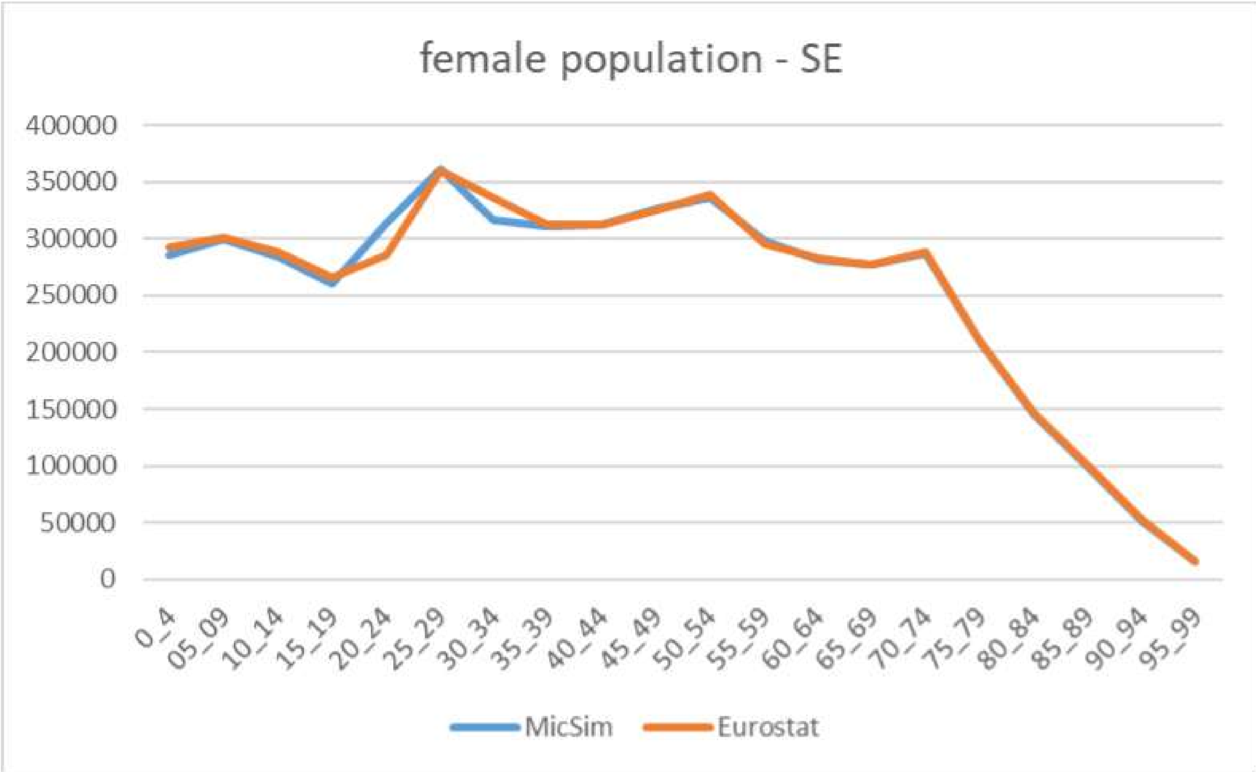
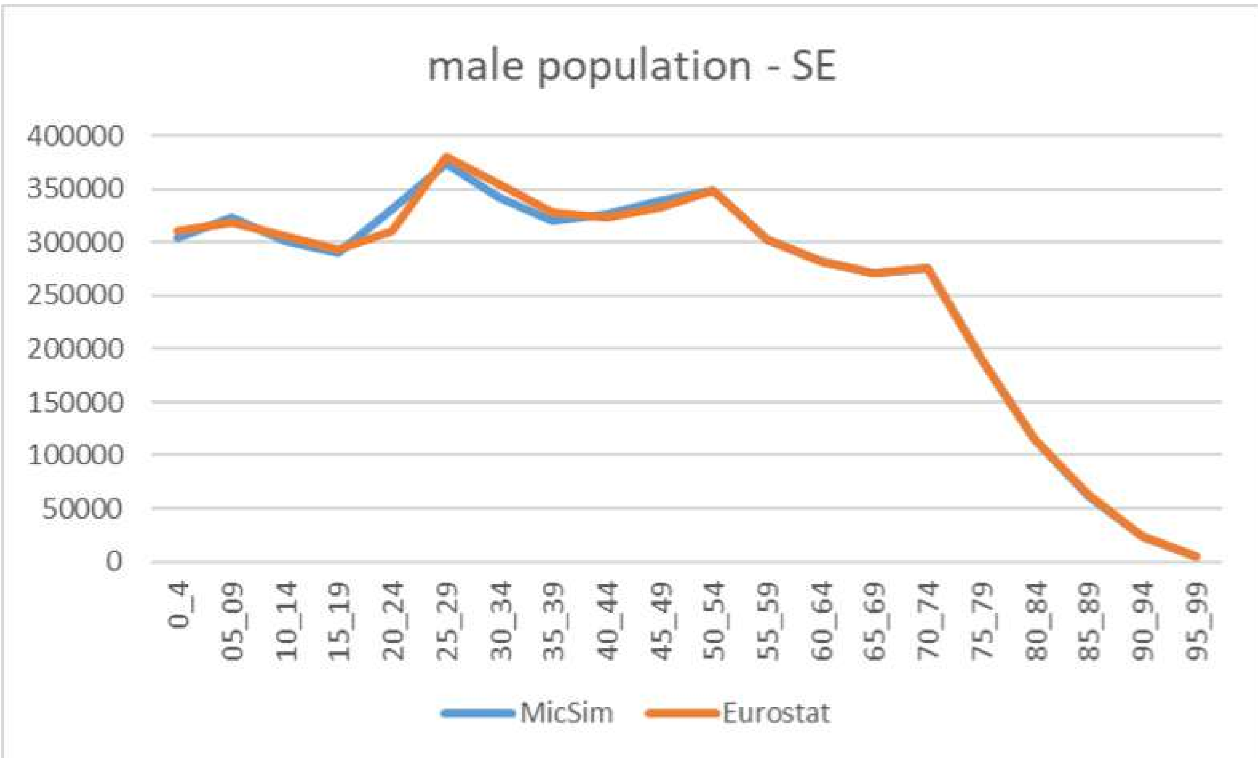


FIGURE A-13. Comparison between predicted male total living population by 5-year age groups (x-axis) in Sweden and official Eurostat data for the time interval 1/1/2014 – 1/1/2019.



Science for policy

The Joint Research Centre (JRC) provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society

EU Science Hub
joint-research-centre.ec.europa.eu

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

@eu_science

Publications Office
of the European Union