Methodological note

**1) Regional population projections by age and sex. Years 2024-2080**

Istat 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 2023 published by Istat in July 2024. Istat is the owner and responsible for the production and dissemination of the projections, as documented in the National Statistical Program. With this new release a new three-year production cycle starts after those of the years 2016-2018 and 2019-2021[[1]](#footnote-1). 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 2024, 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 2024 and 1 January 2080. 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 (2080). 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 (2024). 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 121. They were voluntarily recruited by Istat among the most illustrious experts in demographic-social studies. In particular, there are 69 women and 52 men, mainly employed in universities (68) or in other public research bodies (42). The mean age of the respondents is 51 years while their work experience is on average 23 years.

The survey was carried out in the first quarter of 2023, and its long-term findings are taken as a reference for the entire three-year cycle (i.e. for the forecasts in base ‘22, released last year, for those released here in base ‘23, and for those to be released in 2025 in base ‘24).

**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 2021, 2050 and 2080**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Indicator** | **Total fertility rate** | **Life expectancy at birth – Men** | **Life expectancy****at birth – Women** | **Immigrations (thousand)** | **Emigrations (thousand)** |
| **Year 2021** |
| Observed value | 1,25 | 80,3 | 84,8 | 318 | 158 |
| **Year 2050** |
| Mean assumption | 1,38 | 84,3 | 87,8 | 302 | 136 |
| High assumption | 1,54 | 85,7 | 89,1 | 368 | 169 |
| Variance | 0,016 | 1,239 | 1,106 | 2.613 | 667 |
| **Year 2080** |
| Mean assum. conditional to 2050 mean assumption | 1,45 | 86,2 | 89,6 | 304 | 142 |
| Mean assum. conditional to 2050 high assumption | 1,66 | 88,0 | 91,1 | 389 | 187 |
| High assum. conditional to 2050 mean assumption | 1,68 | 88,2 | 91,5 | 402 | 192 |
| Varianza | 0,058 | 4,586 | 3,689 | 10.302 | 2.774 |
| **Correlation 2050-2080** |
| Correlation coefficient | 0,67 | 0,67 | 0,64 | 0,65 | 0,67 |

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 = 2021", "t1 = 2050", "t2 = 2080", thus generating two sub-intervals, 2021-2050 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 method1.

The last estimation step is aimed at calculating the values of each parameter in the intermediate years of the two sub-intervals 2021-2050 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 2021, 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.

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.

**FIGURE A1. PROBABILISTIC EVOLUTION OF THE TOTAL FERTILITY RATE IN 3,000 SIMULATIONS OBTAINED FROM EXPERT OPIONIONS.** YEARS 2021-2080

**The Limesurvey electronic questionnaire**

To create the electronic questionnaire for collecting data from the experts, the Limesurvey tool was used, an open-source software distributed under the GNU General Public License (GPL) and created on the LAMP platform, for the creation and management of online surveys and questionnaires.

It allows, through an intuitive user interface, the rapid development of web questionnaires and the management of all the subsequent phases of data collection, from the creation of the list of respondents and the related contact methods, to the monitoring of the survey, up to the export of the answers. Various types of questions are allowed, it supports multilingual surveys, and is fully graphically customizable via templates with responsive layouts, i.e. whose content adapts to the dimensions of the browser of the device used.

Istat began using the software more than a decade ago, hosting it and periodically updating it on its servers (exposed and internal), for different types of direct data collection processes in a web environment.

The questionnaire in question, composed of a personal data section and 6 thematic sections (of which 5 reserved for demographic forecasts and 1 for household forecasts, see paragraph 2 below), was implemented faithfully respecting the consistency and validation checks between the values of the questions within the same section and between those belonging to different sections. To this purpose, it was necessary to reprogram (in javascript) the interaction with the user to force him to insert in the various questions, from time to time, an appropriate number of decimal figures, customize the final table on family positions, as well as configure some general graphic aspects.

The list of experts was pre-loaded in the system and each of them was randomly assigned a unique participation code (token), through which the link (URL) for completing the questionnaire is composed. Each participant received an invitation email in their inbox with this link together with a brief information; he had the possibility to access the questionnaire from any browser and from any device (including smartphones) and to complete the questionnaire even in different sessions.

Finally, the system allowed the Limesurvey administrative backend to repeatedly solicit respondents who had not accessed or completed the questionnaire.

**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:

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

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:

$$DN\_{t}^{}=\sum\_{j}^{}DR\_{t}^{j}$$

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

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 (2122, 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 (2023-2080). 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 1977-2021;
* for mortality, the (projection-)probabilities of death by age and sex for the period 1974-2021;
* for internal and international migrations, the changes of residence by age and sex of 2016-2019 and 2021 (without 2020).

**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.

Since the base population of the projections is that recorded as of January 1, 2024, it was necessary put in place some short-term correction of the predicted inputs that affected the first projections years. The correction, in particular, takes advantage of the information from the provisional demographic balance - Year 2024, which Istat released in March 2025[[2]](#footnote-2). With this, in fact, we want to take into account not only the events that characterized the 2024, 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.

From the computational point of view, the review of the short-term assumptions is carried out by applying correction factors. For example, let E𝑗 be the number of demographic events predicted in the first year based on the median scenario in region j. Instead, let Ê 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:

$$r\_{b}^{j}={\hat{E}\_{b}^{j}}/{E\_{b}^{j}}$$

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:

$\hat{f}\_{b,x}^{n,j}=r\_{b}^{j}∙f\_{b,x}^{n,j}$ with x=14, … , 50 e 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 2023, the correction factors were constructed by comparing the data of the provisional demographic balance of each region, released in March 2024 by Istat, to the projections previously produced for that year 3.

For the years after 2024, 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:

$Y^{j}=abs\left(1-r\_{b}^{j}\right)∙ϵ$

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:

$r\_{t}^{j}=\frac{r\_{b}^{j}∙\left(b+Y\_{}^{j}-t\right)+\left(t-b\right)}{Y\_{}^{j}}$ with $t=b, b+1, …, b+Y^{j}-1$

**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.

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 2065, 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 1977-2021 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- 2021. 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 2122. In particular, the convergence constraint provides that, from 2024 to 2122, 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 1974-2021, 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 2080 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 1974-2021 with the same methodology and, subsequently, by making each regional parameter converge to the corresponding national parameter at 2122. 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 2016-2019 and 2021. The year 2020 was deliberately censored to avoid incorporating into the forecasts the effects of the lockdown. 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 (2122) at the same level, which is to the initial half sum of the two values.

Once the totals of inflows and outflows up to 2080 have been determined, the associated age and sex breakdown are derived by applying the Castro-Rogers model (Rogers and Castro, 1981) to the 2016-2019 and 2021 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 2015-2016 and 2021 period, censoring the year 2020 as in the case of international migration. 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 = 2016, ..., 2019, 2021), 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:

$μ\_{x,s}^{i,j}=E\left(m\_{x,s,t}^{i,j}\right)$

$σ\_{x,s}^{i,j}=E\left(m\_{x,s,t}^{i,j}-μ\_{x,s}^{i,j}\right)^{2}$

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 ‘23 and ’24.

First of all, a rather limited difference should be noted between the total base population ‘24 (58 million 971 thousand) and that which had been estimated in the median scenario on the same date by the projections based on ‘23 (58 million 990 thousand).

On the side of projected flows in the common projection section (2024-2080), 20.5 million births is captured in the exercise in base '24 (20.7 in base ’23) while we have the same number of deaths (43.7 million). Even for migrations with foreign countries, the global net migration balance is equal to 9.9 million individuals in both the 2023 and 2024 base forecasts.

The difference between the final populations of the two separate forecast exercises is small (about 300,000 more as of January 1, 2080 for the median base 2023 scenario), confirming the substantial robustness of the base 2022 forecasts, despite the base population change and short-term adjustments on the budgetary components. In this respect, Table A2 highlights how the process of revisiting assumptions for all demographic components affected only the early forecast years.

**TABLE A2. 2021 AND 2022 MEDIAN SCENARIO ASSUMPTIONS FOR THE MAIN DEMOGRAPHIC INDICATORS.** Years 2024, 2030, 2050 and 2080.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Median scenario** | **Total fertility rate** | **Life expectancy at birth – Men** | **Life expectancy at birth – Women** | **Immigration (thousand)** | **Emigration (tohusand)** |
| **Year 2024** |
| Base 2023 | 1.21 | 81.4 | 85.4 | 403 | 142 |
| Base 2024 | 1.18 | 81.7 | 85.6 | 435 | 191 |
| **Year 2030**  |
| Base 2023 | 1.28 | 82.0 | 85.9 | 339 | 141 |
| Base 2024 | 1.26 | 82.2 | 85.9 | 358 | 161 |
| **Year 2050** |
| Base 2023 | 1.38 | 84.3 | 87.8 | 301 | 136 |
| Base 2024 | 1.38 | 84.3 | 87.8 | 301 | 136 |
| **Year 2080** |
| Base 2023 | 1.46 | 86.1 | 89.7 | 305 | 143 |
| Base 2024 | 1.46 | 86.1 | 89.7 | 305 | 143 |

**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 ’22, 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 ’23 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 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).

The UNPD projections 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[[3]](#footnote-3).

The assumptions on fertility are quite similar, although in the medium-long term Eurostat and UNPD produce more favorable forecasts than Istat. The assumptions on survival are also not particularly distant, however Eurostat and especially UNPD highlight very favorable 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. The UNPD projections give a much more pessimistic evolution of the population which approximates the lower limit of Istat confidence interval. 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. 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

**TABLE A3. COMPARISON BETWEEN LATEST MAIN ASSUMPTIONS ON ITALY MADE BY ISTAT (MEDIAN SCENARIO), EUROSTAT (BASELINE) AND UNPD (MEDIUM).** Years 2024, 2030, 2050 and 2080.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Scenario** | **Total fertility rate** | **Life expectancy at birth –****Men** | **Life expectancy at birth –****Women** | **Net migration (thousand)** |
| **Year 2024** |
| Istat Mediano | 1.18 | 81.7 | 85.6 | 244 |
| Eurostat Baseline | 1.25 | 81.6 | 86.0 | 217 |
| UNPD Medium | 1.21 | 81.8 | 85.9 | 95 |
| **Year 2030** |
| Istat Mediano | 1.26 | 82.2 | 85.9 | 197 |
| Eurostat Baseline | 1.28 | 82.4 | 86.8 | 270 |
| UNPD Medium | 1.25 | 82.8 | 86.6 | 44 |
| **Year 2050** |
| Istat Mediano | 1.38 | 84.3 | 87.8 | 166 |
| Eurostat Baseline | 1.37 | 85.0 | 89.0 | 240 |
| UNPD Medium | 1.35 | 85.4 | 89.0 | 43 |
| **Year 2080** |
| Istat Mediano | 1.46 | 86.1 | 89.7 | 163 |
| Eurostat Baseline | 1.48 | 88.1 | 91.9 | 228 |
| UNPD Medium | 1.43 | 88.8 | 92.5 | 39 |

**FIGURE A2. TOTAL POPULATION ACCORDING TO ISTAT, EUROSTAT AND UNPD SCENARIOS.** Years 2024-2080, million.

**2) Households projections, by region. Years 2024-2050**

Household projections show the future trend of the number and type of households that will characterize the population in Italy from 2024 to 2050. These projections derive from the application of a static method, based on propensity rates, applied to the projected population. The purpose is to provide with an integrated system of information that can be useful to several users, both public and private, who deal with goods and services intended for families rather than for individuals. Given the importance of the role of the family, both at the protective level and for individual choices and paths, the demand for information on households arises from planning needs in various areas. First of all, we can consider the decisions to be taken in economic and social policies, such as those relating to housing, social and welfare systems for the young and the elderly. Last, improve the planning of productive strategies of durables goods for households and energy consumption is another potential task.

**Territorial level and time horizon**

Household projections are disseminated at regional and national level. The base population is the one observed on 1.1.2024 while the elaborations cover the period from 2024 to 2050.

**Data**

Several set of data have been implemented. Among them, the official probabilistic projections - base 1.1.2024 of the median scenario to be used as reference for the future evolution of the resident population by sex, age and region; the Base Population Register at 1.1.2023 and 1.1.2024 to estimate the share of population living in institutional cohabitation by sex and single year of age; the Multipurpose Survey "Aspects of daily life", which provide with a long time series (from 2002 to 2024), to derive the family structures by typology and position. The concept of household here in use is the "de facto family” that until 2022 identified a household as "*the set of people linked by ties of marriage, kinship, affinity, adoption, protection, or from emotional ties, cohabitants and having habitual residence in the same Municipality*". In 2023, the definition of household changes slightly with the inclusion of the concept of sharing economic resources: “*it is a group of people who habitually live in the same dwelling and are bound by marriage, civil partnership, kinship, affinity, adoption, guardianship or emotional ties and who share the household's income (by contributing to and/or benefiting from it) and/or daily expenses*”[[4]](#footnote-4).

In addition, since the 2023 edition, the AVQ survey has introduced a new procedure for weighting data to accurately represent the target population. In fact, in order to make the survey results consistent with the evidence annually derived from the Permanent Census of Population and Housing, the calibration constraints for the construction of weights are defined within the system of “Anticipatory Estimates of Demographic and Social Indicators.” Specifically, through this system, which provides the basis for weighting all Istat sample surveys, an estimate of the distribution of households by number of members is constructed in an anticipatory way, since the Permanent Census releases it in final form only the year following the year to which the survey data refer[[5]](#footnote-5).

**Method**

The model is based on an adaptation to the Italian context of the method known as "Propensity model". It is a static method that goes beyond the classic “Headship rate model”, overcoming the concept of 'head of household' and providing a much more detailed set of information. Predictions of the number of future households, their average size and composition, and the population by family role can be easily obtained. The method relies on Propensity rates, defined as the proportion of people of age x in household position i at time t:

$$Propensity Rate\_{x,i,t}=\frac{P\_{x,i,t}}{P\_{x,t}}$$

For example, the propensity for a 30-year-old person to live in a couple with a partner will be given by the ratio between the number of 30-year-old people living in a couple and the total population of 30-year-olds.

The advantages of the method are many: it ties easily to population projections; there is no need to analyse transitions between potential family positions, typical of a dynamic model; it is simple to apply and provides with high detailed results. However, some drawback is also present, which arise mainly from the static nature of the method, do not allowing to reproduce the process of household formation and dissolution. Thus the application of propensity rates to the resident population may in some cases determine inconsistencies in term of global results, for example between sexes or for household positions within age-classes, a problem that it is therefore necessary to solve with ex-post adjustments.

The method consists of 5 steps:

* Step 1. Estimate the base-year and projected population living in households
* Step 2. Calculate household propensity rates
* Step 3. Modelling future trends of household propensity rates
* Step 4. Derive the projected population in the different household positions
* Step 5. Calculate the number, type and size of projected households.

The various steps are explained in detail below.

**Step 1. Estimate the base-year and projected population living in households**

The base-year population is represented by the resident population by gender, age and region at January 1st, 2024, as collected from the last register-based Census. Then, making a preliminary estimate of the population living in households, excluding individuals residing in institutional households (hospitals, barracks, prisons, nursing homes, religious buildings, etc.) is necessary. This operation should then be repeated for every projected year, deducting from the regional projections (whose reference is also in this case the overall resident population) the share of the institutionalized population.

From the Base Population Register as of 1.1.2023 and 1.1.2024, the (average) percentage incidences of the population living in institutions by sex, five-year age group, and region were calculated. Given the substantial stability of the share of this population over time, these percentage values are assumed to be constant throughout the projection time horizon. Applying the 100 complement of these incidences to the total population from 2024 to 2050, we obtained the population living in households by region, sex, and age group (Figure A3).

**FIGURE A3. TOTAL POPULATION AND POPULATION LIVING IN HOUSEHOLD.** Years 2024-2050, median scenario, million.

**Step 2. Calculate household propensity rates**

The second step consists on calculating the propensity rates to live in a given household position by gender and 5-year age groups for the following 10 household positions:

1. lone person;
2. person in a childless couple;
3. person in a couple with at least one child under 20 years of age;
4. person in a couple in couple with all children aged 20 and older;
5. single parent with at least one child under 20 years of age;
6. single parent with all children aged 20 and older;
7. child (living with one parent in a couple or with a single parent);
8. other person living in a family household[[6]](#footnote-6);
9. person in multi-person household (e.g., 2 siblings living together or a divorced individual who has returned home to a parent);
10. person in a household with 2 or more families.

Positions from 2 to 8 refer to individuals in one-family households. People living in households with 2 or more family nucleus have been considered in a separate category, since this typology constitutes a small share of the total number of households (approximately 1.5%).

As mentioned above, propensity rates are constructed as the proportion of persons of age x in category i. In this context, the age variable was considered in five-year classes and the rates were also disaggregated by sex, as the latest variable is very discriminating in household behaviour. Hereinafter, these rates are referred to as *Living Arrangement Propensities* (LAP):

$$Propensity Rate\_{x,i,s,t}=\frac{P\_{x,i,s,t}}{P\_{x,s,t}}=LAP\_{x,i,s,t}$$

where x= five-year age group 0-4, 5-9, ....., 80-84, 85+, i= family position, s=sex, t=time.

LAPs are calculated using data from the *Aspects of Daily Life* (AVQ) survey, along the entire 2002-2024 time series. Since regional estimates by sex and age groups leads to a paucity of data in small regions, it was decided to group regions into "macro-regions”.

A multivariate statistical analysis, including various sociodemographic context factors[[7]](#footnote-7), has generated the following 5 groups of regions:

* Group 1 - North-west (Piemonte, Valle d'Aosta, Lombardia, Liguria);
* Group 2 - Eastern Adriatic (Veneto, Emilia-Romagna, Trentino-Alto Adige, Friuli-Venezia Giulia, Marche);
* Group 3 - Tyrrhenian (Toscana, Lazio);
* Group 4 - South (Campania, Puglia, Calabria, Sicilia);
* Group 5 - Central (Umbria, Sardegna, Abruzzo, Molise, Basilicata).

**Step 3. Assumptions on future trends of household propensity rates**

It is now necessary to make assumptions about the evolution of household propensities from 2023 to 2043. To this end, some modifications to the Propensity rates method have been introduced. The new approach is based on the introduction of a new synthetic indicator, constructed as the sum by age of the LAP, weighted by the years lived at the various ages ($L\_{x}$). This new indicator is named *Total Propensity Rate per household position* (TPT):

$$TPR\_{i,s,t}=\sum\_{x=0-4}^{85+}LAP\_{x,i,s,t}\*L\_{x,s,t}=\sum\_{x=0-4}^{85+}\frac{P\_{x,i,s,t}}{P\_{x,s,t}}\*100\*L\_{x,s,t}$$

where i=household position, s=sex, x=five-year age class, t=time.

$L\_{x,s,t}$ , representing the number of years lived in the age class x by sex s in year t, are derived from the projected life tables of the median scenario.

The TPR for a given household position represents how many years on average a generation of individuals expects to live in that position, assuming over the life course the family behaviours and mortality conditions as observed in a given calendar year. It is, therefore, a life expectancy in that family status, shifted from the cross-sectional to the longitudinal observational dimension. In other words, it takes on the same meaning that better-known cross-sectional indicators have, such as the average number of children per woman, the life expectancy at birth or the total marriage rate.

If in 2004 a man counted on living as a single person an average of 5.8 years (out of a total life expectancy of 77.6), in 2024 the expected time in this state rises to 11.7 years (out of a total of 81.4). In contrast, as a result of declining birth rates, in 2004 women expected to live 14.0 years (out of a total of 83.2) as a person in a couple with at least a child under 20, but in 2024 this expected time has fallen to 13 years (out of a total life expectancy that has since risen to 85.5 years). As a final example, the time in "child" status has increased from 30.6 to 31.1 years for males and from 27.8 to 28.8 for females, due to the prolonged stay of young people within the family of origin (Table A4).

**TABLE A4. TOTAL PROPENSITY RATES BY HOUSEHOLD POSITION AND SEX.** Years 2004, 2014 and 2024.

|  |  |  |  |
| --- | --- | --- | --- |
| **Household position** | **MALES** |  | **FEMALES** |
| **2004** | **2014** | **2024** |  | **2004** | **2014** | **2024** |
| Lone person | 5,8 | 8,6 | 11,7 |  | 10,8 | 12,4 | 14,0 |
| Partner in a childless couple  | 13,5 | 14,5 | 14,3 |  | 12,5 | 13,4 | 13,4 |
| Partner with at least a child <20  | 14,2 | 13,0 | 12,2 |  | 14,0 | 13,3 | 13,0 |
| Partner with all children >=20 years | 8,6 | 7,7 | 6,0 |  | 8,2 | 7,2 | 5,7 |
| Lone parent with at least a child <20 | 0,2 | 0,3 | 0,5 |  | 1,3 | 1,8 | 2,6 |
| Lone parent with all children >=20 years | 0,7 | 0,8 | 1,0 |  | 2,9 | 3,0 | 3,0 |
| Child | 30,6 | 31,1 | 31,1 |  | 27,8 | 28,6 | 28,8 |
| Other position in one-nucleus household | 0,9 | 0,9 | 0,8 |  | 1,9 | 1,6 | 1,3 |
| Person in a multi-person household | 1,1 | 1,2 | 1,9 |  | 1,5 | 1,5 | 1,8 |
| Person in a household with 2 or more nuclei | 2,2 | 2,1 | 2,0 |  | 2,3 | 2,3 | 2,0 |
| **Total** | **77,6** | **80,3** | **81,4** |  | **83,2** | **85,0** | **85,5** |

\*Since the Aspects of Daily Life survey was not conducted in 2004, values for this year were estimated as the average of the years 2003 and 2005.

In order to hypothesize future trends in propensities, we proceeded to project the Total Propensity Rates by single family position, and then to estimate its distribution broken down by age group ($LAP\_{x,i,s,t}$) in each projected year. Predicting total intensity in a first step made it possible, on the one hand, to more easily translate the assumptions about family behaviour and, on the other, to keep together the trends in the various household positions. These latter, if projected separately by single age group, would be more difficult to control with the risk of obtaining unreliable results (e.g., a higher rate for the “child family position” at intermediate ages than at younger ones).

The final goal of Step 3, which is to define the projected LAPs from 2023 to 2043 by region, was achieved by first performing the projection in the 5 established territorial groups (Step 3.1), and then moving from these to a regional detail (Step 3.2).

**Step 3.1 Projecting LAPs in the 5 territorial groups**

The total intensity of each family position and sex ($TPR\_{i,s,t}$) was predicted by trend extrapolation over the period 2002-2024, using time series analysis models. ARIMA, Random walk with drift or Linear Trend type models were applied for each family position and sex (Table A5).

**TABLE A5. PREDICTIVE MODELS OF TOTAL PROPENSITY RATES BY HOUSEHOLD POSITION AND SEX**

(Prevailing model among the 5 territorial groups) \*

|  |  |  |
| --- | --- | --- |
| **Household position** | **MALES** | **FEMALES** |
| Lone person | RWD || ARIMA(1,1,0) | RWD||ARIMA(0,1,0) |
| Partner in a childless couple  | ARIMA(1,1,0) | Linear Trend |
| Partner with at least a child <20  | RWD | ARIMA(2,2,0) |
| Partner with all children >=20 years | RWD||ARIMA(0,1,0) | RWD||ARIMA(0,1,0) |
| Lone parent with at least a child <20 | RWD||ARIMA(0,1,0) | RWD |
| Lone parent with all children >=20 years | RWD||ARIMA(0,1,0) | ARIMA(1,0,0) |
| Child | ARIMA(1,0,0) | RWD |
| Other position in one-nucleus household | ARIMA(2,1,0) | ARIMA(1,0,1) |
| Person in a multi-person household | RWD | Linear Trend |
| Person in a household with 2 or more nuclei | RWD | RWD||ARIMA(0,2,0) |

\*RWD=*Random Walk with Drift model*; ARIMA=*AutoRegressive Integrated Moving Average model*.

An example of the application of the above models is shown in Figure A4, where the behaviour of the Northwest spatial group for the main family positions is depicted. More generally, with regard to each area of the country, variations in time spent in various household position are assumed, resulting in:

* an increase of “single people”;
* a decrease of “partners with children”;
* a slight increase of “partners without children”;
* an slight increase of people in “child” position;
* a slight increase of “lone parents”, especially fathers;
* a slight decrease in “other people” living in households with one family nucleus and a slight increase for people in households consisting of two or more family nuclei.

**FIGURE A4. TOTAL PROPENSITY RATES BY HOUSEHOLD POSITION AND SEX. North-west.** Years 2002-2050.

**Males**

**Females**

Estimation of the predicted distribution by age, i.e. $LAP\_{x,i,s,t}$ from 2024 to 2050, was obtained using predicted TPR, predicted years lived $L\_{x}$, and observed distributions from AVQ survey data over the three-year period 2022-24. For this latter purpose, the average 2022-24 distributions by household position were weighed by two coefficients: one to account for the predicted TPR in year t relative to that in the 2022-24:

$$WP\_{s,i,t}=\frac{TPR\_{s,i,t}}{TPR\_{s,i,2022-24}} t=2024, …, 2050$$

and a coefficient expressing changes in mortality over time:

$$WL\_{x,s,t}=\frac{L\_{x,s,2022-24}}{L\_{x,s,t}} t=2024, …, 2050$$

Therefore, household propensities throughout the projection horizon were calculated using the formula:

$$LAP\_{x,s,i,t}=LAP\_{x,s,i,2022-24}\*WP\_{s,i,t}\*WL\_{x,s,t} t=2024, …, 2050$$

where: x=age groups 0-4, ... ,85+, s=sex, i=household position.

Since the method of estimation did not assume annual variation about the age distribution of LAPs, it was implicitly assumed that behaviours in terms of family choices would maintain in the future an age-group distribution proportional to that found in the AVQ survey in the three-year period 2022-2024. Last, the sum of LAPs by household position in each age group approximates but does not always equal the value of 100, so ex-post adjustments were necessary.

**Step 3.2. Projecting LAPs in the regions**

In order to project households at the regional level, it is necessary to ensure that each region has its own socio-demographic specificity within the projecting group to which it belongs. Considering that, a regional correction factor has been defined to be applied to the LAP projections of the various territorial groups in order to obtain those specific to each region:

$$FC\_{r,i}=\frac{TPR\_{2022-24, i, r}}{TPR\_{2022-24,i, G}}$$

where i=household position, r=region, G=group to which region r belongs.

The projected LAPs for the spatial groups are then multiplied by the regional correction factor calculated in this way, determining the series of regional LAPs from 2024 to 2050. For example, for the single male person, the detected TPR in Piemonte is 11.9 while in group 1 it is 11.1. The correction factor is therefore equivalent to 1.07 in this case. This means that since Piemonte has a higher TPR than the group to which it belongs, an adjustment must be made by multiplying all LAPs at different ages and projected years by 1.07, increasing the level slightly.

**Step 4. Obtain the projected population by single household position.**

In this step, the regional propensities are applied to the projected population living in households, as it was obtained in Step 1. The projected population in the different household states by sex, age group, and region from 2024 to 2050 is then derived.

**Step 5. Calculate the number, type, and size of projected households**

The projected number of households is obtained directly from the population separated by family position, gender, and age, as:

* each "single person" represents 1 household (coefficient=1);
* persons in a couple constitute 0.5 of a family (coefficient=0.5);
* each "single parent" represents 1 family (coefficient=1);
* “multi-person households" are obtained by dividing the number of persons living in multi-person households by the average size of this type of household, substantially stable over time and equal to about 2.1 members (coefficient=2.1);
* households "with 2 or more families” are obtained by dividing the number of persons living in households with 2 or more families by the average size of this type of household, which assumes time series coefficient values between 5.1 and 5.4 depending on the territorial group of reference.

Applying these coefficients to the population of sex s and age x yields as a final product the number of households by the family types of interest.

The average number of members is then calculated by dividing the population living in the household to the number of households. It can be disaggregated for total households and those with at least one family (excluding single persons and multi-person households).

For dissemination purposes, multi-person households and those with 2 or more families are considered together in the "other type of household" item.

**3) Data dissemination and terms of use**

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

The dissemination of the **population projections** is implemented into the three following sections: population structure by sex and single age group; components of the population change; main demographic indicators. Each table shows the values ​​of the median scenario and the lower and upper limits of the 90%, 80% and 50% confidence intervals.

The components of the population balance include:

* population at start and end of the year, total variation
* births, deaths, natural change
* in-migration and out-migration, net migratory balance
* interregional changes of residence, net internal migration balance.

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

Regarding the demographic indicators, the tables include:

* birth, mortality and natural growth rates
* immigration rate from abroad, emigration for abroad and net migration rate with abroad
* internal immigration rate, internal emigration rate and net internal migration
* total net migration rate and total growth rate
* mean age of the population
* % of population 0-14 years, 15-64 years, 65 years and over, 85 years and over
* structural dependency ratio, elderly dependency ratio and aging index
* total fertility rate
* life expectancy at birth and at 65 years of age by sex.

The dissemination of data relating to **household projections** is articulated into three sections including tables that can be processed on the structure by sex, five-year age group and household position of the population, on the distribution of household by type and on the average number of household members. All results refer to the median scenario.

The reproduction of the information contained in this note and in the databases dati.istat.it and demo.istat.it is left free, provided that the Istat source is quotes.

Istat periodically produces population projection under the activity line “System of Population Estimates and Projections,” in accordance with the National Statistical Program, project “Population Projections” (PSN code IST-01448).

**References**

1. Billari, F.C., Corsetti G., Graziani R., Marsili M. e Melilli E. (2014), *Towards stochastic forecasts of the Italian population: an experiment with conditional expert elicitations*. Proceedings of the 6th Eurostat/UNECE Work Session on Demographic Projections, pagg. 326-338, Istat, 2014.
2. Billari, F.C., Corsetti G., Graziani R., Marsili M. e Melilli E. (2014), *A stochastic multi-regional model for Italian population projections*. Budapest, 25-28 giugno 2014, European Population Conference. <http://epc2014.princeton.edu/papers/140361>.
3. Corsetti G., Marsili M. (2012), *A stochastic population projection from the perspective of a national statistical office*. European Population Conference. Stoccolma, 13-16 giugno 2012, EAPS. <http://epc2012.princeton.edu/papers/120635>.
4. Corsetti G., Marsili M. (2013), *Previsioni stocastiche della popolazione nell’ottica di un Istituto nazionale di statistica*. Rivista di statistica ufficiale, n. 2-3, p. 5-29, Istat.
5. Istat (1989), *Previsioni della popolazione residente per sesso, età e regione – Base 1.1.1988*, Note e Relazioni, n.4.
6. Istat (1989), *Previsioni del numero di famiglie italiane dal 1995 al 2020*. Notiziario. Serie 4. Foglio 41 – Anno X No. 19. Dicembre 1989.
7. Istat (1997), *Previsioni della popolazione residente per sesso, età e regione – Base 1.1.1996*, Informazioni, n. 34.
8. Istat (2001), *Previsioni della popolazione residente base 1° gennaio 2000*, Statistiche in breve, [www.istat.it](http://www.istat.it).
9. Istat (2003), *Previsioni della popolazione residente per sesso, età e regione, base 1.1.2001*, Informazioni n.13.
10. Istat (2006), *Previsioni demografiche nazionali 1° gennaio 2005-1° gennaio 2050*, [www.istat.it](http://www.istat.it), Nota informativa, 22 marzo 2006.
11. Istat (2008), *Previsioni demografiche 1° gennaio 2007-1° gennaio 2051*, [www.istat.it](http://www.istat.it), Nota Informativa, 19 giugno 2008.
12. Istat (2011), *Il futuro demografico del paese - Previsioni regionali della popolazione residente al 2065*, Statistiche Report, [www.istat.it](http://www.istat.it), 28 dicembre 2011.
13. Istat (2017), *Il futuro demografico del Paese – Previsioni regionali della popolazione residente al 2065*, Statistiche Report, [www.istat.it](http://www.istat.it), aprile 2017.
14. Istat (2018), *Il futuro demografico del Paese – Previsioni regionali della popolazione residente al 2065 – base 1.1.2017*”, Statistiche Report, [www.istat.it](http://www.istat.it), maggio 2018.
15. Istat (2019), *Il futuro demografico del paese - Previsioni regionali della popolazione residente al 2065 – base 1.1.2018*, Nota metodologica, [www.istat.it](http://www.istat.it), ottobre 2019.
16. Istat (2021), *Previsioni della popolazione residente e delle famiglie – base 1.1.2020*, Statistiche Report, [www.istat.it](http://www.istat.it), novembre 2021.
17. Istat (2022), *Previsioni della popolazione residente e delle famiglie – base 1.1.2021*, Statistiche Report, [www.istat.it](http://www.istat.it), settembre 2022.
18. Istat (2023), *Previsioni della popolazione residente e delle famiglie – base 1.1.2022*, Statistiche Report, [www.istat.it](http://www.istat.it), settembre 2023.
19. Istat (2024), *Indicatori demografici – Anno 2023*, Statistiche Report, [www.istat.it](http://www.istat.it), marzo 2024.
20. Marsili M. (2007), *Demographic projections: the impact of net international migration on population ageing in Italy*, Atti del Convegno Intermedio della SIS 2007 “Rischio e Previsione”, Università Ca’ Foscari, Venezia, 6-8 giugno.
21. Marsili M. (2020) *Scenari demografici, previsioni per l’uso*, Atti della 13° Conferenza nazionale di statistica, Dall’incertezza alla decisione consapevole: un percorso da fare insieme, Roma, 4-6 luglio 2018, pagg. 246-252, Istat, 2020.
22. Terra Abrami V. (1998), *Le previsioni demografiche*, Il Mulino, Bologna.
23. UNECE (2018), *Recommendations on Communicating Population Projections*, United nations economic commission for Europe, United Nations, New York and Geneva, agosto 2018.
1. The 2019-base forecasts, theoretically belonging to the second production cycle, skipped due to the unavailability in the required timeframe of the base population, for the first time in history derived from the new Permanent Population Census, as well as the need to produce a historical population reconstruction for the period 2002-2018 and to recalculate all the demographic indicators of reference for the forecasts (in particular, probability of death and specific fertility rates). [↑](#footnote-ref-1)
2. Cfr.: Istat, Indicatori demografici – anno 2024, https://www.istat.it/comunicato-stampa/indicatori-demografici-anno-2024/ [↑](#footnote-ref-2)
3. 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). [↑](#footnote-ref-3)
4. The change of AVQ's definition of household in 2023 is a consequence of the amendment of the Eurostat Regulation on Social Statistics, in which this definition is adopted. [↑](#footnote-ref-4)
5. This major process innovation is based on a methodology that takes advantage of the current historical series about the distribution of households by number of members from census sources and, through an appropriate extrapolative nowcasting model, is able to estimate number of households and its breakdown for the year for which the information is needed. [↑](#footnote-ref-5)
6. Family refers to people who form a couple or a parent-child relationship. It means a married couple, civilly united or cohabiting, without children or with never married children, or even a single parent together with one or more children who have never been married. Within a household there may be one or more families (family households), but there may also be none, as in the case of households formed by a single member (one person household) or several isolated members (multi-person household). [↑](#footnote-ref-6)
7. In order to identify homogeneous groups of regions, united by common family structures and similar evolution over time, a dynamic principal component analysis was carried out using the STATIS methodology. The analysis examined the main socio-demographic variables at the regional level in the years 2002-2019, including: fertility rates, mean age at birth, average size of families, separation and divorce rates, female employment rates, internal and foreign migration rates, quotes of some family types (single people, couples with and without children, single parents, etc.). The procedure was optimized by eliminating the variables with low latent variability explained by the axis. [↑](#footnote-ref-7)