Contents

[Demographic data 1](#_Toc127802550)

[Census data 1](#_Toc127802551)

[Household positions 1](#_Toc127802552)

[Parent-child match 4](#_Toc127802553)

[Birth trajectory 4](#_Toc127802554)

[Partner match 5](#_Toc127802555)

[Household compositions: NFRA, other and collective household 6](#_Toc127802556)

[Household transition rates 7](#_Toc127802557)

[Birth-related household transitions 9](#_Toc127802558)

[Mortality rates 9](#_Toc127802559)

[Migration rates 10](#_Toc127802560)

[Population sample 11](#_Toc127802561)

[Microsimulation 14](#_Toc127802562)

[Fertility 14](#_Toc127802563)

[Household transitions 19](#_Toc127802564)

[Migration 23](#_Toc127802565)

[Mortality 25](#_Toc127802566)

[Event-log 25](#_Toc127802567)

[Code 26](#_Toc127802568)

Demographic data

Census data

We use Belgian census data from January 1st 2011 and January 1st 2012. The following individual variables are extracted from the census:

* ID
* ID of household of residence (2011 and 2012)
* Date of birth
* Sex
* LIPRO household position (2011 and 2012)
* ID of parents (ID\_DEMO\_PAR\_1\_C, ID\_DEMO\_PAR\_2\_C)

Household positions

All individuals in the population have a LIPRO household position for 2011 and 2012, which describes the relation an individual has to other household members and/or the type of household [1]. The LIPRO typology contains the following categories:

* CMAR: Child in family with married parents
* CUMR: Child in family with cohabiting parents
* C1PA: Child in one-parent family
* SING: Single (one-person household)
* MAR0: Married, living with spouse but without children
* MAR+: Married, living with spouse and one or more children
* UNM0: Cohabiting, no children present
* UNM+: Cohabiting with one or more children
* H1PA: Head of one-parent family
* NFRA: Non-family related adult (adult living with MAR0/MAR+/UNMO/UNM+/H1PA)
* OTHR: Other (e.g. multiple single adults living together)
* Collective: Member of collective household

The LIPRO household positionsonly consider one family nucleus per household and all other household members are assigned LIPRO positions in relation to that family nucleus. In case of a multigenerational household with grandparents, parents and children, the grandparents would typically have the LIPRO position *MAR+/UNM+*, their child would be *CMAR/CUMR* and the partner of their child and their grandchildren would be *NFRA.* In order to allow for more than one family nucleus and at the same time disregard the distinction between marriage and cohabitation, the LIPRO household positionsare modified in the new variable *household position.* Each category of this variable is described below as well as the included LIPRO positions. Each category name is followed by an abbreviation, which will be used onwards. For some categories, LIPRO positions are only included in a given household position if certain conditions involving other variables are fulfilled.

* Child *(child)*
  + Individual living in parental household without own children or partner
  + LIPRO positions: CMAR, CUNM, C1PA
  + LIPRO positions with conditions:
    - NFRA/OTHR if individuals is younger than 16
* Union without child (*union*)
  + Individual living together with partner and without children
  + LIPRO positions: MAR0, UNM0
  + LIPRO positions with conditions:
    - CMAR/CUNM/C1PA/NFRA if individual is in a union in 2012 (2011) with a partner that also was a household member in 2011 (2012)
    - CMAR/CUNM/C1PA/NFRA/OTHR/H1PA if two individuals living in the same household are parents to the same child (no longer in household)
* Union with child (*union+)*
  + Individual living together with partner and child
  + LIPRO positions: MAR+, UNM+
  + LIPRO positions with conditions:
    - CMAR/CUNM/C1PA/NFRA if individual is in union in 2012 (2011) with a partner that also was a household member in 2011 (2012) and a (step-)child is present in the household (further description in section *Partner match*)
    - CMAR/CUNM/C1PA/NFRA/OTHR/H1PA if two individuals living in the same household are parents to the same child and a (step-)child is present in the household (further description in section *Partner match*)
* Single-person household (*single)*
  + Individual living in a one-person household
  + LIPRO position: SING
  + LIPRO positions with conditions:
    - All LIPRO positions if household of size 1
* Non-family related adult (*NFRA*)
  + Individual living without own family nucleus but living in same household as unrelated family nucleus
  + LIPRO position: NFRA
* Other
  + Individual living together with other unrelated individuals
  + LIPRO position: OTHR
* Collective household (*collective)*
  + Individual living in collective household (prison, special care facility, nursing homes, student accommodation,…)
  + LIPRO position: Collective
  + LIPRO positions with conditions:
    - All LIPRO positions if household members have category ‘collective’
* Single parent (*single+)*
  + Individual living together with their child but without a partner in the household
  + LIPRO position: H1PA
  + LIPRO positions with conditions:
    - NFRA/OTHR/CMAR/CUNM/C1PA/MAR+/UNM+ if individual without a partner or with a missing partner match (further description in section *Partner match*) lives in the same household as their child
  + Single parents living in their own parental household are referred to as *single+\** in cases where the distinction is necessary
* Union in multigenerational household (*multi\_U*)
  + Individual living together with partner, child and family nucleus of child (e.g. grandchild and/or partner of child)
* Single in multigenerational household (*multi\_S)*
  + Individual without partner living with child and family nucleus of child (e.g. grandchild and/or partner of child)

In Table S1, the categories in the new variable *household position* are broken down by *LIPRO household position* for the census population. Differences between the two variables are especially seen for the LIPRO positions *NFRA* and *OTHR.*

*Table S1: Household position vs. LIPRO household position. Source: Belgian census 2011.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **LIPRO household position** | | | | | | | | | | | | | |
| **Household position** |  | SING | MAR0 | MAR1 | CMAR | UNM0 | UNM1 | CUMR | H1PA | C1PA | NFRA | OTHR | Collective |
| Child | 28 | 2 | 0 | 2065579 | 406 | 83 | 460206 | 1 | 693979 | 62715 | 18581 | 656 |
| Collective | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 0 | 39 | 127021 |
| Multi\_S | 0 | 0 | 2 | 32 | 6 | 32 | 22 | 15128 | 72 | 22396 | 130 | 3 |
| Multi\_U | 0 | 112 | 41438 | 23 | 26 | 4638 | 20 | 17 | 52 | 7363 | 48 | 0 |
| NFRA | 0 | 0 | 0 | 18 | 0 | 0 | 13 | 0 | 8 | 69709 | 0 | 0 |
| Other | 0 | 6 | 0 | 0 | 120 | 15 | 0 | 0 | 0 | 0 | 125519 | 0 |
| Single | 1589161 | 0 | 0 | 0 | 35 | 24 | 3 | 4 | 3 | 42 | 68 | 1510 |
| Single+ | 0 | 2 | 6 | 9916 | 405 | 199 | 1888 | 440286 | 8534 | 4434 | 12442 | 3 |
| Union | 0 | 1915841 | 579 | 2472 | 488299 | 287 | 55 | 2 | 1462 | 5424 | 366 | 0 |
| Union+ | 0 | 6571 | 2161670 | 9338 | 1069 | 544300 | 593 | 939 | 5464 | 17169 | 4127 | 0 |

Parent-child match

The variables *ID\_DEMO\_PAR\_1\_C* and *ID\_DEMO\_PAR\_2\_C* are used to create the variables *ID mother* and *ID father* by looking up the sex of the individual*.* For reasons of simplicity, only one parent in same-sex couples is registered as parent. For individuals with household position *child*, the parent ID is considered to be incorrect and set to NA if there are less than 12 years between parent and child. Moreover, the parent IDs are modified if both parents do not live in the same household as their child (e.g. foster care) or both IDs are missing. In those cases, the parental role is assigned to a household member who is at least 14 years older than the child. In case multiple household members fulfil that requirement, the one with an age difference to the child closest to 29 years (mean age of Belgian females at first birth in 2016[[1]](#footnote-1)) is chosen. The parent IDs were adjusted for 63,075 individuals. For a small number of children (1,438), no suitable parent was found in the household and the individuals were instead assigned to randomly chosen households of females in a union of at least 30 years of age.

Birth trajectory

Since the variables *ID\_DEMO\_PAR\_1\_C* and *ID\_DEMO\_PAR\_2\_C* contain the IDs of an individual’s parents, they can be used to compute birth trajectories. Birth trajectories are only computed for the female population. A frequency table containing the IDs of mothers is used to compute the number of births, which is then added to the information of the mother. Moreover, the children’s birth dates (sorted by date) are added to the information of their mother. The birth date of the youngest child is used to create the variable *index birth.* There will be some discrepancies between the computed and the actual birth trajectories because some females have given birth to individuals that are no longer in the population (stillbirths, emigrants). Moreover, the variables *ID\_DEMO\_PAR\_1\_C* and *ID\_DEMO\_PAR\_2\_C* also contain missing values.

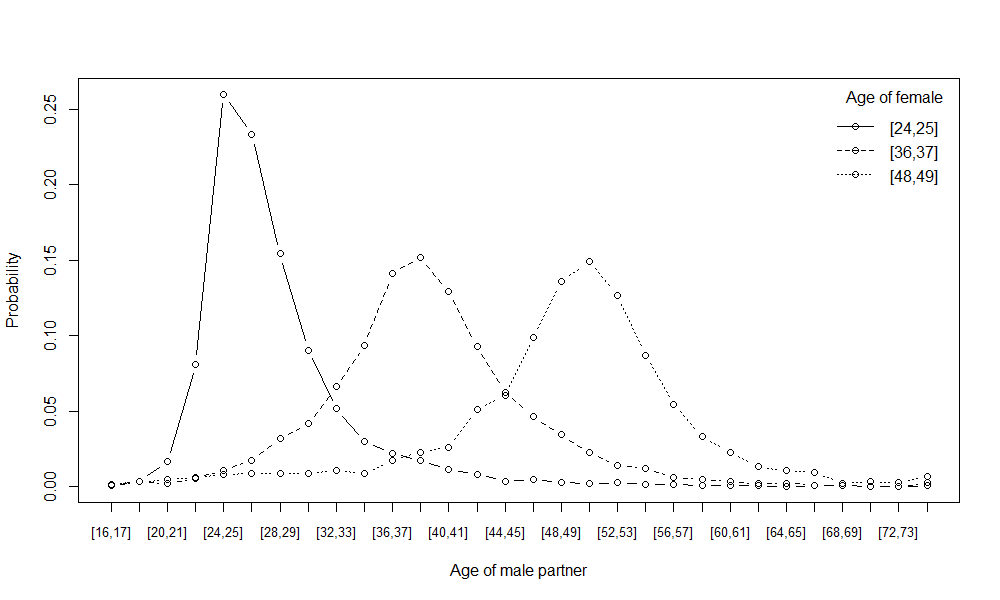
Partner match

The original LIPRO household positions only allow for one union per household. For that reason, individuals from the same household and with the LIPRO household position *MAR0/MAR1/UNM0/UNM1* are assumed to be in a union. However, errors in the LIPRO household positions cause a smaller number of unions between parent and child (521), which are changed to more appropriate categories in the variable *household position*. Moreover, no partner is initially found for 1,700 individuals because no one else in the household has LIPRO household position *MAR0/MAR1/UNM0/UNM1*. 1,396 of these individuals are matched with a household member with a missing value as LIPRO household position, if they are of opposite sex and the age difference is less than 15 years. In case of multiple matches, the individual with the most similar age is chosen as partner. The remaining individuals were changed to *single parent* (157) and *other* (147).

Unions are also detected through the parental IDs. Individuals who are parents to the same child and living in the same household are assumed to be a couple and the original LIPRO household position is changed to the category *union* in the new variable household position. Finally, individuals with LIPRO positions *CMAR/CUNM/C1PA/NFRA* in 2011 and *MAR0/MAR1/UNM0/UNM1* in 2012 are considered to already be in a union in 2011 and are matched with a household member for whom the same applies if they live in the same household in 2011 as well as 2012.

For reasons of simplicity, same-sex unions are changed to opposite-sex unions and the oldest individual is assumed to be the male. This implies that we assume demographic events for individuals in a same-sex union to be similar to those of opposite-sex unions. It should be noted that unions formed by individuals living in different households are not detected.

Based on the union formations taking place in 2011, the age distribution of the male partners is computed for each age group of the female partners as shown in Figure S14. The partners tend to be of similar age or the male partner being slightly older than the female partner. The distributions at the youngest and oldest ages are less smooth due to the small number of unions at those ages. The age distributions will be used as input in the microsimulation.



*Figure S14: Age-specific probability of male entering union conditional on the age of female partner.*

Household compositions: NFRA, other and collective household

Household size distributions are computed for individuals with household position *NFRA* and *other* as shown in Figure S15 and Figure S16*.* These are used in the microsimulation to allocate individuals to new households.

|  |  |
| --- | --- |
| *Figure S15: Household size distribution for individuals with household position 'NFRA'* | *Figure S16: Household size distribution for individuals with household position 'other'* |

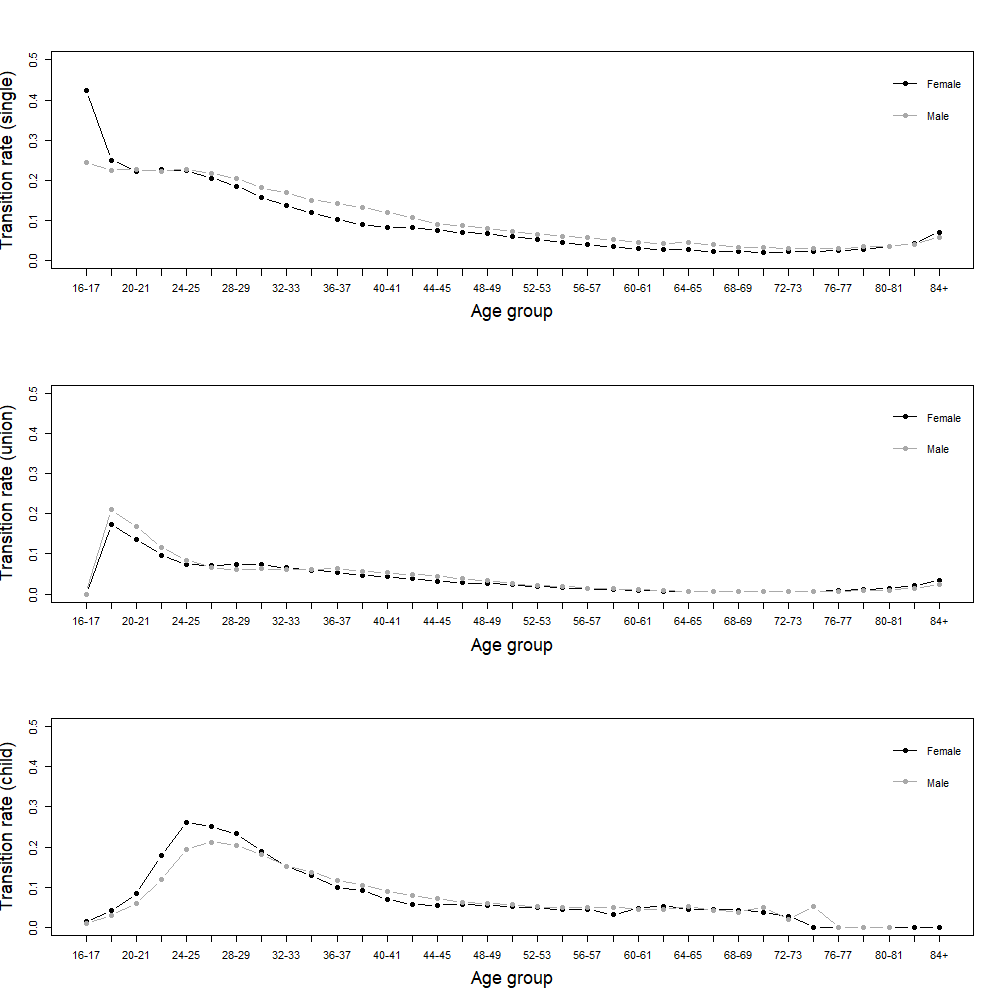
For individuals in collective households, the household size distribution is computed by age group and shown in Figure S17. This distribution is used to allocate individuals to collective households in the microsimulation.

|  |
| --- |
| *Figure S17: Household size distribution by age group for individuals in 'collective' households* |

Household transition rates

Household transitions are based on the individual household positions observed on January 1st 2011 and January 1st 2012. Some individuals are not present in the population in 2012 and the household position is for that reason missing (*NA*). We also set ‘indirect’ transitions resulting from other demographic events (fertility, mortality, migration) to *NA* because the household transitions related to these events are considered when the process is modelled. This includes changes in the household position *child* due to the death or emigration of the parents. Union dissolutions resulting from the death or emigration of the partner are also considered to be indirect transitions, as well as transitions to and from *single parent* resulting from giving birth or the last child leaving the parental household. The individuals with missing household positions are assumed to have left the population and risk-set in the middle of 2011. The rate at which individuals leave the household position *i* is computed as follows:

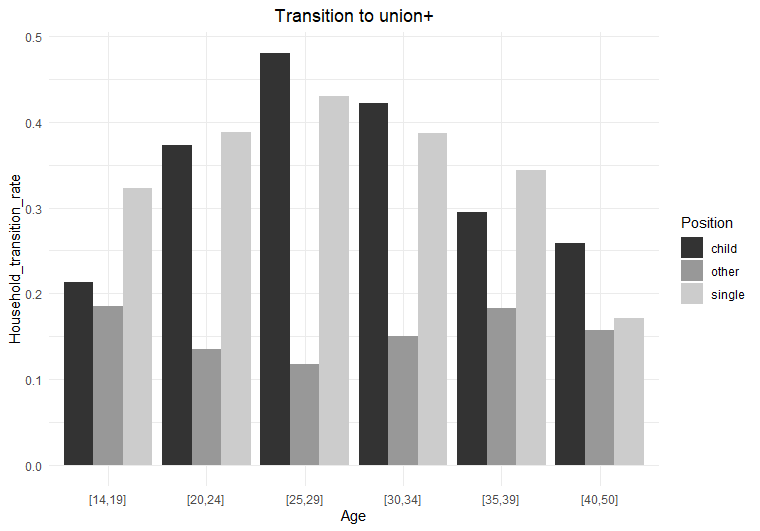
where . The set A contains all household positions different from *i*. The parameters *a* and *s* indicate age group in 2011 and sex, respectively. is the number of individuals with household position *i* in 2011 (risk set), is the number of individuals from the risk set with household positions different from i in 2012 and is the number of individuals from the risk set with a missing household position in 2012 (indirect transition). Transition rates are computed separately for the females who gave birth in 2011.



*Figure S18: Household transition rates by age and sex for positions single, union and child.*

Birth-related household transitions

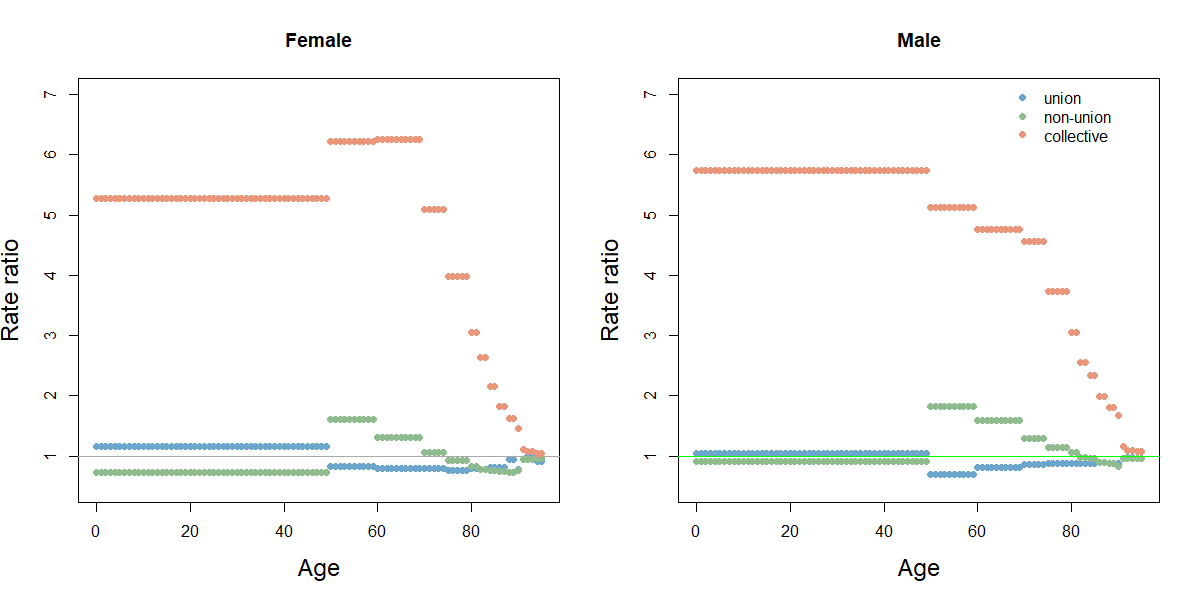
The birth-related household transitions are based on the household positions of the female population with births observed in 2011. The household positions on January 1st 2011 and 2012 are compared and transition rates are computed for the age groups <20, 20-24, 25-29, 30-34, 35-39, >=40. The rates of the transitions from *child, other* and *single* to *union+* are shown in Figure S19. It should be noted that household transitions may already have taken place during the pregnancy and potentially in 2010 for females giving birth in the first three quarters of 2011. Transitions taking place in 2010 are unknown because we are only using the household positions on January 1st 2011 and 2012. Thus, the birth-related household transitions are probably underestimated to some degree.



*Figure S19: Birth-related household transition from child, other, single to union+.*

Mortality rates

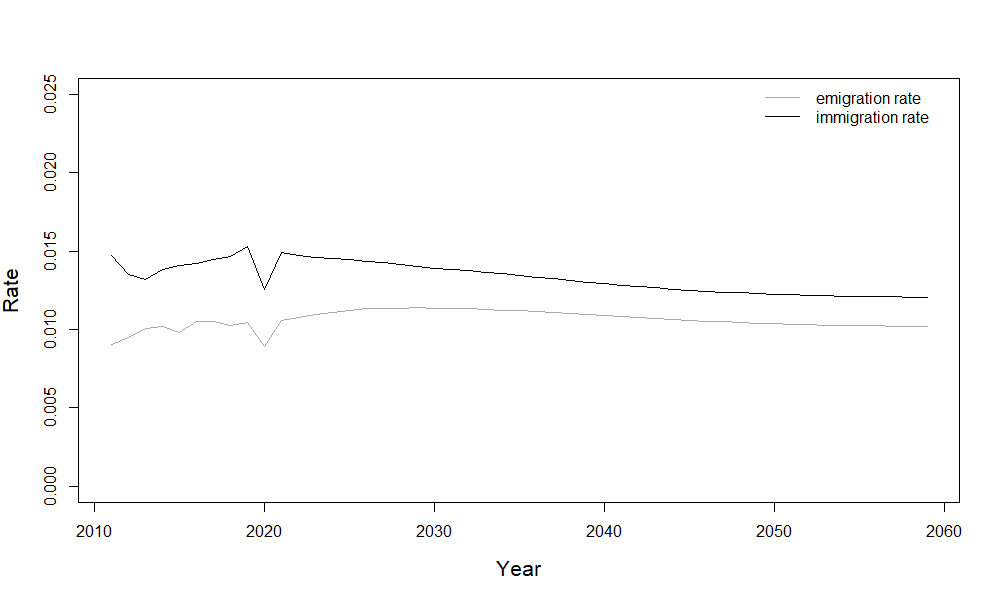
By breaking the central rate of mortality (mx) observed in the census of 2011 further down by the household positions union(+), non-union (single/child/NFRA/other) and collective, rate ratios can be computed by dividing the age-, sex- and household-specific mortality rates by the overall mortality rate for the given age and sex. The rate ratios are computed for the age groups 0-49, 50-59, 60-69, 70-74, 75-79, 80-84, 85-90 and 91+ as shown in Figure S20. We multiply the rate ratios from 2011 by the observed and projected age- and sex-specific mortality rates from Statistics Belgium (StatBel) from 2011 to 2050 in order to simulate mortality over time.



*Figure S20: Mortality rate ratios by sex, household position and age group.*

Migration rates

Observed (2011-2020) and projected (2021-2050) immigration and emigration rates are retrieved from StatBel (see Figure S21) .



*Figure S21: Observed (2011-2021) and projected (2022-2050) migration rates. Source: StatBel.*

Distributions of age, sex and household positions of immigrants as well as emigrants are computed based on the observed immigration in 2010-2011 and emigration in 2011-2012 (See Figure S22-Figure *S25*).

|  |  |
| --- | --- |
| *Figure S22: Age-sex distribution of immigrants 2011* | *Figure S23: Household positions of immigrants 2011. (NA: missing value)* |

|  |  |
| --- | --- |
| *Figure S24: Age-sex distribution of emigrants 2011* | *Figure S25: Household positions of emigrants 2011* |

Population sample

A sample of 500,000 households was drawn from the census based on the household size distribution. Households with more than 230 household members were excluded, as this is a limitation in the infectious disease model. All members of the sampled households, 1,118,021 individuals (ca. 10% of census population), make up the sample population. The household size distribution, household position distribution and age distribution overall and by household size are shown for the sample and census in Figure S26 to Figure S35.

|  |  |
| --- | --- |
| *Figure S26: Household size distribution.* | *Figure S27: Household position distribution (collec.:collective household, multi\_S: single in multigenerational household, multi\_U: oldest union in multi-generational household, NFRA: non-family related adult, single+: single parent, union+: union with child).* |

|  |  |
| --- | --- |
| *Figure S28: Age distribution* | *Figure S29: Age distribution household size 1* |

|  |  |
| --- | --- |
| *Figure S30: Age distribution household size 2* | *Figure S31: Age distribution household size 3* |

|  |  |
| --- | --- |
| *Figure S32: Age distribution household size 4* | *Figure S33: Age distribution household size 5* |

|  |  |
| --- | --- |
| *Figure S34: Age distribution household size 6* | *Figure S35: Age distribution household size 7+* |

The parent(s) of sampled individuals may not be included in the sample population if child and parent(s) live in different households. In those cases, an individual is assigned ‘replacement’ parents. Females in the sample population of the same age and household position as the actual mother in the full census make up a pool of candidate ‘replacement’ mothers and one is randomly drawn and assigned to the individual. The partner of the assigned mother is assigned as ‘replacement’ father. For individuals with only a father in the full census, the same procedure is carried out but based on the characteristics of the father. The birth trajectories of the ‘replacement’ parents are not updated.

|  |
| --- |
| Figure S36: Model sequence |

Microsimulation

We simulate the Belgian population from 2011 to 2050 using microsimulation. The initial population is based on the sample of the census from January 1st 2011 (see section *Population sample*). The population is updated in discrete time steps by simulating the demographic processes sequentially as visualised in Figure S36. We are using yearly time steps, but this can be adjusted to smaller intervals. Each event is assigned a date within the time step at random. In case an individual experiences multiple events within one time step, the dates are adjusted so that the order is respected (e.g. household transition before death). All demographic events are tracked in an event-log.

Fertility

The fertility data contains the birth trajectories of the female population up to December 31st 2011 (described in section *Birth trajectory*). Only females of age 14-50 are included. Based on the birth trajectories, the age of females at the index birth (most recent birth) is computed as well as the duration between the index birth and June 30th 2011 (if applicable). For females giving birth in 2011, the duration between the birth taking place in 2011 and the index birth is used. The household positions of the females on January 1st 2011 are also included in the data.

Generalised additive models (GAMs) are fitted to the fertility data [2,3]. In a GAM, like in a generalised linear model, a link function describes the relationship between the linear predictor and the expected value of the response, but in a GAM, the linear predictor involves a sum of smooth functions of covariates. A smooth function is the sum of a number of basis functions, , weighted by the corresponding regression coefficient, , and can be defined as follows [4]:

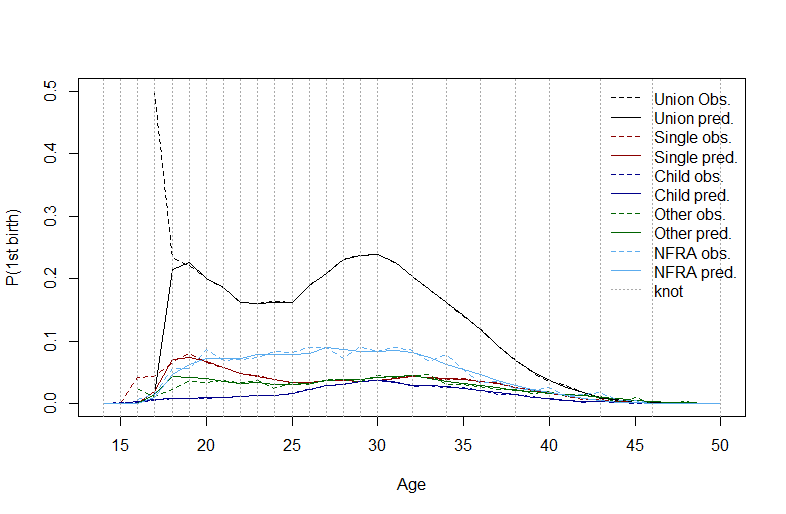
We use cubic regression splines with a penalty modified to shrink towards zero at high enough smoothing parameters (specification in R package mgcv: bs=’cs’), and vary the number and location of knots, which defines the number and interval of the basis functions, in order to obtain an optimal fit. The GAMs are fitted separately for births of order 1, 2, 3 and 4+. The dependent variable is binary (1=birth, 0=no birth) and we use a logit link function.

*Fertility model: First order births*

The following model was fitted to the data of females at risk of a first birth in 2011:

The variable is the age of the female in the middle of the time step and dummy variables *(child, union, single, NFRA, other)* are used to represent the household position at the beginning of the time step. The terms represent the smooth functions.

In Figure S37, the predicted and observed probabilities of a first birth are shown for each household position. The location of the knots are shown as vertical dotted lines. Overall, the predicted and observed probabilities are very similar except at the very young ages where births are less common.



*Figure S37: Probability of 1st birth by household position (obs.: observed, pred.: predicted).*

*Fertility model: Second and higher order births*

For second and higher order births, models were not only fitted separately for each parity but also for the household positions *union (with child)* and *single (parent)*. A small number of females at risk of second and higher order births had different household positions (e.g. *child*) and were assigned to *single (parent).* The following model is fitted separately for each parity and household-position:

where is the birth parity and is a dummy variable which is 1 for parity 3 and higher and 0 for all other parities. The variable is the age of the female at index birth and is the duration since index birth. These two variables are fitted with a tensor product smooth, , which can be formulated as follows:

where and are basis functions, and are the corresponding basis dimensions and is a vector of coefficients. The predicted and observed probabilities for second and third births by household position are shown in Figure S38 to Figure S49 for different ages at index birth. The knot locations are indicated by vertical red dotted lines.

|  |  |
| --- | --- |
| *Figure S38: Probability of 2nd birth for females in union with age at index birth of 20 years.* | *Figure S39: Probability of 2nd birth for females in union with age at index birth of 25 years.* |
| *Figure S40: Probability of 2nd birth for females in union with age at index birth of 30 years.* | *Figure S41: Probability of 2nd birth for females not in union with age at index birth of 20 years.* |
| *Figure 42: Probability of 2nd birth for females not in union with age at index birth of 25 years.* | *Figure 43: Probability of 2nd birth for females not in union with age at index birth of 30 years.* |
| *Figure S44: Probability of 3rd birth for females in union with age at index birth of 27 years.* | *Figure S45: Probability of 3rd birth for females in union with age at index birth of 30 years.* |
| *Figure S46: Probability of 3rd birth for females in union with age at index birth of 33 years.* | *Figure S47: Probability of 3rd birth for females not in union with age at index birth of 27 years.* |
| *Figure S48: Probability of 3rd birth for females not in union with age at index birth of 30 years.* | *Figure S49: Probability of 3rd birth for females not in union with age at index birth of 33 years.* |

*Fertility simulation*

In the microsimulation, females of age 14-50 years with the household positions *union, union+*, *single, single+, child, NFRA* or *other* are considered eligible for giving birth. This means that females in collective households and the oldest generation in multi-generational households are excluded. Furthermore, females with a duration since index birth of less than 12 months (pregnancy of nine months and period of lactational amenorrhoea of three months) are also excluded from giving birth in the given time step.

The eligible females are assigned a probability for giving birth by applying the fertility model corresponding to their parity and household position and converting the log odds to probabilities. The fertility models are independent of time, however, the resulting birth probabilities are adjusted according to a time-trend in order to obtain an overall fertility level (i.e. total fertility rate) that resembles the observed or projected rates from Belgian Statistical Office (StatBel) and the Belgian Federal Planning Bureau for the given year [5]. Consequently, we are assuming that the relative change over time in the birth probabilities is equal across age groups, household positions and birth trajectories. Since our data on birth trajectories only covers one year (2011), it was not straightforward to develop more detailed and realistic time-trends (e.g. by age and birth trajectory).

If a female’s probability of giving birth is higher than or equal to a randomly drawn number between 0 and 1, she is assigned a newborn child and her birth history is updated accordingly. The birth event is assigned a random date within the time step. The newborns are added to the population and the household of the mother. If the mother has a partner, the partner is assigned as the father to the child.

The females with a birth event are eligible for a birth-related household transition conditional on their household position before the birth (data described in section *Birth-related household transitions*). Some transitions are ignored because they are rare or for simplicity (e.g. union dissolution). The applicable household positions are shown in white cells in Table S1. Females with household position *child* can stay in their parental household, but the household position changes to *single+\** and the household becomes a multi-generational household. However, birth events among females with household position *child* can also involve a move to a new household either as *single+* or *union+*.

The transition to *union+* involves a partner search based on the age and household position of the female as well as the male candidates. This process is described in further detail in the section *Union formation*. Females with position *single(+)* either stay in their household and keep the household position or move to *union+* with a matched partner. Births to females with household position *NFRA* or *other* always involve a move to a new household and a change to household position *single+* or *union+*. Females in a union and with a birth event stay in the union and get the household position *union+*. The household transitions are assigned the same event date as the birth. Moreover, household IDs and household positions are also updated for the individuals indirectly affected by birth-related household transitions (e.g. new partner, grandparents in multi-generational households).

*Table S1: Birth-related household transitions (white=applicable, black=non-applicable)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **To**  **From** | Child | Single+\* (parental household) | Single+ | NFRA | Other | Union+ |
| Child |  |  |  |  |  |  |
| Single (+) |  |  |  |  |  |  |
| NFRA |  |  |  |  |  |  |
| Other |  |  |  |  |  |  |
| Union (+) |  |  |  |  |  |  |

Household transitions

After updating the individuals affected by birth-related households transitions, the household positions of the remaining population are updated. This may involve a transition between existing households or the creation of a new household. Individual household transitions take place with a probability conditional on current household position, age and sex. Only individuals of age 16 and older can change household position. Some transitions are not considered eligible because they are rare or for the sake of simplicity. Eligible transitions are shown in the white cells in Table S2, black cells indicate non-applicable transitions and grey cells are transitions resulting from other demographic events (e.g. child birth, child leaving parental household). The household position *single+\** was added in Table S2 to make a distinction between single parents who live in their parental household (*single+\*)* and single parents that do not live in their parental household *(single+)*.

*Table S2: Household transitions (white=applicable, black=non-applicable/no transition, grey=indirect/other demographic events)*

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **To**  **From** | Child | Union | Union+ | Single | Single+ | Single+\* | Other | NFRA | Collec. | Multi\_U | Multi\_S |
| Child |  |  |  |  |  |  |  |  |  |  |  |
| Union |  |  |  |  |  |  |  |  |  |  |  |
| Union+ |  |  |  |  |  |  |  |  |  |  |  |
| Single |  |  |  |  |  |  |  |  |  |  |  |
| Single+ |  |  |  |  |  |  |  |  |  |  |  |
| Single+\* |  |  |  |  |  |  |  |  |  |  |  |
| Other |  |  |  |  |  |  |  |  |  |  |  |
| NFRA |  |  |  |  |  |  |  |  |  |  |  |
| Collec. |  |  |  |  |  |  |  |  |  |  |  |
| Multi\_U |  |  |  |  |  |  |  |  |  |  |  |
| Multi\_S |  |  |  |  |  |  |  |  |  |  |  |

*Union formation*

Household transitions to *union(+)* are disregarded if the individual is male. Instead, the process of union formation starts when a female experiences a transition to *union(+)* and the males are indirectly assigned as partners. For each female entering a union, we search for a match in a pool of male candidates. A male is only considered a candidate if he is not already in a union and is 16 years or older. The age of the future partner of a given female is drawn from the age distribution of male partners estimated by the age of females entering a union in 2011 (described in section *Partner match*). Then the pool of male candidates is narrowed down to those with an age corresponding to the drawn value. Each remaining candidate is assigned a match probability corresponding to their probability of a household transition to *union* given their age and initial household position. A partner is then assigned to the female by drawing from the pool of candidates conditional on the match probabilities. The newly formed union is assigned to a new household as well as children previously living with any of the partners.

*Union dissolution*

Union dissolution takes place if a female with household position *union(+)* is assigned a transition to *child*, *single(+), other, NFRA* or *collective* and if a female with household position *multi\_U* is assigned a transition to *single* or *collective.* Union dissolution initiated by a household transition of the male partner is disregarded, like for union formation. However, when a female leaves the union and thereby the household according to our assumptions, a new household position is also drawn for the male partner based on the corresponding household transition rates. In case of a dissolution of *union+*, the children are assigned a ‘main parent’, which they follow to their new household. The probability of becoming the main parent is assumed to be 0.25 and 0.75 for the father and mother, respectively. The eligible household transitions from the position *union+* are conditional on whether the individual is assigned the role of main parent. If the main parent is moving to their own parental household, the new household position is *single+\** and not *child*. Moreover, a transition to *single, other, NFRA* or *collective* is changed to *single+* if the individual is assigned the role of main parent. Finally, for reasons of simplicity it is not possible to move directly from one union to another.

A union dissolution is also triggered by a transition from *union* to *collective,* because we only consider unions between individuals living in the same household, although the relation in many cases probably continues.

*Transition to parental household*

Individuals can move back to the household of their parent(s) if one or both are present in the population and the parent is not living in a collective household. The household position changes to *child*, when moving back to the parental household, also for adult individuals. This transitions is not eligible for the oldest generation in multi-generational households for whom a parental household rarely is present in the population. Single parents and individuals in a union cannot move into their parental household, unless it is in combination with a union dissolution. However, the parents of an individual with household position *union(+)* or *single+* can move into their child’s household (see section *Transition to multi-generational household*).

*Transition to multi-generational household*

Multi-generational households can be created by two different types of transitions. As already mentioned, individuals can move from a union and back to their parental household together with the children, for whom they are the main parent, but without the former partner. In that case the individual’s household position changes to *single+\*.* Moreover, the household position of the oldest generation in the household (the grandparents) indirectly changes to *multi\_U* or *multi\_S,* depending on whether they are in a union or single. However, individuals can also experience a direct transition to the position *multi\_U/multi\_S*, if they have a child with household position *union(+)* or *single+.* In that case, the older generation moves into the household of their child’s family nucleus. This type of transitions does not affect the household position of the younger generation, except for single parents, which go from *single+* to s*ingle+\**.The oldest generation in multi-generational households follow their child in case of future household transitions (e.g. union formation or dissolution), but can also leave the household and change household position to *union* or *single* depending on whether it involves a union dissolution.

*Transition to household positions other and NFRA*

Transitions to *other* and *NFRA* are only eligible if the individual is not a single parent or assigned the role of main parent after union dissolution. Moreover, the oldest generation in multi-generational households are also not eligible for those household transitions. For each individual with a transition to *other,* a target household size is drawn from the household size distribution of *other* households observed in 2011 (see Figure S16). The group of individuals assigned to a target size of two are matched in pairs based on their age and assigned a new household ID. For an individual with a target household sizes larger than two, an existing *other* household is assigned so that the resulting household size matches the target size of the individual. The household ID of the individual is updated accordingly and the household position changes to *other*. The household positions of the new household members are not affected by the transition.

For individuals with a transition to *NFRA*, a target household size is drawn from the household size distribution of non-family related adults observed in 2011 (see Figure S15). Candidate households are found by identifying individuals with household position *single+* and *union(+),* which are considered to be in a family nucleus. Conditional on the target household size, a household ID is drawn from the pool of candidates and assigned to the individual so that the resulting household size matches the target size. The individual is assigned the household position *NFRA* and the new household ID. The household positions of the new household members are not affected by the transition.

*Transition to collective household*

In the first time-step, all collective households are assigned a maximum capacity which is calculated as follows:

where is the initial household size of collective household *i*. This implies that the maximum capacity is set to the initial size of households with less than 10 household members, while households with 10 or more individuals are assigned an extra capacity corresponding to 15% of the initial size. The median age of the household members in each collective household is computed in each time-step.

Transitions to a *collective* household are only eligible if the individual is not a single parent or assigned the role of main parent after union dissolution. For those individuals with an eligible transition to *collective* household, a target household size is drawn from the age-specific household size distribution for collective households observed in 2011 (see Figure S17). The individuals are assigned to existing collective households with a median age of ±15 years of their own age and with a size corresponding to their target household size. When the household capacity is met, no more individuals are assigned. In case some individuals have not been assigned a household, new collective households are created with a maximum capacity corresponding to the target household size.

*Transition to single/single parent*

All individuals are eligible for a transition to *single* or *single parent.* The final household position depends on whether the individual has the role as main parent. In both cases, a new household is created.

*Indirect transitions*

Household positions can also change indirectly as a result of household transitions among household members. This for example takes place when a child is born, a child leaves the parental household or the spouse of an individual dies.

Migration

*Immigration*

The number of immigrants in a given time step is computed by multiplying the population on January 1st by the immigration rate for the given simulation year (see Figure S21). Age and sex of the immigrants are drawn from the distributions in Figure S22. Some immigrants are assigned the household position *child* based on the age-specific distribution observed in 2011. Female immigrants without a household position are assigned the positions *union, union+* and *single+* conditional on the age-specific distribution of household positions observed for female immigrants in 2010. Females in *union(+)* are matched with male immigrants using the same union formation procedure as in the section *Union formation*.

Females with household position *union+* and *single+* are assigned to children as mothers based on an allocation procedure. For a given child, candidate mothers are identified by computing the age difference between the female and the child. Females are added to the pool of candidates if the age difference is between 14 and 45 years. The female candidates are assigned a probability of being the mother of the given child, which is based on the observed age difference between mothers and children immigrating in 2011 and the age of the child (age group 0-9, 10-19, 20+). A female candidate is then drawn from the pool and assigned as mother to the given child. In case the pool of candidates is empty, females who already have been assigned as mothers to other children are added to the pool again and can in that way be mothers to multiple children. In case the pool of candidates remains empty, the upper limit for the age difference between mother and child is removed. The children are added to the household of the mother and her partner, if any, which is assigned as father of the child.

Some children in the immigrant population are assigned native parents according to the age-specific proportions observed in the immigrant population in 2011.

The remaining immigrants without a household position, are assigned the positions *single, NFRA* and *other* according to the distribution of household positions observed in 2011 as seen in Figure S23. Immigrants with household position *NFRA* are assigned to a household in the native population in the same manner as described in section *Transition to household positions other and NFRA*.

Immigrants are assigned a date of entering the population which is the same for all members of a given household. Immigrants are not at risk of a household transitions in the time-step of entry into the population but they are at risk of a birth event.

*Emigration*

The number of emigrants in a given time step is computed by multiplying the population on January 1st by the emigration rate for the given simulation year (see Figure S21). Emigrants are split into two groups, individuals emigrating alone and whole households emigrating. Each group accounts for 43% and 57% of the emigrants, respectively.

Individuals are eligible for emigrating alone if they are 16 years or older and have the household position *child, single, other* or *NFRA.* Eligible individuals are chosen for emigration conditional on the household position and age-sex distribution observed for emigrants in 2011 (see Figure S24Figure *S25*).

Whole households are eligible for emigration with the exception of collective households. Eligible households are assigned an age-composition-specific probability of emigrating which is based on the observed composition of households that emigrated in 2011. While taking the probabilities into account, households are drawn until the estimated number of emigrants in the given time step has been met.

Emigrants are assigned an event date within the time step for leaving the population, which can only take place after any other demographic events of the given individual. Household members emigrating together are assigned the same event date. Emigrants are removed from the population and cannot re-enter the population.

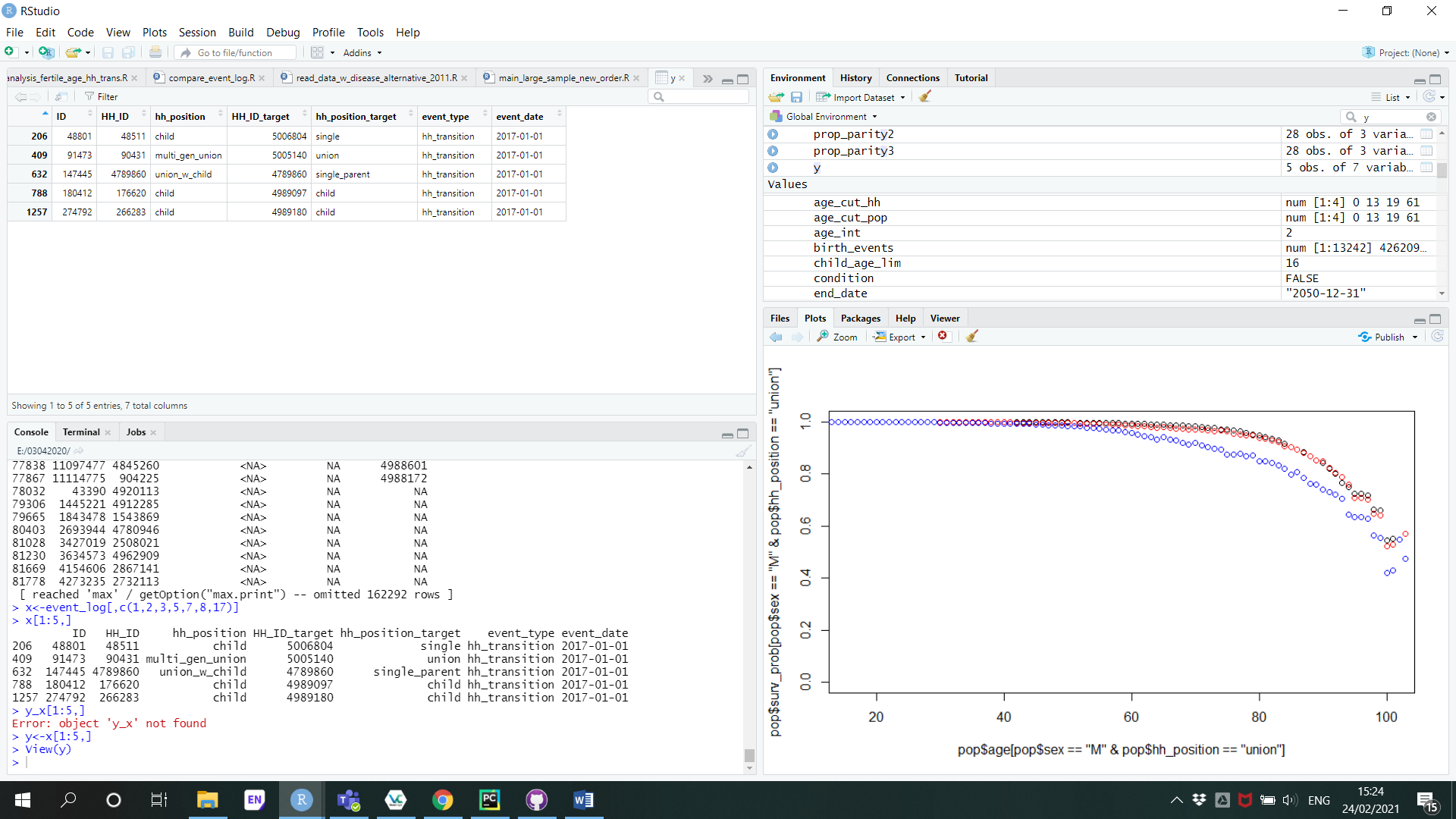
Mortality

In each time step, all individuals have a probability of dying based on age, sex- and household-position-specific mortality probabilities (qx). These probabilities are computed by first converting the observed/projected age-sex-specific mortality probabilities from StatBel to mortality rates, which are then multiplied by the household-specific rate ratios (see section *Mortality rates*). Finally, the mortality rates are converted back to mortality probabilities and assigned to the individuals according to age, sex and household position. A random number between zero and one is drawn for each individual and if the random number is lower than the mortality probability, the individual dies within the time step. A date of death is assigned, which will always be after any other demographic events of the given individual. The dead individuals are removed from the population.

The partner of a dead individual is at risk of a household transition. Moreover, children of a dead individual and with household position c*hild* follow/move to the other parent and in case both parents are dead and the child is younger than 16, it is adopted by another family.

Event-log

All demographic events in the microsimulation are tracked in an event-log (see example in Figure S50). This includes the following variables: ID, household ID (initial and target), household position (initial and target), type of event and date of event. In this way, the evolution of the population can easily be re-created and used for other applications, e.g. as input in models for infectious disease transmission.



*Figure S50: Example of event-log*

Code

The model is implemented in *R* and the code is available from the GitHub repository: <https://github.com/signemoegelmose/demographic_microsimulation_EXTERNAL>

References

1. van Imhoff E, Keilman N. LIPRO 2.0: an application of a dynamic demographic projection model to household structure in the Netherlands. NIDI/CBGS Pubications nr. 23. Amsterdam/Lisse: Swets & Zeitlinger; 1991.

2. Hastie T, Tibshirani R. Generalized Additive Models. London: Chapman and Hall; 1990.

3. Hastie T, Tibshirani R. Generalized Additive Models. Stat Sci. 1986;1(3):297–318.

4. Wood SN. Generalized Additive Models. An introduction with R. Second edi. CRC Press. Taylor & Francis Group; 2017.

5. Vandresse M. Modelling fertility for national population projections. The case of Belgium. 2020. (Economic analyses and forecasts). Report No.: 3–20.

1. https://ec.europa.eu/eurostat/documents/2995521/8774296/3-28032018-AP-EN.pdf/fdf8ebdf-a6a4-4153-9ee9-2f05652d8ee0 [↑](#footnote-ref-1)