ANTHROPOLOGIE • LXII/2 • pp. 83–95 • 2024

PATRIK GALETA

DEMRAT: AN R PACKAGE FOR PREDICTING GROWTH AND FERTILITY RATES IN SKELETAL SAMPLES USING AGE-AT-DEATH RATIOS

ABSTRACT: The growth and fertility rates of past populations can be estimated by analyzing the age-at-death distribution of skeletal samples. The procedure involves regressing growth or fertility rate on the age-at-death ratio, which is a proxy that captures the number of skeletons in two broad age-at-death categories (e.g., D5+/D20+). Galeta and Pankowská (2023, doi: 10.1371/journal.pone.0286580) recently developed a new prediction algorithm. They proposed to estimate growth and fertility rates using a unique prediction formula for each skeletal sample. Each formula is based on a unique reference set of simulated skeletal samples that match the size of the target real skeletal sample. The simulated skeletal samples are generated from populations with similar mortality levels to those assumed in the time period represented by the target skeletal sample. A correct setting of the sample size and the level of mortality increases the accuracy of the estimate. The approach, however, is computationally intensive because it involves generating many simulated reference skeletal samples. In this paper, we present the demrat package, written in the R programming language, which automates the simulation. The functions of the package provide a complete workflow from a real skeletal sample to the prediction of demographic rates. In addition, we offer a web application that allows non-R users to deploy predictions using the demrat package with a user-friendly, point-and-click graphical interface. Although the demrat package allows for estimating demographic rates for a single skeletal sample, we recommend predicting demographic rates in a larger set of skeletal samples and producing smoothed general demographic trends over large areas and/or long periods of time.

KEY WORDS: R package – Age-at-death ratio – Growth and fertility rate – Paleodemography – Stochastic error

INTRODUCTION

Predicting growth and fertility rates from age distribution of deaths in skeletal samples is of increasing interest to palaeodemographers (McFadden 2021). The approach is based on the observation that the age distribution of deaths reflects the level of fertility rather than mortality (Johansson, Horowitz 1986, Milner *et al.* 1989,

Received 27 November 2023; Accepted 13 March 2024. Available online 7 May 2024.

DOI: https://doi.org.10.26720/anthro.24.02.16.1

Sattenspiel, Harpending 1983). A population with high fertility rates tends to have a larger number of children, which in turn contributes to a higher proportion of total deaths (McFadden 2021).

AGE-AT-DEATH RATIOS AND PREDICTION OF GROWTH AND TOTAL FERTILITY RATE

The proportion of children and juveniles in the sample of deaths (skeletons) is quantified by a simple proxy indicator – age-at-death ratio – the ratio of the number of skeletons in two broad age-at-death categories. For example, the D5+/D20+ age-at-death ratio (Galeta, Pankowská 2023a) divides the sum of individuals over 5 years (D5+) with the number of adult deaths (D20+). Several other age-at-death ratios have been introduced over the years. The most popular are the D5–19/D5+ ratio (called P index, Bocquet-Appel 2002, 2011), the D0–14/D ratio (McFadden, Oxenham 2018a, 2018b) or the D30+/D5+ ratio (Buikstra *et al.* 1986). All age-at-deaths ratios closely correlate with growth and fertility rates.

The major advantage of age-at-death ratios in estimating growth and fertility levels is that they overcome the major challenges of paleodemographic analysis: the error in estimating age-at-death from the skeleton and the under-enumeration of pre-adult individuals in skeletal samples (Boldsen *et al.* 2022, Galeta, Pankowská 2023a, Taylor *et al.* 2023). The ratios minimize the error in estimating age-at-death from the skeleton because they combine all skeletons into broad age-at-death categories. For instance, the D5+/D20+ ratio assigns all adults to the D20+ category, regardless of whether they are young, middle-aged, or elderly. The ratios are reliable proxies even in skeletal samples with poor preservation of children, because most of them exclude the deaths of newborns and the youngest children. Although some areas show excellent preservation of pre-adult skeletons (e.g., Southeast Asia, McFadden, Oxenham 2018a, Taylor *et al.* 2023), the majority of skeletal assemblages are characterized with the under-enumeration of children skeletons due to several factors including differential burial, decomposition, and recovery (Guy *et al.* 1997, Manifold 2013, Pokines, De La Paz 2016).

The methodology for predicting growth and total fertility rates from skeletal samples involves regressing a growth or fertility rate on the age-at-death ratio (Bocquet-Appel 2002, Galeta, Pankowská 2023a). The prediction methodology consists of two stages (*Figure 1*). Stage 1 involves constructing a reference set of populations with a known growth or fertility rate and age-at-death ratio. In Stage 2, a regression equation is calculated based on the reference set from Stage 1. This equation is then used to predict the unknown growth or total fertility rate from the age-at-death ratio calculated in the target skeletal sample.

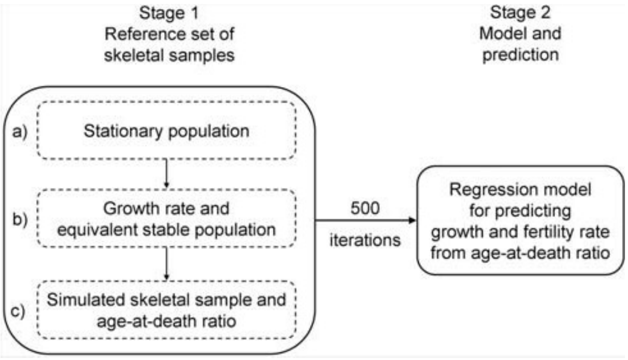


FIGURE 1: Outline of the algorithm for predicting the demographic characteristics of a population represented by a skeletal sample using age-at-death ratios (see, Galeta, Pankowská 2023a for details).

ROLE OF MORTALITY IN REFERENCE POPULATIONS FOR PREDICTION

The published age-at-death ratio methods differ mainly in the construction of the reference set of populations and their age distribution of deaths (Stage 1). Buikstra *et al.* (1986) constructed the reference set based on 24 model life tables (Coale, Demeny 1966) with the life expectancies between 18 and 80 years. Each table was subjected to the annual growth rate between −1 and 5% (in 0.5% increments), resulting in a final set of 312 stable reference populations. Bocquet-Appel (2002) favored the use of 45 life tables of real historical and modern populations with life expectancies at birth between 18 and 38 years subjected to the annual growth between −2.5 and 2.5% (in increments of 0.25%, i.e., 504 life tables in total). Finally, McFadden and Oxenham (2018b) opted to use crude numbers of deaths of 52 modern countries from the 1960 United Nation Database. The life expectancies at birth in these countries ranged between 42 and 74 years and their populations grew between −0.3 and 6.4% p.a.

Setting the level of mortality in reference populations is crucial for reliable prediction of growth or fertility rates since the hazard of dying at age x determines the number of deaths in age groups that are used to calculate the age-at-death ratio (Galeta, Pankowská 2023a). *Figure 2* demonstrates how setting mortality (life expectancy) affects the final prediction of annual growth rate on the example of Neolithic skeletal sample from Nitra, Slovakia (Bocquet-Appel 2002, Table 1, pp. 640–641). The D5+/D20+ age-at-death ratio in the Nitra skeletal sample is 1.28. The annual growth rates corresponding to this ratio may vary significantly depending on the reference level of mortality used in Stage 1 (*Figure 1*) of the prediction model. If the life expectancy of reference populations is set between 20 and 25 years, the model predicts an annual growth rate of 0.3%. However, if the prediction model is based on reference populations with a life expectancy between 40 and 45 years, the model leads to a much higher growth rate prediction of 1.8%. In general, the lower the mortality rate (i.e., the higher the life expectancy) in the reference population, the higher the predicted growth rate, assuming that fertility is held constant.

Figure 2 illustrates why the prediction of growth or total fertility rate from the age-at-death ratio may be unreliable if mortality levels differ between the reference populations and the population represented by the target skeletal sample. For example, prediction equations published by Bocquet-Appel (2002) or McFadden and Oxenham (2018b) would likely overestimate growth rates for prehistoric populations whose life expectancies probably did not exceed 25 years (Bocquet-Appel 2008) or 30 years (Gage, DeWitte 2009) because both prediction equations are based on reference populations with higher life expectancies (18–38 years in the former and 42–74 years in the latter method).

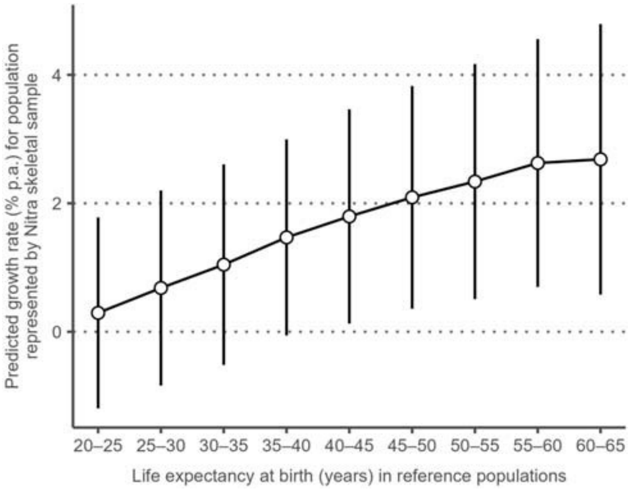


FIGURE 2: Effect of mortality level (life expectancy) of reference populations on predicted annual growth rate. The prediction is based on a skeletal sample from the Nitra site (n=72 skeletons, the D5+/D20+ ratio is 1.28, Bocquet-Appel 2002, Table 1, pp. 640–641) and computed using the `diest` function of the `demrat` package.

Another problem with existing age-at-death methods is that they do not account for the existence of stochastic errors that may affect the age distribution of deaths. Skeletal samples typically consist of only a few tens or hundreds of skeletons, and a small number of accidental deaths can significantly affect the age-at-death ratio and predictions of growth or total fertility rate (Galeta, Pankowská 2023a).

A NEW PREDICTION ALGORITHM

Recently, Galeta and Pankowská (2023a) presented a new method for prediction growth, birth, and total fertility rate from age-at-death ratio in skeletal samples, which accounts for both the effect of reference level of mortality and stochastic variation in small samples. In contrast to the previous authors (e.g., Bocquet-Appel 2002, Buikstra *et al.* 1986, McFadden, Oxenham 2018a, 2018b), Galeta and Pankowská avoided offering a universal prediction formula that would be applicable to skeletal samples of all sizes representing all time periods. Instead, they suggested calculating a specific prediction formula for each skeletal sample. The formula is based on a set of reference skeletal samples that are customized for each skeletal sample analyzed. The reference skeletal samples are equal in size to the target skeletal sample (taking into account stochastic error) and are derived from populations whose mortality is comparable to that assumed for the time period represented by the target skeletal sample (minimizing the effect of mortality). It means, for example, that predicting total fertility rate of Neolithic population represented by the Nitra skeletal sample with 47 adult skeletons involves building the set of (500–1000) reference skeletal samples with 47 adult skeletons that are randomly drawn from the reference populations with life expectancy between 18 and 25 years, the level expected for Neolithic period (Bocquet-Appel 2008).

The new prediction algorithm involves generating a set of reference skeletal samples (Stage 1 in *Figure 1*), calculating a unique regression formula, and predicting growth or total fertility rate from the age-at-death ratio in the real skeletal sample under study (Stage 2 in *Figure 1*) (see, Galeta, Pankowská 2023a for details). Since a unique prediction model is constructed for each target skeletal sample, the prediction algorithm requires computationally demanding operations.

The most computer-intensive part is generation of a large number of reference skeletal samples (Stage 1 in *Figure 1*). Individuals of each reference skeletal sample (i.e., deaths) are generated by uniform random sampling from the cumulative age distribution of deaths (Paine, Harpending 1998). The cumulative age distribution of deaths is based on Coale and Demeny abridged model life tables (Coale *et al.* 1983), which are first smoothed using Siler's five-parameter competing-risk mortality model, resulting in 63 complete life tables with the life expectancy between 18 and 80 years. Such a complex simulation is difficult to perform manually, so automation is desirable.

In this study, we introduce an R package called `demrat` that allows the prediction of growth, birth, and total fertility rate using the D5+/D20+, D3+/D20+, or D1+/D20+ age-at-death ratios calculated in a real skeletal sample using the methodology described in Galeta and Pankowská (2023a). The package allows for flexible settings of the input parameters of the prediction model, namely (a) mortality and (b) growth rate levels of the reference populations, and (c) the number of reference skeletal samples generated. Here, we describe the main functions of `demrat` package and provide a complete workflow from a real skeletal sample to the prediction of demographic rates. The package is designed for R users and prediction is made in R environment. For non-R users, we provide a web application that allows users to deploy predictions with a point-and-click graphical user interface.

INSTALLATION AND LOADING `demrat` PACKAGE

The `demrat` package is available on GitHub (https://github.com/galetap/demrat). The latest version is installed using `devtools` R package as follows:

```
if(!require(devtools))
install.packages("devtools")
devtools::install_github("galetap/demrat")
```

The package is installed once, but needs to be loaded in R/RStudio in every session by calling:

```
library(demrat)
```

INPUT DATA

The `demrat` package accepts data in two formats; in raw data and summary data format. The package includes examples of both data formats. Both datasets are adopted from Bocquet-Appel's (2002, Table 1, pp. 640–641) set of European Meso- and Neolithic skeletal samples.

The raw data format consists of age-at-death estimates of individual skeletons from one or more skeletal samples. Data must contain at least four variables with the following names: Site, Culture, Age_min, and Age_max. The latter two variables represent the lower and upper limits of the age-at-death estimates, respectively (e.g., 4 and 8 for skeleton with the age-at-death estimate of 6±2 years). Note that the age-at-death intervals are closed on the left and open on the right. Age_min of 4 and Age_max of 8 means that individual died between 4.0 and 7.9 years of age. The example dataset is stored in the `BAraw` R object and contains age-at-death estimates for 5,115 skeletons from 68 skeletal samples of Bocquet-Appel's (2002) dataset.

The summary data format consists of summary information about the skeletal sample(s). Data must have at least four variables with the following names: Site, Culture, D20_, and one out of the following three variables D5_D20_, D3_D20_, or D1_D20_. D20_ is the number of adult skeletons (20+) in the sample. D5_D20_, D3_D20_, and D1_D20_ are the D5+/D20+, D3+/D20+, and D1+/D20+ age-at-death ratios, respectively. The example dataset is stored in the `BA` R object and contains summary statistics for 68 skeletal samples of Bocquet-Appel's (2002) dataset.

MAIN FUNCTIONS OF `demrat` PACKAGE

The list of functions of the `demrat` package is provided in *Table 1*. The functions provide a complete analysis of the skeletal sample(s) from input data (age-at-death estimates of individuals) to outputs (predicted growth, birth, and total fertility rates). The functions allow for flexible input parameter settings, such as the mortality and growth rate levels of reference populations and the number of simulated reference skeletal samples.

COMPLETE WORKFLOW WITH `demrat` PACKAGE

The following text provides step-by-step instructions on how to use `demrat` package to import input skeletal dataset, to convert them to the format used by the package, and to estimate growth, birth, and total fertility rates.

TABLE 1: The list of main functions of `demrat` package.

Function	Description
`dr`	Translate raw data format to summary data format. Raw data format contains age-at-death estimates for each skeleton. The function converts raw data format to the summary data format by computing number of deaths in several age-at-death categories and six age-at-death ratios.
`diest`	Estimates demographic rates for skeletal sample(s). The function estimates growth, crude birth, and total fertility rates based on age-at-death ratios by Site and Culture. The function accepts data in the summary format only. The function can be applied to one or more skeletal samples. The function also allows for the setting of the mortality level of reference populations and generates reference skeletal samples with the same number of adults as in the skeletal sample under study.
`plot_diest`	Plots prediction model for a single skeletal sample. The function visualizes the prediction model and can only be applied to a single skeletal sample. The prediction model is plotted as a scatterplot with the age-at-death ratio (independent variable) on the X axis and a demographic variable (dependent variable) on the Y axis. The scatter shows the set of simulated reference skeletal samples (points), regression line used for a prediction, the age-at-death ratio, and predicted value for a given target skeletal sample.
`age_div`	Proportionally distributes an individual (skeleton) into 100 one-year categories. Age-at-death estimates from skeleton cross cut several age categories (4-8, 20-40 years). The function allows for distributing individuals into 100 parts, representing years of age between 0 and 99. For example, an individual with an age-at-death estimate of 6±2 years has 0.25 entries for the year of 4, 5, 6, and 7, and zero elsewhere (0-3 and 8-99 years). This is a helper function that can also be useful, for example, in computing life tables.
`simdr_CD`	Creates a reference set of simulated skeletal samples. The function is a helper that creates a reference set of skeletal samples drawn from Coale and Demeny's (1983) model life tables. Arguments can be set as in the `diest` function. The outputs are primarily used within the diest function but can also be used elsewhere, such as to assess the relationship between demographic variables and age-at-death ratios.

In the examples below, we consistently use `tidyverse`-style R code (Wickham, Grolemund 2017). Before proceeding with the examples below, make sure to install the `tidyverse` package if you have not already done so, and load the package with the following code:

```
install.packages("tidyverse")
library(tidyverse)
```

LOADING INPUT DATA

To import skeletal data to R/RStudio, users can choose from a variety of input file formats. For instance, they can use the `read_csv` function of the `readr` `R` package (https://readr.tidyverse.org/reference/read_delim.html) to import data from a CSV file or the `read_excel` function of the `readxl` R package (https://readxl.tidyverse.org/reference/read_excel.html) to import data from an MS Excel file.

CONVERTING DATA FROM THE RAW TO THE SUMMARY DATA FORMAT (`dr` FUNCTION)

The `demrat` package for estimating demographic rates from skeletal samples requires data in summary data format, which consists of age-at-death ratios computed at the level of skeletal sample. However, skeletal data are usually collected in the raw data format, which provides age-at-death estimates for individual skeletons. The `demrat` package provides a `dr` function that easily converts data from raw to summary data format. The function computes numbers of skeletons in ten age-at-death categories (D0, D1+, D3+, D5+, D15+, D0–14, D5–14, D5–19, D20+, and n, the total number of skeletons) and six age-at-death ratios: D1+/D20+, D3+/D20+, and D5+/D20+ (Galeta, Pankowská 2023a), JI (the Juvenility index, D5–14/D20+, Bocquet-Appel, Masset 1982), the P index (D5–19/D5+, Bocquet-Appel 2002), and D0–14/D (McFadden, Oxenham 2018a). All summary information is computed separately by Site and Culture. For example, `BAraw` embedded input data file that is in raw format can be converted to the summary data format by running:

```
dr(BAraw)
```

ESTIMATING GROWTH, BIRTH, AND TOTAL FERTILITY RATE (`diest` FUNCTION)

Estimation of demographic rates from age-at-death ratios using the `demrat` package is performed by the `diest` function. The function expects data in summary format with at least four mandatory variables (see Input data chapter above) and can be applied for single or multiple skeletal samples. Note that the estimating algorithm involves simulating many reference skeletal samples, which may take several minutes for samples with a large number of skeletons and/or estimation for many skeletal samples.

The demographic estimation for two skeletal samples of Bocquet-Appel (2002) dataset is done using the following code:

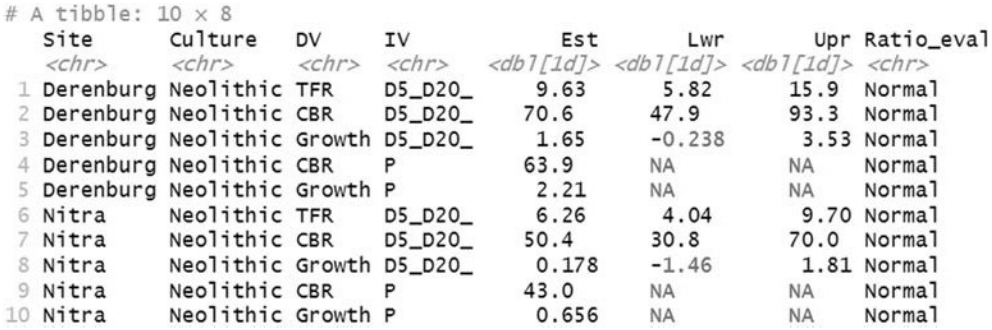
```
BA %>%
  dplyr::slice(22,44) %>%
  diest()
```

The result is a tibble (a dataframe in `tidyverse` syntax) shown in *Figure 3*. Each row in the table presents an estimate of a demographic indicator (`DV` column, dependent variable) predicted based on the corresponding age-at-death ratio (`IV` column, independent variable) for a single skeletal sample and time period (`Site` and `Culture` columns). Predicted values are given in `Est` column and lower and upper limits of 95% prediction interval are provided in `Lwr` and `Upr` columns, respectively. Prediction intervals can be computed only for the D5+/D20+, D3+/D20+, and D1+/D20+ age-at-death ratios. The structure of the resulting tibble enables easy manipulation with the result, e.g., filtering rows, mutating columns, exporting predictions from R/RStudio to a spreadsheet.

In addition to predictions of demographic rates, the `diest` functions also evaluates whether the age-at-death ratio measured in a target skeletal sample may be biased (see, `Ratio_eval` column in *Figure 3*). The `diest` function compares the observed age-at-death ratio value with its distribution in reference skeletal samples. If the observed value of the ratio is not likely for skeletal samples of the given size and demographic regimes expected in the simulation (see below), the function returns "`Out of limit`" value. Age-at-death ratio may be biased due to cultural, taphonomic, or excavation factors. In these cases, prediction should be avoided.

The `diest` function has eleven input parameters (see below for default values) that allow the prediction model to be customized to the user's preferences.

The setting of the two input parameters (`e0_min` and `e0_max`) should be carefully considered. The parameters determine the level of mortality (life expectancy) of the reference populations and should be set at a level that is appropriate for the time period from



```
# A tibble: 10 × 8
   Site      Culture    DV     IV               Est         Lwr         Upr Ratio_eval
   <chr>     <chr>      <chr>  <chr>    <dbl[1d]>   <dbl[1d]>   <dbl[1d]> <chr>
 1 Derenburg Neolithic  TFR    D5_D20_       9.63        5.82        15.9  Normal
 2 Derenburg Neolithic  CBR    D5_D20_      70.6        47.9         93.3  Normal
 3 Derenburg Neolithic  Growth D5_D20_       1.65       -0.238        3.53 Normal
 4 Derenburg Neolithic  CBR    P            63.9        NA           NA    Normal
 5 Derenburg Neolithic  Growth P             2.21       NA           NA    Normal
 6 Nitra     Neolithic  TFR    D5_D20_       6.26        4.04         9.70 Normal
 7 Nitra     Neolithic  CBR    D5_D20_      50.4        30.8         70.0  Normal
 8 Nitra     Neolithic  Growth D5_D20_       0.178      -1.46         1.81 Normal
 9 Nitra     Neolithic  CBR    P            43.0        NA           NA    Normal
10 Nitra     Neolithic  Growth P             0.656      NA           NA    Normal
```

FIGURE 3: A screenshot showing the prediction of growth, crude birth, and total fertility rate for two skeletal samples from the Bocquet-Appel (2002) dataset (Derenburg, Germany and Nitra, Slovakia) using the `diest` function from the `demrat` package. The arguments of the `diest` function are set to default values.

which the target skeletal sample is derived. For example, the default settings (`e0_min=18` and `e0_max=25`) are reasonable for Neolithic skeletal samples (Bocquet-Appel 2008, Gage, DeWitte 2009). Inappropriate values of these two arguments lead to unreliable prediction of demographic rates (see *Figure 2* and chapter Role of mortality in reference populations for prediction).

Two other arguments (`growth_min` and `growth_max`) specify annual growth rates (in %) of reference population (*Figure 1*, Stage 1b). The default settings (`growth_min=-3` and `growth_max=3`) seems to be a reasonable limits for most past human populations.

The `samples` argument specifies the number of simulated reference skeletal samples. The default setting (`samples=100`) is suitable for preliminary simulations because it saves computational time. To improve the reliability of the prediction model, we recommend increasing the number of simulated skeletal samples to 500 in the final predictions by setting `samples=500`. Preliminary simulations have revealed that increasing the number of simulated skeletal samples above 500 does not affect the final predictions, but slows down the computation significantly.

The Boolean (true/false) argument `sss` determines whether prediction algorithm involves creating simulated skeletal samples (*Figure 1*, Stage 1c). If `sss=FALSE`, the reference consists of population data and the step in the algorithm where the simulated skeletal samples are created (Stage 1c), is skipped. Reference populations then consist from large (infinite) number of skeletons and the prediction does not reflect the impact of stochastic error on the distribution of deaths. If `sss=TRUE`, the prediction algorithm involves Stage 1c step in *Figure 1* and it is based on a set of simulated skeletal samples with the same number of adult skeletons (D20+) as the target skeletal sample. The default is `sss=TRUE`.

The confidence level of prediction is set by the `pred_level` argument. The default is `pred_level=0.95`, i.e., the `diest` function computes 95% prediction intervals.

The remaining arguments of the `diest` function make the results more user-friendly without altering the results themselves: argument `IV` allows only selected age-at-death ratios (independent variables) to be used in the prediction (`IV="All"` by default), and argument DV allows only selected demographic (dependent) variables to be estimated (`DV="All"` by default).

The `summary` argument affects the level of detail shown in the results. The default setting is `summary=TRUE`, which only shows the predicted values of demographic variables with their prediction intervals (*Figure 3*). If the argument is changed to `summary=FALSE`, the function returns complete results including regression model characteristics (variance explained, regression coefficients, etc.), complete reference set of simulated skeletal samples for each model, and other details.

While the `diest` function can estimate demographic variables for a single skeletal sample, we highly recommend analyzing a larger set of skeletal samples. This approach produces smoothed general trends in demographic rates over large areas and/or long time periods, which is more reliable than predicting demographic characteristics of individual sites (Milner, Boldsen 2023). An example is study by Galeta and Pankowská (2023b) describing changes in growth and fertility levels in the Middle Ages in Central Europe. The study is based on evidence of almost 13 thousands skeletons from 59 medieval skeletal samples dated between 500 and 1000 AD.

PLOTTING REGRESSION MODEL USED FOR DEMOGRAPHIC PREDICTION (`plot_diest` FUNCTION)

The `plot_diest` function visualizes results of demographic prediction. *Figure 4* shows the prediction model for Neolithic skeletal sample from Nitra (Bocquet-Appel 2002, Table 1, pp. 640–641). The plot has the D5+/D20+ age-at-death ratio (independent variable, `IV="D5_D20_"`) on the X axis and annual growth rate (dependent variable, `DV="Growth"`) on the Y axis. The scatter shows the reference set of 200 simulated skeletal samples (points, `samples=200`), regression fit used in the prediction, the D5+/D20+ ratio (1.28), and predicted growth rate (0.2% p.a.) for Nitra skeletal sample.

The `plot_diest` function can be applied to a single skeletal sample only. The function accepts data in the summary data format. By default, the plot shows a regression model to predict the growth rate from the D5+/D20+ ratio. The `plot_diest` function has the same arguments as the `diest` function with same default values. The arguments can be set in the same way as in the `diest` function.

SUPPLEMENTARY FUNCTIONS

The `demrat` package involves two supplementary functions that are primarily used by the main functions of the package, but can also be used independently in specific situations.

The `age_div` function proportionally divides age-at-death estimates of one or more individuals into 100 age categories (0–99 years). These parts can then be combined back into broader categories according to the user's needs. In age-at-death ratio studies, the broader categories include, for example, D5+, D20+ age groups. In life table analysis, individuals are grouped into, for example, 10-year categories such as 20–29, 30–39 years of age. The "division method" performed by the `age_div` function is based on the fact that the age-at-death estimate from the skeleton is imprecise (to be accurate) and is reported as an interval estimate (e.g., 6±2 years or 4.0–7.9 years). In palaeodemographic analysis, it is more reliable to not reduce the interval to a point estimate (6 years), but to use the entire interval.

The last function of `demrat` package presented here is `simdr_CD`. The function generates simulated skeletal samples based on mortality regimes given in Coale and Demeny model life tables (Coale *et al.* 1983). By default, the `simdr_CD` creates a set of 100 skeletal samples (argument `sss=TRUE`, `samples=100`) with 50 adult skeletons each (argument `D20_raw=50`) generated from Coale and Demeny West model life tables with life expectancy at birth between 18 and 25 years (arguments `e0_min=18` and `e0_max=25`) that are exposed to an annual growth between −3 and 3% (arguments `growth_min=-3` and `growth_max=3`). The function returns demographic characteristic of the populations from which each sample was drawn (life expectancy at birth, crude birth rate, total fertility rate, the proportion of individuals alive at 27.5 years) and the summary characteristics of age distribution of deaths

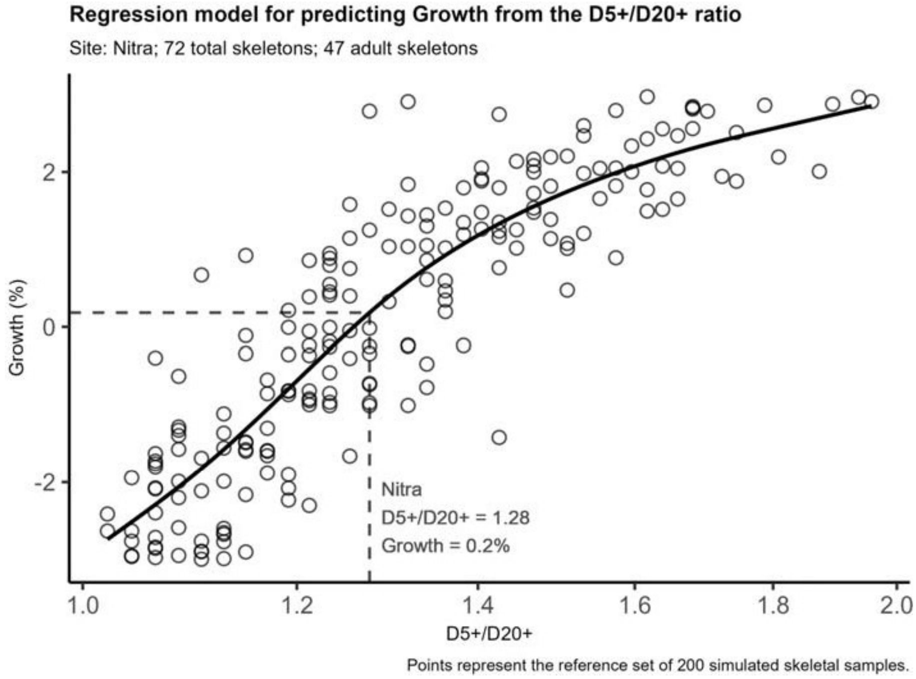


FIGURE 4: Result of the `plot_diest` function for Neolithic skeletal sample from Nitra (n=72 skeletons, Bocquet-Appel 2002, Table 1, pp. 640–641). The plot displays the D5+/D20+ ratio and annual growth rate (% p.a.) in the reference set of 200 simulated skeletal samples (points) with the same number of adult skeletons as in the Nitra sample, the regression fit (solid line) used in prediction, and the D5+/D20+ ratio (1.28) and predicted growth rate for population represented by the Nitra skeletal sample (0.2% p.a.). Simulated skeletal samples are drawn from populations with life expectancies at birth ranging from 18 to 25 years and subjected to the annual growth rates between −3 and 3%.

in the sample: number of individuals in several age-at-death categories (D0, D1+, D3+, D5+, D15+, D0-14, D5-14, D5-19, D20+, and n, the total number of skeletons) and six age-at-death ratios (D1+/D20+, D3+/D20+, D5+/D20+, the Juvenility index (D5-14/D20+), the P index (D5-19/D5+), and the D0-14/D ratio).

The `simdr_CD` function is used to generate reference skeletal samples for `diest` and `plot_diest` functions, but can be used outside the age-at-death ratio analysis to demonstrate the relation between demographic variables. *Figure 5*, for example, shows the relationship between total fertility and the growth rate at two different levels of life expectancy at birth. The scatterplot shows that in populations with low mortality (circles, life expectancy between 42 and 74 years), zero growth can be maintained by having 2-3 children per woman. In populations with high mortality (crosses, life expectancy between 18 and 25 years), the same level of growth is achieved by a woman giving birth to an average of 6-7 children.

Figure 5 further emphasizes that the estimation of growth or fertility rates from skeletal samples requires an algorithm that allows customizing the mortality level in the reference populations. Methods for calculating age-at-death ratios published to date have relied on reference populations with fixed mortality levels. However, when applied to skeletal samples from time periods with lower or higher mortality levels, these methods may provide unreliable results.

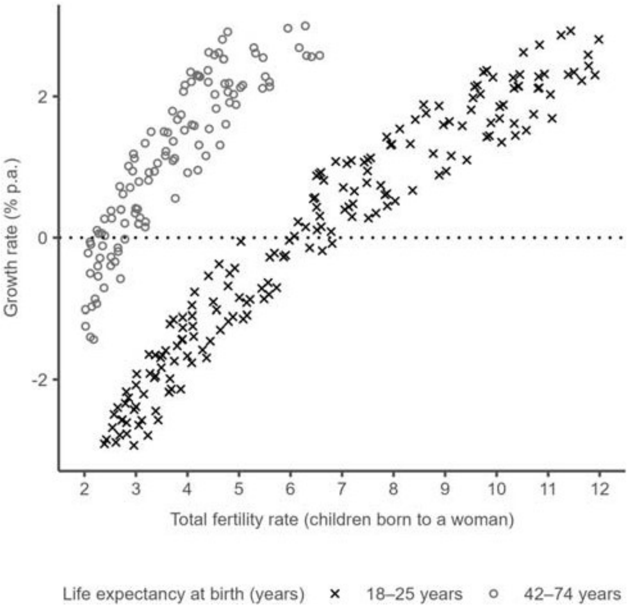


FIGURE 5: Relationship between annual growth rate (%) and total fertility rate (TFR, number of children born to a woman) based on the `simdr_CD` function from the `demrat` package. The plot shows 200 simulated populations (points) that experienced the growth between −3 and 3% p.a. and with life expectancies at birth between 18 and 25 years (crosses) and between 42 and 74 years (circles).

WEB APPLICATION FOR PREDICTING DEMOGRAPHIC RATES FOR NON-R USERS

Web (shiny) application enables non-R users to deploy demographic predictions with `demrat` package with a point-and-click interface. The shiny application is available on https://galetap.shinyapps.io/demrat/. Screenshot of the application is shown in *Figure 6*.

The shiny application consists of three tabs. The first tab (labeled Read me) provides help to the application. The second tab (labeled Load data) is used for uploading a data file. Data can be imported in MS Excel format (.xls or .xlsx) with the data placed on the first sheet. Data are accepted in both raw and summary data format (see Input data chapter). The user can also use sample data to explore how the application works. The third tab (labeled Prediction) makes the demographic predictions using `demrat` package. The application enables the setting of several input parameters that correspond to arguments of `diest` and `plot_diest` functions. The user can set the number of reference skeletal samples that are generated, the life expectancies at birth of reference populations, and their growth rate range. The tab also enables the selection of demographic variables to be predicted and the selection of age at death ratios to be used in the prediction. The Prediction tab displays demographic predictions in a tabular format that can be exported to an MS Excel file. If only one skeletal sample is selected, the Prediction tab also displays a plot with the prediction model.

CASE STUDY: ESTIMATING FERTILITY LEVEL IN NEOLITHIC EUROPE

In 2002, Bocquet-Appel published an influential study introducing the model of demographic changes during the Neolithic transition. Based on evidence from 68 skeletal samples from Europe and North Africa, he argued that the population of early farmers grew rapidly at the onset of the Neolithic period, which demographically distinguished them from the stationary population of the last Mesolithic foragers.

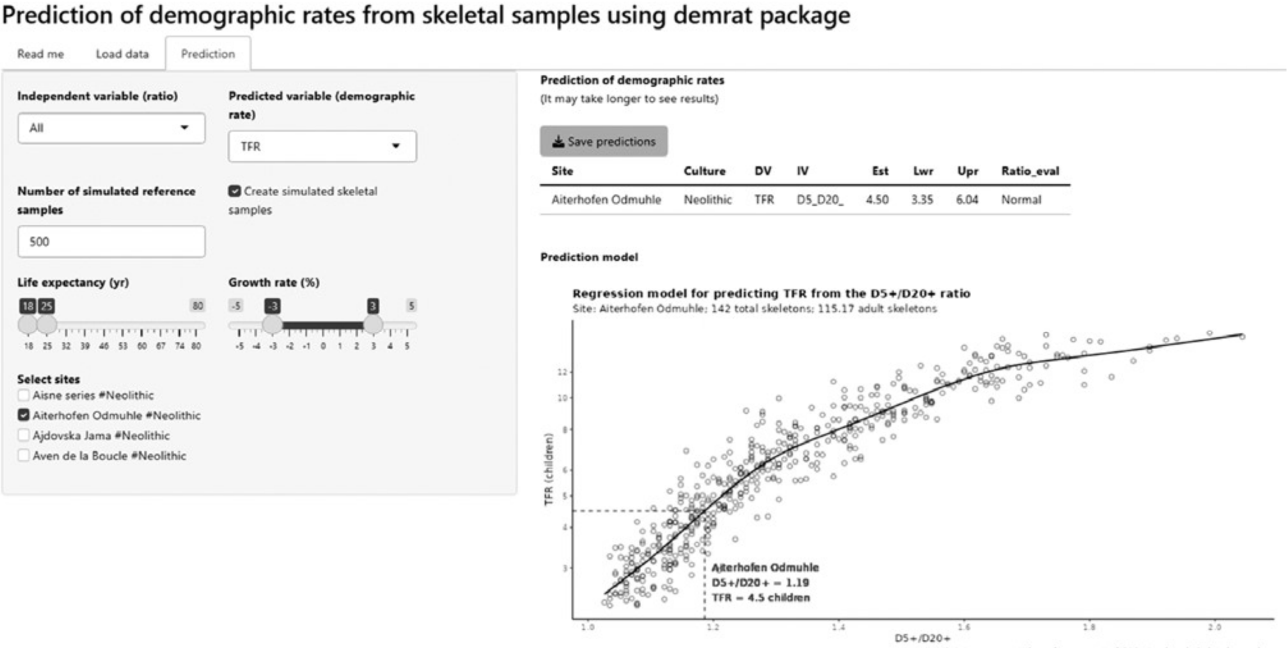
Appendix 1 provides an annotated R code that reanalyzes the Bocquet-Appel's (2002) dataset. The reanalysis takes advantage of the new estimating algorithm and uses `demrat` package to predict total fertility rates from skeletal samples. The R code produces a chronological profile of the total fertility rate (*Figure 7*), which describes changes in fertility levels at the beginning of the Neolithic period in Europe. *Figure 7* suggests that early Neolithic mothers gave births to around seven children on average.

The R code covers all stages of the demographic estimation. It begins with estimating age-at-death for individual skeletons from 68 samples (`BAraw` sample data in raw format) and then calculates age-at-death ratios using the `dr` function. Finally, a prediction model is built for each skeletal sample using the `diest` function. These models are then used to estimate total fertility rates. The remaining R code generates the chronological profile plot.

Note that R code in *Appendix 1* and *Figure 7* provides an illustrative example of the use of the `demrat` package on a large number of skeletal samples and should not be used as evidence for the model of Neolithic demographic transition. Bocquet-Appel's (2002) dataset has been criticized. Some skeletal samples are very small; consisting of fewer than 30 skeletons. In small samples, distribution of deaths may be highly unreliable due to stochastic error and other biases.

CONCLUSION

The `demrat` package provides a set of functions in the R programming language for predicting growth, birth, and total fertility rates from age-at-death distributions of skeletal samples. The package utilizes a recently published methodology (Galeta, Pankowská 2023a) that overcomes two limitations of previous approaches: effect of mortality and stochastic variation on prediction. The prediction algorithm generates of a large set of reference skeletal samples and the `demrat` package simplifies the process. The functions of the package provide a complete workflow from an age-at-death distribution in real skeletal samples to the predictions of demographic rates. Demographic prediction using the `demrat` package is also available for non-R users, who can upload their skeletal data into the free on-line application, modify input parameters of the prediction model and get results in a user-friendly graphical interface.



Prediction of demographic rates from skeletal samples using demrat package

Independent variable (ratio): All

Predicted variable (demographic rate): TFR

Number of simulated reference samples: 500

☑ Create simulated skeletal samples

Life expectancy (yr): 18 – 25

Growth rate (%): -3 – 3

Select sites
☐ Aisne series #Neolithic
☑ Aiterhofen Odmuhle #Neolithic
☐ Ajdovska Jama #Neolithic
☐ Aven de la Boucle #Neolithic

Prediction of demographic rates
(It may take longer to see results)

Site	Culture	DV	IV	Est	Lwr	Upr	Ratio_eval
Aiterhofen Odmuhle	Neolithic	TFR	D5_D20_	4.50	3.35	6.04	Normal

Prediction model

Regression model for predicting TFR from the D5+/D20+ ratio
Site: Aiterhofen Odmuhle; 142 total skeletons; 115.17 adult skeletons

Points represent the reference set of 500 simulated skeletal samples.

FIGURE 6: Screenshot of a shiny web application (https://galetap.shinyapps.io/demrat/) that allows non-R users to predict growth, crude birth, and total fertility rate from skeletal samples using the `demrat` package in a graphical web interface.

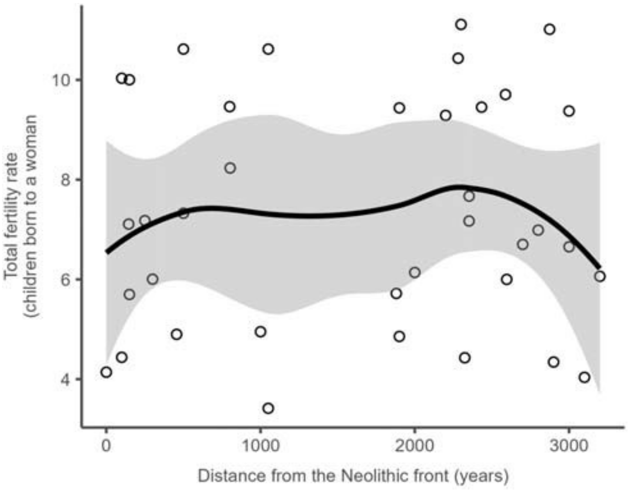


FIGURE 7: The total fertility rate profile in the Neolithic of Europe. The plot is based on 37 skeletal samples from the Bocquet-Appel Neolithic dataset (2002, Table 1, pp. 640–641) with more than 50 total skeletons and reliable value of the D5+/D20+ ratio. Total fertility estimates are calculated using the `diest` function of the `demrat` package assuming life expectancy at birth between 18 and 25 years. Note that the profile is made to demonstrate the `demrat` package and the profile may be unreliable (see text).

ACKNOWLEDGEMENT

This research was supported by a grant from the Czech Science Foundation [grant number GA19-17810S].

REFERENCES

BOCQUET-APPEL J.-P., 2002: Paleoanthropological traces of a Neolithic demographic transition. *Curr. Anthropol.* 43: 637–650. https://doi.org/10.1086/342429

BOCQUET-APPEL J.-P., 2008: Explaining the Neolithic demographic transition. In: J.-P. Bocquet-Appel, O. Bar-Yosef (Eds.): *The Neolithic demographic transition and its consequences*. Pp. 35–55. Springer, New York.

BOCQUET-APPEL J.-P., 2011: When the world's population took off: the springboard of the Neolithic Demographic Transition. *Science* 333, 6042: 560–561. https://doi.org/10.1126/science.1208880

BOCQUET-APPEL J.-P., MASSET C., 1982: Farewell to paleodemography. *J. Hum. Evol.* 11: 321–333. https://doi.org/10.1016/S0047-2484(82)80023-7

BOLDSEN J. L., MILNER G. R., OUSLEY S. D., 2022: Paleodemography: from archaeology and skeletal age estimation to life in the past. *Yearb. Biol. Anthropol.* 178, S74: 151–180. https://doi.org/10.1002/ajpa.24462.

BUIKSTRA J. E., KONIGSBERG L. W., BULLINGTON J., 1986: Fertility and the development of agriculture in the prehistoric Midwest. *Am. Antiq.* 51: 528–546. https://doi.org/10.2307/281750

COALE A. J., DEMENY P. G., 1966: *Regional model life tables and stable populations*. Princeton University Press, Princeton.

COALE A. J., DEMENY P. G., VAUGHAN B., 1983: *Regional model life tables and stable populations*. Academic Press, New York.

GAGE T. B., DEWITTE S., 2009: What do we know about the agricultural demographic transition? *Curr. Anthropol.* 50, 5: 649–655. https://doi.org/10.1086/605017

GALETA P., PANKOWSKÁ A., 2023a: A new method for estimating growth and fertility rates using age-at-death ratios in small skeletal samples: the effect of mortality and stochastic variation. *PLoS One* 18, 6: e0286580. https://doi.org/10.1371/journal.pone.0286580

GALETA P., PANKOWSKÁ A., 2023b: Population dynamics in the Middle Ages in Central Europe: Reconstruction based on age-at-death distributions of skeletal samples. *J. Archaeol. Sci.* 156: 105816. https://doi.org/10.1016/j.jas.2023.105816

GUY H., MASSET C., BAUD C.-A., 1997: Infant taphonomy. *Int. J. Osteoarchaeol.* 7: 221–229. https://doi.org/10.1002/(SICI)1099-1212(199705)7:3<221::AID-OA338>3.0.CO;2-Z

JOHANSSON S. R., HOROWITZ S., 1986: Estimating mortality in skeletal populatios: influence of the growth rate on the interpretation of levels and trends during transition to agriculture. *Am. J. Phys. Anthropol.* 71: 233–250. https://doi.org/10.1002/ajpa.1330710211

MANIFOLD B. M., 2013: Differential preservation of children's bones and teeth recovered from early medieval cemeteries: possible influences for the forensic recovery of non-adult skeletal remains. *Anthropol. Rev.* 76, 1: 23–49. https://doi.org/10.2478/anre-2013-0007

MCFADDEN C., 2021: The past, present and future of skeletal analysis in palaeodemography. *Proc. R. Soc. Lond. B Biol. Sci.* 376, 1816: 20190709. https://doi.org/10.1098/rstb.2019.0709

MCFADDEN C., OXENHAM M. F., 2018a: The D0-14/D ratio: a new paleodemographic index and equation for estimating total fertility rates. *Am. J. Phys. Anthropol.* 165, 3: 471–479. https://doi.org/10.1002/ajpa.23365

MCFADDEN C., OXENHAM M. F., 2018b: Rate of natural population increase as a paleodemographic measure of growth. *J. Archaeol. Sci.: Rep.* 19: 352–356. https://doi.org/10.1016/j.jasrep.2018.03.012

MILNER G. R., BOLDSEN J. L., 2023: Population trends and the transition to agriculture: global processes as seen from North America. *Proc. Natl. Acad. Sci.* 120, 4: e2209478119. https://doi.org/10.1073/pnas.2209478119

MILNER G. R., HUMPF D. A., HARPENDING H. C., 1989: Pattern matching of age-at-death distributions in paleodemographic analysis. *Am. J. Phys. Anthropol.* 80: 49–58. https://doi.org/10.1002/ajpa.1330800107

PAINE R. R., HARPENDING H. C., 1998: Effect of sample bias on paleodemographic fertility estimates. *Am. J. Phys. Anthropol.* 105: 231–240. https://doi.org/10.1002/(SICI)1096-8644(199802)105:2<231::AID-AJPA9>3.0.CO;2-X

POKINES J. T., DE LA PAZ J. S., 2016: Recovery rates of human fetal skeletal remains using varying mesh sizes. *J. Forensic Sci.* 61: S184–S189. https://doi.org/10.1111/1556-4029.12922

SATTENSPIEL L., HARPENDING H., 1983: Stable population and skeletal age. *Am. Antiq.* 48: 489–498. https://doi.org/10.2307/280557

TAYLOR B. R., OXENHAM M., MCFADDEN C., 2023: Estimating fertility using adults: a method for under-enumerated pre-adult skeletal samples. *Am. J. Biol. Anthropol.* 181, 2: 262–270. https://doi.org/10.1002/ajpa.24739

WICKHAM H., GROLEMUND G., 2017: *R for data science: import, tidy, transform, visualize, and model data*. O'Reilly Media, Sebastopol.

Patrik Galeta[1,*]
[1] Department of Anthropology,
University of West Bohemia,
Sedláčkova 15, 306 14 Pilsen,
Czech Republic
E-mail: galeta@ff.zcu.cz

* Corresponding author

Appendix 1. R code for estimating total fertility rate (TFR) in Neolithic using the `demrat` package. Input age-at-death estimations from skeletons are based on Bocquet-Appel's dataset (2002, Table 1, pp. 640–641).

```
# Load R packages
# Install packages with install.packages() first if they are not installed
library(demrat)
library(tidyverse)

# Estimating TFR using demrat package
est <-
  # Opening Bocquet-Appel (2002) dataset in raw format
  # (age-at-death estimation for each individual)
  BAraw %>%
  # Converting raw to summary data format
  # (age-at-death ratios for each skeletal sample)
  dr() %>%
  # Adding dt variable (distance from the Neolithic front) to the dataset
  left_join(BA %>% select(Site, dt), by = "Site") %>%
  # Estimating TFR based on the D5+/D20+ ratio, adding dt and n to the results
  diest(samples = 500,
        IV = "D5_D20_", DV = "TFR",
        e0_min = 18, e0_max = 25,
        extra_var = c("dt", "n"))

# Creating chronological profile of TFR during the Neolithic transition
est %>%
  # Filtering Neolithic samples with "normal" value of age-at-death ratio
  # and with more than 50 skeletons
  filter(Ratio_eval=="Normal",
         n>=50,
         dt>=0) %>%
  # Plot
  ggplot(aes(x = dt, y = Est)) +
  # Points (samples)
  geom_point() +
  # Adding loess fit
  geom_smooth(col = "black") +
  # Changing axis titles
  labs(x = "Distance from the Neolithic front (years)",
       y = "Total fertility rate\n(children born to a woman") +
  # Adjusting axis breaks
  scale_x_continuous(breaks = seq(-3000, 3000, 1000)) +
  scale_y_continuous(breaks = seq(2, 12, 2))
```