Deliverable M.4.1

Provisional demographic outline

Lorenzo Turci
Arnaud Bringé
Elisabeth Morand
Sophie Pennec
Laurent Toulemon
Rodolfo Baggio
Francesco Billari

INED

FP7-244557

Revision: 1

22/11/2010

Contents

Provisional demographic outline

Lorenzo Turci
Arnaud Bringé
Elisabeth Morand
Sophie Pennec
Laurent Toulemon

INED
133 boulevard Davout
75020, Paris (France)

Teleph.: +33 (0) 1 56 06 20 00
e-Mail-Address: lorenzo.turci@ined.fr

Rodolfo Baggio
Francesco Billari

Bocconi University
via G. Röntgen 1
20136, Milan (Italy)

Teleph.: +39 02 58 36 35 35
e-Mail-Address: rodolfo.baggio@unibocconi.it

22/11/2010

Abstract

The deliverable M.4.1 of work package 4 describes the demographic module proposed for the SustainCity project. It is an external module written with the help of Modgen (a programming language based on C++ and developed by Statistics Canada to support the creation of dynamic microsimulation models), that simulates the future of individuals and their households over time. Events that may eventually be faced by individuals are simulated such as birth, death, union formation, union dissolution, leaving the parental home...

Keywords

Demographic microsimulation model; ModGen; microsimulation; demography; microdata

Preferred citation style

Turci, L., A. Bringé, E. Morand, S. Pennec, L. Toulemon, R. Baggio and F. Billari (2010) Provisional demographic outline, *SustainCity Deliverable*, **4.1**, INED, Paris.

1 General features of the module

The proposed model is a closed cross-sectional discrete demographic model (Morand et al.,2010. The demographic model starts with a population of persons and households at an initial point in time t_0 (t is measured in years). Every member of the population ages from one year to the next, from t_0 to the end of the simulation period.

Annual probabilities. The ageing process depends on annual transition probabilities which vary with several covariates (most often only sex and age). Probabilities may change with time, i.e. the transition probabilities will be estimated for each year, based on existing projections. The user may choose whether to leave transition probabilities constant, to add a time covariate (yearly dummies), or to produce a set of probabilities for each year. Moreover, some events may depend not only on their probability but also on the features of other members of the population (e.g., agent-based model logic for the mating process).

Covariates. Individual probabilities are estimated as sex- and age- (or duration-)specific rates, and parameters indicating the relative risks. Some interactions can be included, but the model has to be as simple as possible: events already experienced define the current state of each individual. Transition probabilities (for each event) are defined for each individual, depending on sex, age and other covariates. A demographic covariate is included in the microsimulation model if it has an impact for other UrbanSim modules or if it is correlated with some of the transition probabilities within the demographic module. Some events may concern groups of individuals. Union formation, separation, and fertility processes are women driven. Migrations are made by all the members of a household together. For each transition, a set of basic covariates is defined, as a minimum model. Additional covariates may be included in order to increase the relevance of the estimation.

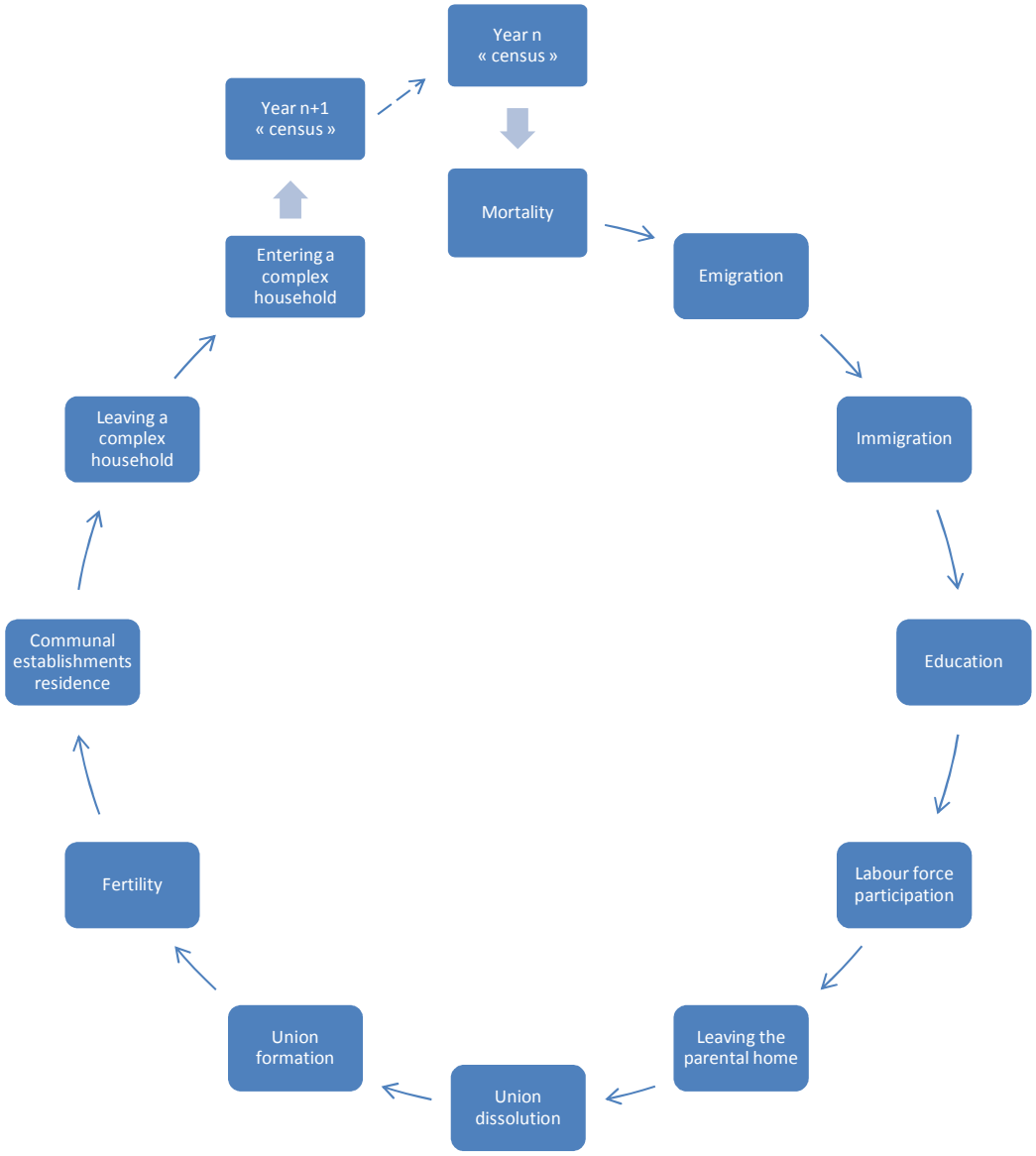
For the sake of simplicity, age as a covariate used to estimate transition probabilities is increased every year, and remains constant over the year (period-cohort parallelograms in the Lexis diagram), even if the software treats age as a continuous variable.

Social heterogeneity is accounted for with two covariates: level of education, divided in three groups, low (primary), middle (secondary) and high (university); place of birth (born in the region, born in the country, born abroad). For immigrants duration since arrival can be used as a covariate. Labour force participation can take two values: "in the labour force" (working or unemployed) or "not in the labour force" (inactive or retired). In an extended version of the demographic module, unemployed and retired categories could be isolated.

Individuals and households. Each individual belongs to one and only one household. One reference person is defined for each household. The precise definition will be decided with the users. See e.g. http://www.insee.fr/en/methodes/default.asp?page=definitions/pers-ref-menage-exp-prin-rrp.htm for the French definition. The demographic module does not include any information on dwellings. The location of households within dwellings will be part of the Location module.

Order of events. The events run in the order presented here: a person who dies during the year will not face any other event e.g. the birth of a child can occur only to women alive and living as a couple at the beginning of the year etc. Figure 1 shows the order of events:

Figure 1 One-year event modelling order, starting at "Year n "census"



Output. The microsimulation leads to two sets of results. First, an overall population is produced each year, as if there was an annual census. Any data can then be extracted from this population, at individual or household level. Second, for each year, based on the initial population and the transition probabilities, the population experiences some events which can be tracked. The output of the demographic model can be at individual or household level. It will be a type delimited file containing all the useful variables for Urbansim.

Calibration. Some calibration can be made in order to post-stratify the population (individuals or households) by size or according to any other variable, in order to get a population consistent with possible exogenous constraints. This calibration process is not explicitly implemented in the demographic module. It must be done by trial and error, as there can be many ways to post-stratify a population. For instance, if the population growth is larger than expected, it can be made smaller, either directly through lower fertility, more severe mortality, less immigrants or more emigrants, or indirectly with lower union formation probabilities or larger union dissolution risks (if couples have a higher fertility than people living alone, or if people living as a couple emigrate less than people living by their own). Different calibrations may have different side-effects: a decline in fertility will lead to a small decrease in the number of households, at least in the long run, while an increase in union dissolutions will lead to less births, but also to an increase in the number of households, at least in the short run.

Programming language, interface. We will create a separate demographic module using the Modgen language. It is a language developed and maintained by Statistics Canada for creating microsimulation models. This language is based on C++ and is free of charge. It has been used for 15 years and different types of models have been developed using this language.

Relation with the other modules. The demographic module will be built as an external module (as agreed on the 12/11/2010 meeting). The Ined team will test it and make a sensitivity analysis before delivering it. Then, the Berkeley team will examine the algorithms and data structures used in order to decide whether to develop an interface to read tables of results (a person and a household file per year) or create a Python version of the module to be embedded in the UrbanSim software.

2 Individual Events

The events modelled are the following: death, emigration (exits from the region), immigration (entries into the region), education attainment and enrolment, professional status, leaving of the parental home, union dissolution (dissolution of de facto union and divorce), union formation (cohabitation and marriage), birth of a child, and communal establishment residence. The demographic module simulates the evolution of the population size through entries (births, immigration) and exits (death, emigration). Interactions between agents are also taken into account: some events occurring to an individual may have an effect on some other individuals. For instance, when a person dies, his/her spouse, if any, becomes widow(er) and children, if any, become orphans; moreover the size and composition of the household change.

All the events that involve a couple/household decision will be woman driven. Hence, the probabilities of these events will be estimated only for women. For instance, the probability of entering a union will be applied to women not living in union, then a process of mate matching to men not in union will take place in order to determine the best suitable partner through an optimisation technique under constraints.

2.1 Death

2.1.1 Basic covariates

Death probabilities by sex and age are taken from country-level data. Region-specific data are not used, as they can be misleading if migrations are not independent from death risk. If available, relative risks by marital status are used.

2.1.2 Additional covariates

The level of education and communal establishment residence (non-private dwellings) are used as additional covariates. As noted in Sec. 1, it will be possible for users to simulate time trends. As far as mortality is concerned, for example, it is easy to simulate death probabilities which decline over time.

2.1.3 Other individuals concerned in case of death

There can be at most three "external" consequences of a person's death: the union in which (s)he is in, if any, is ended; her/his partner's marital status is set to "widow(er)" if he/she was married; the link with her/his children, if any, is severed and their status is set to "orphan".

2.2 Emigration

The demographic module takes into account immigrations into the region and emigrations from the region. Domestic migrations are simulated in the localisation module.

Emigration and immigration are separated in two parts: migrations to and from abroad, migrations to and from other parts of the country. Migrations are a household-level phenomenon: all the members of a household migrate together, and a specific union-formation process is put in place for immigrant women (see Sec. 2.3).

2.2.1 Basic covariates

Emigration is a one-way process. The probability that a household emigrates is estimated separately for four categories of households:

- Households with a couple and children. The emigration probabilities vary with the number of children and age of last child.
- Other households with no couple and children (lone parent families). The emigration probabilities vary with the number of children and age of last child.
- Households with at least a man and no children (childless households including a man). The emigration probabilities vary with the age of the (oldest) man, and the number of household members.
- Households with women, without men or children (childless households with no man). The emigration probabilities vary with the age of the (oldest) woman and the number of household members.

A specific calibration will be needed in order to insure the consistency of emigration data (most often composed by individuals by sex and age) and emigrations of households.

Probabilities will be estimated separately for emigration to another region of the country or abroad.

2.2.2 Additional covariates

Optional covariates include place of birth, especially for emigration abroad, and level of education of the reference person of the household, if suitable data are available.

2.3 Immigration

Two pools of immigrants are built from the most recent census or survey data on immigrant population, as far as possible on recent immigrants (who arrived in the region not long before data collection). One group includes immigrants from abroad, one group immigrants from other regions of the country. Immigrants come as a household (one-person or multiple-person households); among this group, some immigrants, especially women with children, have a very high probability to enter a union (family-based migration): see Sec. 2.8.

When only net migration by age is available, a specific calibration must be performed in order to ensure the consistency between entries and exits, both at the household and global level (i.e. net migration). Calibration is made on entry and exit probabilities, as well as on the household situation of recent immigrants.

2.3.1 Basic covariates

The number of immigrants by sex and age is used. The simplest way to obtain the immigrants pool is to clone some households from the existing population. When available, recent immigrants at census allow to fill in the pools of potential immigrants. The potential immigrant group has to be grouped into households; covariates must be imputed for immigrants if they are not available.

2.3.2 Additional covariates

The simplest way is to clone some households from the existing population, chosen among the households with a recent immigrant. If these data are not available, the potential immigrants have to be cloned from the existing population (by level of education and age).

2.4 Education attainment

Education is used in different ways in the model. The demographic module uses level of education as a proxy for socio-economic status. It can also be used by other module if needed. A second variable, age at end of studies, is used in order to take into account the fact that some people leave the education system at different ages despite having the same level of education. This may affect the family structure and therefore the localisation process.

Two ways to simulate education can be thought of:

1. the model follows each child and determines each year whether (s)he enters the next level, remains at the same level, or leaves the education system;

2. level of education and age at end of studies are directly imputed once and for all.

The first option seems more attractive but it needs much more data to be performed, especially if social reproduction is needed (i.e. children of parents with high level of education more likely to have a high level of education). The second option is chosen for the sake of simplicity and because the simulation of education is not our main purpose.

2.4.1 Basic covariates

For the level of education, the basic variables used are parents' level of education and sex.

To directly input the age at end of studies, instead, the level of education and sex are used.

For the localisation module, the distribution of children by type of school is needed (kindergarten, primary school, secondary school, tertiary education). Age of entering in each type of institution will be imputed according to each agent's level of education and age at end of studies.

2.5 Labour force participation

Regarding labour force participation, two statuses are possible: in labour force (working and looking for a job) and not in labour force (retired, not working).

2.5.1 Basic covariates

Probabilities to enter into, or exit from, the labour force by sex and age are considered, together with age at end of studies.

2.5.2 Additional covariates

Optionally, it is possible to make a finer classification of professional activity, by dividing the population in four statuses: working, unemployed, not working and retired. If data are available, relative risks by couple situation (in a couple or no) and enrolment (student or not) are used, as well as level of education.

2.6 Leaving the parental home

2.6.1 Basic covariates

The basic covariates include the probabilities of leaving parental home by sex and age.

2.6.2 Additional covariates

Marital status, student status, working status can be additionally used, if data are available. In order to take into account the fact that co-residence with parents may end with a union, people who just entered a union leave the parental home with a higher probability.

2.7 Union dissolution

2.7.1 Basic covariates

As all events involving couple decisions, union dissolutions are woman driven. Therefore, female union dissolution probabilities by age drive this process. In case of disruption, children stay with the mother, women keep the household and men move to another household.

2.7.2 Additional covariates

Optional variables include marital status, parity, age of youngest child, duration of latest union and level of education.

2.8 Union formation

Union formation is women driven and follows a two-step process:

1. a draw among non-partnered women determines who enters a union;
2. a partner market is used to match a non-partnered man to each woman entering a union.

2.8.1 Basic covariates

The probability of entering a union varies by age, number of children living with the mother and marital status.

2.8.2 Additional covariates

"Living with parents" dummy to create complex households can be added. Newly immigrant women (especially those with children) have a higher probability to enter a union (family-based migration). Women who just left the parental home may also have a higher probability to enter a union.

2.8.3 Mate matching

For each woman entering a union, a suitable partner must be found. The model is closed: therefore partners belong to the population; every non-partnered man is eligible.

Two different techniques can be used to build such a partner market: stable marriage optimisation or stochastic matching. Some studies have shown that the stochastic process gives a better distribution of the pairs than the optimisation process. The mate matching process is performed as follows.

For each woman selected as entering in union, a subset of around 50 potential partners is chosen randomly among the non-partnered men. For each potential pairs, a distance function is calculated. Then, a stochastic matching routine is used to determine which of these partners will be chosen. The “born overseas - just arrived" category is used in order to take into account family-based migration, i.e. those entering as spouse and children of a region resident. Variables used to calculate the distance function are: age, level of education, place of birth (born in the region; born in the country but not in the region under study; born overseas), former marital status, duration since arrival in the region.

2.9 Marriage

Marriage can be either direct, when the partners have not been living together before, or indirect, i.e. the legalisation of a de facto relationship.

For direct marriages, two alternatives are possible:

1. applying the same process as entering a union by applying a probability of direct marriage and a marriage market to choose a husband from;
2. considering that the probability of entering a union covers both de facto and de jure unions and then selecting a percentage of unions as direct marriages.

For the sake of simplicity, we take the second approach.

2.9.1 Basic covariates

For those already in de facto union, a transition probability to marry will be applied, by age and union duration (in order to take into account "direct" marriages). Having a child (age 0) can also be related to a higher probability of marriage.

2.9.2 Additional covariates

Four additional covariates can be used: level of education, number of children, year of union, place of birth.

2.10 Birth of a child

For this module, two hypotheses are made:

- only women aged from 15 to 49 can have children;
- women must be in couple to have children.

2.10.1 Basic covariates

The basic specification includes female fertility by age and birth order.

2.10.2 Additional covariates

Additional variables are: union duration, time elapsed since last birth (age of youngest child), and level of education. Similarly to the mortality module, it is possible to simulate fertility time trends.

2.11 Communal establishments/non-private dwellings

For communal establishments simulation, entry and exit probabilities by sex and age are considered. No further details other than living in non-private dwellings will be simulated. The location module, and not the demographic module, will determine the type of communal establishment in which these persons, given their age and sex, will be located.

Persons entering a non-private dwelling are teenagers and young adults entering boarding school or student hall, or young workers homes, and older people entering long term care unit or nursing homes.

Exit from non-private dwellings will affect mainly young adults, those ending their education process.

2.11.1 Basic covariates

The basic covariates are sex, age and marital status.

2.11.2 Additional covariates

Additionally, immigration status can be used to take young newly immigrants into account (those arrived at the beginning of the year), who have a large probability to move to a communal establishment, especially if they are students.

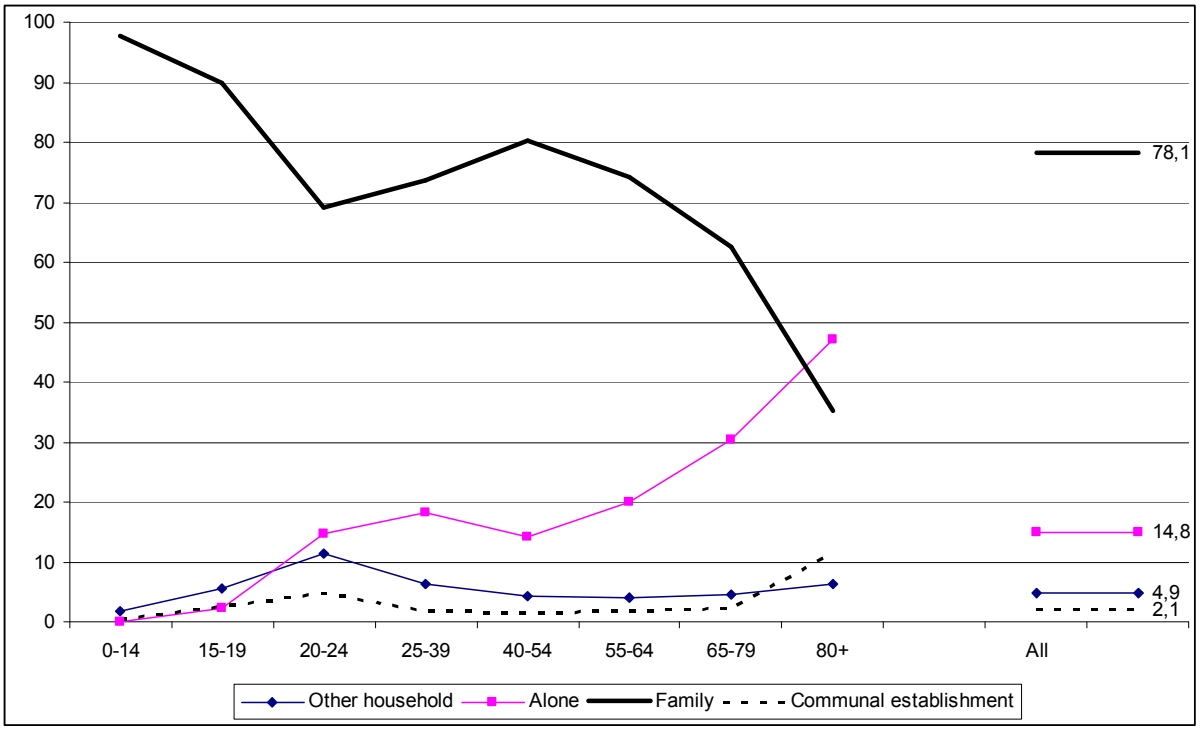
People living alone have a higher probability to enter a communal establishment. People quitting a communal establishment move to live by themselves when they quit.

2.12 Additional events and statuses

2.12.1 Complex households

According to the 2006 census, in Île-de-France 22% of inhabitants do not live within a family: 15% are living alone in their household, 4.9% are living in a household with friends, relatives, etc. and 2.1% are living in a communal establishment (two thirds of them in a medium-stay hospital or extended care facilities).

Figure 2 Distribution of Île-de-France inhabitants in 2006 by family and household situation



Source: http://www.recensement.insee.fr/accesTableauxDetailles.action

As these numbers show, it is important to simulate complex households if the aim is to design a realistic model. Therefore, three different types of transition to complex households are envisaged:

1) Non-partnered childless persons who enter a flat-sharing household. Three possibilities: they can move only to one-person household; to a multiple-person household but without couple; to multiple household but without children; or to all kinds of household. (Probably it would be good to check which one of these option is more important in the data).

2) New-formed couples who stay in the parental household of one of the two partners. This process is assumed to be one-way only, i.e. the couple never leaves this household.
3) Old widowed parents who move to the household of one of the daughters/sons. Also this transition is supposed to be one-way only.
4) Homosexual couples will not be explicitly simulated: they will be implicitly simulated at point 1, grouped with non-family households.

2.12.2 Individuals and households

Individuals are born within their parental household. They leave the parental household in order to live by themselves. In case of a parental disruption the children stay with their mother. The man joins the woman's household when they enter a union, and leaves the household in case of disruption. Each individual can enter into or leave a communal establishment at any age.

Leaving a complex household: household disaggregation. In addition to these demographic events, people who are not living as a couple, nor living with parents, may quit the household in order to live alone.

Entering a complex household: household aggregation. In addition, household aggregation take place. Practically, people living alone have a probability to join another household. In case of aggregation, a matching process leads to the choice of the household the person joins. For instance, older persons may often join adults living as a couple (mostly their children), while young adults may more often join other young adults of the same age. People who just left a communal establishment may have a higher probability to aggregate.

Thus two events are created, related to leaving or entering into a complex household.

2.13 Leaving a complex household

People who are not living as a couple, nor living with parents, may quit the household in order to live alone.

Basic covariates

The basic covariates include the probabilities of leaving a complex household by sex and age.

Additional covariates

The presence of a child in the household may be taken into account.

2.14 Entering a complex household

People who are living alone can enter another household. Entering a complex household follows a two-step process:

1. a draw among people living alone determines who enters another household;
2. a household market is used to match an existing household to each person entering a household.

Basic covariates

The probability of entering a household varies by sex, age, and the occurrence of a child living in the area.

Additional covariates

The sex and marital status of the children may be taken into account (presence of a daughter or daughter-in-law in the area).

Individual matching and household choice

For each person entering a household, a suitable one must be found. The matching process is similar to the mate matching process: people entering a complex household join another individual of the population.

For each person selected as entering a complex household, a subset of around 50 potential mates is chosen randomly among the population. For each potential pair, a distance function is calculated. Then, a stochastic matching routine is used to determine which of these partners will be chosen. The “born overseas - just arrived” category is used in order to take into account family migration and other grouping of immigrants. Variables used to calculate the distance function are: parent-child relationship, age (with large difference being allowed, as older persons may join younger related people, and young adults may join older ones), size of the household (in order to avoid a large probability for large households). Then the person enters into the chosen person’s household.

3 Variable needs for the demographic module

3.1 Starting population (in 1990 and, if possible, in 2000)

The starting population must include individuals with their personal and household ID and all the necessary covariates at both levels. This implies to find a data source with all the variables needed or if necessary to impute some of the variables.

Individual data are preferred.

Table 1 Stock variables needed

	Basic covariates	**Additional covariates**
Sex and age	- Sex: female or male - Age	
Fertility	- Number of children in the household - Parity - Age of the youngest child	- Age of the oldest child
Union formation and dissolution	- Marital status: never been in union/ single (ever been in union)/ in de-facto union/ married (in de-jure union)/ divorced/ widowed - Former marital status - Year of union	- If in union: union duration and order
Migrations	- Place of birth: "city"/ country/ abroad - Immigration data: previous residence ("city"/ country/abroad)	- Duration since arrival in the region
Education	- Level of education - Parents' level of education - Enrolment - Age at end of studies, when applicable - Distribution of children by	- Level of education of the reference person of the household

	type of school	
Labour force participation	- Professional status: in the labour force or not in the labour force	- Detailed professional activity: working/ unemployed/ not working/ retired
Identification	- Household ID - Partner, father and mother ID, if they are in the household	- Parents alive - Parents live in the city
Communal establishments	- Type of household: private or communal establishment	- Type of communal establishment

3.2 Ending population (in 2000 and, if possible, in 2005)

Same as for the starting population, in order to check the calibration.

3.3 Events and transition probabilities

Probabilities by sex and age are the best. Events by sex and age are second best. Probabilities and events can be given for the period just after of just before the first period (starting population). Time trends are welcome if they can easily be projected.

Table 2 Flow variables needed

	Basic Covariates	**Additional Covariates**
Mortality	- Death probabilities by sex and age (country-level)	
Emigration	- Emigration probability to abroad and to another region of the country by type of household, number of children and age of last child (if applicable); or number of household members and age of the oldest man (or woman if inapplicable)[1]	- Place of birth - Level of education of the reference person

[1] See Sec. 2.2.1 for a more detailed explanation.

Immigration	- Number of immigrants by sex and age	
Education	- Age at end of studies, by sex	
Labour force participation	- Probabilities of entry and exit into/out of the labour force, by sex and age	- Unemployment rates by sex and age - Retirement rates by sex and age - Employment, retirement and entries/exits into/out of work by marital status and level of education
Leaving the parental home	- Probability to leave the parental home by sex and age	- Probability to leave the parental home by sex, age, level of education and enrolment
Union formation and dissolution	- Female union probability by age and union status - New unions by each partners' age, education and place of birth - New unions by number of children living with the mother, marital status and level of education - Subsequent unions by parity - Marriage probabilities by age - Marriage probabilities by presence of a child in the union - Marriages by previous cohabitation - Female union dissolution probability by age	- Female union dissolution probability by other covariates than age (e.g. parity, union duration)
Fertility	- Female fertility by age - Fertility by birth order (and age)	- First births probability by union duration - First birth probability by union duration - High order births by duration since last birth (age of the last child)
Communal establishments	- Estimates of population in communal establishment	- Probabilities of entry and exit into/out of communal

establishments by sex and age

4 References

The principles presented in this deliverable are based on the literature review included in the Sustaincity deliverable 2.1 (Morand et al. 2010).

MORAND Elisabeth, TOULEMON Laurent, PENNEC Sophie, BAGGIO Rodolfo, BILLARI Francesco, 2010, Demographic modelling: the state of the art. SustainCity Deliverable 2.1, 38 pages. http://www.sustaincity.org/publications/WP_2.1a_-_Demographic_models.pdf