Methodological note

Istat's regional demographic projections are built with the aim of representing the possible future trend of the population, both in terms of total numbers and in terms of age and sex structure. The information produced represents an important tool to support decisions in economic and social policies, such as those relating to pension, health, education and housing systems. The projections are periodically updated by reformulating the evolutionary assumptions underlying fertility, survival, international and internal migratory movements.

The new set of projections replace those based on 2020 published by Istat in November 2021. Istat is the owner and responsible for the production and dissemination of the projections, as documented in the National Statistical Program. The methodological framework underlying the current exercise is the same as that implemented in the previous three-year cycle, which resulted in the sequential release of the forecasts based on January 1st 2016, 2017 and 2018. This methodology was defined, between 2009 and 2015, by a working group with researchers from Istat and the Luigi Bocconi University of Milan.

The methodological approach, around which the forecasting model works, is of a semi-probabilistic nature. The fundamental characteristic of probabilistic forecasts is to consider the uncertainty associated with the predicted values, determining the confidence intervals of the demographic variables and giving the user the possibility to independently choose the degree of confidence to be assigned to the results.

Compared to the "deterministic" approach, more widely used on an international scale and also adopted by Istat in the past (up to the 2011 based projections), this represents a significant methodological advance. In fact, in the deterministic model the user does not have probability measures associated with the results. Thus, a further advantage of the probabilistic method is the fact that the user can stop to trust uncritically on the work of projection makers, who with the typical "low / high" variants define a priori the alternative boundaries to the variant retained "most likely", generally identified as "main" or "medium" or "central" scenario".

The quantification of uncertainty does not represent the only advantage of the probabilistic model. Another one is the more effective representation of the evolution of a population. In the probabilistic model, in fact, the definable scenarios are infinite on the theoretical level (although in reality, as will be seen later, a finite number is always selected), so assumptions of low survival are mixed with assumptions of high fertility or medium level of migration, or the opposite. Instead, the assumptions of the high/low scenarios of the deterministic approach are defined by pursuing an output oriented logic: the high scenario contemplates assumptions of maximum increase in survival, fertility and migrations, while, on the contrary, the low scenario contemplates only assumptions of minimum. The construction of such opposing scenarios actually captures the goal of determining a future range for the population and its structural components, but based on concomitant assumptions with low chance of occurring.

The subsequent sections contain general information and briefly illustrate the steps that made it possible to build the projections. These sections include information on the following aspects:

- base population
- projection technique
- time horizon
- panel of experts
- expert questionnaire and probabilistic model
- relationship between national and regional projections
- data
- corrective component of nowcasting for short term assumptions
- confidence intervals and median scenario
- regional fertility projections
- regional mortality projections
- regional projections on international migration
- regional projections on internal migration
- comparison with previous projections
- comparison with the projections released by Eurostat and the United Nations

- data dissemination and terms of use
- contact information and personalized data requests.

Base population

The base population is the one broken down by sex, single age group and region as of 1 January 2021, as identified by the last Census of Population and Housing. The population includes all people usually residing in Italy, of any citizenship, while it does not include Italian citizens residing abroad, nor citizens illegally or irregularly present on the national territory who are not enrolled in any municipal register.

Projection technique

Projections are carried out with an iterative technique between 1 January and 31 December of each year, using the so-called "cohort-component" method. To the initial population, in correspondence of each age group, immigrations (from abroad or from other regions) are added while deaths and emigrations are subtracted (for abroad or for other regions), thus obtaining the population alive at the end of the year. Live births in the course of the year have also to be computed and, among them, those still alive as of December 31st, net of deaths and migratory movements that concern them.

For the population (stock), age is defined in completed years on 1 January (from 0 to 110 years and over). The same concept applies for flow data (births, deaths and migratory movements). This allows to identify, always and in any case, the demographic events by single year of birth of the subjects involved, ensuring the required consistency within the population equation.

It is assumed that demographic events may occur linearly at any time of the year. Between death and migration (internally or abroad) it is assumed incompatibility, that is, they cannot involve the same individual in the same year.

Deaths are determined by multiplying the resident population by age group on 1 January by the respective (projection-)probabilities of death, i.e. those involving subjects belonging to the same cohort.

Births in a given year are achieved in three steps. In the first, the average number of women for each fertile age (obtained as the average of the populations of that age at the beginning and end of the year) is multiplied by the respective fertility rate. In the second, the sum of the births by age of the mother is calculated, obtaining the total number of births in the year. In the third, births are broken down by sex using the fixed ratio of 106 male births per 100 female births.

Projections have a territorial profile and are built in the logic of the multi-regional model, a model which, with particular regard to internal migratory flows, simultaneously and coherently works the distinct territorial units of reference. The model on internal migration starts from the construction of a multi-regional matrix of migration probabilities by region of origin, region of destination, sex, and age. This matrix, applied to the population at risk of migration, identifies a coherent series of immigrants and emigrants in each forecasting year.

Time horizon

Projections cover the period between 1 January 2021 and 1 January 2070. The main purpose is to provide with information on the future development of the population in the short term (2030), and therefore to provide with information in the medium (2050) and long term (2070). With regard to this latter time reference information, data should be used with caution since the results become the more uncertain the further we go from the base year (2021). This risk is the more concrete the more attention is paid to the smaller territorial units, as in the case of some Italian regions.

Panel of experts

A panel of national experts supported Istat in formulating the demographic assumptions for Italy as a whole. The assumptions for the regions, on the other hand, were handled by Istat on the basis of a specific "bridge" methodology between the national and regional assumptions. The experts who replied to the questionnaire (with CAWI technique), providing with useful and complete information to define the assumptions, were 86. They were voluntarily recruited among the participants in the 13th edition of the Population Study Days organized by the Italian Association for Population Studies (AISP), which took place in Milan between 24 and 26 January 2019 at Bocconi University. In particular, there are 50 women and 36 men, mainly employed in universities (21 in Northern Italy, 11 respectively in Central and Southern Italy and 10 in foreign universities) or in other public research bodies (24). The mean age of the respondents is 44 years while their work experience is on average 16 years.

In all the phases that involved the building of the methodological framework underlying the projections, Istat made use of the concrete cooperation of Francesco Billari and Rebecca Graziani of the Bocconi University in Milan.

Expert questionnaire and probabilistic model

The probabilistic method adopted is based on the opinions of experts (expert-based model) to define the future evolution of the most important demographic indicators. It falls within the broader class of random scenario models. This model, used for the definition of probabilistic scenarios at a national level, is based on the elicitation of a series of parameters from which the future stochastic evolution of each demographic component is derived. Experts are asked to provide values at a given year "t" with regard to a series of summarized demographic indicators, conditional on the values assumed by the same indicators in instants of time prior to year "t" (Billari, Graziani and Melilli, 2012).

The method has the advantage of being simple and flexible. In fact, in the questionnaire, the necessary demographic components are summarized through the following indicators: the average number of children per woman; life expectancy at birth by sex; immigration and emigration from abroad. The other information necessary for the production of the projections, such as that regarding the age-breakdown of demographic events, is purposely kept out and subsequently processed in order to make the questionnaire and the forecasting model itself parsimonious.

TABLE A1. MEAN VALUES, VARIANCES AND CORRELATIONS UNDER ASSUMPTIONS OF THE EXPERTS BY DEMOGRAPHIC INDICATOR. Years 2019, 2050 and 2080

Indicator	Total fertility rate	Life expectancy at birth - Men	Life expectancy at birth - Women	Immigrations (thousand)	Emigrations (thousand)
			Year 2019		
Observed value	1.27	81.1	85.4	333	180
			Year 2050		
Mean assumption	1.51	84.7	88.1	256	131
High assumption	1.75	86.3	89.5	343	172
Variance	0.034	1,441	1,309	4,593	1,042
			Year 2080		
Mean assum. conditional to 2050 mean	1.55	87.1	90.0	240	127
Mean assum. conditional to 2050 high	1.74	88.4	91.0	305	158
High assum. conditional to 2050 mean	1.75	88.5	91.5	348	186
Variance	0.044	2,180	2,002	7,523	2,675
			Correlation 2050-2080		
Correlation coefficient	0.68	0.66	0.54	0.51	0.46

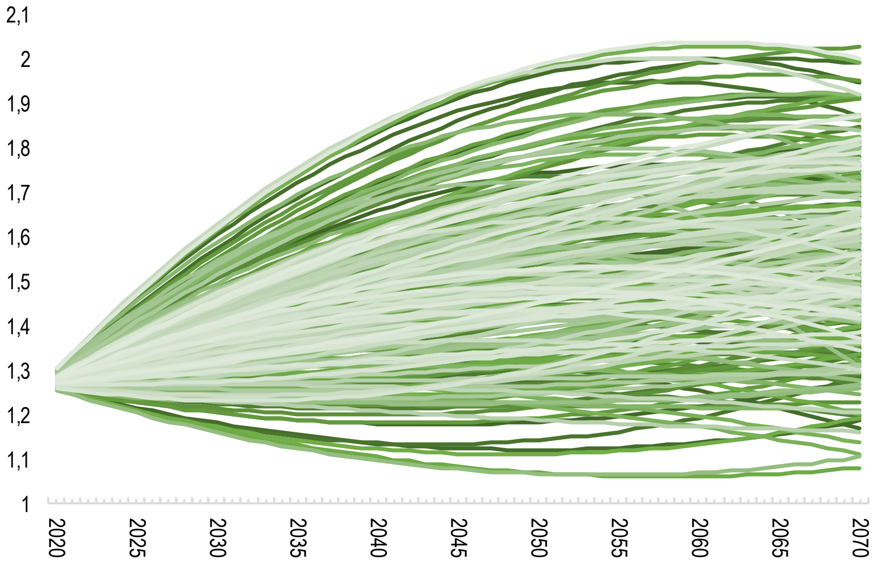
Two time points are considered for each demographic indicator: an intermediate year "t1" and a year "t2" corresponding to the last forecasting year. In the questionnaire submitted to the experts, "t0 = 2019", "t1 = 2050", "t2 = 2080", thus generating two sub-intervals, 2019-2050 and 2050-2080. Expressing the value of life expectancy at birth in the year 2080, given its expected value in 2050, is a practical example of how the mechanism works.

The demographic indicators are assumed, for the sake of simplicity, independent of each other (for example, the total fertility rate is not influenced by the level of migration and vice versa), although the model allows in its generalized version the possibility of interacting among them. It is also assumed that the pair of elicitations at 2050 and 2080 of a given indicator has a bivariate normal distribution.

Under these conditions, once the mean values provided by the experts have been obtained, it is possible to estimate the variance associated with each of the two future time instants as well as the correlation between the first and the second period (Table A1). On the basis of the corresponding bivariate normal distributions, 3,000 simulations were then carried out with the Markov Chain Monte Carlo method[1].

The last estimation step is aimed at calculating the values of each parameter in the intermediate years of the two sub-intervals 2019-2050 and 2050-2080. This activity is carried out, for each of the 3,000 simulations, by interpolation with quadratic curves, passing through the known points corresponding to the years 2019, 2050 and 2080. Thus, the definition of 3,000 stochastic curves for each demographic indicator has been achieved at national level. As an example, Figure A1 describes the bundle of trajectories relating to the number of children per woman, obtained from the procedure described above.

FIGURE A1. PROBABILISTIC EVOLUTION OF THE TOTAL FERTILITY RATE IN 3,000 SIMULATIONS OBTAINED FROM EXPERT OPIONIONS. Years 2020-2070



The choice to consider a number of 3,000 simulations is the result of a compromise between two needs, both strategic: 1) faithfully representing the uncertainty of demographic events; 2) optimizing the machine times for processing the projections. The latter, despite today's availability of increasingly powerful and sophisticated hardware / software tools, represents a technical aspect which is anything but secondary, given the huge amount of data processed.

[1] The choice fell on the years 2050 and 2080 in order to identify two time intervals of similar length. Although the availability of information collected from the experts allows us to extend the time horizon up to 2080, it was decided to stop the iterative simulation exercise up to 2070 in order to derive a forecasting period of precise 50 years.

Relationship between national and regional projections

The probabilistic model releases a set of 3,000 national simulations for each summary demographic indicator. Since the objective of the Istat projections is also to give indications at a territorial level, so continuing the longstanding tradition of the multi-regional model, a "bridge" procedure has been implemented between the definition of national and regional inputs. The approach pursued is therefore top-down on the side of the assumptions building while, as will be seen later, it is bottom-up on the side of the production of final outputs.

The main action is to derive 3,000 regional stochastic scenarios from the 3,000 national ones. The first operation in this sense is to elaborate an intermediate deterministic forecast, applying the multi-regional cohort component model. From this forecast, obtained by extrapolating the regional trends considered most probable for each component (see following paragraphs), the same summary indicators of the previously described stochastic model are obtained, i.e. average number of children per woman, male and female life expectancy at birth, migratory movements with foreign countries. Such a first intermediate forecast, unique and deterministic, essentially resembles that which in a deterministic approach would be labelled with the term "central scenario".

The transition from the regional deterministic model to the regional stochastic model is achieved by multiplying, and repeating the procedure 3,000 times, the regional deterministic forecast for the relationship between the national stochastic and the deterministic forecast. In formula, indicating with "n" the generic simulation (n = 1, ..., 3,000), with "j" the regional territorial code, with DR the deterministic regional forecast, with SR the stochastic one, with DN and SN, respectively, the national deterministic and stochastic forecast, we have:

$$\mathrm{SR}_{t,n}^{j} = \mathrm{DR}_{t}^{j} \times \frac{\mathrm{SN}_{t,n}}{\mathrm{DN}_{t}}$$

thus linking, in each simulation, the vector of regional values to the national stochastic reference value. Note that with regard to the synthetic indicators of immigration and emigration from abroad, we have:

$$\mathrm{DN}_{t} = \sum_{j} \mathrm{DR}_{t}^{j}$$

$$\mathrm{SN}_{t,n} = \sum_{j} \mathrm{SR}_{t,n}^{j}$$

Once the synthetic stochastic indicators have been obtained at the regional level, we move on to the construction of the inputs necessary for the application of the cohort-component method, i.e. the (projection-)probabilities of death by age and sex, the age specific fertility rates and the distribution of immigrants / emigrants from abroad by age and sex. The procedure therefore associates each summary indicator with its own sex-age breakdown. The latter, not treated in a stochastic way, is the one that derives from the regional deterministic model and, from simulation to simulation, adapted to the specific synthetic stochastic indicator.

The coupling of the 3,000 death probability vectors (each vector develops a number of elements equal to the "number of regions X age classes X sex X forecast years") with the 3,000 fertility vectors, and the same number on immigration and emigration from abroad and, finally, with the 3,000 O / D probability matrices of internal migration, it is randomly executed.

After introducing a corrective nowcasting component (see next paragraph) relating to the very first years of forecasting, the cohort component model is then run 3,000 times, thus obtaining the required outputs: population by age and sex, demographic flows by age and sex, plus the series of demographic indicators to support the analysis, from generic rates (birth, mortality, etc.) to structural indicators (mean age, dependency rations, etc.).

The results at national level (as well as those at geographical area level) in the context of each regional simulation are obtained by sum (bottom-up approach). Therefore, the amount of the expected population, deaths, and migrations, classified by age and sex, and births by age of the mother that are determined at the national level are the sum of the forecast regional trajectories. The assumed national levels relating to the summary indicators placed into dissemination, for example regarding life expectancy or the average number of children per woman, are recalculated ex-post on the basis of these regional summaries.

It should be noted that the stochasticity introduced at the regional level, borrowed top-down from the national one and limited only to summary indicators, may result not sufficient to reproduce the randomness of the various demographic events. This is particularly true in small areas where uncertainty tends to be relatively higher. For this reason, although the number of simulations still offers ample guarantee of representativeness of the variability on a regional scale, it is more appropriate to speak of a semi-stochastic approach when referring to regional projections.

A second observation concerns the fact that in the Istat model a generalized statistical treatment of the covariance between the Regions is excluded (for example: the forecast of increase / decrease in fertility in a given region how much it conditions or how much is in turn conditioned by the forecast of increase / decrease in another). To this solution, also excluded for reasons of parsimony of the model, another one was preferred, that of territorial convergence. In fact, the initial deterministic regional model, subsequently transformed into a stochastic model through the procedure described above, is built on the assumption of very long-term convergence (2120, well beyond the last year of projections) between the regions for each fundamental demographic component. This implies that the 3,000 regional stochastic scenarios represent 3,000 different hypotheses of convergence of demographic behaviours among regions.

The main hypothesis underlying the convergence is that the socio-economic and cultural differences currently existing between the regions are destined to disappear in the long term. Therefore, their progressive cancellation would also involve a generalized rapprochement of demographic behaviours. The idea of convergence is not new in demography and there are many examples of demographic forecasts that follow it (Eurostat and the UN, in particular), including past Istat ones. In Istat projections, convergence is understood as the progressive shift of a given demographic behaviour towards a very distant point in the future which represents the instant of full regional convergence (in the sense that at that point the values would be identical for the different regions), but that in reality it is far from being reached within the time horizon considered (2020-2070). In fact, it is correct in this circumstance to speak more of a model of semi-convergence than of a model of full convergence.

Data

The assumptions defined at the regional level in the preliminary deterministic model, before the transition to the stochastic model, were obtained by extrapolating future trends from the analysis of the observed time series. In particular, these assumptions were defined using the following data series:

- for fertility, the mother's age-specific rates for the period 1977-2019;
- for mortality, the (projection-)probabilities of death by age and sex for the period 1974-2019;
- for internal and international migrations, the changes of residence by age and sex of 2015-2019.

Corrective component of nowcasting for short term assumptions

Before being launched at full capacity along the time horizon with the cohort-component method, the probabilistic projections incorporate a corrective nowcasting factor (from the term nowcast = forecast of the present). With this operation we intend to ensure that the forecast relating to the very first years is as much in line with the trend that emerged in the last period or in the last historical year (jump-off effect). This type of operation is particularly suitable in years characterized by sudden, and as such unpredictable, changes in the demographic situation. This is the case, as happened in 2020 and to a lesser extent in 2021, of the effects caused by the Covid-19 pandemic on all components of the demographic change. Not only, albeit primarily, on mortality, but also on birth rates and internal and international changes of residence. No forecasting model applied to the historical series mentioned in the previous paragraph could have been able to accurately predict the shock caused by the pandemic. Certainly not the over 740 thousand deaths found for all causes in 2020 (about 100 thousand more than expected), but not even the further decline in births (405 thousand) in the context of an overall picture already compromised by the well-known contraction of reproductive behaviours. Nor, finally, could the strong contraction recorded in migration be foreseeable, as a result of the measures undertaken at national level to contain the spread of the virus (lockdown).

Since the base population of the projections is that recorded as of January 1, 2021, it was necessary put in place some short-term correction of the predicted inputs that affected the first projections years. With this, in fact, we want to take into account not only the exceptional events that characterized the 2021, but also the subsequent years within which it is assumed that the pandemic effects may end and allow the short term inputs to be in line with medium and long term ones[2].

From the computational point of view, the review of the short-term assumptions is carried out by applying correction factors. For example, let E_b^j be the number of demographic events predicted in the first year based on the median scenario in region j. Instead, let $\widehat{E}_b^j$ be the observed value of such events or, in the absence of the actually observed value, the best estimate that can be obtained (for example, using nowcasting procedures or similar statistical models). The ratio:

$$r_b^j = \widehat{E}_b^j / E_b^j$$

represents the correction factor to be applied to the statistical measures that give rise to type "E" events in year "b" for region j. For example, if these events were the total number of births then the quantity:

$\hat{f}_{b,x}^{n,j} = r_b^j \cdot f_{b,x}^{n,j}$ with x=14, … , 50 and n=1, … , 3000

represents the series of fertility rates by age of the mother (n-th simulation) corrected for year "b". Similar considerations apply to the determination of the correction coefficients relating to mortality and migratory movements.

As regards 2021, the correction factors were constructed by comparing the data of the provisional demographic balance of each region, released in March 2022 by Istat, to the projections previously produced for that year[3].

For the years after 2021, the correction factors are applied for a limited period of the time horizon, processing weights that progressively tend to one. In particular, the number of years for which the correction factor is applied to the series of interest is obtained from:

$$Y^j = \text{abs}\left(1 - r_b^j\right) \cdot \epsilon$$

with ϵ arbitrary quantity, appropriately chosen to ensure that, on regional average, the number of years to guarantee the return from short-term to medium-long term projections does not exceed 5 years. At this point, the levels of the correction factors for the years following "b", for a total of "Y" years, are given by:

$$r_t^j = \frac{r_b^j \cdot \left(b + Y^j - t\right) + (t - b)}{Y^j} \quad \text{with } t = b, b+1, \ldots, b + Y^j - 1$$

Confidence intervals and median scenario

Once the calculation procedure inherent to the 3,000 regional simulations has been launched, uncertainty is calculated for all possible information levels, from the predicted population to the flow data, also broken down by age and sex. These margins of uncertainty depend in turn on the uncertainty inherent in the future levels of mortality, fertility and migration that are also made available. The dissemination of the results contemplates the release of only the confidence intervals of 90%, 80% and 50% but it is possible to define intervals on any scale of interest. The confidence interval provides information on how likely it is that a given demographic indicator falls within predetermined limits. From this point of view it should be remembered that this probability itself represents a forecast, as it is based on hypotheses whose validity is uncertain. Furthermore, in no case should the extremes of the confidence interval be interpreted as extreme limits, upper or lower, of future demographic behaviour.

[2] Furthermore, bearing in mind the iterative calculation mechanism offered by the cohort component method, i.e. a mechanism of continuous stock-flow interaction over time, the correction imposed in the first years also affects the outcome of all subsequent years, up to 2070.

[3] Cfr.: Istat, la dinamica demografica – anno 2021, https://www.istat.it/it/archivio/267834.

The construction of a confidence interval is here based on the determination of the percentiles in the distribution of the 3,000 simulations. For example, the 90% confidence interval for a given indicator is determined by considering the distribution values that fall between the 5th and 95th percentiles. It is also recalled that the uncertainty always refers to the domain of the specific estimated parameter. The limits of the confidence interval for a given hierarchical level are estimated on their own, and not constructed by summation of limits obtained at a hierarchically lower level of disaggregation. The criterion is also applied in non-territorial hierarchical contexts; for example in the composition by age of the population or in that by sex.

The "median scenario" was built with the aim of defining a "punctual" forecast that can be adopted as the most likely reference of future demographic evolution. This scenario corresponds to a 3001-th simulation, obtained by construction, but which in fact was not detected in the observation field of the 3,000 simulations. Its set of assumptions is identified by taking as a reference the median value between all the simulations at the level of the individual demographic components (fertility, mortality, migration) within the possible combinations of the covariates age, region and year of forecast. For example, the median scenario specific fertility rate at the age of 32 in the Tuscany region, in the year 2040, is identified as the median value with these characteristics identified among all the simulations. The same specific rate but at the following age, or in the following year, is identified with the same procedure but it probably arises from a different simulation. For the identification of the median scenario on mortality and migration, the procedure is identical but with the additional covariate of sex. Furthermore, for internal migrations, the selection also includes the region of origin and destination.

The scenario is therefore “median” from the side of the fundamental inputs. From the point of view of the outcome (population and expected flows) that this scenario generates once the procedure for cohort-components has been launched, for the typical properties of the median it returns values very close to the median ones.

Regional fertility projections

For regional fertility the projections concerned the classic parameters of intensity and age-breakdown, i.e. the average number of children per woman and the distribution of specific fertility rates by age of the mother.

The average number of children per woman was represented using ARIMA models (n, p, k), searching, separately for each region, the one most suitable for predicting the future intensity of reproductive behaviour. On the basis of the 1977-2019 historical series the predominantly model was an ARIMA (2,0,0) with intercept.

The fertility age profile was modelled using a quadratic splines function system (Schmertmann, 2003). This model functionally describes the curve of specific fertility rates standardized as a function of three parameters: the age of onset of the fertile age α; the age P in which fertility reaches its maximum level; age H, subsequent to P, in which fertility is halved compared to the maximum level. By specific standardized fertility rate we mean the specific fertility rate normalized to the unit, where the value one corresponds to the maximum value observed within its age distribution.

The quadratic splines model fits five second-degree polynomials to the fertility curves. The final function is continuous with the first derivative also continuous. Moreover, thanks to suitable mathematical restrictions it is uniquely determined by the three parameters [α, P, H] mentioned above.

In practice, the prediction of the specific fertility rate is transformed into the prediction of the three parameters (through ARIMA models) which express it functionally, once the series has been estimated in the period 1977-2019. To do this, a hypothesis of convergence between the Italian regions was adopted, assuming that the territorial differences in terms of reproductive behaviour tend to decrease in the long term. From an operational point of view, full convergence was set in 2120. In particular, the convergence constraint provides that, from 2020 to 2120, the parameters of the regional vector [α, P, H] converge linearly to the values of a hypothetical national vector, specially designed for the operation.

Regional mortality projections

Regional mortality projections were produced using the Lee-Carter model (1992) in the variant proposed by Lee-Miller (2001), a model in which the adjustment procedure leads the fitted probabilities of death to reproduce precisely the observed level of life expectancy at birth, rather than the total deaths observed as in the original version. Furthermore, here the model is applied to death probabilities rather than to mortality rates of the original formulation.

The model approximates the logarithmic form of the probability of death using three synthetic parameters, one of which is related to the trend [k(t)] and two related to the age distribution [(a(x), b(x)].

As for fertility, also for mortality the construction of the model originates from the definition of a provisional reference scenario at national level. The forecast is determined by projecting into the future the only national trend parameter k(t), whose series is identified over the period 1974-2019, while the parameters a(x) and b(x) remain invariant over time in this phase. In particular, due to its substantial linearity, the k(t) parameter was projected to 2070 with a random walk with drift.

The assumptions at the regional level are derived from the provisional national reference scenario, by first estimating the regional values of the three parameters in 1974-2019 with the same methodology and, subsequently, by making each regional parameter converge to the corresponding national parameter at 2120. Therefore, as a consequence of the convergence process and unlike the classical approach of the Lee-Carter model, here the regional parameters a(x) and b(x) are also varied over time.

Regional projections on international migration

In order to capture the most recent trends, the regional projections of migratory flows with abroad focus the analysis only on the last five years, namely 2015-2019. This need, considering the complexity of predicting international migratory flows by resorting to analysis of long historical series, leads to use a very simple model. Without forgetting that at this level of operations it is a question of structuring an intermediate deterministic model, whose values are subsequently calibrated on the intensities produced by the expert-based stochastic model.

In the first year of the projection, the total values of immigration and emigration from abroad are set equal to the mean value observed over the last five years. In accordance with the general convergence framework of the deterministic model, it is therefore assumed that in each region inflows and outflows converge linearly in the long term (2120) at the same level, which is to the initial half sum of the two values.

Once the totals of inflows and outflows up to 2070 have been determined, the associated age and sex breakdown are derived by applying the Castro-Rogers model (Rogers and Castro, 1981) to the 2015-2019 series. With this model it is shown that the characteristic age profile of migrants can be described, regardless of the intensity of the phenomenon, by a mathematical function composed of 4 additive components and up to 11 predictive parameters. These parameters, whose estimate in the observed period is produced thanks to a generalized procedure for non-linear models (category in which the Castro-Rogers function fully falls), are kept constant in the forecast period. The conclusive result is therefore that the global intensity of migratory flows with abroad may vary over time but on the basis of a constant composition by age.

Regional projections on internal migration

Interregional migrations are developed according to a multidimensional approach, which allows to simultaneously consider the areas of origin and destination of migratory flows, to define the entrances in a given area as the sum of the exits with that destination from all the other areas of the system. The system is by construction consistent for all the forecast years since the marginal row and column of the O/D matrix, corresponding respectively to the inflows and outflows in/from each region, give the same sum, corresponding to the amount total of movements within the national territory.

The probability of migration specific for age (110), sex (2), region of origin (21) and destination (21) represents the basic component of the O/D matrix composed of 110x2x21x21 = 97020 cells for each calendar year. The probabilities are estimated on the basis of the levels observed in the individual years of the 2015-2019 period. The probability vectors thus obtained, at the level of each annuity, are subsequently modelled using the Castro-Rogers function.

Therefore, indicating with

$m_{x,s,t}^{i,j}$

a generic (projection-)probability of migration for an individual of age "x" and sex "s" between the region "i" and the region "j" relating to the year "t" (t = 2015, ..., 2019), is assumed that this represents a normal random variable with an average equal to the mean value of the five-year period and variance equal to the variance detected in the five-year period:

$$\mu_{x,s}^{i,j} = E\left(m_{x,s,t}^{i,j}\right)$$

$$\sigma_{x,s}^{i,j} = E\left(m_{x,s,t}^{i,j} - \mu_{x,s}^{i,j}\right)^2$$

From the above mentioned random variables, 3,000 values are randomly extracted for each of the 97,020 elements of the O/D matrix, thus giving rise to the random creation of 3,000 different matrices. The O/D matrix relating to the median stochastic scenario is identified by taking as a reference the median value between all the simulations within the possible combinations of the covariates sex, age, region of origin and region of destination. This median matrix is also used with the preliminary purpose of producing the deterministic forecast of the population, prior to the transition to the actual stochastic model (see previous paragraph on the relationship between national and regional projections).

Note that in the context of each simulation (including the median scenario) the O/D matrix is assumed to be invariant over time. The hypothesis underlying the model is based, in fact, on maintaining a propensity for mobility that remains constant throughout the time horizon. This implies that internal migratory flows evolve over time only because of the variations affecting level and age structure of the population exposed to the risk of migration.

Comparison with previous projections

An assessment of the change that occurred between the last two rounds can be made by comparing the median scenarios of the projections based on 2020 and 2021.

First of all, a rather limited difference should be noted between the total base population 2021 (59 million 236 thousand) and that which had been estimated in the median scenario on the same date by the projections based on 2020 (59 million 249 thousand).

TABLE A2. 2020 AND 2021 MEDIAN SCENARIO ASSUMPTIONS FOR THE MAIN DEMOGRAPHIC INDICATORS.
Years 2021, 2030, 2050 and 2065.

Median scenario	Total fertility rate	Life expectancy at birth – Men	Life expectancy at birth – Women	Immigration (thousand)	Emigration (tohusand)
		Year 2021			
Base 2020	1.21	80.1	84.9	271	130
Base 2021	1.25	80.0	84.6	286	129
		Year 2030			
Base 2020	1.37	82.2	86.2	279	146
Base 2021	1.37	82.2	86.2	281	145
		Year 2050			
Base 2020	1.50	84.7	88.1	258	131
Base 2021	1.50	84.7	88.1	258	131
		Year 2070			
Base 2020	1.55	86.5	89.5	244	126
Base 2021	1.55	86.5	89.5	244	126

On the side of the expected flows in the period of common projection (2020-2065), a slightly better assessment can be seen in the projections based on 2021, where for example 19.6 million births were expected against the 19.5 million of the previous exercise. Even the other components of the population change, although not such as to overturn the results emerged from the projections based on 2020, are more favourable for the exercise. The latter, in fact, has 69 thousand fewer deaths in the common projection period, 65,000 more immigrants from abroad and 19,000 fewer emigrants to abroad.

The difference between the final populations of the two distinct projections is also small (as of January 1, 2070, just 136 thousand units more for the median scenario 2021 based), confirming the substantial stability of the projections based on 2020, despite the change in the base population and short-term adjustments to the balance components. From this point of view, table A2 highlights how the process of reviewing the assumptions for all demographic components only affected the first years of forecasting.

Comparison with the projections released by Eurostat and the United Nations

In order to compare the projections produced by Istat with those of other bodies, it makes sense to take as a reference the projections released by Eurostat and the United Nations Population Division (UNPD). For years, the statistical institute of the European Union has been carrying out the task of producing demographic forecasts on a regular basis for all member countries. The latest release is based on 2019, whose main reference scenario is the so-called baseline scenario. The UNPD, in turn, also produces demographic projections on a regular basis through the World Population Prospects, which include all the countries of the globe. In this latter case, the latest available exercise is based on 2021 and the main reference scenario is the so-called medium variant.

It should be noted in the introduction that, despite the comparability on the level of projective technique, the exercises produced by the two international organizations present some methodological differences compared to the Italian one. Among these, in the first place, the fact that the Eurostat projections are based on 1 January 2019, that is, they project a population that is not in line with the results of the 2020 census and, in particular, with respect to the latter, a significantly higher population. Secondly, it should be mentioned that the two international models examined here are uninational, i.e. they project the resident population in Italy as a whole without taking into account the demographic development of the regions.

Table A3 shows the main scenario assumptions compared. As regards migratory flows, the comparison is limited to net migration as both Eurostat and UNPD build the assumptions directly on this indicator (without distinction between immigrants and emigrants).

For all the demographic components, the assumptions are initially very different between Istat and Eurostat/UNPD. This is due to the fact that, unlike the Istat scenario, Eurostat projections do not take into account the 2020 demographic shock produced by the Covid-19 pandemic. The UNPD projections, in turn, present very limited assumptions in terms of net migration, not only in the initial projection period but over the entire projection horizon. In the medium and long term the assumptions continue to be rather differentiated between the various producer bodies. In particular, with regard to migratory movements, where compared to a UNPD that is rather cautious about Italy, Eurostat is opposed with a much more optimistic vision. This evidence is partly due to the Eurostat methodology, which, in addition to predicting the underlying evolution of net migration, incorporates an additive replacement-migration component into the model[4].

The assumptions on fertility are quite similar, although in the medium-long term Eurostat and UNPD produce less favourable forecasts than Istat. The assumptions on survival are also not particularly distant, however Eurostat and especially UNPD highlight very favourable expectations about the lengthening of life expectancy, which are only partially glimpsed in the Istat model.

The development of the different demographic assumptions therefore gives rise to differences in terms of expected results which, as regards the evolution of the total population, can be appreciated in Figure A2. Unfortunately, the initial difference due to the different population bases adopted in the Eurostat projections

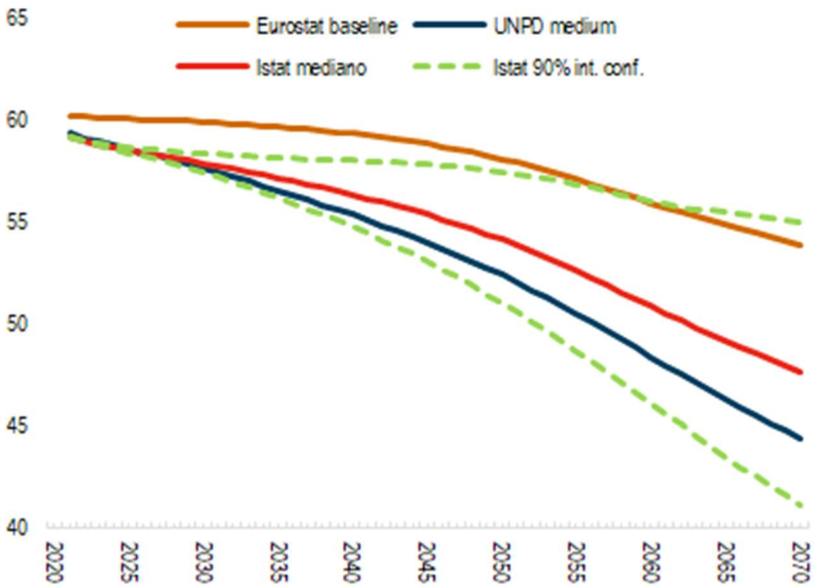
[4] This component assigns in each forecast year an additional quota of net migrants in the measure equal to 10% of the reduction found in the population of working age (15-64 years).

make the comparison with the Istat scenario rather spurious. In turn, the UNPD projections, although aligned from the start with those of Istat, give a much more pessimistic evolution of the population.

TABLE A3. COMPARISON BETWEEN LATEST MAIN ASSUMPTIONS ON ITALY MADE BY ISTAT (MEDIAN SCENARIO), EUROSTAT (BASELINE) AND UNPD (MEDIUM). Years 2021, 2030, 2050 and 2070.

Scenario	Total fertility rate	Life expectancy at birth – Men	Life expectancy at birth – Women	Net migration (thousand)
		Year 2021		
Istat Mediano	1.25	80.0	84.6	157
Eurostat Baseline	1.33	81.4	85.8	178
UNPD Medium	1.28	80.5	85.1	28
		Year 2030		
Istat Mediano	1.37	82.2	86.2	136
Eurostat Baseline	1.37	82.6	86.9	224
UNPD Medium	1.35	83.2	87.1	58
		Year 2050		
Istat Mediano	1.50	84.7	88.1	127
Eurostat Baseline	1.45	84.9	89.0	214
UNPD Medium	1.44	85.8	89.6	58
		Year 2070		
Istat Mediano	1.55	86.5	89.5	118
Eurostat Baseline	1.52	87.0	90.9	207
UNPD Medium	1.48	88.2	91.9	58

FIGURE A2. TOTAL POPULATION ACCORDING TO ISTAT, EUROSTAT AND UNPD SCENARIOS. Years 2021-2070, million.



Nonetheless, the evolutionary trajectory of the population is consistent between the three scenarios. In fact, all of them foresee a progressive decline of the population which tends to worsen in the medium-long term. The Eurostat scenario, given the significant impact of a more sustained net migration, is particularly optimistic. Up to the point of maintaining a population even wider than the upper limit of the 90% confidence interval of Istat projections for most part of the time horizon. The UNPD scenario, on the other hand, tends to be about halfway between the Istat median scenario and its lower confidence interval.

Data dissemination and terms of use

The detailed picture of the assumptions underlying the projections and the main results can be consulted both on the general internet site dati.istat.it (topic: Population and families> Demographic projections) and on the thematic site demo.istat.it.

Data dissemination is divided into three sections including tables on the structure by sex and individual age group of the population, on the components of population change and on the main supporting demographic indicators. Each table shows the values of the median scenario and the lower and upper limits of the confidence intervals at 90%, 80% and 50%.

The components of the population change include:

- population at start and end of the year, total growth;
- live births and deaths, natural growth;
- immigrants and emigrants with abroad, net migration with abroad;
- Interregional immigrants and emigrants, net interregional migration.

The data described above and those relating to the age distribution of the population are rounded to the nearest unit.

As regards the demographic indicators, the tables include:

- live birth rate, mortality rate and natural growth rate;
- immigratory, emigratory and net migration rate with abroad;
- immigratory, emigratory and net migration rate with internal regions;
- total net migration and total growth rate;
- mean age of the population;
- % of population aged 0-14, 15-64, 65 years and more, 85 years and more;
- dependency ratio, elderly dependency ratio, aging index;
- total fertility rate;
- life expectancy at birth and at 65 years by sex.

Istat periodically produces demographic projections as part of the line of activity "Population estimates and projections", in accordance with the provisions of the National Statistical Program, "Demographic projections" project (PSN code IST-01448).

For customized data requests, in addition to disseminated information, it is necessary to contact the Cont@ct Centre at https://contact.istat.it/.

Glossary

Age specific fertility (rate): the ratio of the number of live births to women between the ages of x and x + 1 and the average number of women of that age in a given year.

Average number of children per woman: the number of children a woman would have if she was subjected to the fertility calendar (in the form of age-specific fertility rates) of a given calendar year during her reproductive life span.

Birth (rate): ratio between the number of live births in the year and the average amount of the resident population, multiplied by 1,000.

Cohort component (model): the continuous calculation algorithm that in iterative mode simulates the evolution of the fundamental population equation by age group, allowing to determine the demographic flows and to obtain the surviving population at the end of each year.

Couple: two people linked by an emotional and sentimental relationship. Can be formed by opposite or same sex people. The bonds between people in couples can be formal (de jure couple: married, civilly united or de facto cohabiting pursuant to law 76/2016) or informal (de facto couple).

Death: the cessation of any sign of life at any time after the vital birth.

Demographic projection: elaboration that shows the future development of a population when certain assumptions are made regarding the future course of mortality, fertility and migration.

Deterministic demographic projection: elaboration on the future development of a population, summarized in a single series of values obtained from a single set of demographic assumptions, which does not report any measure regarding the uncertainty usually associated with the results.

Dependency ratio: ratio between the population of inactive age (0-14 years and 65 years and over) and the population of active age (15-64 years), multiplied by 100.

Elderly dependency ratio: ratio between the population aged 65 and over and the population aged 15-64, multiplied by 100.

Emigration for abroad (rate): the ratio between the number of emigrations to abroad and the average amount of the resident population, multiplied by 1,000.

Families with nucleus: includes couples with children, couples without children, single parents, families with two or more nucleus.

Families without nucleus: people living alone or multi-person families; this latter do not constitute a family unit even if composed of several people.

Family: group of people linked by ties of marriage, kinship, affinity, adoption, protection, or by emotional ties, cohabitants and having habitual residence in the same Municipality. It can also be constituted by a single person.

Family nucleus: set of people who form a couple or a parent-child relationship. It means a married couple, civilly united or cohabiting, with or without children, or even a single parent together with one or more children who have never been married. Within a family there may be one or more family nucleus, but there may also be none, as in the case of families formed by an isolated member (single-component families) or several isolated members (other resident persons).

Family typology: classification based on the presence or absence of at least one nucleus and by type of nucleus.

Immigration from abroad (rate): the ratio between the number of immigrations from abroad and the average amount of the resident population, multiplied by 1,000.

Internal emigration (rate): the ratio between the number of internal emigrations and the average amount of the resident population, multiplied by 1,000.

Internal immigration (rate): the ratio between the number of internal immigrations and the average amount of the resident population, multiplied by 1,000.

Internal migration balance: difference between the number of registrations for change of residence from another Municipality and the number of de-registrations for change of residence to another Municipality.

Internal net migration (rate): the difference between the internal immigration rate and the internal emigration rate.

Life expectancy at age "x": the average number of years that a person of completed age "x" can count to survive in the hypothesis that, in the course of his subsequent life, he was subjected to the risks of mortality by age (from age "x" up) of the year of observation.

Life expectancy at birth: the average number of years that a person can count to live from birth in the hypothesis that, in the course of his existence, he was subjected to mortality risks by age of the year of observation.

Live birth: the product of conception which, once expelled or completely extracted from the maternal body, regardless of the duration of gestation, breathes or manifests other signs of life.

Mean age: mean age of the population at a certain date expressed in years and tenths of a year.

Mean age at birth: the mean age at birth of mothers expressed in years and tenths of a year, calculated considering only live births.

Migratory balance with abroad: difference between the number of registrations for change of residence from abroad and the number of de-registrations for change of residence to abroad.

Mortality (rate of): ratio between the number of deaths in the year and the average amount of the resident population, multiplied by 1,000.

Natural balance: difference between the number of births and the number of deaths.

Natural growth (rate): the difference between the birth rate and the death rate.

Net migration with abroad (rate): the difference between the immigration rate from abroad and the emigration rate with abroad.

Old age (index): ratio between the population aged 65 and over and the population aged 0-14, multiplied by 100.

Predictive (or confidence) interval: an interval associated with a random variable yet to be observed, with a specific probability that the random variable falls within it.

Probabilistic demographic projection: elaboration on the future development of a population, summarized in a set of values or in a probability distribution, in which the variables used are of a random nature that cannot be predicted with certainty and in which not all assumptions are equally probable.

Probability of death: the probability that an individual of precise age x will die before the birthday x + 1.

Projection: development expected in the future.

Projection probability of death: the probability that an individual of age x (in years completed on 1st January) will not survive within the year.

Projection probability of interregional migration: the probability that an individual of age x (in years completed on January 1st) moves residence between two regions before the end of the year.

Range: measure of the variability of a quantitative phenomenon defined by the difference between its maximum and the minimum value.

Registration and de-registration for transfer of residence: registration concerns people who have moved to a Municipality from other Municipalities or from abroad; the de-registration concerns people who have moved to another municipality or abroad.

Resident population: constituted in each Municipality (and similarly for other territorial divisions) of people with habitual residence in the Municipality itself. Persons temporarily residing in another Municipality or abroad, for the exercise of seasonal occupations or for reasons of limited duration, do not cease to belong to the resident population.

Scenario approach: the description of the context, even conceptual, in which the population is projected. In a deterministic approach it usually refers to the main or central assumption. In a stochastic it can refer to the assumption identified as mean or median.

Simulation: the quantitative implementation of a single set of demographic assumptions to be launched in the cohort-component model in order to obtain a single set of demographic projections.

Total balance: sum of the natural balance and the total migratory balance.

Total growth (rate of): the sum of the total net migration rate and the natural growth rate.

Total migratory balance: the sum of the migration balance with abroad and the internal migration balance.

Total net migration (rate): the sum of the net internal migration rate and the net migration rate with abroad.