HAL
open science

Cause-of-Death Mortality: What Can Be Learned From Population Dynamics?

Séverine Arnold, Alexandre Boumezoued, Héloïse Labit Hardy, Nicole El Karoui

To cite this version:

Séverine Arnold, Alexandre Boumezoued, Héloïse Labit Hardy, Nicole El Karoui. Cause-of-Death Mortality: What Can Be Learned From Population Dynamics?. 2015. hal-01157900

HAL Id: hal-01157900

https://hal.science/hal-01157900

Preprint submitted on 28 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cause-of-Death Mortality: What Can Be Learned From Population Dynamics?

Séverine Arnold [1], Alexandre Boumezoued [2], Héloïse Labit Hardy [1], Nicole El Karoui [2]

May 28, 2015

Abstract

This paper analyses cause-of-death mortality changes and its impacts on the whole population evolution. The study combines cause-of-death analysis and population dynamics techniques. Our aim is to measure the impact of cause-of-death reduction on the whole population age structure, and more specifically on the dependency ratio which is a crucial quantity for pay-as-you-go pension systems. Whereas previous studies on causes of death focused on mortality indicators such as survival curves or life expectancy, our approach provides additional information by including birth patterns. As an important conclusion, our numerical results based on French data show that populations with identical life expectancies can present important differences in their age pyramid resulting from different cause-specific mortality reductions. Sensitivities to fertility level and population flows are also given.

Keywords: Cause-of-Death Mortality, Population Dynamics, Birth and Death Process, Individual-Based Model, Age Pyramid, Age Dependency Ratio, Macroscopic Behavior

[1]Département de Sciences Actuarielles (DSA), Faculté des HEC, Extranef, Université de Lausanne, 1015 Lausanne, Suisse.
Email: severine.arnold@unil.ch, heloise.labithardy@unil.ch

[2]Laboratoire de Probabilités et Modèles Aléatoires (LPMA), UMR CNRS 7599 Université Paris 6, 4 Place Jussieu, 75005 Paris, France.
Email: alexandre.boumezoued@upmc.fr, nicole.el_karoui@upmc.fr

1 Introduction

The studies on cause-of-death mortality have provided a better understanding of the level and the evolution of aggregate mortality rates over the years. At the same time, population dynamics models have been developed in mathematical ecology and demography to model the evolution of the population age structure over time by including both death and birth patterns. Whereas studies on cause-of-death mortality focused on the impact on mortality indicators, this paper aims to combine these two fields in order to answer the following question: what is the impact of changes in cause-of-death mortality on the whole population age structure ?

To address such question, the first step is the modeling of cause-of-death mortality. Each observed death is recorded with a primary cause, such as for example cancer, accident, or a disease related to circulatory or respiratory system. For modeling cause-of-death mortality, one famous framework in actuarial science and biostatistics is the competing risks model. In this framework, each cause of death is modeled at the individual level by a clock, and the death occurs when the first clock rings. In other words, the lifetime of an individual is modeled as the minimum between competing cause-specific lifetimes. Unfortunately, those cause-specific durations are not observable in the data: in practice we only observe the minimum between the durations related to each cause. Since information on causes of death durations are not observable, assumptions about the dependence between causes of death have to be set in order to capture the marginal distributions and to model changes in cause-of-death mortality. The seminal work of Chiang (1968) on the competing risks theory considered independent causes of death and led to further studies during the last decades. The competing risks framework has been investigated with several dependence structures between the competing lengths of time (see e.g. Carriere (1994), Kaishev et al. (2007) and Dimitrova et al. (2013)). For our study, we focus on the competing risks framework under the independence assumption, although we also briefly survey the debate on the dependence structure. This independence assumption is much used (see e.g. Prentice et al. (1978), Tsai et al. (1978), Wilmoth (1995) and Putter et al. (2007)), and allows us to already highlight interesting effects of cause removal on the population dynamics.

The previous studies in the cause-of-death literature focused on the impact of cause-of-death removal or reduction on mortality indicators: e.g. life expectancy for Prentice et al. (1978), survival curve for Dimitrova et al. (2013) and death probabilities for Alai et al. (2015). In this paper, the aim is to go further and study the impact on the population age pyramid (the number of individuals by age-class) by means of a population dynamics model. The age pyramid is a crucial quantity of interest since it gives informations on how age-classes evolve and interact with each others. To our knowledge, this has not been addressed in the current literature.

To do so, we rely on both deterministic and stochastic population dynamics models. Historically, the first models that addressed the deformation of the age structure of a population over time were deterministic models. The most famous one is called the McKendrick-Von Foerster model (see McKendrick (1926) and Von Foerster (1959)). Based on death rates by age, and also birth rates, they are able to capture the time evolution of the number of individuals by age-class. For demographic purposes, dealing with gender classification has become natural as census and national data make this distinction between males and females for a long time. It is also crucial since age but also gender is known to have a real impact on mortality. To this aim, there is now a sizable literature on two-sex population deterministic models both with and without age structure (see e.g. Keyfitz (1972) and Iannelli et al. (2005)). The deterministic point of view is useful to understand the evolution of the "average" age pyramid, and can provide first interesting insights about the impact of cause-of-death reductions on the whole population, based on a fast numerical discretization scheme. But in practice, we are not only interested in the age pyramid, but also in non-linear quantities such as the age dependency ratio. This is defined as the number of people aged above 65 divided by the number of individuals aged between 15 and 65. This indicator is a crucial quantity of interest for pure demographic analysis but also for pay-as-you-go pension systems. The deterministic evolution of the population and the age dependency ratio, interesting in itself, is not realistic for populations with finite size in which individuals have random lifetimes and times of birth. What is then the population dynamics ? First insights tell us that the average stochastic dependency ratio, as a non-linear quantity, cannot be equal to the ratio of the average population. Therefore a proper stochastic modeling framework is needed. In fact, in a large population asymptotics, the stochastic behavior can be approximated by its deterministic counterpart, making this approach micro-macro consistent. In other words, at the microscopic scale, the population behaves stochastically whereas if the number of individuals is "large", the behavior is described by deterministic equations that are classical in demography. The micro-macro consistency is at the core of what are called individual-based models, that have been developed in mathematical ecology (see Fournier and Méléard (2004) and Champagnat et al. (2006)), in particular to understand the evolutionary mechanism. In addition, these models have been developed to take ages into account (see Ferriere and Tran (2009)), which is crucial for our purpose.

These processes have been little used for demographic applications compared to issues linked to biology or evolutionary theory. In particular, to our knowledge none of them address the impact of cause-of-death removals on the population age structure. Importantly for our purpose, we are inspired by the work of Bensusan et al. (2010–2015) (see also Bensusan (2010)) who have extended and used population dy-

namics models in view of human population applications, including pension systems and longevity risk in insurance. Among the few applications to demographic purposes, note also that an alternative named Markovian binary tree is considered in Hautphenne and Latouche (2012) for demographic applications. Markovian binary trees differ from individual-based models as they involve discrete age-classes and a random aging pattern to preserve the Markov property. The key feature of our approach is to take into account aging in continuous time and to use a microsimulation procedure for the stochastic scenarios. It is interesting to note that microsimulation procedures are used in a discrete time setting by several public institutions to test economic and retirement policies (see e.g. the review by Li and O'Donoghue (2013)). This includes for example the French National Institute for Statistics and Economic Studies (INSEE) with the model *Destinie* (see Bonnet et al. (1999) and Duée (2005)). Apart from government bodies, microsimulation approaches have also been used in the demographic and statistical literature. As an example, Hyndman and Booth (2008) used discrete-time stochastic simulations to compute the population evolution and the age pyramid, while aggregate mortality, fertility, and migration rates are forecasted.

The paper is organized as follows. In Section 2, we first describe the competing risks framework for the modeling of cause-of-death mortality, and secondly introduce the deterministic and stochastic population dynamics approaches. Section 3 details the data we use to carry out our study: the World Health Organization data for the cause-of-death mortality rates and the data from INSEE for the birth rates and the initial age pyramid. Section 4 contains our numerical results regarding deterministic and stochastic patterns of the age pyramid and the dependency ratio, as well as sensitivities to fertility scenarios and also to population flows computed from the data. As an important conclusion, our numerical results show that under same life expectancy improvement, the age pyramid can evolve in many ways if different causes of death are reduced.

2 The Model

As usual when working with human population, we are concerned with population structured by gender and age, so we make the distinction between the age pyramid of females and the age pyramid of males. It is well known that gender and age induce significant differences on mortality. In the model, each individual, female or male, is exposed to d competing causes of death, such as cancers, diseases of circulatory system, diseases of respiratory system or accidents, well described in Section 3, but may die from any single one of the d causes. After a brief description of the classical competing risks framework in survival analysis, we introduce the debate on the

dependence structure between causes of death, and describe the different points of view proposed in the literature. Afterwards, we present the meaning of removal of a cause of death. The population dynamics modeling is described secondly.

2.1 Cause-of-Death Mortality

Competing risks framework Standard survival data measure the length of time from the time origin until the occurrence of some events; for human population, one is interested in the duration between the time of birth and the time of death (*frailty time* in other framework). This is called the *lifetime* and is represented by a random variable τ^ϵ, with survival function $S^\epsilon(a) = \mathbb{P}(\tau^\epsilon > a)$, for an individual with gender ϵ ($\epsilon = f$ for female or $\epsilon = m$ for male). This is the probability for an individual with gender ϵ to survive until age a. As usual, the information is often expressed in term of *hazard rate* (also called *force of mortality* in actuarial sciences), defined as the probability to die before age $a + da$ for an individual alive at age a:

$$\mu^\epsilon(a) = \frac{\mathbb{P}(a < \tau^\epsilon \leqslant a + \mathrm{d}a \mid \tau^\epsilon > a)}{\mathrm{d}a} = -\frac{\mathrm{d}\ln(S^\epsilon(a))}{\mathrm{d}a}. \tag{2.1}$$

At this elementary stage, as in Elandt-Johnson (1976), we assume that in each group, females or males, each individual with gender ϵ is assigned a vector of potential lifetimes $(\tau_1^\epsilon, \tau_2^\epsilon, ..., \tau_d^\epsilon)$ corresponding to the d causes of death. The d causes of death are modeled by d lengths of time, that is to say d competing clocks, denoted $\tau_1^\epsilon, \tau_2^\epsilon, ..., \tau_d^\epsilon$. The death occurs when the first event corresponding to one of the d causes of death occurs: that is when the first clock rings. In other words, the lifetime of an individual τ^ϵ is defined as the minimum between the different lengths of time: $\tau^\epsilon = \min_{1,\dots,d} \tau_i^\epsilon$. In terms of survival function, we have:

$$S^\epsilon(a) = \mathbb{P}(\tau_1^\epsilon > a, \cdots, \tau_d^\epsilon > a). \tag{2.2}$$

To study the lifetimes associated to causes of death, we need the corresponding marginal survival functions. In particular, to model changes in cause-of-death mortality, as a reduction of death due to cause i, the survival function has to be expressed with the marginal distribution $S_i^\epsilon(a) = \mathbb{P}(\tau_i^\epsilon > a)$ and these have to be estimated from the data. In practice, when looking at a sample of lifetimes, we observe for a given individual, the age at death and the cause from which this individual dies. In the competing risks framework, this means that we observe the minimum of the d competing clocks, and that we know what is the clock which rang first. However, we do not observe the clocks for the other causes of death. Since marginals are not observable, dependence assumptions between causes of death have to be set.

On the dependence structure It is well known that, without additional assumption, it is impossible to identify the dependence structure of the joint distribution and their marginal from the data (Tsiatis (1975)). A considerable literature has been devoted to the competing risk model for cause-of-death mortality, starting from the independent case (Chiang (1968)) to more complex dependence structures. Carriere (1994), Kaishev et al. (2007) and Dimitrova et al. (2013) have worked on the theory of copula fonctions to model dependence: they have been interested in modeling density function and survival curve. Alai et al. (2015) have studied causal mortality shocks with a multinomial logistic model. They have studied and modeled the death probabilities. Also models incorporating observed individual risk factors (see e.g. Manton (1986a), Rosen (2006) and Girosi and King (2006)) and unobserved ones (frailty models, see e.g. Vaupel and Yashin (1983) and Manton et al. (1986b)) have been developed. More recently, cointegration analyses have been applied to causes of death studies (see Arnold and Sherris (2013), Arnold and Sherris (2015a), Arnold and Sherris (2015b)). In that framework, cointegration is used to model dependence between the cause-of-death mortality rates.

Nonetheless, the independence assumption between cause-of-death durations is widely used (see e.g. Prentice et al. (1978), Tsai et al. (1978) and Wilmoth (1995), Putter et al. (2007)). Even if we agree with many authors on the partial irrelevance of this assumption, that can unfortunately not be tested in practice, it contributes to considerably simplify the further developments of this study. In particular, the marginals are directly estimated from the data (see the discussion below). Moreover, under the independence assumption, we can already highlight interesting effects of cause removal on the population dynamics.

Independence assumption Under the independence assumption between causes of death, the survival function defined in Equation (2.2) can be expressed with the marginals: $S^\epsilon(a) = \prod_{i=1}^d S_i^\epsilon(a)$. In this way, the hazard rate of the individual lifetime is the sum of the hazard rates by cause: $\mu^\epsilon(a) = \sum_{i=1}^d \mu_i^\epsilon(a)$, where μ_i^ϵ is such that $\mathbb{P}(\tau_i^\epsilon > a) = \exp\left(-\int_0^a \mu_i^\epsilon(s)\mathrm{d}s\right)$. The rate $\mu_i^\epsilon(a)$ is called the *net* force of mortality for cause i, gender ϵ and age a. Formally, this corresponds to the probability to die instantaneously if only cause i exists, given the survival at age a:

$$\mu_i^\epsilon(a) = \frac{\mathbb{P}(a < \tau_i^\epsilon \leqslant a + \mathrm{d}a \mid \tau_i^\epsilon > a)}{\mathrm{d}a}. \tag{2.3}$$

As already described, the cause-specific durations are not totally observable in practice. However, from the data one can estimate the distribution of τ_i^ϵ for all individuals that died from cause i, because in this specific case the duration τ_i^ϵ is completely observed. In other words, it is possible to estimate the quantity $\mathbb{P}(\tau_i^\epsilon \in [a, a + \mathrm{d}a), \tau^\epsilon = \tau_i^\epsilon)$. By independence between the cause-specific durations,

this quantity is equal to

$$\mathbb{P}(\tau_i^\epsilon \in [a, a + \mathrm{d}a), \tau^\epsilon = \tau_i^\epsilon) = S_i^\epsilon(a)\mu_i^\epsilon(a)\mathrm{d}a \prod_{k \neq i} S_k^\epsilon(a) = S^\epsilon(a)\mu_i^\epsilon(a)\mathrm{d}a.$$

This way, it is possible to recover the *net* force of mortality by

$$\mu_i^\epsilon(a) = \frac{\mathbb{P}(\tau_i^\epsilon \in [a, a + \mathrm{d}a), \tau^\epsilon = \tau_i^\epsilon \mid \tau^\epsilon > a)}{\mathrm{d}a}.$$

In the actuarial literature, the right hand side of the previous equation is called the *crude* force of mortality. Under the independence assumption, it can be directly estimated form the data. The fact that *net* and *crude* mortality forces are equal under the independence assumption is a fundamental result which considerably simplifies the estimations.

Cause-of-death removal To address the impact of cause-of-death reduction or removal, a rigorous definition is needed. In the competing risks literature, two possible definitions have been introduced and tested: *ignore* or *eliminate*, see e.g. Elandt-Johnson (1976) and Dimitrova et al. (2013).
(i) To ignore The first one says that if cause k is ignored, then the lifetime becomes $\tau_{(-k)}^\epsilon = \min_{1 \leqslant i \leqslant d, i \neq k} \tau_i^\epsilon$. In this case, the survival function is replaced by $\mathbb{P}(\tau_{(-k)}^\epsilon > t)$. Under the *ignoring* definition, the random variable $\tau_{(-k)}^\epsilon$ can be interpreted as the lifetime in the world where cause k does not exist.
(ii) To eliminate The other definition is the *elimination*: in this case, the lifetime is now conditioned to the fact that τ_k^ϵ grows to infinity, that is the survival function is replaced by $\mathbb{P}(\tau^\epsilon > t \mid \tau_k^\epsilon \to +\infty)$. This models the fact that the cause still remains, but its associated lifetime is postponed to an arbitrary value. Despite their difference in terms of interpretation, in the case where the τ_i^ϵ are independent, the two definitions are equivalent. This is another fundamental consequence of the independence assumption. In our case, we will use the term *remove* to refer to both definitions.
We will also refer to cause-of-death reduction: we say that cause k is reduced by a factor $\alpha^\epsilon \in [0, 1]$ for individuals with gender ϵ if the death rate $\mu^\epsilon(a)$ becomes:

$$(1 - \alpha^\epsilon)\mu_k^\epsilon(a) + \sum_{1 \leqslant i \leqslant d, i \neq k} \mu_i^\epsilon(a).$$

Debate on temporality of death rates It is well known that mortality is improving over the years. This is captured for example by the increase in life expectancy which reflects the fact that individuals belonging to recent generations are living longer. In particular, the time evolution of mortality differs across age classes. In this paper, as a first step, we consider time-invariant death rates (and also fertility

rates), although the modeling framework can be described with time varying demographic rates. Whereas the purpose of our paper is not to forecast cause-specific mortality rates, we are interested to forecast the population itself with current fertility and cause-of-death mortality rates, and possibly population flows, in order to analyse the deformation of the age structure in this special setting. This is known as the stable population framework which is standard in demography. Although, it does not provide realistic population forecasts, this framework already allows to understand and interpret some short and long term impacts of a change in current cause-of-death mortality on the population age pyramid.

2.2 Population Dynamics

Cause-of-death mortality rates are useful to understand the distribution of individual lifetimes in a static population made of both men and women. However, this does not tell something about the time evolution of the whole population, in particular its interdependence with birth patterns and mortality reduction.

Two-sex birth-death process with age structure The simple one-sex population model is the well known birth-death process, in which birth and death rates are constant. While applied demographers primarily worked with one-sex population models, many sociological applications required two-sex models that explicitly address matching of females and males into couples. Some important applications such as "marriage squeeze", i.e. the impact on marriage of an imbalance between males and females (see e.g. Schoen (1983), Guilmoto (2012)), further investigated age-structured models that keep track of the age distributions of single females and single males as well as the joint distribution of ages across couples. In fact, dealing with gender classification has become natural as census and national data make this distinction for a long time. There is now a sizable literature on two-sex population models both with and without age structures (see e.g. Keyfitz (1972), Iannelli et al. (2005)). However, these two-sex population models do not embed cause-of-death analysis, which is the purpose of the present paper which focuses on both deterministic and stochastic two-sex birth-death processes with age structure and cause-of-death mortality. In order to understand the behavior of the population over time, in particular that of the age pyramid, we first need to specify how individuals give birth. The age pyramid (see Figure 3 for an illustration) is a crucial quantity of interest for pure demographic analysis, but also public pension pay-as-you-go systems. As for private insurers, this is crucial to understand how national mortality data is computed, especially how a biased knowledge on birth patterns and corresponding age pyramid can lead to misevaluation of reference mortality rates (see Cairns et al. (2014)). Further investigation on the impact of (heterogenous) birth

patterns is carried out in Bensusan et al. (2010–2015).

Female fertility In our model, each individual with gender ϵ and age a gives birth to new individuals (of age 0) with rate $b^\epsilon(a)$. In a stochastic setting, the first time of birth T_b^ϵ, for an individual with gender ϵ, is given by its marginal distribution $\mathbb{P}(T_b^\epsilon > a) = \exp\left(-\int_0^a b^\epsilon(s)\mathrm{d}s\right)$. As this individual can give birth to more than one individual, the times of birth occur as the jumps of a non-homogenous Poisson process with rate $a \mapsto b^\epsilon(a)$, until the parent dies. At birth, the new individual of age 0 is a female (f) with probability p and a male (m) with probability $1-p$. In our application, we assume that only a female can be the cause of the arrival of a new individual in the population. This choice is driven by the lack of data on male birth rates, in particular the age of the father at birth of his children. So the birth rate for any individual of gender ϵ and age a can be written $b^\epsilon(a) = b(a)\mathbf{1}_f(\epsilon)$, where $\mathbf{1}_f(\epsilon)$ is the indicator function defined as $\mathbf{1}_f(f) = 1$ and $\mathbf{1}_f(m) = 0$. In this model, only each female has an intrinsic birth rate $b(a)$, so we do not take into account the impact of the relative number of males and females in the population on the birth patterns. This has been the topic of several models and debates in demographic studies. In particular, this could be of interest to model events such as shocks in birth patterns prior and posterior to world wars. Also, the general setting can be useful to test several assumptions on the link between the birth rate and the whole population. This issue is left for further research.

Deterministic population model How does the population evolve over the years? In particular, what is the dynamics of the age pyramid over time? The first models which tried to address these questions were deterministic population models. They go back to McKendrick (1926) and Von Foerster (1959) and focus on the quantity $g(\epsilon, a, t)$ of individuals with gender ϵ and age a at time t (see also Iannelli et al. (2005) and the brief review in Tran (2008)). The population structured by gender is described by the vector

$$\mathbf{g}(a,t) = \begin{pmatrix} g(f,a,t) \\ g(m,a,t) \end{pmatrix}.$$

First, individuals with non-negative ages evolve according to deaths over time, that is

$$(\partial_a + \partial_t)\mathbf{g}(a,t) = -Q(a)\mathbf{g}(a,t), \tag{2.4}$$

where $Q(a) = \begin{pmatrix} \mu^f(a) & 0 \\ 0 & \mu^m(a) \end{pmatrix}$. This equation corresponds to the transport component: it states that individuals are aging over time and that the number of individuals is decreased by the number of deaths. Now, at a given time t, the number

of newborn is given by the sum over all newborns of all females with any age, and added to the gender class depending on the probability at birth, that is

$$\mathbf{g}(0,t) = \left(\int_{\mathbb{R}_+} g(f,a,t)b\,(a)\,\mathrm{d}a\right)\begin{pmatrix} p \\ 1-p \end{pmatrix}. \tag{2.5}$$

This is called the renewal component, which sets some boundary condition on the space of ages. Initially, at time 0, the age pyramid is given by some initial condition $\mathbf{g}(a,0) = \mathbf{g}_0(a)$ for each $a \geqslant 0$. Equations (2.4) and (2.5) form what is called the McKendrick-Von Foerster equation. Note that this formulation exhibits an autonomous behavior of the female population, and that the renewal of the male population can be computed based on the female population at time t. Note also that we deal with "quantities" of individuals as $g(f,a,t)$ and $g(m,a,t)$ are real numbers: this represents the average population quantity. A first advantage of the deterministic equation is to be numerically very fast to compute through an explicit discretization scheme, and thus provides first insights on the average dynamics.

The numerical scheme The deterministic equation can be computed in the following way. Let Δt be the discretized time step. Let also $\mathcal{G}$ denote the grid $\{0, \Delta t, 2\Delta t, ...\}$ on R_+, and $\mathcal{G}_+$ denote the grid $\mathcal{G}$ without 0. We start with an initial population $\mathbf{g}_0(a)$ for $a \in \mathcal{G}$, with $\mathbf{g}_0(a)$ the quantity of individuals of age in $[a, a+\Delta t[$ for each gender ϵ in the initial population. The explicit numerical scheme can be written

$$\begin{aligned}
&\mathbf{g}(a+\Delta t, t+\Delta t) - \mathbf{g}(a,t) = -Q(a)\mathbf{g}(a,t)\Delta t,\ \forall a \in \mathcal{G}_+ \\
&\mathbf{g}(0,t) = \sum_{a\in\mathcal{G}_+} g(f,a,t)b\,(a)\,\Delta a \begin{pmatrix} p \\ 1-p \end{pmatrix},\ \text{where } \Delta a = \Delta t, \\
&\mathbf{g}(a,0) = \mathbf{g}_0(a),\ \forall a \in \mathcal{G}.
\end{aligned} \tag{2.6}$$

To clarify the simulation steps, we describe how to proceed recursively. Suppose that the population at time t has been computed. From the first line of (2.6), for $a \in \mathcal{G}$, one can compute

$$\mathbf{g}(a+\Delta t, t+\Delta t) = (I - \Delta t Q(a))\,\mathbf{g}(a,t),$$

where I denotes the 2×2 identity matrix. Then, one gets from the second line of (2.6), where t is replaced by $t+\Delta t$,

$$\mathbf{g}(0,t+\Delta t) = \sum_{a\in\mathcal{G}_+} g(f,a,t+\Delta t)b\,(a)\,\Delta t \begin{pmatrix} p \\ 1-p \end{pmatrix}.$$

This leads to the whole population at time $t+\Delta t$.

From deterministic to stochastic population The deterministic equation is useful to understand the evolution of the "average" age pyramid, in the sense that $g(\epsilon, a, t)$ represents the average quantity of individuals with gender ϵ and age a at time t. But in practice, we are not only interested in the age pyramid, but also in non-linear quantities such as the dependency ratio. This is defined as the number of people aged above 65 divided by the number of individuals aged between 15 and 65. Let us denote $g_t(a) = g(m, a, t) + g(f, a, t)$ the quantity of people aged a at time t. Let us also introduce the notation $g_t([a_1, a_2)) = \int_{[a_1,a_2)} g_t(a)\mathrm{d}a$ for the quantity of individuals with age $a \in [a_1, a_2)$ at time t. Formally, the dependency ratio computed from the deterministic equation can be defined as

$$r_t = \frac{g_t([65, \infty))}{g_t([15, 65))}.$$

The assumption of a deterministic population pattern is obviously not realistic. In practice, one deals with finite size populations in which individuals have random lifetimes and times of birth. What is then the stochastic population dynamics? Before introducing the stochastic point of view, let us focus on the corresponding stochastic dependency ratio, say R_t. We want to underline that the average dependency ratio $\mathbb{E}[R_t]$ cannot be directly computed from the deterministic equation, as it is not equal to the ratio of the average population, r_t. To further analyse the stochastic behavior and make the link with its deterministic counterpart, let us first describe the stochastic representation.

Analogously to the deterministic counterpart, let us denote $Z_t^\epsilon([a_1, a_2))$ the number of individuals with age $a \in [a_1, a_2)$ and gender ϵ at time t. The difference is that we deal here with (random) number of individuals as we are in the microscopic stochastic setting. The population Z_t^ϵ, which is in fact a measure on the space of ages, evolves over time according to random births and deaths with rates described at the beginning of this Section. Let us introduce the total number $Z_t([a_1, a_2)) = Z_t^f([a_1, a_2)) + Z_t^m([a_1, a_2))$. In this microscopic setting, the stochastic dependency ratio is defined as

$$R_t = \frac{Z_t([65, \infty))}{Z_t([15, 65))}.$$

As mentioned in Bensusan et al. (2010–2015) (see also the discussion in Fournier and Méléard (2004)), as individual birth and death rates do not depend on the whole population, the average population density satisfies $\mathbb{E}[Z_t([a_1, a_2))] = g_t([a_1, a_2))$, provided that such density exists. However, the dependency ratio R_t is a non-linear quantity. In particular, one can write the definition of the covariance to get

$$r_t = \mathbb{E}[R_t] + \mathrm{Cov}\left(R_t, \frac{Z_t([15, 65))}{g_t([15, 65))}\right).$$

That is, the deterministic dependency ratio r_t is the average dependency ratio $\mathbb{E}[R_t]$ plus a corrective term related to the covariance between the dependency ratio and the deviation of the number of people aged [15,65) from its mean. This shows that the stochastic dependency ratio with finite population is obviously not well described by the deterministic equation. Therefore, a proper stochastic framework is needed. In addition, one can be more precise about the link between the stochastic representation and its deterministic counterpart. At the core of what are called individual-based models is the micro-macro consistency: at the microscopic scale, the population behaves stochastically whereas if the number of individuals is "large", the behavior is described by deterministic equations that are classical in demography. In the following, we briefly highlight the large population point of view and then describe the intuitive stochastic simulation procedure.

Link between stochastic and deterministic models One main advantage of such stochastic model is its micro-macro consistency. If the number of individuals increases to infinity, the stochastic dynamics tends to the deterministic pattern described by Equations (2.4)-(2.5). Let us introduce the vector $\mathbf{Z}_t = \begin{pmatrix} Z_t^f \\ Z_t^m \end{pmatrix}$, and recall that each component defines a measure on the space of ages. In particular, $\mathbf{Z}_t([a_1, a_2))$ is the number of individuals for each gender and ages in $[a_1, a_2)$ at time t. To get a large population limit theorem, it is assumed that the size of the initial population tends to infinity proportionally to a scaling parameter $K \geqslant 1$. We consider a sequence of population processes $(\mathbf{Z}_t^K)_{t\in[0,T]}$ with initial population $\mathbf{Z}_0^K$. A non-degenerate process appears in the limit if each individual is weighted $\frac{1}{K}$: the renormalized population process is defined by $\tilde{\mathbf{Z}}_t^K := \frac{1}{K}\mathbf{Z}_t^K$. The convergence to the deterministic limit solution to (2.6) is stated in the following theorem. We state it in a summarized way and refer to Tran (2006) and Bensusan et al. (2010–2015) for the technical details.

Theorem 1. *Assume that there exists a finite and deterministic measure* $\mathbf{g}_0(a)\mathrm{d}a$ *such that* $\tilde{\mathbf{Z}}_0^K$ *converges in distribution to* $\mathbf{g}_0$ *as* $K \to +\infty$ *and that* $\sup_K \mathbb{E}\left[\tilde{\mathbf{Z}}_0^K([0,\infty))^2\right] < +\infty$. *Then the sequence of measure-valued processes* $\left(\tilde{\mathbf{Z}}^K\right)_{K\in\mathbb{N}^*}$ *converges in distribution towards the continuous and deterministic process* $\mathbf{g}_\cdot$*, which satisfies the McKendrick-Von Foerster equation (2.4)-(2.5) (in a weak sense).*

Simulation of the stochastic scenarios We detail here the simulation procedure described in Bensusan et al. (2010–2015) (see also Tran (2006)). The idea of the simulation procedure is to generate candidate times of event for the population, and for each of them, decide if a real change happens or not. At each candidate time, we pick an individual in the population. According to this choice, we decide if

an event (birth or death from a cause) occurs or not. This approach is based on the simulation of a simpler process with more jump times and some of them are deleted with a rule leading to the desired distribution for the population process. This procedure is analogous to the thinning procedure for point processes. This simulation technique is required since demographic rates depend on age which introduce time dependency.

Assumption 1. *(Bounded intensities) Denote $\mathcal{X} = \{f, m\}$ the set of possible gender. There exists positive constants $\bar{b}$ and $(\bar{\mu}_i)_{i=1}^d$ such that for all $(\epsilon, a) \in \mathcal{X} \times \mathbb{R}_+$, $b^\epsilon(a) \leqslant \bar{b}$, $\mu_i^\epsilon(a) \leqslant \bar{\mu}_i$. Let $\bar{\mu} = \sum_{i=1}^d \bar{\mu}_i$.*

We describe below the simulation procedure.

Algorithm 1. *Start at time $T = 0$ with a given population $\mathbf{Z}_T$ of size N_T.*
Step 1.
Simulate a r.v. $\tau \sim \mathcal{E}xp((\bar{b} + \bar{\mu})N_T)$ and set $T + \tau$ as the next candidate time.
Step 2.
Form the population $\mathbf{Z}_{T+\tau}$ by adding τ to all ages in $\mathbf{Z}_T$.
Step 3.
Select an individual (ϵ^I, a^I) uniformly in the population $\mathbf{Z}_{T+\tau}$ and compute his probabilities to give birth (B) and to die for cause i (D_i), denoted respectively p_b and p_i for i in $[\![1, d]\!]$: $p_b = \frac{b^{\epsilon^I}(a^I)}{\bar{b}+\bar{\mu}}$ and $p_i = \frac{\mu_i^{\epsilon^I}(a^I)}{\bar{b}+\bar{\mu}}$. Let also define $p_o = 1 - p_b - \sum_{i=1}^d p_i$ the "no event" probability.
Step 4.
Generate a potential event $\mathcal{E} \in \{B, D_1, ..., D_d, O\}$ with probability $p_b, p_1, ..., p_d, p_o$:
(i) *If $\mathcal{E} = B$ (birth) then generate ϵ' in $\{f, m\}$ with probability $(p, 1-p)$ and add the individual $(\epsilon', 0)$ to $\mathbf{Z}_{T+\tau}$.*
(ii) *If $\mathcal{E} = D_i$ (death for cause i) then remove the individual (ϵ^I, a^I) from $\mathbf{Z}_{T+\tau}$.*
(iii) *If $\mathcal{E} = O$ (no event) then nothing happens.*
Let $T \leftarrow T + \tau$ and go to **Step 1**.

This algorithm is simple and only involves the computation of uniform random variables which can be used to generate $\tau, I, \mathcal{E}$ and ϵ'. Note also that at each candidate time of event, only one individual is checked so that only one intensity function is computed. Remark that for numerical efficiency, the four random variables $\tau, I, \mathcal{E}$ and ϵ' needed can be generated using only one uniform random variable.

3 Data

We choose to apply our model to the French population with birth and death rates for the year 2008. We use data from the World Health Organization (WHO) for the

cause-of-death mortality rates and the French National Institute for Statistics and Economic Studies (INSEE) for the birth rates and the initial age pyramid. The WHO has developed a tool to classify causes of death: the International Classification of Diseases (ICD). The database assembles deaths by cause of death, year, age and gender since approximately 1950 in about 225 countries. In this section, we present the data and the estimation methods for the mortality and fertility rates.

Cause-of-death mortality rates WHO provides the number of deaths by cause, gender and five-year age-class, and the exposure to risk by gender and age-class. We estimate death rates by the number of deaths over the exposure to risk (see e.g. Delwarde and Denuit (2006)). The exposure to risk corresponds to an estimation of the average time during which the population of interest is alive, here, the mid-year population. We denote $D_i^\epsilon(a)$ the number of deaths due to cause i with gender ϵ at age a, and $ER^\epsilon(a)$ the exposure to risk for gender ϵ and age a. Under the assumption of independence between cause-specific durations (see Section 2.1), the *net* force of mortality can be estimated from the data by

$$\hat{\mu}_i^\epsilon(a) = \frac{D_i^\epsilon(a)}{ER^\epsilon(a)}.$$

Data on deaths are provided for age-classes of five years, except classes "0","1-4" and "95+", therefore we assume the mortality to be constant within each age-class. Moreover, the number of death by cause of death are classified until the age-class "95 years and more" while the data on exposures to risk are classified until the age-classes "85 years and more". To use information on the number of deaths for age-classes "85-89", "90-94" and "95 years and more", we approximate the cause-of-death mortality rates using exposures to risk from the Human Mortality Database (HMD) from the corresponding age-classes [4].

In the International Classification of Diseases (ICD) for France, deaths are classified according to 4563 different cause labels. Based on previous papers (Alai et al. (2015), Arnold and Sherris (2015a)), we have chosen to regroup causes of death in six categories: cancers, diseases of the circulatory system, diseases of the respiratory system, external causes, infectious and parasitic diseases (in short: I&P) and all the others causes brought together, see details on Table 1. The three main causes of death are cancers, diseases of circulatory system and the category other causes. More precisely, other causes regroup mainly the diseases of the nervous system (21% of other causes), metabolic diseases (16%), diseases of the digestive system (13%) and mental disorders (12%). Infectious and parasitic diseases, diseases of respiratory system and external causes are less represented in the total deaths but represent an

[4] We compared exposures to risk from HMD and WHO for the entire age range and year 2008. The relative difference between the exposures to risk is lower than 2%.

important part of deaths for some ages. For example, external causes represent 55% of deaths for individuals between ages 10 and 20, and 46% of deaths for individuals between ages 20 and 40. By comparison, death rates for external causes impact more specifically ages from ages 10 to 40, while cancers impact more generally ages above 40.

Cause label	%
Cancers	30 %
Other causes	28 %
Diseases of the circulatory system	27 %
Diseases of the respiratory system	6 %
External causes	7 %
Infectious and parasitic diseases	2 %
Total	100 %

Table 1: Deaths repartition by cause for the French population in 2008

On the whole, the different categories of causes of death impact differently the age-classes and genders. It is illustrated in Figure 1, which represents the death rates by cause, gender and age. As external causes mainly represent the accidents, it is interesting to note that the accident hump of aggregate mortality, which impacts more males, is essentially reflecting these external causes. More generally, it is clear that causes of death impact very different age ranges. This suggests that a drop in some cause-of-death mortality would modify the population age structure in a very specific manner. This is studied in the next section.

Fertility rates Fertility rates by age for females are provided by the INSEE. As for the mortality rates, they are estimated by the ratio between the number of births by the age of the mother at birth and the exposure to risk (see INSEE (2011)). Let us denote $B(f, a)$ the number of births by age a of the mother at birth and $ER(f, a)$ the female exposure to risk at age a. The fertility rate is estimated by the INSEE as follows:

$$\hat{b}(a) = \frac{B(f, a)}{ER(f, a)}.$$

Birth rates by age of the mother at birth in 2008 for France are represented in Figure 2. We observe the particular form of the birth rates: the intensity is maximal around age 30 and the distribution is almost symmetric. On the whole, the age range to give birth is between ages 15 and 50.

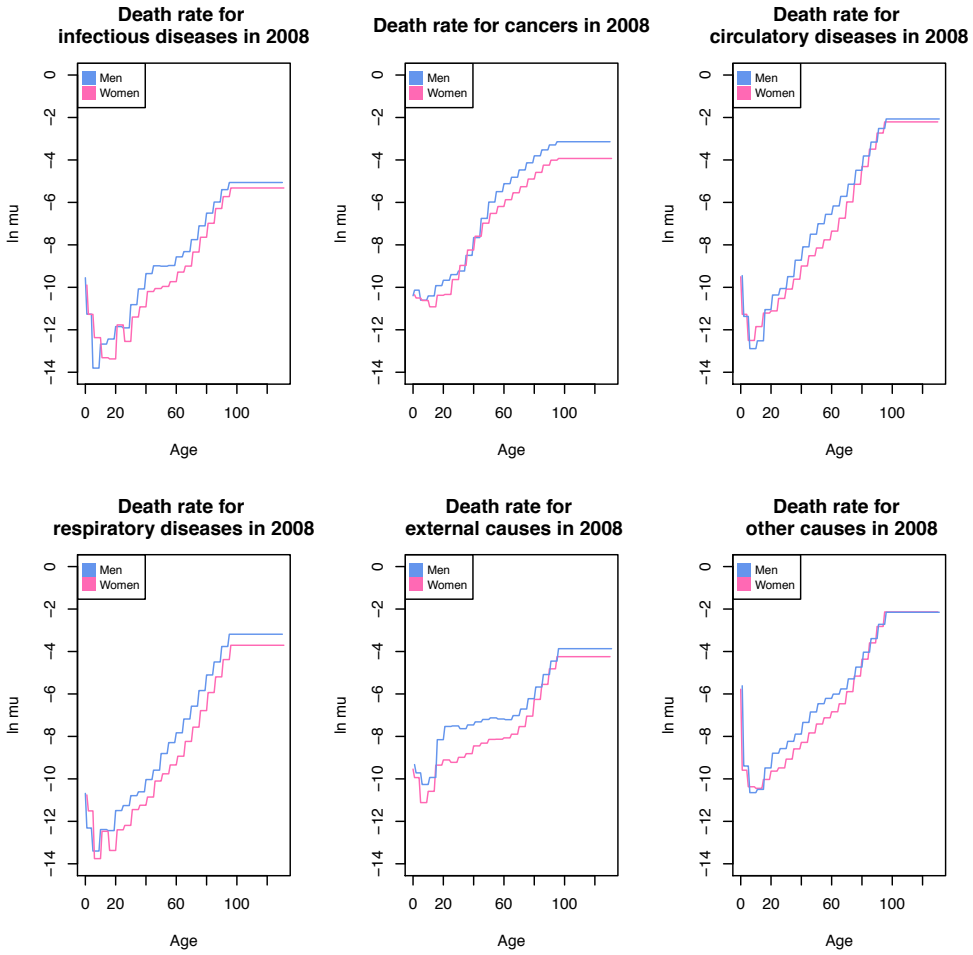


Figure 1: French death rates by cause, gender and age for year 2008

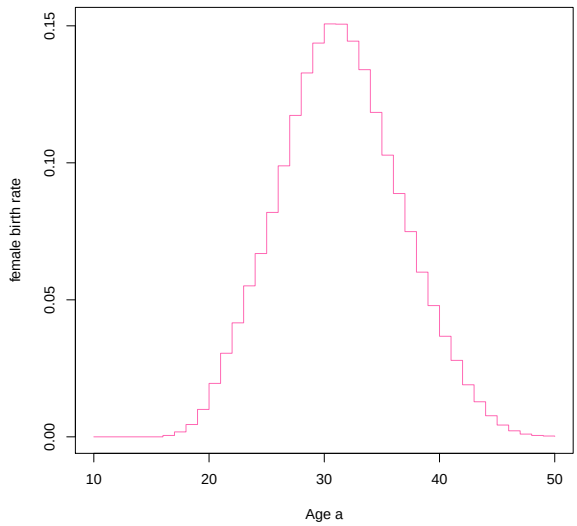


Figure 2: Birth rates by age of the mother in 2008

As detailed in Section 2, a newborn can be a female with probability p and a male with probability $1-p$. The probability p is derived from the data on the French population age pyramid, provided by the INSEE. We estimate the probability p as the ratio between the number of female births in year 2008 and the total number of births in year 2008. We estimate that in France, in 2008, the probability to be born as a female is equal to $\hat{p} = 0.4886$.

4 Results

In this section, we present the results of the simulations of the stochastic scenarios and the deterministic large population model. In the first part, we simulate the population dynamics while considering all causes of death in order to get some reference scenario. The time evolution of the dependency ratio (i.e. the number of individuals aged above 65 divided by the individuals between 15 and 65) already provides many insights regarding the long term impact of the initial population and the level of current birth and death rates. In the second part, we remove each cause of death to test the impact on the population age structure. As expected, the higher the cause-of-death mortality, the bigger the impact on the age pyramid. Finally, in the third part, we study the impact of a cause-specific reduction on the population age structure. We focus in particular on cancer and external causes and investigate how the age pyramid is affected if each cause is modified to get the same improvement in life expectancy.

4.1 Reference Scenario

Stochastic scenario First, we model the population dynamics from 2008 to 2108, considering the general mortality. We perform the simulation of a single stochastic scenario from the microscopic model described in Section 2.2, based on Algorithm 1 and an initial population with size 100 000. On Figure 3, we represent the 2008 initial population with grey lines and the final 2108 population in color (males in blue, females in pink). As explained in Section 3, the initial population is based on the structure of the French population. In the initial population, we observe more individuals between ages 40 and 70 representing the "Baby Boom" generation. In 2108, we see that the population is almost globally uniformly distributed up to age 70, and then decreases in an exponential form. This shape is stable and will be analysed in the deterministic scenarios section. The aging of the population is visible on Figure 4 which represents a path of the age dependency ratio with an initial population of 100 000 individuals. The age dependency ratio is increasing until the individuals from the "Baby Boom" generation die and then becomes stable due to time-independent birth and death rates. With this first analysis, we capture

the order of magnitude of the dependency ratio and we address the impact in time of the "Baby Boom".

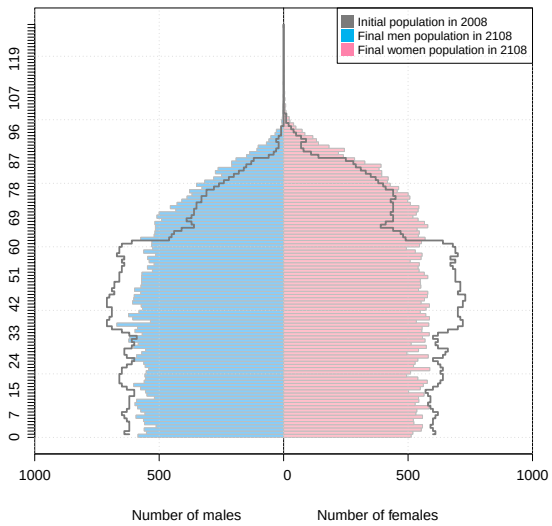


Figure 3: Initial (100 000 individuals) and final population pyramids

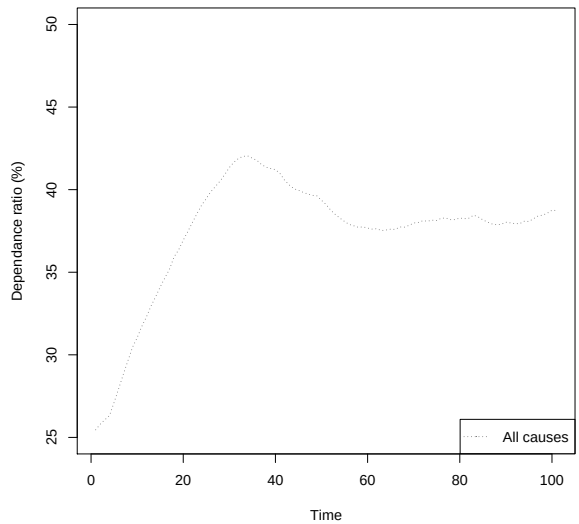


Figure 4: Age dependency ratio from 2008 to 2108

Deterministic scenario The large population equation (2.4)-(2.5) allows to compute the deterministic evolution of the population, in particular the whole age pyramid and the old-age dependency ratio. Here, numerical insights about the stochastic behavior cannot be addressed, but the virtue is that the numerical computation is far faster, as it amounts to discretize a partial differential equation (see Section 2.2). We first plot in Figure 5 the evolution of the dependency ratio for 100 years. This represents, up to the discretization error, the evolution of the dependency ratio under the large population framework. Note that it is the dependency ratio for an "infinite" population, and not the average dependency ratio of any finite-size population. From the large population equation, it is possible to draw the deterministic scenario in the long time. In Figure 6, we represent the deterministic evolution of the dependency ratio over 300 years. Obviously, this time horizon does not make sense in itself, in particular regarding the assumption of time-independent death rates. But it provides the value of the asymptotically stable age dependency ratio. The long-term stabilization of the age pyramid is a well known fact in demography in a time-invariant rates setting. In the Appendix, we sketch the main results on the stable age pyramid in the deterministic model and the closed-formulas for its shape.

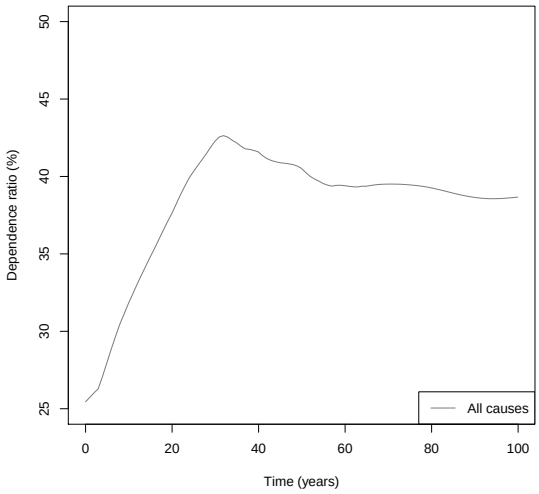


Figure 5: Age dependency ratio from the limit equation from 2008 to 2108

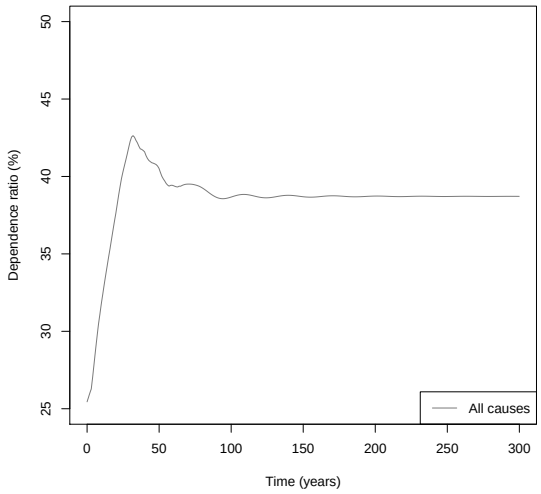


Figure 6: Age dependency ratio from the limit equation from 2008 to 2308

4.2 Cause-of-Death Removal

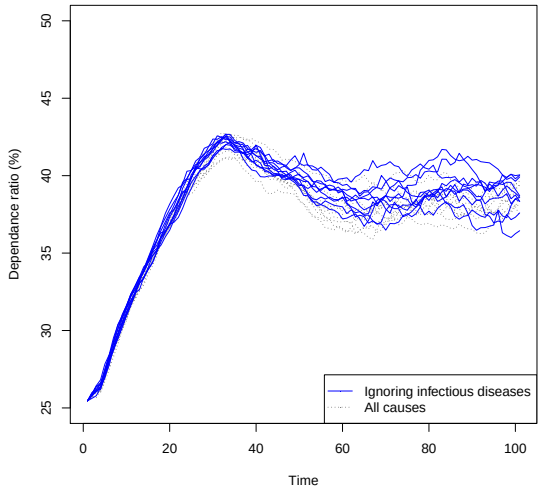
Based on numerical simulations of the microscopic model, we test the impact of removing each of the six causes of death. We now start with an initial population of size 10 000 to allow the computation of several stochastic scenarios and thus gain insights on the demographic fluctuations. The impacts of cause-removal on the old-age dependency ratio are plotted in Figures 7a to 7f. As expected, the causes with low death rates (infectious and parasitic diseases, respiratory diseases, external causes) lead to lower differences with respect to the reference scenarios. Besides, these differences are partly hidden with the demographic stochasticity. On the whole, comparisons between cause removal scenarios are not easy since the importance of each cause is quite different. Moreover, complete removal of one important cause can be considered as unrealistic. In the following, we focus on cancer and external causes and investigate how the age pyramid is affected if each cause is modified to get the same improvement in life expectancy.

4.3 Cause-of-Death Reduction

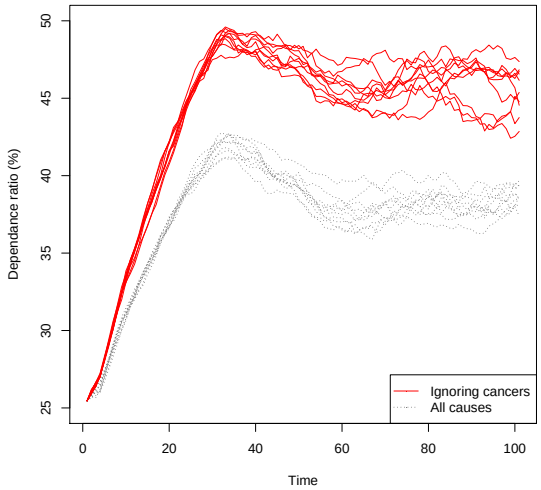
For this numerical experiment, we focus on the two following causes: cancer and external causes. They are known to have a different mortality structure: cancer has an important impact on ages above 40, whereas external causes embed the accident hump and impacts mainly ages around 25 (see Figure 1). Therefore, we want to compare several scenarios:

(i) The first one is the reference scenario, computed with current birth and cause-of-death rates. Life expectancies at birth are 84.4 and 77.7 years for females and

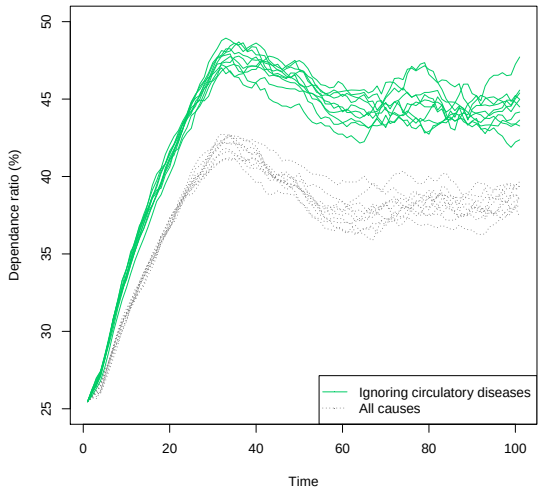
Figure 7: Age dependency ratio from 2008 to 2108



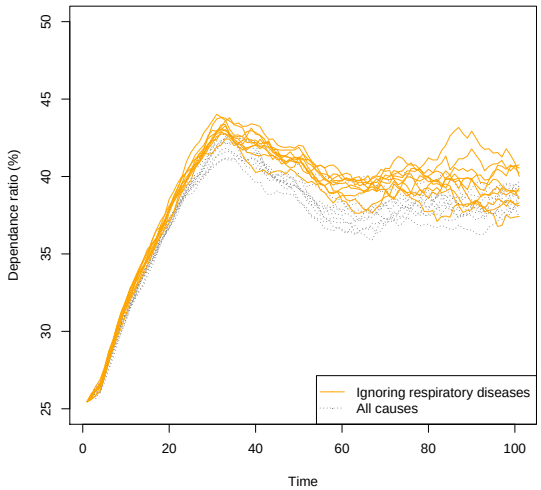
(a) I&P diseases removal



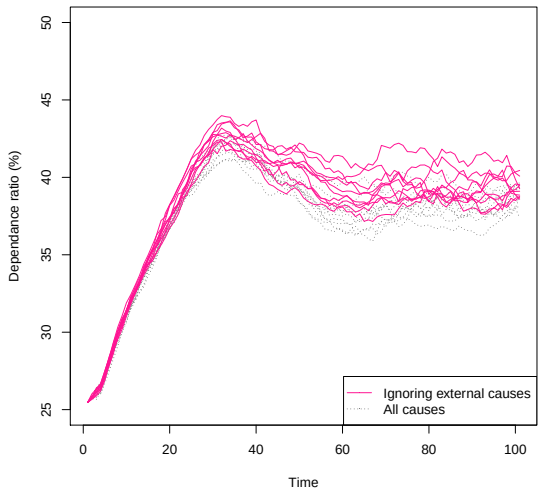
(b) Cancers removal



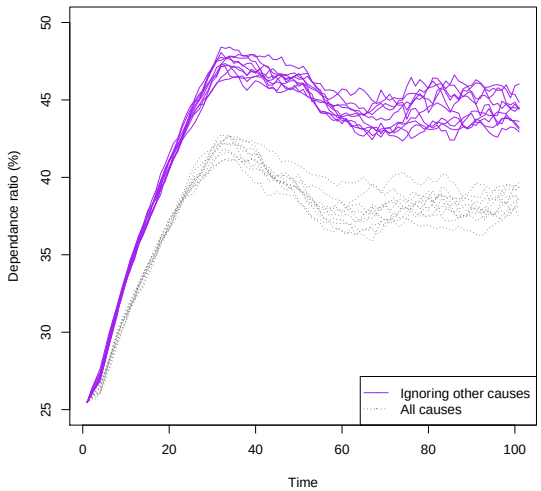
(c) Circulatory diseases removal



(d) Respiratory diseases removal



(e) External causes removal



(f) Other causes removal

males respectively.
(ii) The second one is obtained by removing external causes. Under the assumption of independent causes of death, life expectancies at birth are 85.1 and 79.1 years for females and males respectively.
(iii) The last scenario is obtained by cancer reduction in the following way: each female and male reduction factor, α^f and α^m respectively (see Section 2.1), is computed so as to replicate the same life expectancy as in the scenario of external causes removal. This implies a reduction of cancer death rates by $\alpha^f = 24.6\%$ and $\alpha^m = 35\%$ for females and males respectively.
Comparing (ii) and (iii), we are interested in the evolution of the age pyramid knowing that life expectancies are equivalent in the two scenarios.

Deterministic scenarios The deterministic scenario can be computed using the large population equation. In Figure 8, the evolution of the age dependency ratio is plotted under the three parameter configurations (i), (ii) and (iii). The scenario (ii) removing the external causes is presented, and compared to the scenario (iii) of cancer reduction. Recall that males and females life expectancies are equivalent in both scenarios (ii) and (iii). We are then interested in analysing the impact of each scenario on the population age pyramid. In all cases, the age dependency ratio is higher than the reference scenario. However, the cancer reduction scenario leads to a higher dependency ratio compared to that of external causes reduction. This is explained by the fact that mortality due to cancer impacts mainly higher ages, especially ages at which people do not give birth anymore, so that the effect is only to extend life. As for the mortality due to external causes, it mainly impacts ages at which people can give birth. The removal of external causes has two main effects: as for cancer, the first one is to make people live longer, but at the same time many people who did not die from external causes (e.g. accidents) at young ages will give birth. This population renewal increases the proportion of young people and makes the age dependency ratio lower than that with a reduction in cancer. One can find other demographic studies dealing with the counter-intuitive fact that a reduction in mortality do not necessarily lead to older populations; it can even lead to younger populations (see e.g. Keyfitz (1968)). As we see, the impact on the whole population depends on the ages at which mortality decreases. Different causes impact different ages and thus different populations will result when some causes are reduced or removed.

Remark 1. *Note that external causes impact mainly men, whereas our model assumptions result in a number of births depending only on the female population. In case of removal of external causes, the number of births will increase due to the part of the female population that do not die from external causes, but this could be*

reinforced if a birth rate would have been defined for males.

In Figure 9, the time evolution of the population size under these three scenarios is plotted. It is interesting to note that, even if the age composition becomes stable over time, the population size is still decreasing. This is not consistent with real demographic insights but this can be explained by the fact the migration flows are not taken into account in these simulations. This is tested and discussed in the following subsection. As expected, the population size is higher under cause removal scenarios. In the short term, cancer reduction leads to a bigger population since more individuals survive. However, in the medium-long term, the removal of external causes leads to a higher number of individuals due to the additional effect of the increasing number of births. On the whole, the size evolution depends on the cause which is reduced.

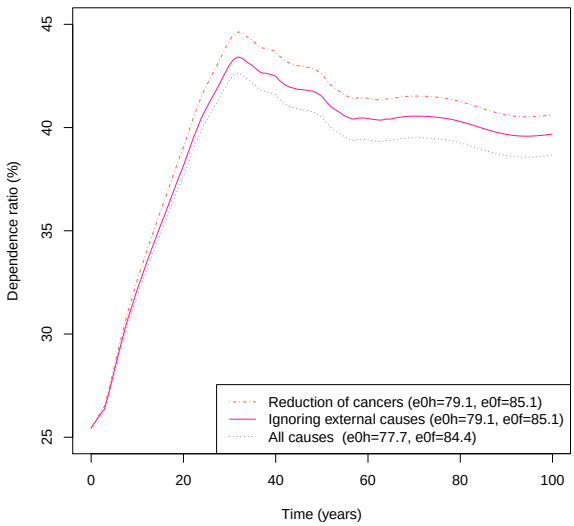


Figure 8: Age dependency ratio over 100 years from the limit equation

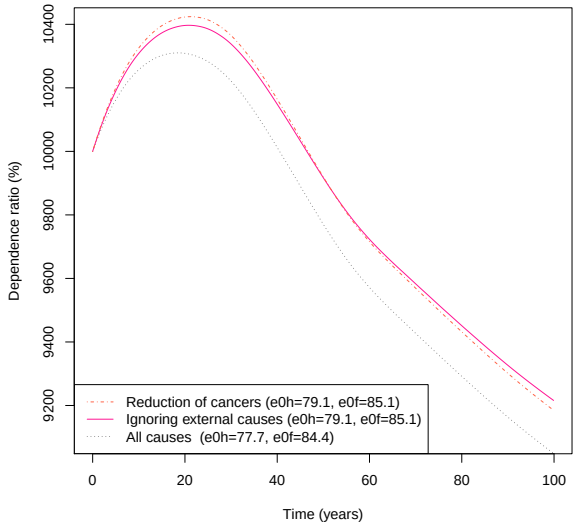


Figure 9: Population size for 100 years from the limit equation

Stochastic scenarios Stochastic scenarios are interesting to get additional insights on the random pattern of the dependency ratio and the magnitude of the demographic noise linked to the finite size of the population. Scenario (ii) of external causes of death removal is depicted in Figure 10, whereas scenario (iii) (cancer reduction) is added in Figure 11. All stochastic scenarios are computed starting with an initial population with size 10 000. For such size, to distinguish both scenarios is a hard task, making previous conclusions less striking with small samples. However, we argue that real national population sizes are more in accordance with "large population" patterns (see Figure 4 for an initial population size of 100 000). Unfortunately, testing it numerically would required computer resources that we do not have, and is beyond the scope of the present paper.

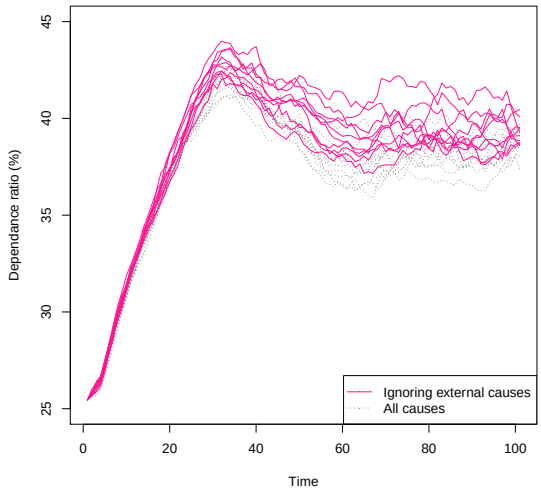


Figure 10: Age dependency ratio from 2008 to 2108: External causes removal (10 scenarios)

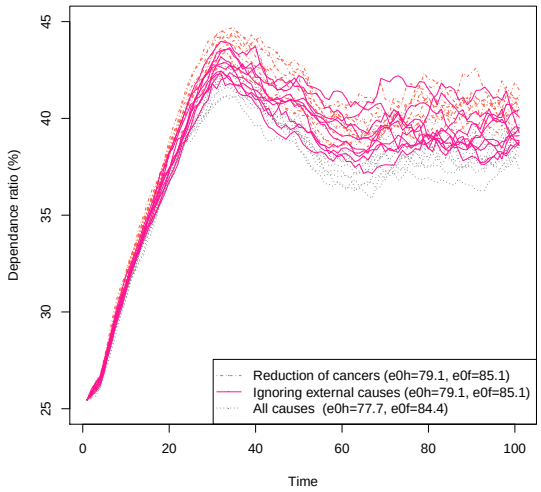


Figure 11: Age dependency ratio from 2008 to 2108: reducing cancers and external causes removal (10 scenarios)

4.4 Sensitivity to Population Flows and Fertility

We address here the sensitivity of our results to several assumptions on migration flows and fertility rates. We emphasize that population flows are computed here in a first basic approach, therefore we keep in mind that what we call "migration" is not measured properly in terms of pure demographic analysis. However, our approach is made so as to replicate the data and allows us to observe several interesting patterns. We first detail how the migration rates are computed and then present nine scenarios with several fertility and migration assumptions. On the whole, our results persist with the inclusion of migration at several levels: the dependency ratios differ between the two scenarios (cancer reduction and external causes removal) even in the short term, making our findings stronger when population flow effects are taken into account.

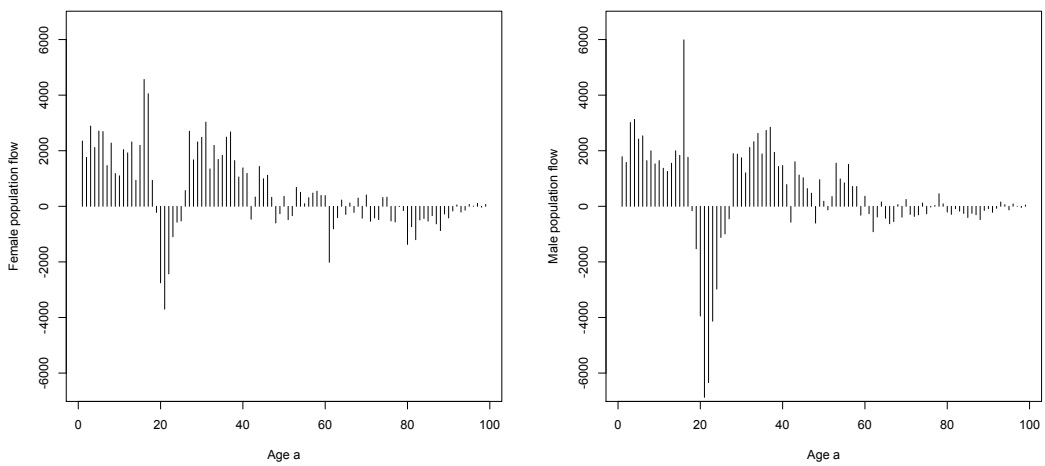
Population flow estimation As for the whole study, we focus on year 2008 for the computation of migration indicators. Migration flows are computed as the difference between the population at the end and the beginning of 2008, augmented by the number of deaths during this year. We approximate the population at the end of the year 2008 by the population at the beginning of year 2009. Let us denote $M^\epsilon(a)$ the population flow per gender ϵ and age a in year 2008. It is computed as follows:

$$M^\epsilon(a) = N^\epsilon_{2009}(a+1) - N^\epsilon_{2008}(a) + D^\epsilon(a),$$

where $N^\epsilon_t(a)$ is the number of individuals with gender ϵ and age a at the beginning of year t, and $D^\epsilon(a)$ is the number of deaths with gender ϵ and age a during the year

2008. Since data on deaths are available until the one-year age-class 98 [1], we perform this computation for ages below 98 and assume that ages greater than 99 do not participate into population flows. This is a reasonable assumption since at old ages, above 95, population flows are negligible. Estimated population flows $M^\epsilon(a)$ by age class can be positive or negative. These are depicted in Figure 12. Let us denote $M^\epsilon_+(a)$ (resp. $M^\epsilon_-(a)$) the population flow if $M^\epsilon(a)$ is positive (resp. negative), so that $M^\epsilon(a) = M^\epsilon_+(a)\mathbf{1}_{M^\epsilon(a)\geqslant 0} - M^\epsilon_-(a)\mathbf{1}_{M^\epsilon(a)\leqslant 0}$. This way, for each age-class, we do not capture the real number of emigrants or immigrants but only the surplus or deficit. By abuse of definition, we will call $M^\epsilon_+(a)$ (resp. $M^\epsilon_-(a)$) the number of immigrants (resp. emigrants). For the entire population in year 2008, population flows implied a population surplus of 97369 individuals (males and females).

Figure 12: Estimated population flows in 2008 for females (left) and males (right)



For the use of the population dynamics model, it remains to transform the population numbers $M^\epsilon_+(a)$ and $M^\epsilon_-(a)$ into intensities. Let us first consider emigration: we assume that the emigrant number $M^\epsilon_-(a)$ is the result of the sum of emigration intensities of each individual in the population. That is, the individual emigration rate $m^\epsilon(a)$ is estimated as $\hat{m}^\epsilon(a) = \frac{M^\epsilon_-(a)}{ER^\epsilon(a)}$, where $ER^\epsilon(a)$ is the exposure to risk at age a for individuals with gender ϵ in the 2008 population. As for immigration, we want to capture a total intensity, say $I^\epsilon_{nat}(a)$, of arrival of an individual with age a and gender ϵ in the reference national 2008 population. Since the intensity $I^\epsilon_{nat}(a)$ represents the average number of arrivals of individuals in age-class a over one year, an estimator for this quantity is nothing but the number of immigrants in the year 2008, that is $\hat{I}^\epsilon_{nat}(a) = M^\epsilon_+(a)$. However, it is natural to think that the immigration flows depend on the size of the population at destination. Therefore we assume a linear link with the initial population, so that if we start with an initial population

[1] INSEE: $http://www.insee.fr/fr/themes/detail.asp?ref_id = ir - irsocsd20133$

with size N^ϵ for gender ϵ, we use the immigration rate $\hat{I}^\epsilon(a) = \frac{N^\epsilon}{N^\epsilon_{\text{nat}}}\hat{I}^\epsilon_{nat}(a)$, where N^ϵ_{nat} is the size of the 2008 national population. Note here that the total immigration rate is adjusted once at the beginning of the simulation and does not vary with time: this is consistent to capture the proper order of magnitude but also not to make it depend on the year to year variations of the population size, since in practice there is no obvious demographic reason to do so.

Large population equation The sensitivities will be studied with the large population equation, including population flows. This is written as

$$(\partial_a + \partial_t)\mathbf{g}(a,t) = -\tilde{Q}(a)\mathbf{g}(a,t) + \begin{pmatrix} I^f(a) \\ I^m(a) \end{pmatrix},$$

$$\mathbf{g}(0,t) = \left(\int_{\mathbb{R}_+} g(f,a,t)b(a)\,\mathrm{d}a\right)\begin{pmatrix} p \\ 1-p \end{pmatrix},$$

where $\tilde{Q}(a) = \begin{pmatrix} \mu^f(a) + m^f(a) & 0 \\ 0 & \mu^m(a) + m^m(a) \end{pmatrix}$.

In this setting, the aging population is decreased by the number of deaths and the number of emigrants, and is increased by the number of immigrants. The dynamics of the births is the same as the case without population flows: the number of newborn is computed as the sum of all newborn of all females, weighted by gender probability at birth. The numerical scheme is a straightforward adaptation of that described in Section 2.2.

Results We present nine deterministic scenarios with several fertility and migration assumptions. The scenario with current fertility and migration is displayed in Figure 13e. In this reference scenario, the age dependency ratio stabilizes around 26% in the mid-term, and then increases to 30% in the long term. Note that previously, without population flows, the dependency ratio was around 40% in the mid and long-term (see Figure 8). In Figure 13e, the two other scenarios (ii) and (iii) are also depicted. Interestingly, cause reduction scenarios exhibit different dependency ratios levels in the short term, therefore making our findings more relevant when population flows replicating the data are taken into account. The same conclusion holds for the other graphs, which present variations of fertility and migration rates (variation of 25% of the current rates). Fertility scenarios can be read from left to right: 75% , 100% and 125% of the current fertility rate, whereas the migration scenarios are displayed from top to bottom: 75%, 100% and 125% of the current migration rates. As expected, the higher are the fertility and migration, the lower is the dependency ratio (see figure 13i). Moreover, we note a sensitivity more important on fertility (left to right) compared to migration (top to bottom). On the

whole, these results including population flows seems to be more in accordance with real demographic insights.

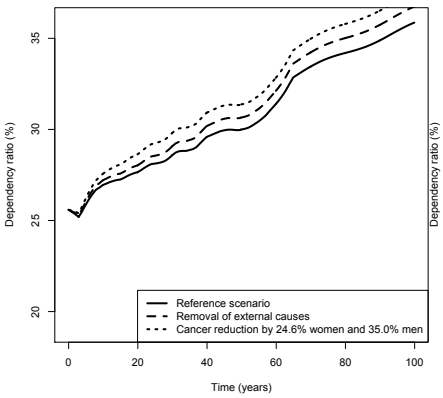
5 Concluding Remarks

In this paper, we studied the impact of a decrease in cause-of-death mortality on the whole population dynamics, including birth patterns and population flows. The study of the whole population structure gives additional information compared to previous studies on causes of death focusing only on mortality indicators. We based our study on a stochastic individual based model which is micro-macro consistent. This allowed us to compute stochastic and deterministic scenarios, and to study both the demographic noise in the scenarios and the stable average age pyramid. Interestingly, as causes of death impact different age ranges, the impact on the age pyramid can be different. Numerical simulations based on French data illustrate that even if life expectancy improves in the same way, the reduction of some causes of death can lead to populations that are older (cancer) or younger (external causes) than expected. One can think of many applications of our model, including the study of pension systems. Indeed, the evolution of the age dependency ratio impacts directly pay-as-you-go pension systems, whose stability is often monitored through the age dependency ratio. This model can also help to study impacts on the population of public awareness campaigns on cancers or road safety for example. Our modeling framework is rather general and flexible, and it is possible at this step to further investigate the impact of several assumptions regarding cause-of-death mortality, fertility and migration. In particular, further improvements of our study could be to take into account different dependence assumptions between causes of death at the individual level.

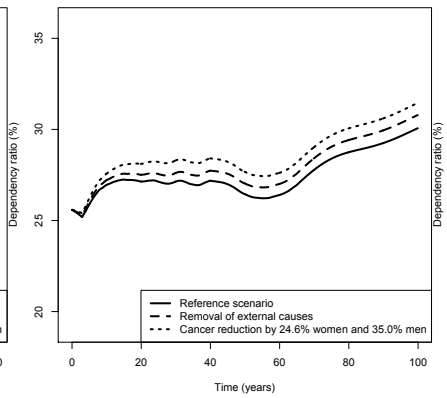
Acknowledgements

The authors thank Frédérick Vautrain, founder of Isthma, for the free use of the tool Deltametric that helped to analyse the data. The authors benefited from the financial support of the "Germaine de Staël" program *Modélisation des causes de décès dans la dynamique des populations* (Gds 2014-14). Alexandre Boumezoued and Nicole El Karoui also benefited from the financial support by ANR project *Dynamic models for longevity with lifestyle adjustments*, Lolita (ANR-13-BS01-0011).

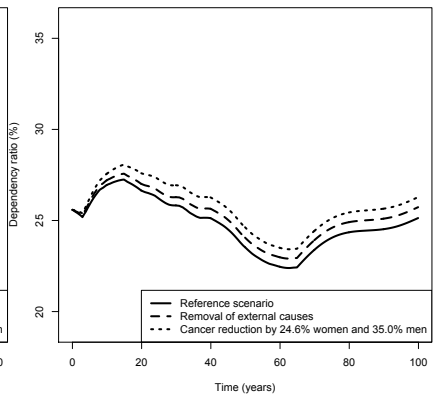
Figure 13: Age dependency ratio from 2008 to 2108 with variation of fertility and migration



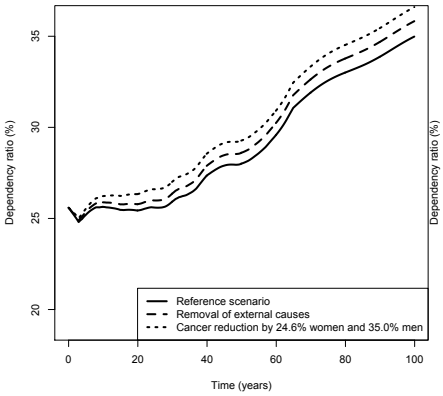
(a) Fertility 75%, Migration 75%



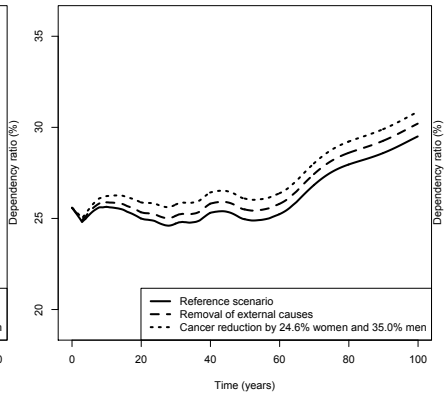
(b) Fertility 100%, Migration 75%



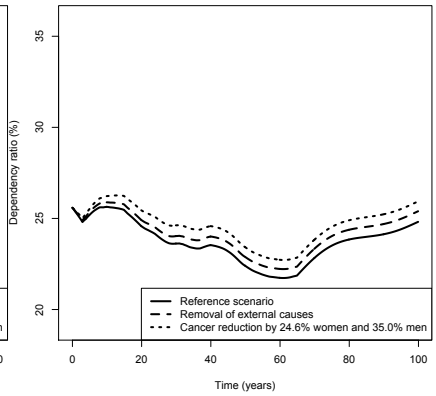
(c) Fertility 125%, Migration 75%



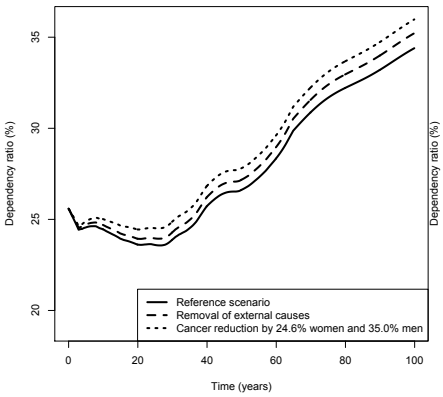
(d) Fertility 75%, Migration 100%



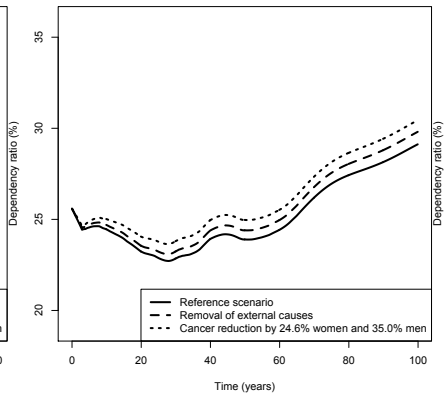
(e) Fertility 100%, Migration 100%



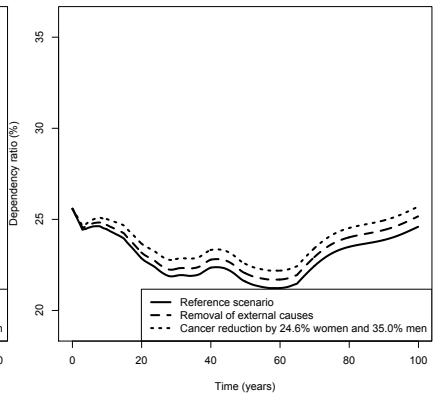
(f) Fertility 125%, Migration 100%



(g) Fertility 75%, Migration 125%



(h) Fertility 100%, Migration 125%



(i) Fertility 125%, Migration 125%

References

Alai, D. H., S. Arnold(-Gaille), M. Sherris. 2015. Modelling Cause-of-Death Mortality and the Impact of Cause-Elimination. *Annals of Actuarial Science, 9(1), 167-186* . 2, 6, 14

Arnold, S., M. Sherris. 2013. Forecasting mortality trends allowing for cause-of-death mortality dependence. *North American Actuarial Journal, 17(4), 273-282* . 6

Arnold, S., M. Sherris. 2015a. Causes-of-death mortality: What do we know on their dependence? *North American Actuarial Journal, in press* . 6, 14

Arnold, S., M. Sherris. 2015b. International cause-specific mortality rates: New insights from a cointegration analysis. *Working paper, HEC Lausanne* . 6

Bensusan, H. 2010. Risques de taux et de longévité: Modélisation dynamique et applications aux produits dérivés et à l'assurance vie. Ph.D. thesis, École Polytechnique. 3

Bensusan, H., A. Boumezoued, N. El Karoui, S. Loisel. 2010–2015. Impact of heterogeneity in human population dynamics. *working paper* . 3, 9, 11, 12

Bonnet, C., C. Burricand, C. Colin, A. Flipo, P.R. Mahieu, B. Sédillot. 1999. Le modèle de microsimulation dynamique: Destinie. *Document de travail INSEE* . 4

Cairns, A. J., D. Blake, K. Dowd, A. Kessler. 2014. Phantoms never die: Living with unreliable mortality data. Tech. rep., Herriot Watt University, Edinburgh. 8

Carriere, J. F. 1994. Dependent decrement theory. *Transactions of the Society of Actuaries* **46** 45–74. 2, 6

Champagnat, N., R. Ferrière, S. Méléard, et al. 2006. Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. *Theoretical population biology* **69**(3) 297–321. 3

Chiang, C. L. 1968. Introduction to stochastic processes in biostatistics. *John Wiley and Sons, New York* . 2, 6

Delwarde, A., M. Denuit. 2006. *Construction de tables de mortalité périodiques et prospectives*. Economica. 14

Dimitrova, D. S., S. Haberman, V. K. Kaishev. 2013. Dependent competing risks: Cause elimination and its impact on survival. *Insurance: Mathematics and Economics* . 2, 6, 7

Duée, M. 2005. La modélisation des comportements démographiques dans le modèle de microsimulation Destinie. *Document de travail INSEE/DESE G* . 4

Elandt-Johnson, R. C. 1976. Conditional failure time distributions under competing risk theory with dependent failure times and proportional hazard rates. *Scandinavian Actuarial Journal* (1) 37–51. 5, 7

Ferriere, R., V. C. Tran. 2009. Stochastic and deterministic models for age-structured populations with genetically variable traits. *ESAIM: Proceedings*, vol. 27. 289–310. 3

Fournier, N., S. Méléard. 2004. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Annals of applied probability* **14**(4) 1880–1919. 3, 11

Girosi, F., G. King. 2006. Demographic forecasting. *Cambridge University Press, Cambridge* . 6

Guilmoto, C. Z. 2012. Skewed sex ratios at birth and future marriage squeeze in China and India, 2005–2100. *Demography* **49**(1) 77–100. 8

Hautphenne, S., G. Latouche. 2012. The markovian binary tree applied to demography. *Journal of mathematical biology* **64**(7) 1109–1135. 4

HMD. 2014. The human mortality database: French data. URL `http://www.mortality.org/cgi-bin/hmd/country.php?cntr=FRA&level=2`.

Hyndman, R. J., H. Booth. 2008. Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting* **24**(3) 323–342. 4

Iannelli, M., M. Martcheva, F. A. Milner. 2005. *Gender-structured population modeling: mathematical methods, numerics, and simulations*, vol. 31. Siam. 3, 8, 9, 30

INSEE. 2011. Sources et méthodes: Les indicateurs démographiques. URL `http://www.insee.fr/fr/methodes/sources/pdf/Indicateurs_demographiques.pdf`. 15

INSEE. 2014. Fécondité selon l'âge détaillé de la mère. URL `http://www.insee.fr/fr/themes/tableau.asp?reg_id=0&ref_id=bilandemo2`.

Kaishev, V. K., D. S. Dimitrova, S. Haberman. 2007. Modelling the joint distribution of competing risks survival times using copula functions. *Insurance: Mathematics and Economics* . 2, 6

Keyfitz, N. 1968. Changing vital rates and age distributions. *Population Studies* **22**(2) 235–251. 21

Keyfitz, N. 1972. The mathematics of sex and marriage. *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, vol. 4. University of California Press Berkeley, 89–108. 3, 8

Li, J., C. O'Donoghue. 2013. A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation* **6**(2) 3–55. 4

Manton, K. G. 1986a. Past and future life expectancy increases at later ages: Their implications for the linkage of morbidity, disability, and mortality. *Journal of Gerontology, 41(5), 672-681* . 6

Manton, K. G., E. Stallard, J. W. Vaupel. 1986b. Alternative models for the heterogeneity of mortality risks among the aged. *Journal of the American Statistical Association, 81(395), 635- 644* . 6

McKendrick, A.G. 1926. Application of mathematics to medical problems. *Proc. Edin. Math. Soc.* **54** 98–130. 3, 9

Prentice, R. L., J. D. Kalbfleisch, A. V. Peterson Jr, N. Flournoy, VT. Farewell, NE. Bres-

low. 1978. The analysis of failure times in the presence of competing risks. *Biometrics* 541–554. 2, 6

Putter, H, M Fiocco, RB Geskus. 2007. Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine* **26**(11) 2389–2430. 2, 6

Rosen, M. 2006. Forecasting life expectancy and mortality in sweden, some comments on methodological problems and potential approaches. technical report 4. Tech. rep., Social Insurance Studies from the Swedish Social Insurance. 6

Schoen, R. 1983. Measuring the tightness of a marriage squeeze. *Demography* **20**(1) 61–78. 8

Tran, V. C. 2006. Modèles particulaires stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques. Ph.D. thesis, Université de Nanterre-Paris X. 12

Tran, V. C. 2008. Large population limit and time behaviour of a stochastic particle model describing an age-structured population. *ESAIM: Probability and Statistics* **12**(1) 345–386. 9

Tsai, S. P., E. S. Lee, R. J. Hardy. 1978. The effect of a reduction in leading causes of death: Potential gains in life expectancy. *American Journal of Public Health* . 2, 6

Tsiatis, A. 1975. A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* . 6

Vaupel, J. W., A. I. Yashin. 1983. The deviant dynamics of death in heterogeneous populations. technical report rr-83-001. Tech. rep., International Institute for Applied Systems Analysis (IIASA). 6

Von Foerster, H. 1959. *The Kinetics of Cellular Proliferation*. Grune & Stratton. 3, 9

WHO. 2014. The world health organization: the who mortality database. URL `http://www.who.int/healthinfo/statistics/mortality_rawdata/en/`.

Wilmoth, J. R. 1995. Are mortality projections always more pessimistic when disaggregated by cause of death? *Mathematical Population Studies* . 2, 6

Appendix

We present the asymptotic behavior of the deterministic age pyramid (see Iannelli et al. (2005)). Let us first focus on the dynamics of the number of females derived from Equations (2.4)-(2.5):

$$
\begin{aligned}
&(\partial_a g + \partial_t g)(f,a,t) = -\mu^f(a)\, g(f,a,t) \\
&g(f,0,t) = p \int_{\mathbb{R}_+} g(f,a,t) b(a)\, \mathrm{d}a \\
&g(f,a,0) = g_0(f,a).
\end{aligned}
$$

In the deterministic model, the number of children per female is

$$R_0^f = \int_{\mathbb{R}_+} b(a) e^{-\int_0^a \mu^f(s)\mathrm{d}s} \mathrm{d}a.$$

Then, the mean number of children per individual is given by $R_0 = pR_0^f$. With our data, in the reference scenario, we find $R_0^f = 1.97$ and $R_0 = 0.964$. Note that the first value R_0^f is of the same order of magnitude as the fertility statistics for France. Note also that we are in the case $R_0 < 1$ so that the average population size decreases after some time and goes to zero as time goes to infinity. In Figure 9, one can see that the speed of decrease for the population size depends on the cause of death which is under reduction. But up to some scaling, one can be more precise about the age pyramid long term behavior. Let us introduce the characteristic equation in λ,

$$p \int_{\mathbb{R}_+} b(a) e^{-\lambda a - \int_0^a \mu^f(s)\mathrm{d}s} \mathrm{d}a = 1.$$

The unique real solution λ^* (say) is called the population intrinsic growth rate. Since we have $R_0 < 1$, then $\lambda^* < 0$. Numerically, we find $\lambda^* = -0.00123$ (reference scenario) and we are in the sub-critical regime. It can be interpreted as follows: this is the constant to be added to the death rate to get exactly one children per individual, in other words to make the dynamics critical. One can now focus on the long term behavior of $e^{-\lambda^* t} g(f, a, t)$: this is known as the Sharpe-Lotka theorem. It states that the long-term age pyramid density for the female population can be written, up to a renormalizing constant, as

$$\phi(f, a) = e^{-\lambda^* a - \int_0^a \mu^f(s)\mathrm{d}s}.$$

Now, from Equations (2.4)-(2.5), it can be shown that $e^{-\lambda^* t} g(m, a, t)$ converges up to a normalizing constant to $(1-p)\int_{\mathbb{R}_+} \phi(f, a) b(a) \mathrm{d}a$, and since $g(m, a, t+a) = g(m, 0, t) e^{-\int_0^a \mu^m(s)\mathrm{d}s}$, we formally get

$$e^{-\lambda^*(t+a)} g(m, a, t+a) = e^{-\lambda^* t} g(m, 0, t) e^{-\lambda^* a - \int_0^a \mu^m(s)\mathrm{d}s},$$

which converges as $t \to +\infty$ up to a constant to $e^{-\lambda^* a - \int_0^a \mu^m(s)\mathrm{d}s}$. In other words, the age pyramid density for males is given up to a constant by

$$\phi(m, a) = e^{-\lambda^* a - \int_0^a \mu^m(s)\mathrm{d}s}.$$

So the age pyramid for males and female is governed both by mortality and natality, since the solution λ^* depends on these two. To conclude, the overall average age pyramid density is given up to a renormalizing constant by

$$p\phi(m, a) + (1-p)\phi(f, a).$$

The comparison between the average age pyramid at time 300 years and the theoretical age pyramid is given in Figure 14.

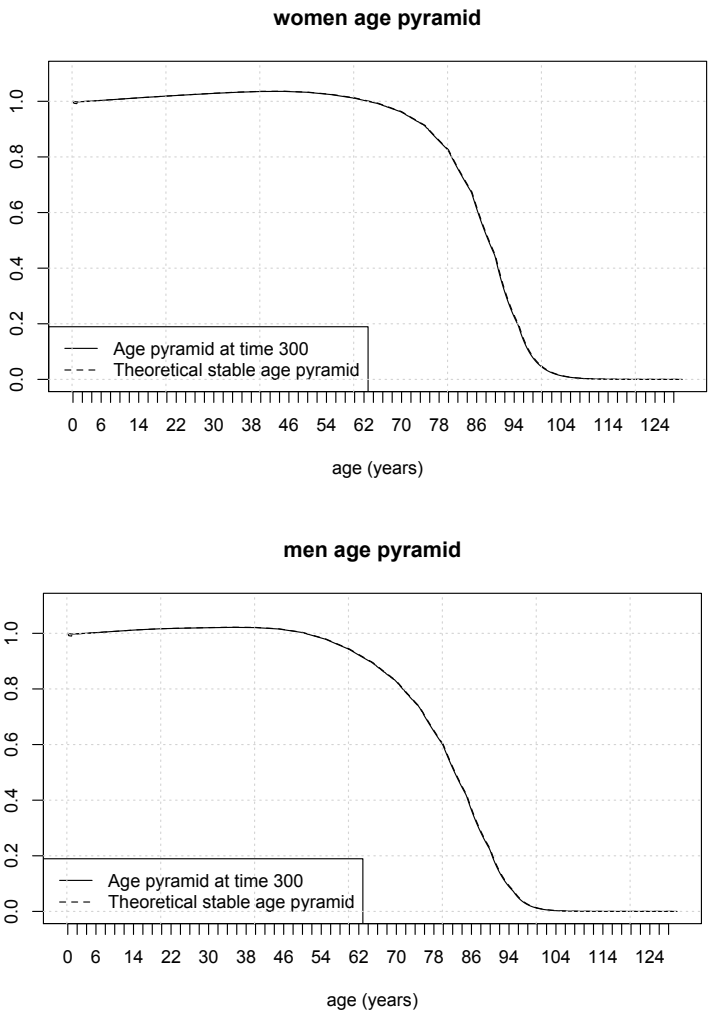



Figure 14: Comparison between the theoretical age pyramid and the age repartition at time 300 for females and males.