Nouveaux paradigmes en dynamique de populations hétérogènes : modélisation trajectorielle, agrégation, et données empiriques

Sarah Kaakai

▶ **To cite this version:**

Sarah Kaakai. Nouveaux paradigmes en dynamique de populations hétérogènes : modélisation trajectorielle, agrégation, et données empiriques. Probability [math.PR]. Université Pierre et Marie Curie - Paris VI, 2017. English. NNT : 2017PA066553 . tel-01897474

HAL Id: tel-01897474

https://theses.hal.science/tel-01897474v1

Submitted on 17 Oct 2018

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.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PIERRE ET MARIE CURIE

Spécialité : **Mathématiques Appliquées**

École doctorale de Sciences Mathématiques de Paris-Centre

Présentée par

Sarah KAAKAI

Nouveaux paradigmes en dynamique de populations hétérogènes :

Modélisation trajectorielle, Agrégation, et Données empiriques

New paradigms in heterogeneous population dynamics:

Pathwise modeling, aggregation, and empirical evidence

Résumé

Cette thèse porte sur la modélisation probabiliste de l'hétérogénéité des populations humaines et de son impact sur la longévité. Depuis quelques années, de nombreuses études montrent une augmentation alarmante des inégalités de mortalité géographiques et socioéconomiques. Ce changement de paradigme pose des problèmes que les modèles démographiques traditionnels ne peuvent résoudre, et dont la formalisation exige une observation fine des données dans un contexte pluridisciplinaire. Avec comme fil conducteur les modèles de dynamique de population, cette thèse propose d'illustrer cette complexité selon différents points de vue: Le premier propose de montrer le lien entre hétérogénéité et non-linéarité en présence de changements de composition de la population. Le processus appelé Birth Death Swap est défini par une équation dirigée par une mesure de Poisson à l'aide d'un résultat de comparaison trajectoriel. Quand les swaps sont plus rapides que les évènements démographiques, un résultat de moyennisation est établi par convergence stable et comparaison. En particulier, la population agrégée tend vers une dynamique non-linéaire. Nous étudions ensuite empiriquement l'impact de l'hétérogénéité sur la mortalité agrégée, en s'appuyant sur des données de population anglaise structurée par âge et circonstances socioéconomiques. Nous montrons par des simulations numériques comment l'hétérogénéité peut compenser la réduction d'une cause de mortalité. Le dernier point de vue est une revue interdisciplinaire sur les déterminants de la longévité, accompagnée d'une réflexion sur l'évolution des outils pour l'analyser et des nouveaux enjeux de modélisation face à ce changement de paradigme.

Abstract

This thesis deals with the probabilistic modeling of heterogeneity in human populations and of its impact on longevity. Over the past few years, numerous studies have shown a significant increase in geographical and socioeconomic inequalities in mortality. New issues have emerged from this paradigm shift that traditional demographic models are not able solve, and whose formalization requires a careful analysis of the data, in a multidisciplinary environment. Using the framework of population dynamics, this thesis aims at illustrating this complexity according to different points of view: We explore the link between heterogeneity and non-linearity in the presence of composition changes in the population, from a mathematical modeling viewpoint. The population dynamics, called Birth Death Swap, is built as the solution of a stochastic equation driven by a Poisson measure, using a more general pathwise comparison result. When swaps occur at a faster rate than demographic events, an averaging result is obtained by stable convergence and comparison. In particular, the aggregated population converges towards a nonlinear dynamic. In the second part, the impact of heterogeneity on aggregate mortality is studied from an empirical viewpoint, using English population data structured by age and socioeconomic circumstances. Based on numerical simulations, we show how a cause of death reduction could be compensated in presence of heterogeneity. The last point of view is an interdisciplinary survey on the determinants of longevity, accompanied by an analysis on the evolution of tools to analyze it and on new modeling issues in the face of this paradigm shift.

A mes parents Hafid et Sylvie,
A Céline,

No man is an island,
Entire of itself;
Every man is a piece of the continent,
A part of the main.

John Donne

Table of contents

List of figures

List of figures

Chapter 1

Introduction

In over two centuries, the world population has been transformed dramatically, under the effect of considerable changes induced by demographic, economic, technological, medical, epidemiological, political and social revolutions. The age pyramids of ageing developed countries look like "colossus with feet of clay", and the complexity of involved phenomena make the projection of future developments very difficult, especially since these transitions are unprecedented.

The problem does not lie so much in the lack of data or empirical studies. For several years now, a considerable amount of data have been collected at different levels. A number of international organizations[1] have their own open databases, and national statistical institutes[2] have been releasing more and more data. On top of that, more than fifty public reports are produced each year. The private sector is also very active on these issues, especially pension funds and insurance companies which are strongly exposed to the increase in life expectancy at older ages.

However, the past few years have been marked by a renewed demand for more efficient models. This demand has been motivated by observations of recent demographic trends which seem to be in contradiction with some firmly established ideas. New available data seem to indicate a paradigm shift over the past decades, toward a more complex and individualized world. Countries which had similar mortality experiences until the 1980s now diverge, and a widening of health and mortality gaps inside countries has been reported by a large number of studies. These new trends have been declared as key public issues by several organizations, including the WHO in its latest World report on ageing

[1] For instance, the United Nations (UN) (unstats.un.org), the World Bank (WB) (data.worldbank.org) or the World Health Organisation (WHO) (who.int/gho).

[2] such as the Institut National de la Statistique et des Etudes Economiques (INSEE) and the Institut National d'Etude Demographique (INED) in France, or the Office for National Statistics (ONS) in the United Kingdom.

and health (World Health Organization (2015)), and the National Institute on ageing in the United States, which created in 2008 a panel on Understanding Divergent Trends in Longevity in High-Income Countries, leading to the publication of a comprehensive report in National Research Council and Committee on Population (2011). There is thus an important need for finer grained models capable of integrating the population heterogeneity, interactions at several scales or variability of the environment.
In the first part of the introduction of this thesis, we first give a general background on the evolution of human longevity. In the second part, we give a more detailed description of the issues which motivated this thesis, followed by an outline of the different approaches used in this thesis. A summary of the results of the thesis is given in the last part of the introduction.

1.1 General background

As stated by Vaupel and coauthors in their famous 2009 article, "*The remarkable gain of about 30 years in life expectancy in western Europe, the USA, Canada, Australia, and New Zealand—and even larger gains in Japan and some western European countries, such as Spain and Italy—stands out as one of the most important accomplishments of the 20th century* " (Christensen et al. (2009)). For more than 150 years, record female life expectancy has grown at the pace of almost 2.5 years per decade (Oeppen and Vaupel (2002)). In France, the investigations of Louis Henry in the field of Historical Demography have allowed the construction of historical mortality tables going back to the mid-eigteenth century, thanks to the remarkable work of Vallin and Meslé (2001). Based on their estimations, life expectancy at birth was estimated to be about be 25 years of age in the mid-eighteenth century. By 1830, it had attained 40 years and remained stable until the 1870s. Less than a century later, in 1950, life expectancy had increased of more than 30 years, to 69 years for females and 63 for males. Today, life expectancy in France is estimated to have reached the age of 85 for females, and 79 for males[3]. Obviously, these dramatic changes have been accompanied by major societal and economic changes, and the future consequences of such levels of longevity are the source of numerous and urgent debates, both in the research community and in the civil society.

The demographic transition The sustained improvement of the duration of life, together with fertility decline, is part of the larger process of the demographic transition,

[3]Source: Cambois et al. (2009), Institut National de la Statistique et des Etudes Economiques (INSEE) (https://www.insee.fr/fr/statistiques/2416631).

which has been ongoing in most developed country since the mid-eighteenth century. These major demographic changes have been both the source and the consequence of massive social and economic upheavals which include the industrial revolution, rampant urbanization, the increase of living standards and educational levels, together with greater social and political equality, especially between men and women with the entry of women in the labor force. The demography transition was "*the most important social and economic change to take place in Europe in centuries*" (Reher (2011)). In countries which experienced the historic transition[4], important public health and medical advances have significantly contributed to mortality declines during the twentieth century[5] (Cutler et al. (2006)). At the same time, the reduction of mortality at younger ages, which took place before the beginning of fertility declines, led to a spectacular growth of the population and the increase of the proportion of individuals of working-age until some time between the late 1950s and the early 1980s. Consequently, these economically advantageous age-structures have contributed to create favorable conditions for the development of national pension schemes, in which the large working-age population could pay pension benefits for a small group of elderly people (Reher (2011)). In turn, the improvement of social welfare has probably contributed to further mortality declines.

New challenges brought by the demographic transition But today, the impact of longevity improvements is also producing new issues and challenges at multiple societal levels. Population ageing poses a major challenge to the sustainability of intergenerational risk sharing mechanisms such as pay-as-you-go systems and public health systems. In the recent years, pension funds and governments have been closing Defined Benefit retirement plans or shifting them toward a Defined Contribution system[6]. This shows an indicator of a transfer of the demographic risk back to the policyholders (Barrieu et al. (2012)). As also stated in Barrieu et al. (2012), the insurance industry is facing challenges linked to increasing longevity, in the form of greater regulatory capital and the need to transfer part of their risk to reinsurers and financial markets. The increase in longevity is also generating a whole new paradigm in the way the life course is perceived, and the timescale of many life course strategies have deeply changed. The traditional three broad periods of life (childhood, adulthood and old age) have been replaced by a four period model, with the third age period being divided in two stages, "young old" and "oldest old", the

[4]The historic transition affected most of European countries and countries with European roots (Argentina, Uruguay, the United States, Canada, New Zealand (Reher (2011)).

[5]We will return to the demographic transition and the importance of public health in Chapter 5.

[6] A Defined Benefit retirement plan means that pension benefits are predetermined, based on paramaters such as the level of income. In a Defined Contribution plan, contributions rates are predetermined, but future pension benefits are not known.

latter being characterized by reduced autonomy and greater medical needs (Christensen et al. (2009)). The notion of age itself has changed. At a fixed age, individuals seem to have rejuvenated: individuals now "become old at older ages" than before (d'Albis and Collard (2013)). As retirement age has been raised in several countries, there is also a need to redefine and redistribute work within these populations which are simultaneously ageing and rejuvenating, as argued in Vaupel and Loichinger (2006).

An heterogeneous evolution of longevity If the increase in record life expectancy has been rather stable, a more detailed analysis of underlying processes is by no means simple. Historically, the decline in mortality was not uniform in age, and while life expectancy at birth increased by 30 years in England between 1841 and 1950, life expectancy at age 10 increased of only 15 years and life expectancy at 65 remained virtually constant until 1950. Studies on socioeconomic inequalities in mortality, which is one of the main topics of this thesis, can be traced back to the early eighteenth century. For instance, Villermé (1830) was one of the first people to exhibit a link between mortality rates and socioeconomic status[7]. More recently a growing amount of evidence seems to indicate that along with the increase of life expectancy, developed countries have also experienced a widening of socioeconomic gaps in health and mortality since the second half of the twentieth century (Cutler et al. (2006); Elo (2009); National Research Council and Committee on Population (2011); Olshansky et al. (2012)). For instance, Olshansky et al. (2012) in the United States have found out that the life expectancy gap (at birth) between females with less than 12 years of education and females with more than 16 years of education grew from 7.7 years in 1990 to 10.3 years in 2008 (13.4 to 14.2 for males). Life expectancy of less educated white females and males even decreased during this 18-year period.

The pervasive effect of socioeconomic factors on health and mortality is also at the source of economic and social issues, and has become an important public problem for many countries and international organizations. For instance the World Health Organization (WHO) named the reducing of inequities one of the key issues for public health action in their last report on ageing and health (World Health Organization (2015)). Beside reducing the fundamental inequity in the correlation of the duration of life with one's income or social status, they recommended to target policies overcoming these inequities, in the sense that "*strategies must look not just improve conditions for the best-off or the average older person. Attention must also be given to [..] narrowing the gaps in*

[7]For this purpose, he studied deaths in Paris per borough and did what is now called a data linkage with tax authorities statistics, by comparing mortality rates to rates of non-taxable households per borough (see Mireaux (1962) for more details).

the total inequalities observed among older individuals". To that matter, understanding the social determinants at the root of these inequalities is critical for designing efficient policies. However, underlying factors responsible for socioeconomic differences in health and mortality are still not clearly understood. As a consequence, policy recommendations can differ substantially, according to the theory taken into account in order to model pathways involved into translating socioeconomic status into mortality outcomes. These issues are discussed in more detail in Chapter 5.
Socioeconomic differences in mortality can also impact the equity of national pension schemes by allocating more resources to individuals higher in the socioeconomic ladder, already benefiting from better health and longevity (Villegas and Haberman (2014)). Furthermore, not taking heterogeneity into account might lead to significant errors when assessing pensions liabilities or regulatory capital. A striking and simple example is the evolution of the socioeconomic composition of cohorts of seniors. As individuals in the same cohort grow older, low-income pensioners die at earlier ages, so that average benefits increase and costs rise (Edwards and Tuljapurkar (2005)). This phenomenon is typically nonlinear, and is one of many issues generated by the interactions between the evolution of an heterogeneous population and aggregate indicators, to which we will come back to in Chapter 4.

Two centuries of interdisciplinary literature All these challenges have generated a considerable amount of research, producing an interdisciplinary literature in fields ranging from mathematics to history, and also including actuarial science, biodemography, biology, computer science, demography, economics, epidemiology, medical research, public health or sociology[8]. In 1825, B. Gompertz presented to the Royal Society of London his " law of human mortality", describing age specific mortality rates as an exponential function of age, with only two parameters (Gompertz (1825), see also Kirkwood (2015) for a commentary on Gompertz's original article). Gompertz, who was an English actuary, was actually interested in improving the calculation of rates for the selling and purchasing of annuities. Almost two centuries later, his law of mortality is still regarded as a kind of "fundamental law of mortality", and seems to hold for a wide range of species. On the other hand, a justification for the existence of this law has not been widely established yet. This illustrates how little we still know on the evolution of human longevity and its societal impacts, as many questions are still open and the source of heated debates. In the new context of open data, population data have also been increasingly released by

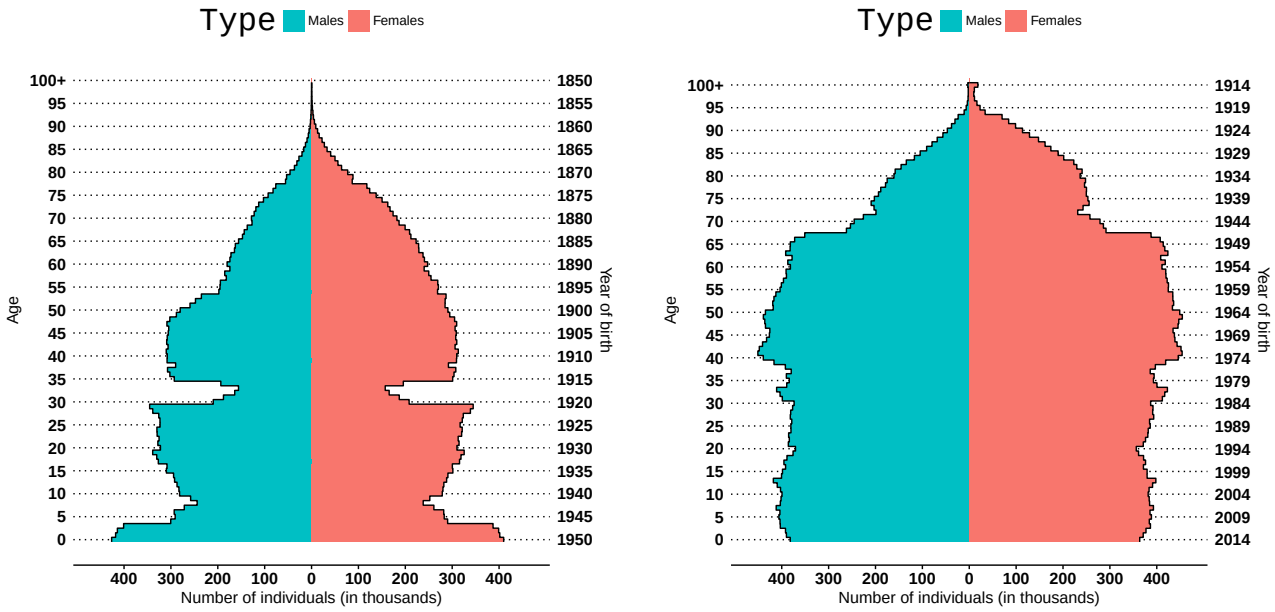
[8] This list is not exhaustive, and scientific fields such as biodemography or gerontology are at the crossroad of several fields.

governmental statistical institutions and international agencies. For instance, The United Nations, the World Bank or the WHO now all have their own online open database. These data allow new questions to be explored and shed a new light on issues at stake. Thus, the multiplicity of approaches and data accounts for both the richness and the difficulty of this subject. This thesis lies within the broader scope of understanding these approaches and their link to social issues, in order to provide a theoretical and simulation framework to compare and question some of the common practices. In the remainder of this section, we will detail more precisely the questions which served as guidelines for this thesis, and then move to a brief description of the different approaches used to address the concerned issues.

1.2 Motivations of the thesis

What can population dynamics do for longevity ? The aggregate longevity observed at the macroscopic level of a national population is the result of complex non-linear demographic mechanisms. In presence of heterogeneity, significant longevity variations could be induced by changes in the cohorts' composition or size, caused for instance by changes in the fertility processes or interactions between individuals. Even the estimation of quantities such as annual death probabilities can be a complex task. Usually, these quantities are estimated on samples as large as possible, due to the rare occurrence of death events. However, an increase in the sample population size might also mean an increase of its heterogeneity. The more heterogeneous a population is, the further from average may individuals behave, thus increasing the variance of estimators and creating a trade-off on the population size.
Standard statistical mortality models such as the Lee-Carter model and its extensions (Lee and Carter (1992); Renshaw and Haberman (2006)) or the Cairns-Blake-Dowd (CBD) model (Cairns et al. (2006)) are based on the modeling of age-specific mortality rates as time series, in order to make projections of future mortality rates. More recently, a new modeling method for mortality rates has been proposed by Ludkovski et al. (2016), based on Gaussian Processes models, and able to quantifies uncertainty associated with smoothed historical experience. The dramatic changes in the demographic and societal structure of populations in developed countries question however the ability of historic data to be a "good guide to the future". In Metropolitan France for instance, the population rose from 41.7 millions in 1950 to 64 millions in 2014, an increase of over 54%. 11% of the population was more 65 years old in 1950, in comparison with over 18%

today[9].The age pyramids of Metropolitan France in 1950 and 2014 are represented in Figure 1.1. Those evolutions have changed the way seniors are perceived, which in turn might have influenced the evolution of their mortality. The composition of the population has also changed significantly. For instance, the proportion of individuals of age 30-45 with no diploma or only a primary school diploma has dropped from about 40% in 1980 to less than 15% in 2010[10]. There is thus an inherent complexity in comparing inside the same time series populations so different, in order to make robust projections.
Furthermore, insurers or pensions funds are mainly interested in mortality at older ages, typically above 65, and sometimes consider these ages only in their models. Limiting data to this age class constitutes however a substantial loss of information, by failing to capture information on younger cohorts (social composition, smoking habits...) and which can give valuable insights into the future, since " today's youths are tomorrow's seniors".



(a) 1950 (b) 2014

Fig. 1.1 Age pyramid of metropolitan France (Source: INSEE)

Heterogeneous population dynamics For these reasons, this thesis focuses on the probabilistic modeling of heterogeneous population dynamics rather than on the statistical projection of mortality rates, by taking on an approach similar to that of N.

[9]Source: INSEE.
[10]Source: Echantillon Démographique Permanent (EDP), INSEE. The EDP is the french longitudinal study.

El Karoui and coauthors (Bensusan (2010) and Boumezoued (2016)). We are interested in the mathematical modeling of complex population dynamics, as an experimenting and simulation tool to generate scenarios, rather than to make realistic predictions.
As for all human systems, the study of human population dynamics is complex due to the very nature of underlying mechanisms. Phenomena are often non-stationary, heterogeneous, and often include interactions taking place at different scales and with sometimes opposite effects. Due to these difficulties, producing a pertinent modeling directly at the macro-level appears to be a very complicated, if not impossible, task. Hence, demographic models have increasingly shifted towards a finer-grained modeling of the population in the last decades. Understanding the aggregated dynamic is thus a major challenge brought by these non-linear "micro" models. One way of proceeding is to reduce the complexity of the aggregated population dynamic in order to obtain a tractable model.
With the rise of available data and computing power, so-called Microsimulation models have been developed in social sciences for the past decades (Li, J. and OḊonoghue, C. (2013)). Used by government and institutional bodies[11], they provide a simulation tool in order to address a broad variety of questions, ranging from evaluating the impact of policy changes and demographic shocks to the study of kinship structure. These models are mostly data-driven and their description often relies on a simulation algorithm. These features constitute an important limitation to their implementation, which needs a considerable amount of data (Silverman et al. (2011)). Furthermore, the complexity of microsimulation models can be significantly limited by computational costs. For instance, inter-individual interactions are often limited in microsimulation models, due to specification problems caused by data limitation or unobservable hidden processes (Zinn (2017)), as well as too high computational costs and time. Capturing the influence of "micro" behaviors at the aggregated level is also a difficult task, in a purely data-driven approach which does not allow for testing behavioral changes. A robust approach to the mathematical modeling of these complex dynamics could help us reach greater understanding on how heterogeneity and interactions operate at the macro-levels. In addition, this approach could also serve as a means to escape the tyranny of data described by Silverman et al. (2011), in order to alleviate "*some of the burdens of the time-consuming and combinatorially expensive data collection required to continue in the traditional fashion*".

[11]For instance, the MiCore tool have been developed as part of the European project Mic-Mac (2005-2009). Another widely used microsimulator is SOCSIM, which originates from a collaboration between Peter Laslett, Eugene Hammel and Kenneth Wachter.

In very different fields, recent advances in probability, mathematical biology and ecology have contributed to the development of a new mathematical framework for individual-based stochastic population dynamics (see e.g. Champagnat et al. (2006); Fournier and Méléard (2004); Méléard and Tran (2009)). These models have been applied for the study of human population dynamics in Bensusan (2010) and Boumezoued (2016). In particular, in order to model the dynamic evolution of heterogeneity inside a population, the latter included change of characteristics for individuals, such as changes of occupational class or marital status. However, these changes are described within a linear framework, while the behavior of individuals is often influenced by interactions with others. Thus, modeling the dynamic changes of the population composition within a non-linear framework and analyzing the aggregated dynamics produced by such models remains an important challenge.
Finally, the individual scale is not always the best chosen granularity for our purpose. Indeed, longitudinal data can be scarce and it can be sometimes more interesting to group individuals into larger risk classes. We will return to these questions in Chapter 2 and 3.

Guidelines of the thesis To sum up, let us state the main line of questioning that have served as general guidelines for this thesis:

- *How can we define a general framework for the modeling of stochastic heterogeneous populations? How does (socioeconomic) heterogeneity impact aggregated dynamics? What approximations can be made in order to reduce the complexity of the studied evolution?*
- *What ingredients are needed for a more realistic theoretical modeling of human population dynamics, yielding populations with real-life characteristics?*
- *How can we provide an analytical and simulation framework serving as an experimental laboratory in order to support decision making? How can we test the validity or potential consequences of existing theories and common practices?*

The complexity and the scope of these questions have led us to recognize the necessity of adopting an integrated approach, both theoretical and empirical, as neither theoretical modeling nor data seem to be sufficient to provide satisfactory explanations. On the one hand, theoretical modeling can be used to test the validity of some theories, when empirical studies can't. Human social experiences are mostly non reproducible, and the ability to modify individuals' behaviors or risk exposures is often limited and ethically

challenging (National Research Council and Committee on Population (2011)). Moreover, theoretical modeling allows us to derive and/or justify approximations in order to simplify complex dynamics. On the other hand, the study of data throws light on important problems that need to be taken into account for a relevant modeling. The ability of data to point us in the right direction has been greatly increased in the current context of important data releases, which have in turn generated a need to provide a theoretical framework to study and deal with them. The fundamental interdisciplinary nature of the study of human populations prevents us from adopting too naive a modeling approach. Opportunities for mathematical developments arise in multiple directions when taking a cross-disciplinary approach, especially since theoretical models are able to incorporate qualitative data, for instance by defining specific interactions or behavior rules.

In this thesis, the modeling of heterogeneous population dynamics is addressed to from three different points of views: The first part (Chapter 2 and 3) is devoted to the modeling of non-linear changes of composition in a stochastic heterogeneous population model, and to the analysis of the aggregated dynamics produce by such models in presence of several timescale. The second part (Chapter 4) studies the impact over time of socioeconomic heterogeneity on aggregated longevity from an empirical point of view, based on recent English data by level of deprivation. The third and last part (Chapter 5) is a cross-disciplinary survey of selected topics on the evolution of human longevity and demographic models. The following sections give a overview of the specific motivations and result of each part of the thesis.

1.3 Part I: Overview and summary of the results

The results presented in Chapter 2 and 3 have been obtained as part of an joint work with Nicole El Karoui.

1.3.1 Introduction

A number of empirical studies have highlighted diverging trends in lifespan since several decades, with growing differences within and among countries. If these inequalities are now well documented, the interpretation of empirical measurements made at different points in time is much more complex within this new heterogeneous framework. There is a real difficulty in comparing several subgroups at different points in time, due in part to time dependent phenomena changing dynamically the composition of the subgroups. For instance, while increases of geographical differences in health and mortality have been reported by several studies, it remains difficult to interpret mechanims underlying the observed inequalities: they could either be interpreted as true "trends" between different areas, or as the consequence of non observable changes of composition inside the areas, generated by changing patterns of internal migration. Dowd and Hamoudi (2014) take the example of migration between rural and urban areas in the US. Before the 70s, migrations from a rural to an urban county might have concerned more socioeconomically disadvantage individuals, while the reverse might have been true for individuals born during the post 1955 baby boom and moving during the late 70s/early 80s. Thus, these changing social patterning of migration could have generated an "artefactual trend" in mortality differentials between rural and urban areas.

Estimating flows of population between several areas or changes in the socioeconomic characteristics of a population is often a challenge. This may be due to a lack of information on incoming flows in cross-sectional data (we don't know "where" people are coming from), or to sample size of longitudinal study that are to small to provide reliable estimates. Furthermore, individuals are not independent, and their behavior is influenced by interactions with others, resulting in the introduction of *non-linearity* or *density dependence* in the population dynamics. For instance, the ability for individuals to change of characteristics could be influenced by the number of individuals in the population. In that case, the composition of a particular area would be influenced not only by the social patterning of migration but also by the intensity of the incoming flow of individuals.

Thus, changes of the population composition add significant complexity to the population dynamics, and one of the great challenges posed by this rising heterogeneity is to

understand how these changes affect the population and its longevity on a aggregated level. Indeed, the dynamic modeling of the population heterogeneity can produce unexpected or counter-intuitive effects at the aggregated level, which cannot be directly modeled by traditional "macro" demographic models.

The first part of this thesis focuses on the stochastic modeling of compositional changes in an heterogeneous population structured by discrete subgroups, and the study of the aggregated dynamic produced by such dynamics. Due to the complexity already introduced by changes of composition, the age-structure of the population is however not taken into account in the modeling. In this sens, the model does not focus particularly on human populations. For instance, similar models have been studied by Auger and coauthors in a deterministic framework, motivated by the study of spatial ecological systems (see e.g. Auger et al. (2000, 2012); Marvá et al. (2013))[12].

In the classical framework of Markov multi-type *Birth Death* processes, heterogeneous populations are usually described by demographic events, that is by births (or entry) and deaths occurring in the subgroups. Here, changes in the composition of the population are also taken into account. They are described by so-called *swap events*, corresponding to the move of an individual from one subgroup to another. Furthermore, the population evolution here is not assumed to be Markovian, in order to take into account additional randomness expressing the variability of the environment and time dependence (for instance the reduction of the mortality intensity over time). The heterogeneous population dynamics is thus called a *Birth Death Swap (BDS)* system.

These features lead us to adopt a point of view different than usual, by representing the population using multivariate counting processes on a state space different than the population state space. This point of view has proven to be very effective, as we detail in the outline below. In particular, in this representation, no specific assumptions are made to explain mechanisms generating swap or demographic events, so that the studied dynamics are very general.

The complexity generated by the presence of swap events makes it difficult to apprehend the population dynamic directly on an aggregated level. However, a separation of timescale can be observed between swap and demographic events in many cases. In particular, when changes in the composition of the population are fast in comparison with the demographic timescale, a simpler approximation of the aggregated population dynamics can be derived. The separation of timescale allows us to obtain an averaging result, reducing the Birth Death Swap system to an "autonomous" multi-type Birth Death process, with averaged intensities: due to swap events, non-linearities emerge at

[12]Examples of such models are detailed at the beginning of Chapter 2.

the aggregated level. Classical averaging results such as those obtained by Kurtz (1992) or Yin and Zhang (2012) cannot be applied here. In order to overcome this difficulty, we rely strongly on the pathwise representations of BDS system introduced in Chapter 2, as well as on *the stable convergence* of concerned processes.

Outline The general model is presented in the first section of Chapter 2. Our presentation gives an algebraic decomposition of the population, based on the multivariate process counting the number of events occurring in the population, called the *jumps counting system*.
The second section is dedicated to the *pathwise representation* of the jumps counting system, as the solution of a multivariate Stochastic Differential System (SDE) driven by an extended Poisson measure. The existence of the jumps counting system is derived from a more general result obtained on the construction of multivariate counting systems by *strong domination* by a non-exploding process. The existence of BDS systems is then obtained as a simple corollary. The construction by strong domination offers several advantages. In particular, the existence of the jumps counting system is obtained under weaker assumptions than usual sub-linear growth or Lipschitz assumptions of the intensities, and "free" tightness properties are derived from the construction in Chapter 3. In the last section of Chapter 2, an alternative construction of BDS systems is presented, called the *Birth Death Swap decomposition algorithm*. The decomposition algorithm allows us to disentangle swap events from demographic events, for instance when they are supposed to have their own timescale.

In Chapter 3, the BDS system is studied in the presence of two timescales, in which swap events are assumed to occur at a much faster timescale than demographic events. The model in presence of two timescales is described in the second section. In the third section of Chapter 3, a general identification result for the processes counting the number of demographic events is proven, in the limit when swap events become instantaneous with respect to demographic events. At the limit, the intensity of demographic events are averaged against *stable limits* of the population, when seen on a suitable space.
This result is then applied in the last section, in which a convergence result is obtained in the particular case of deterministic swap intensities, but with general birth and death intensities. In particular, we show that the aggregated population converges to an autonomous (non Markov) *non-linear* Birth Death process, with birth and death intensities which have been averaged against stationary distribution of pure Swap processes. In order to prove the result, we heavily rely on the BDS decomposition algorithm of Chapter 2.

1.3.2 Pathwise constructions of Birth Death Swap systems (Chapter 2)

Birth Death Swap systems

Let us introduce the stochastic model which serves as a basis for the results presented in Chapter 2 and 3. The population is structured in p discrete subgroups, and the state of the population is described at time t by the random vector $Z_t = (Z_t^1, .., Z_t^p)$, where Z_t^i is the number of individuals in subgroup i. The temporal evolution of the population is thus described by the $\mathbb{N}^p$-valued càdlàg process $Z = (Z_t)_{t\geq 0}$, called the population process.

As stated above, we consider two different types of events: *demographic events* - corresponding to a birth or death in a given subgroup - and *swap events* - corresponding to the move of an individual from one subgroup to another. To our knowledge, the terminology Birth Death Swap has been introduced in discrete time by Huber (2012), for the very different purpose of generating random variables as stationary distribution of pure jumps processes.

The evolution of the population is described by listing the events which change the composition of the population, rather than describing what happen to individuals. For instance, no information is given on the origin of a birth. A birth event can be either endogenous to the population - the individual has parents in the population whose characteristics determine his subgroup - or exogenous - the individual is an immigrant. This approach, quite different from the usual description of individual based models, has the advantage of providing important flexibility in the modeling. In particular, dynamics including strong non-linearity due to complicated interactions can be described very simply.

Jumps counting process In Section 2.2, an algebraic representation of the population dynamic is given, based only on the study of the number of demographic and swap events. All processes are defined on a given probability space $(\Omega, \mathcal{G}, \mathsf{P})$, equipped with equipped with a filtration $(\mathcal{G}_t)$ verifying the usual assumptions of right-continuity and completeness.

We first define a unified description of the different types of events, particularly useful in the rest of the chapter. The set of all types of events is defined by a set $\mathcal{J}$ of cardinal $p(p+1)$ (p birth events, p death events, and $p(p-1)$ swaps). With each type of event is associated a corresponding jump $\phi(\gamma)$ of the population process:

- If $\gamma = (b, j)$ is an event of type "birth in subgroup j", an individual is added to subgroup

j and the corresponding jump is $\phi(b, j) = \mathbf{e}_j$, where $\mathbf{e}_j = (0, .., 1_j, 0..)$.
- If $\gamma = (d, i)$ is an event of type "death in subgroup i", an individual is removed to subgroup i and the corresponding jump is $\phi(d, i) = -\mathbf{e}_i$. Observe that this event cannot occur if the subgroup is empty.
- Finally, if $\gamma = (i, j)$, $i \neq j$, is a swap event from i to j, an individual is simultaneously removed to subgroup i and added to subgroup j. This is also only possible if the subgroup i is not empty, and the resulting change in the population is $\phi(i, j) = \mathbf{e}_j - \mathbf{e}_i$.

As a càdlàg pure jump process, the population process Z can be written as the sum of its jumps and by distinguishing the different types of events we obtain that

$$Z_t = Z_0 + \sum_{\gamma \in \mathcal{J}} \phi(\gamma) N_t^\gamma, \quad \text{with } N_t^\gamma = \sum_{0<s\leq t} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}}. \tag{1.3.1}$$

For each type of event $\gamma \in \mathcal{J}$, the counting process N^γ is the process which counts the number of events of type γ which happened in the population. By assumptions, two events cannot occur at the same time and the processes N^γ have no common jumps. Thus, the process $\mathbf{N} = (N^\gamma)_{\gamma \in \mathcal{J}}$ indexed by $\mathcal{J}$ is a well defined multivariate counting process, called the *jumps counting process* of Z. The previous affine relation can be rewritten as $Z_t = Z_0 + \phi \odot \mathbf{N}_t$. A matrix interpretation of ϕ is given in the beginning of Chapter 2.
Actually, the jumps counting process is no other than a rewriting of the jump measure of Z. This algebraic representation can thus be applied to any pure jump process generated by a finite or countable type of jumps. For instance, a similar representation is used in Anderson and Kurtz (2015) in the particular case of Continuous Time Markov Chains for the modeling of chemical reaction network. However, this representation is to our knowledge less usual for population dynamics.

In many cases, the choice of time 0 for the initial condition Z_0 is very arbitrary, and there is a real interest in relaxing this condition. For instance, Massoulié (1998) defines the initial condition as the state at time 0 of some $\mathbb{N}^p$-valued process $(\xi_t)_{t\leq 0}$. In our setting, we consider a generalized initial condition starting at a random date $\tau \geq 0$ from state $\zeta_\tau \in \mathbb{N}^p$, defined by the *entry process* $\xi_t = \zeta_\tau \mathbb{1}_{\{\tau \leq t\}}$. The population is then rewritten as,

$$Z_t = \xi_t + \phi \odot \mathbf{N}_t, \quad N_t^\gamma = \sum_{\tau < s \leq t} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}}.$$

Population system Due to the many tools available for the study of counting processes, we are interested in the reverse approach, that is the construction of a population process from an entry process and a multivariate counting process. Given a couple $(\xi, \mathbf{N})$, a population could be defined by the affine relation (1.3.1). However, such a process is

not necessarily a well defined population process since its components can take negative values. A necessary and sufficient condition for the population process to be well-defined is actually the support condition (1.3.2), which ensures that no death or swap event can occur from an empty population (support condition), and that $\mathbf{N}$ does not increase on the set $\{t;\ \xi_{t-} = 0\}$ (starting condition).

Definition 1.3.1 (Population system with random departure (Definition 2.2.1))**.**
a) *Let $(\xi_t = \zeta_\tau \mathbb{1}_{\{\tau \leq t\}})$ be an $\mathbb{N}^p$-valued entry process. A $p(p+1)$-multivariate counting process $\boldsymbol{N}$ indexed by $\mathcal{J}$ is called a jumps counting process starting from ξ iff*

$$\begin{cases} \textit{Starting condition} & \mathbb{1}_{\{\xi_{t-}=0\}} d\boldsymbol{N}_t = 0 \\ \textit{Support condition} & \mathbb{1}_{\{\xi^i_{t-} + (\phi \odot \boldsymbol{N})^i_{t-} = 0\}} dN^{i,\beta}_t = 0 \quad \forall i \in \mathcal{I}_p, \quad \forall \beta \in \mathcal{I}^{(i)} \end{cases} \tag{1.3.2}$$

b) *The companion population process of $(\xi, \boldsymbol{N})$ is defined by $Z_t = \xi_t + \phi \odot \boldsymbol{N}_t$. In paticular, Z is a well-defined population process with the jumps counting process $\boldsymbol{N}$ and initial condition ξ. The triplet $(\xi, \boldsymbol{N}, Z)$ is called a population system.*

The populations in which no demographic events occur are called *swap processes* and will play a very important role in the following. As for population systems, a swap system is defined by a triplet $(\xi, \mathbf{N}^{\text{sw}}, X)$, where the swap jumps counting process $\mathbf{N}^{\text{sw}}$ is now indexed by the set of swap events denoted by $\mathcal{J}^{\text{sw}}$. The restriction of the function ϕ to swap events is denoted by ϕ^s, and the companion swap process is defined by $X_t = \xi_t + \phi^s \odot \mathbf{N}^{\text{sw}}_t$.

Several useful transformations on population systems can be directly obtained from this algebraic representation of population systems. For instance a population system $(Z_0, \mathbf{N}, Z)$ starting from 0 in state Z_0 can be decomposed at a random time τ. The population system stopped at time τ is $(Z_0, \mathbf{N}^\tau_t = \mathbf{N}_{t\wedge\tau}, Z^\tau_t = Z_{t\wedge\tau})$. We can also define a population process starting from τ in state Z_τ,

$$\xi^\tau_t(Z) = Z_\tau \mathbf{1}_{\{\tau \leq t\}}, \quad \mathbf{N}^{\tau+}_t = \textstyle\int_0^t \mathbb{1}_{]\tau,\infty)}(s) \mathrm{d}\mathbf{N}_s, \quad Z^{\tau+}_t = \xi^\tau_t(Z) + \phi \odot \mathbf{N}^{\tau+}_t.$$

The population process can be decomposed in $Z = Z^\tau + \phi \odot \mathbf{N}^{\tau+}$ and $\mathbf{N} = \mathbf{N}^\tau + \mathbf{N}^{\tau+}$.

Birth Death Swap intensity By describing the events changing the population composition rather than the behavior of individuals, a very flexible algebraic description of the population is obtained, based on the jumps counting system. Thanks to this general representation, Birth Death Swap systems are defined in the last part on the first section, by transferring the support conditions onto the intensity of jumps counting process. In order for the population system to become a *Birth Death Swap system*, additional assumptions are made on the multivariate intensity of the jump counting

system.
Since only the companion population process Z is usually observed, the intensity process is assumed to be a functional of the population process rather than of the jump counting process itself. To go further and take into account some additional time-dependent randomness, the multivariate intensity process is also assumed to depend in a predictable way on additional randomness.

Definition 1.3.2 (BDS intensity functional and BDS system (Definition 2.2.2))**.**
a) *A* BDS intensity functional $\boldsymbol{\mu}(\omega, t, z) = \boldsymbol{\mu}(t, z) = (\mu^{\gamma}(t, z))_{\gamma \in \mathcal{J}}$ *is a multivariate* $\mathcal{G}$*-predictable non-negative functional depending on* $z \in \mathbb{N}^p$*, satisfying*

$$\boldsymbol{\mu}(t, \boldsymbol{0}) \equiv 0 \quad \textit{and} \quad \sum_{i \in \mathcal{I}_p} \sum_{\beta \in \mathcal{I}^{(i)}} \mu^{i,\beta}(t, z) \mathbf{1}_{\{z^i = 0\}} \equiv 0, \quad \mathrm{d}t \otimes \mathrm{dP}\, a.s. \tag{1.3.3}$$

b) *A Birth Death Swap (BDS) system of intensity functional* $\boldsymbol{\mu}$ *is a population system* $(\xi, \boldsymbol{N}, Z)$ *such that the jumps counting process* $\boldsymbol{N}$ *is a multivariate counting process of* $\mathcal{G}_t$*-intensity* $\boldsymbol{\lambda}_t = \boldsymbol{\mu}(t, Z_{t-}) = \boldsymbol{\mu}(t, \xi_{t^-} + \phi \odot \boldsymbol{N}_{t^-})$.

Linear intensities: For each type of event γ, $\mu^{\gamma}(t, Z_{t^-})$ is the intensity corresponding to the occurrence of the event of type γ in *all* the population, and not the *rate* at which the event of type γ can occur to *one individual*. A direct interpretation of intensities in term of individual rates can be derived in the case of linear intensities. For instance, if the death intensity in subgroup i is $\mu^{(d,i)}(\omega, t, z) = d^i(\omega, t) z^i$, the interpretation is that all individuals in subgroup i die independently at death rate $d^i(\omega, t)$. A similar interpretation can be given for linear swap or birth intensities.
Markov BDS: When the BDS intensity functional is an homogeneous deterministic function $\boldsymbol{\mu}(z)$, the BDS system is a Continuous Time Markov Chain (CTMC), and can be described using classical tools of CTMC.

Birth Death Swap Differential equation

In section 2.3 of Chapter 2, we tackle the issue of BDS systems and their pathwise realization. The two main results of this section are in the spirit of the recent renewed interest for pathwise representations of point processes, which has led to the development of a consequent body of literature in various domains. These representations are based on the pathwise realization of point processes as solutions of Stochastic Differential Equations (SDE) driven by Poisson measures. In particular, solutions are obtained from the thinning of an "augmented Poisson measure".
We apply these representations in order to realize the jumps counting processes $\mathbf{N}$ as

solution of multivariate SDEs driven by a Poisson measure. Focusing on the jumps counting process allows us to adopt the point of view of point processes, within a general framework similar to that of Massoulié (1998).
Non-explosion is often central in the analysis of SDEs driven by Poisson measures, and in our setting, solutions are considered to be well-defined iff they stay finite in finite time with probability one.
The main point of the Section 2.3 is that the existence of BDS systems is derived from a more general result which is first obtained, on the construction of multivariate counting processes by *strong domination* by a non-exploding process. In particular, some of the usual assumptions on the intensity functional that are Lipschitz or sublinear growth conditions can be relaxed by using this result.

Construction of multivariate counting processes by strong domination Theorem 1.3.1 is the first main result of Section 2.3, and concerns the construction of solutions of SDEs driven by a Poisson measure by strong domination by a multivariate counting process driven by the same Poisson measure. Before stating the result, we give in this introduction a brief overview of the thinning of Poisson measure, which is described in more details in Subsection 2.3.1 of Chapter 2.
Given a multivariate Poisson measure $\bar{Q}(\mathrm{d}t, \mathrm{d}\theta) = (Q^\imath(\mathrm{d}t, \mathrm{d}\theta))_{\imath=1\dots\rho}$ with components of intensity $q(\mathrm{d}t, \mathrm{d}\theta) = \mathrm{d}t \times \mathrm{d}\theta$, and a ρ-multivariate predictable intensity process $(\bar{\lambda}_t)$, a multivariate Cox process of $\mathcal{G}_t$-intensity λ_t can be obtained by *thinning and projection* of $\bar{Q}$, $\bar{Q}^\lambda_t = \int_0^t \int_{\mathsf{R}+} \mathbb{1}_{\{0<\theta\leq\lambda_s\}} \bar{Q}(\mathrm{d}s, \mathrm{d}\theta) = \int_0^t \bar{Q}(\mathrm{d}s,]0, \lambda_s])$. Keeping the mark θ is sometimes interesting, and a random measure $\bar{Q}^\Delta(\mathrm{d}t, \mathrm{d}\theta)$ can be defined as the restriction of $\bar{Q}$ to the random set $\Delta = \{(t, \theta);\, 0 < \theta \leq \lambda_t\}$: $\bar{Q}^\Delta(\mathrm{d}t, \mathrm{d}\theta) = \mathbb{1}_\Delta(t, \theta) Q(\mathrm{d}t, \mathrm{d}\theta)$. Q^Δ is a random measure of random intensity measure $q^\Delta(\mathrm{d}t, d\theta) = \mathbb{1}_\Delta(t, \theta)\, \mathrm{d}t \times d\theta$. When the intensity $\bar{\lambda}$ of the multivariate counting process $\bar{Q}^\lambda$ is a predictable functional of $\bar{Q}^\lambda$ itself, the thinning equation becomes a stochastic differential equation, driven by the Poisson measure $\bar{Q}$.
A ρ-multivariate counting process $\bar{Y}^\alpha$ is said to *be strongly dominated* by $\bar{Y}^\beta$, $\bar{Y}^\alpha \prec \bar{Y}^\beta$ iff $\bar{Y}^\beta - \bar{Y}^\alpha$ is a multivariate counting process, or equivalently iff all jumps of $\bar{Y}^\alpha$ are jumps of $\bar{Y}^\beta$. In Theorem 2.3.1, the solution $\bar{Y}^\alpha$ (of intensity functional $\bar{\alpha}(t, y)$) of a multivariate SDE is built by strong comparison with a dominating process $\bar{Y}^\beta$ (of intensity functional $\bar{\beta}(t, y)$), under the following assumption:

$$\alpha^\imath(t, \tilde{y}) \leq \beta^\imath(t, \bar{y}), \quad \forall 1 \leq \imath \leq \rho,\ \tilde{y} \leq \bar{y} \in \mathbb{N}^\rho.$$

Theorem 1.3.1 ((Theorem 2.3.1)). *Let $\bar{Q}(\mathrm{d}t, d\theta) = (Q^\imath(\mathrm{d}t, d\theta))_{\imath\in E}$ be a multivariate Poisson measure, and $\bar{\alpha}$ and $\bar{\beta}$ two predictable ρ-dimensional intensity functionals defined*

on $\mathcal{Y} = \mathbb{N}^\rho$, where $\bar{\alpha}$ is assumed to be dominated by $\bar{\beta}$ ($\bar{\alpha} \leq \bar{\beta}$).
Assume the existence of a unique non-exploding solution $\bar{Y}^\beta \in \mathbb{N}^\rho$ of the multivariate SDE:

$$d\bar{Y}_t^\beta(\bar{y}) = \bar{Q}(dt,]0, \bar{\beta}(t, \bar{y} + \bar{Y}_{t-}^\beta(\bar{y}))]), \tag{1.3.4}$$

Then, for all $\tilde{y} \leq \bar{y}$, there exists a unique (non-exploding) solution to the equation,

$$d\bar{Y}_t^\alpha(\tilde{y}) = \bar{Q}\Big(dt,]0, \bar{\alpha}(t, \tilde{y} + \bar{Y}_{t-}^\alpha(\tilde{y}))]\Big) \tag{1.3.5}$$

Furthermore, $\bar{Y}^\alpha(\tilde{y})$ is strongly dominated by $\bar{Y}^\beta(\bar{y})$: $\bar{Y}^\alpha(\tilde{y}) \prec \bar{Y}^\beta(\bar{y})$.

Sketch of the proof: The thinning procedure is well-adapted to solve this problem, and for Cox processes the answer is immediate. If two Cox processes $Q_t^{\lambda^i} = Q(]0,t] \times]0, \lambda_t^i])$, $i = 1, 2$ have ordered intensities $\lambda_t^1 \leq \lambda_t^2$, then the thinning construction using the same Poisson measure for both processes directly yields that $Q^{\lambda^1} \prec Q^{\lambda^2}$. The key to the proof is to rewrite Q^{λ^1} as $Q_t^{\lambda^1} = Q(]0,t] \times]0, \lambda_t^1 \wedge \lambda_t^2]) = Q^{\Delta_2}(]0,t] \times]0, \lambda_t^1])$. This means that Q^{λ^1} can be obtained by *thinning of* Q^{Δ_2} instead of Q. In particular, all jump times of Q^{λ^1} are jump times of Q^{λ^2}, and $Q^{\lambda^1} \prec Q^{\lambda^2}$.
The direct application to general multivariate counting processes is not straightforward, since the order $\bar{\alpha} \leq \bar{\beta}$ on the intensity functionals does not necessary imply a natural order on the stochastic intensities $\alpha(t, \tilde{y} + \bar{Y}_{t-}^\alpha)$ and $\beta(t, \bar{y} + \bar{Y}_{t-}^\beta)$. The key idea of the proof is however similar to the case of Cox processes. The idea is to study a slightly different version of (1.3.5), by replacing $\bar{Q}$ with the measure $\bar{Q}^{\Delta_\beta}$:

$$d\tilde{Y}_t^\alpha = \bar{Q}^{\Delta_\beta}(\mathrm{d}t,]0, \bar{\alpha}(t, \tilde{y} + \tilde{Y}_{t-}^\alpha)]) = \bar{Q}(\mathrm{d}t,]0, \bar{\alpha}(t, \tilde{y} + \tilde{Y}_{t-}^\alpha) \wedge \bar{\beta}(t, \bar{y} + \bar{Y}_{t-}^\beta)]), \tag{1.3.6}$$

$\tilde{Y}^\alpha$ is obtained by thinning of Q^{Δ_β}, which guarantees the strong domination of $\tilde{Y}^\alpha$ by $\bar{Y}^\beta$. The existence and uniqueness of solutions of (1.3.6) is easier to prove, since the jump times of $\bar{Q}^{\Delta_\beta}$ can be enumerated in increasing order, which is not the case for $\bar{Q}$. The proof is concluded by showing the equivalence of Equation (1.3.5) and (1.3.6).
Theorem 1.3.1 generalizes the results of Rolski and Szekli (1991) for the comparison of Cox processes, and of Bhaskaran (1986) for the comparison of pure birth Markov processes. The proof of 1.3.1 can actually be extended to the case of Point processes with continuous marks, and with intensity functional depending on the past of the process. This is the subject of a ongoing work with N. El Karoui.

Dominating processes Once given this result, the key is to find a class of non-exploding multivariate counting processes large enough to serve as reasonable dominating

processes. The most simple example of such processes is the one dimensional Markov pure birth process, also called online Markov pure birth process, of intensity function here denoted by $Kg(y)$. It is well-known that if the function g verifies the Feller criterion

$$\sum_{j=1}^{\infty} \frac{1}{g(j)} = \infty, \tag{1.3.7}$$

the process does not explode in finite time. The multivariate case is an easy extension when the multivariate intensity function is a function of the size $\bar{y}^\natural = \sum_1^\rho y^\imath$ of the birth process, $K\bar{g}(\bar{y}) = (Kg^\imath(\bar{y}^\natural))$. Non-explosion is guaranteed if all functions $g^\imath$ satisfy the Feller criterion.
However, the domination by a multivariate Markov birth process is often not satisfactory. The assumption can be relaxed by using *Cox Birth processes* as dominating processes. Cox Birth processes are multivariate counting processes with product intensity $k_t\bar{g}(\bar{y}^\natural)$. They are obtained by replacing the constant K of Markov Births with a locally bounded predictable process (k_t), i.e bounded by a sequence (K_p) along a nondecreasing sequence of stopping times (S_p) going to ∞. The existence and non-explosion of Cox Birth processes is a corollary of Theorem 1.3.1, since the solutions of (1.3.5) with intensity functional $\bar{\alpha}^p(t,\bar{y}) = (k_t \wedge K_p)\bar{g}(\bar{y}^\natural)$ are dominated by the Markov birth intensity $\bar{\beta}(t,\bar{y}) = K_p\bar{g}(\bar{y}^\natural)$ and do not depend on p on the interval $[0, S_p]$.

Birth Death Swap multivariate SDE In the last part of the section, we come back to the study of BDS systems defined in 1.3.2, and introduce the so-called Birth Death Swap multivariate SDE:

Definition 1.3.3 (BDS multivariate SDE (Definition 2.3.1))**.**
Let $\boldsymbol{Q} = (Q^\gamma)_{\gamma\in\mathcal{J}}$ be a multivariate Poisson measure, $\boldsymbol{\mu}(t,z)$ a BDS intensity functional and (ξ_t) be an entry process. The Birth Death Swap multivariate SDE associated with the entry process ξ and intensity functional $\boldsymbol{\mu}$ is defined by

$$\mathrm{d}\boldsymbol{N}_t = \boldsymbol{Q}(\mathrm{d}t,]0, \boldsymbol{\mu}(t, \xi_{t-} + \phi \odot \boldsymbol{N}_{t-})]), \quad \textit{with } Z_t = \xi_t + \phi \odot \boldsymbol{N}_t. \tag{1.3.8}$$

If $\boldsymbol{N}$ is a solution of (1.3.8), then $(\xi, \boldsymbol{N}, Z)$ is a BDS system of entry process ξ and BDS intensity functional $\boldsymbol{\mu}$.

In order to obtain the existence and uniqueness of solutions (1.3.8), we give sufficient conditions on the birth intensities in order to obtain a solution of (1.3.8) by strong domination with a non-exploding process. In order to use the point process point of view of Theorem 1.3.1, the BDS intensity functional is expressed in terms of the jumps

counting process, rather than in terms of the population process. $\mathbb{N}^p$-valued vectors are written as $z = \xi + \phi \odot \boldsymbol{\nu}$, with $\boldsymbol{\nu} = (\boldsymbol{\nu}^b, \boldsymbol{\nu}^d, \boldsymbol{\nu}^s)$ a vector indexed by $\mathcal{J}$. The BDS intensity functional $\boldsymbol{\mu}$ can be rewritten as functional of $\boldsymbol{\nu}$ instead of z, $\boldsymbol{\lambda}(t, \boldsymbol{\nu}) = \boldsymbol{\mu}(t, \xi_t + \phi \odot \boldsymbol{\nu})$. The following assumption on the birth intensity is made:

$$\textit{Cox Birth Hyp} \quad \forall i \in \mathcal{I}_p,\ \lambda^{(b,i)}(t, \boldsymbol{\nu}) = \mu^{(b,i)}(t, \xi_t + \phi \odot \boldsymbol{\nu}) \leq k_t\, g^{(b,i)}(\xi_t^\natural + \boldsymbol{\nu}^{b,\natural}) \quad (1.3.9)$$

where $k_t \mathbf{g}^b = (k_t g^{(b,i)})$ is a p-Cox birth intensity functional.
No further assumptions are made on the swap and death intensities, which are naturally dominated by an increasing function of the number of births:

$$\textit{Swap and death inequality} \quad \lambda^{(i,\beta)}(t, \boldsymbol{\nu}) = \mu^{(i,\beta)}(t, \xi_t^\natural + \boldsymbol{\nu}) \leq \hat{\mu}^{(i,\beta)}(t, \xi_t^\natural + \boldsymbol{\nu}^{b,\natural}) \quad (1.3.10)$$

The dominating process is the $p(p+1)$ multivariate counting process $\mathbf{G} = (\mathbf{G}^\gamma)_{\gamma \in \mathcal{J}}$ defined in the following two steps:
The first step is to introduce the p-Cox Birth process $\mathbf{G}_t$ solution of the following multivariate SDE:

$$\mathrm{d}\mathbf{G}_t^b = \mathbf{Q}^b(\mathrm{d}t,]0,\ k_t\, \mathbf{g}^b(\xi_t^\natural + \mathbf{G}_{t^-}^{b,\natural})]). \quad (1.3.11)$$

The second step consist in introducing "swap and death coordinates" to $\mathbf{G}^b$, defined by the p and $p(p-1)$ multivariate Cox processes:

$$\mathrm{d}\mathbf{G}_t^d = \mathbf{Q}^d(\mathrm{d}t,]0, \hat{\boldsymbol{\mu}}(t, \xi_{t^-}^\natural + G_{t^-}^{b,\natural})]), \quad \mathrm{d}\mathbf{G}_t^s = \mathbf{Q}^s(\mathrm{d}t,]0, \hat{\boldsymbol{\mu}}(t, \xi_{t^-}^\natural + G_{t^-}^{b,\natural})]). \quad (1.3.12)$$

The$(p(p+1)$ multivariate counting process $\mathbf{G} = (\mathbf{G}^b, \mathbf{G}^d, \mathbf{G}^s)$ is non-exploding and is called the *dominating process*.
Finally, the existence of the BDS multivariate SDE is obtained at the end of Section 2.3, under the Cox Birth domination assumption:

Theorem 1.3.2 (Theorem 2.3.2)**.** *Assume that the Cox Birth assumption (1.3.9) is verified: $\boldsymbol{\mu}^b(t, z) \leq k_t\, \boldsymbol{g}^b(z^\natural)$ where the components of $\boldsymbol{g}^b$ are non-decreasing and satisfy the Feller criterion (1.3.7). Moreover, assume that k_t, $\boldsymbol{\mu}^s(t, K)$ and $\boldsymbol{\mu}^d(t, K)$ are locally bounded in time for any K.*
Then, there exists a unique solution to Equation (1.3.8),

$$d\boldsymbol{N}_t = \boldsymbol{Q}(dt,]0, \boldsymbol{\mu}(t, \xi_{t-} + \phi \odot \boldsymbol{N}_{t-})]), \quad \textit{with } Z_t = \xi_t + \phi \odot \boldsymbol{N}_t.$$

The triplet $(\xi, \boldsymbol{N}, Z)$ is a well-defined BDS system of BDS intensity functional $\boldsymbol{\mu}$ and entry process ξ. Furthermore, $\boldsymbol{N}$ is strongly dominated by $\boldsymbol{G}$, $\boldsymbol{N} \prec \boldsymbol{G}$.

BDS decomposition algorithm

There are many advantages in the construction by domination, in particular since the BDS jumps counting process **N** can be localized by a sequence of increasing stopping times which do not depend on the process itself. However, the same property can be a drawback when simulating the BDS system by strong domination. Indeed **G** can have much more jumps that **N**, making the simulation inefficient.
In Section 2.4 of Chapter 2, an alternative construction of the solution of (1.3.8) is presented, called the *Birth Death Swap decomposition algorithm* and based on the disentanglement of swap and demographic events. The decomposition algorithm is better suited to the simulation of BDS systems when swap and demographic intensities are of a very different nature, for instance when they are supposed to have their own timescale. The disentanglement of swap and demographic events will also be instrumental in the proof of Theorem 3.5.1 in Chapter 3.

1.3.3 Averaging of BDS systems in the presence of two timescales (Chapter 3)

BDS system in the presence of two timescales

In Chapter 3, the BDS system built in the previous chapter is studied in the presence of two timescales. Swap events are assumed to happen at a much faster timescale than demographic events. Intuitively, this means that demographic events happen with intensities of order "$O(1)$", while swap events occur with greater intensities of order "$O(\frac{1}{\epsilon})$" depending on a small parameter ϵ.

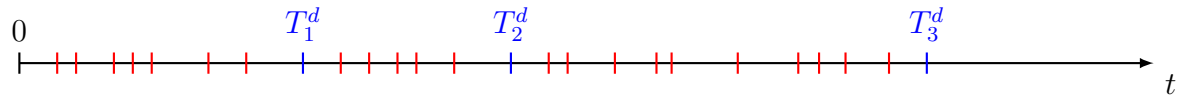


Fig. 1.2 Example of distribution of swap events and demographic events

More formally, the BDS system now depends on a small parameter ϵ, and is denoted by $(\xi, \mathbf{N}^\epsilon, Z^\epsilon)$. The BDS multivariate SDE (1.3.8) becomes:

$$Z_t^\epsilon = \xi_t + \phi^s \odot \mathbf{N}_t^{s,\epsilon} + \mathbf{N}_t^{b,\epsilon} - \mathbf{N}_t^{d,\epsilon}, \tag{1.3.13}$$
$$\mathrm{d}\mathbf{N}_t^{\mathrm{dem},\epsilon} = \mathbf{Q}^{\mathrm{dem}}(\mathrm{d}t,]0, \boldsymbol{\mu}^{\mathrm{dem}}(t, Z_{t^-}^\epsilon)]), \quad \mathrm{d}\mathbf{N}_t^{s,\epsilon} = \mathbf{Q}^s(\mathrm{d}t,]0, \frac{1}{\epsilon}\boldsymbol{\mu}^s(t, Z_{t^-}^\epsilon)]).$$

The swap part and the demographic part of the counting system behave very differently in this separation of timescale framework. On the one hand, the swap counting system $\mathbf{N}^{s,\epsilon}$ evolves on a fast timescale and depends on the small parameter ϵ through the population process *and* its intensity functional $\frac{1}{\epsilon}\boldsymbol{\mu}^s(t,z)$. Thus, $\mathbf{N}^s$ explodes in finite time when $\epsilon \to 0$. On the other hand, the demographic counting system $\mathbf{N}^{\mathrm{dem},\epsilon}$ only depends on ϵ through Z^ϵ. The population process Z^ϵ is now a two-time scale $\mathcal{G}_t$-adapted process, with swap events generated by the multivariate counting process $\mathbf{N}^{s,\epsilon}$ occurring on a much faster timescale than demographic events generated by $\mathbf{N}^{\mathrm{dem},\epsilon}$.
The aim of Chapter 3 is to study the convergence of the BDS system in the limit when swap events become instantaneous with respect to demographic events ($\epsilon \to 0$).

Population process: Due to the explosion of swap events, the family of population processes $(Z^\epsilon)_{\epsilon>0}$ is not tight in the space of càdlàg pure jump processes on $\mathbb{N}^p$. However, the tightness of (Z^ϵ) in a weaker framework is obtained at the beginning of the chapter, by not seeing the population process as a dynamic process anymore, but rather as an $\mathbb{N}^p$-valued random variable $Z^\epsilon(\omega, s)$ defined on the product space $\Omega \times \mathsf{R}^+$. The key to this result is once again to control the size of the population.

Demographic counting system: In the two timescales framework, the dominating process $\mathbf{G}^\epsilon$ as defined in Chapter 2 only depends on ϵ through its swap components, $\mathbf{G}^\epsilon = (\mathbf{G}^{\text{dem}}, \mathbf{G}^{s,\epsilon})$. This consequence of the construction by strong domination is critical, since this means that the family of demographic counting systems $(\mathbf{N}^{\text{dem},\epsilon})$ is uniformly strongly dominated by $\mathbf{G}^{\text{dem}}$. A key consequence of this property is that the family $(\mathbf{N}^{\text{dem},\epsilon})$ of demographic counting systems is tight in the space of $2p$-multivariate counting processes.

In Section 3.4 of Chapter 3, a general identification result is proven. The intensity of *stable limits* of the demographic counting systems is identified with the demographic intensity functional $\boldsymbol{\mu}^{\text{dem}}$ which is averaged against *stable* limits of the population processes (viewed as $\mathbb{N}^p$ random variables). The notion of *stable convergence* is critical here. Due to the non-Markov framework of BDS systems, classical averaging results based on the weak convergence, such as in Kurtz (1992), cannot be applied here. In particular, no further assumptions are needed on the dependence of the BDS intensity functional $\boldsymbol{\mu}(\omega, t, z)$ on (ω, t), as it is often required.

This result is then applied in Section 3.5, in which a convergence result is obtained in the particular case of deterministic swap intensities. In order to prove the result, we will heavily rely on the BDS decomposition algorithm of Chapter 2.

Recall on the stable convergence

Stable convergence Section 3.3 is dedicated to an overview on the stable convergence. In order to state the results of the chapter, let us recall some basic principles of the stable convergence.

A sequence $(Y_n(\omega))$ of $\mathcal{X}$-valued random variables (r.vs) defined on the *same* probability space $(\Omega, \mathcal{G}, \mathsf{P})$ is said to converge stably iff there exists a random probability kernel $\Gamma(\omega, \mathrm{d}x)$ from $(\Omega, \mathcal{G})$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that for all bounded $\mathcal{G}$-measurable random variable $K(\omega)$ and bounded continuous function f:

$$\mathsf{E}[K(\omega) f(Y_n)] \to \mathsf{E}[K(\omega) \textstyle\int_{\mathcal{X}} f(x) \Gamma(\omega, \mathrm{d}x)] = \mathsf{E}[K(\omega) \Gamma(\omega, f)].$$

In particular, (Y_n) converges in distribution to the probability measure defined by $\mu(f) = \mathsf{E}[\Gamma(f)]$. The stable convergence is actually a stronger mode of convergence than the convergence in distribution, since it allows to keep information on the initial probability space. The previous definition, however, is not necessarily the most useful in the applications.

The stable convergence can be seen as an extension of the convergence in distribution to the convergence of probability measures defined on an enlarged product space $(\bar{\Omega}, \bar{\mathcal{G}}) = (\Omega \times \mathcal{X}, \mathcal{G} \otimes \mathcal{B}(\mathcal{X}))$, with fixed marginals on Ω equal to P. The set of such probability

measures is called the set of *rules* and is denoted by $\mathcal{R}(\mathsf{P}, \mathcal{X})$. Furthermore, any rule $\mathsf{R} \in \mathcal{R}(\mathsf{P}, \mathcal{X})$ can be disintegrated in $\mathsf{R}(\mathrm{d}\omega, \mathrm{d}x) = \mathsf{P}(\mathrm{d}\omega)\Gamma(\omega, \mathrm{d}x)$, with Γ a random probability kernel from $(\Omega, \mathcal{G})$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.
The extended class of test functions for the convergence of rules is the space $\mathcal{C}_{bmc}(\Omega \times \mathcal{X})$ of $\bar{\mathcal{G}}$-measurable and bounded functionals H, continuous in x for any ω but without any regularity in ω. Thus, a sequence of rules (R^n) is said to converge stably to $\mathsf{R} \in \mathcal{R}(\mathsf{P})$ iff

$$\mathsf{R}^n(H) \text{ converges to } \mathsf{R}(H), \quad \forall H \in \mathcal{C}_{bmc}(\Omega \times \mathcal{X}). \tag{1.3.14}$$

It is actually enough to use product functionals $H(\omega, x) = K(\omega)f(x)$ as test functions.
A rule R^Y can be associated with every $\mathcal{X}$-valued random variable, $\mathsf{R}^Y(\mathrm{d}\omega, \mathrm{d}x) = \mathsf{P}(d\omega)\delta_{Y(\omega)}(\mathrm{d}x)$. Thus, when $\mathsf{R}^n = \mathsf{R}^{Y_n}$, the stable convergence of (R^n) to R is equivalent to the stable convergence of (Y^n) to Γ. Furthermore, the stable limit of (Y_n) can be *realized* on the extended space $(\bar{\Omega}, \bar{\mathcal{G}}, \mathsf{R})$. The r.vs (Y_n) can be naturally extended on $(\bar{\Omega}, \bar{\mathcal{G}}, \mathsf{R})$ by setting $Y_n(\omega, x) = Y_n(\omega)$, and the stable convergence of (R^n) to R can be interpreted as the stable convergence of (Y_n) to the canonical variable Υ on the extended space $(\bar{\Omega}, \bar{\mathcal{G}}, \mathsf{R})$.
The stable convergence can also be interpreted as a mode of convergence for the sequence of *random kernels* (δ_{Y_n}):

$$\forall f \in C_b(\mathcal{X}), \quad \Gamma^{Y_n}(f) = f(Y_n) \text{ converges to } \Gamma(f) \text{ weakly in } \mathbb{L}^1(\Omega, \mathcal{G}, \mathsf{P}),$$

where we recall that a sequence of r.vs (ξ_n) converges to ξ *weakly* in $\mathbb{L}^1(\Omega, \mathcal{G}, \mathsf{P})$ iff for any bounded random variable H, $\mathsf{E}[H\,\xi_n] \to \mathsf{E}[H\,\xi]$.

Stable relative compactness One of the most interesting property of the stable convergence is that relative compactness for the stable topology is "free", in the sense that if a sequence (Y_n) converges in distribution, then there exists a subsequence of (Y_n) which converges stably to a rule R, whose marginal on $\mathcal{X}$ is the limit distribution. More formally,

The sequence (R^{Y_n}) is relatively compact in $\mathcal{R}(\mathsf{P})$ iff (Y_n) is tight.

For the sake of simplicity, the abuse of notation "(Y_n) or (Γ^{Y_n}) is stably relatively compact" is also used.

Application to the population process

The given probability space As stated at the beginning of this section, the family of population processes (Z^ϵ) is not tight in the space of $\mathbb{N}^p$-valued càdlàg processes, due to the explosion of swap events. The idea is to not see the population processes as dynamic

processes anymore, but rather as $\mathbb{N}^p$-valued r.vs, defined on the product-space:

$$(\tilde{\Omega}, \tilde{\mathcal{G}}) = (\Omega \times \mathsf{R}^+, \mathcal{G} \times \mathcal{B}(\mathsf{R}^+)), \text{ with } Z^\epsilon(\tilde{\omega}) = Z^\epsilon(\omega, s) = Z^\epsilon_s(\omega).$$

The space $(\tilde{\Omega}, \tilde{\mathcal{G}})$ is equipped with the probability measure $\tilde{\mathsf{P}}(\mathrm{d}\omega, \mathrm{d}s) = \mathsf{P}(\mathrm{d}\omega) \otimes \lambda^e(\mathrm{d}s)$. In this new space, for any $\tilde{\mathcal{G}}$-measurable functional $H(\tilde{\omega}, z)$,

$$\tilde{\mathsf{E}}[H(\tilde{\omega}, Z^\epsilon(\tilde{\omega}))] = \mathsf{E}[\textstyle\int_{\mathsf{R}^+} H(\omega, s, Z^\epsilon_s(\omega))\lambda^e(\mathrm{d}s)] = \mathsf{E}[\textstyle\int_{\mathsf{R}^+} e^{-s} H(\omega, s, Z^\epsilon_s(\omega))\mathrm{d}s],$$

and the distribution measure of Z^ϵ is the probability distribution on $\mathbb{N}^p$, defined by $\tilde{\mu}^\epsilon(f) = \mathsf{E}[\int_{\mathsf{R}^+} e^{-s} f(Z^\epsilon_s)\mathrm{d}s]$.

Tightness and stable relative compactness Since the demographic counting systems $(\mathbf{N}^{\mathrm{dem},\epsilon})$ are uniformly strongly dominated by $\mathbf{G}^{\mathrm{dem}}$, the size of the population $Z^{\epsilon,\natural} = N^{b,\natural,\epsilon} - N^{d,\natural,\epsilon}$ is bounded by $G^{b,\natural}$, $\forall \epsilon > 0$. Thus, the $\mathbb{N}^p$ r.vs $Z^\epsilon(\tilde{\omega})$ are uniformly bounded by $\mathbf{G}^b(\tilde{\omega})$. It follows, that $(\tilde{\mu}^\epsilon)$ is tight, or equivalently that (Z^ϵ) is stable relatively compact in $\mathcal{R}(\tilde{\mathsf{P}}, \mathbb{N}^p)$. Up to a subsequence, the family $(Z^\epsilon(\tilde{\omega}))$ converges stably to a random kernel $\tilde{\Gamma}(\omega, s, dz)$ from $\tilde{\Omega}$ to $\mathbb{N}^p$, and for all bounded measurable functionals $H(\omega, s, z)$:

$$\tilde{\mathsf{E}}[H(\omega, s, Z^\epsilon)] \underset{\epsilon \to 0}{\longrightarrow} \tilde{\mathsf{E}}[\int_{\mathbb{N}^p} H(\omega, s, z)\tilde{\Gamma}(\omega, s, \mathrm{d}z)] = \mathsf{E}[\int_0^\infty \int_{\mathbb{N}^p} H(\omega, s, z)\tilde{\Gamma}(\omega, s, \mathrm{d}z)\lambda^e(\mathrm{d}s)]. \tag{1.3.15}$$

The usual notation $\tilde{\Gamma}(\omega, s, H) = \int_{\mathbb{N}^p} H(\omega, s, z)\tilde{\Gamma}(\omega, s, \mathrm{d}z)$ is used. We also show that the kernel $\tilde{\Gamma}(\omega, s, dz)$ inherits the $(\mathcal{G}_t)$-adaptation of the population processes Z^ϵ.
When applying the previous equation to the demographic intensity functional μ^{dem}, the following result is obtained:

Proposition 1.3.1. *(Proposition 3.3.1) Under the dominating assumption and up to a subsequence, the predictable compensator $\boldsymbol{A}^\epsilon_{t\wedge\tau}$ of the demographic counting system $\boldsymbol{N}^{\mathrm{dem},\epsilon}_{t\wedge\tau}$ converges weakly in $\mathbb{L}^1(\Omega, \mathcal{G}, \mathsf{P})$ to $\boldsymbol{A}_{t\wedge\tau}$ for all $t \geq 0$; where is $\boldsymbol{A}$ a continuous non decreasing process whose density is $\boldsymbol{\mu}^{\mathrm{dem}}$ averaged again the optional kernel $(\tilde{\Gamma}(., t, dz))$, that is:*

$$\boldsymbol{A}^\epsilon_{t\wedge\tau} = \int_0^t \boldsymbol{\mu}^{\mathrm{dem}}(s, Z^\epsilon_s)\mathrm{d}s \; cv \; weakly - \mathbb{L}^1 \; to \quad \boldsymbol{A}_{t\wedge\tau} = \int_0^{t\wedge\tau} \tilde{\Gamma}(s, \boldsymbol{\mu}^{\mathrm{dem}}(s, .))\mathrm{d}s \tag{1.3.16}$$

Convergence of the demographic counting systems

In Section 3.4 of Chapter 3, a general identification result is proven concerning the stable limits of the demographic counting system $\mathbf{N}^{\mathrm{dem},\epsilon}$, solution of:

$$\mathrm{d}\mathbf{N}^{\mathrm{dem},\epsilon}_t = \mathbf{Q}^{\mathrm{dem}}(\mathrm{d}t,]0, \boldsymbol{\mu}^{\mathrm{dem}}(t, Z^\epsilon_{t^-})]) \prec \mathrm{d}\mathbf{G}^{\mathrm{dem}}_t.$$

Since $\mathbf{N}^{\mathrm{dem}}$ only depends on ϵ through the population process Z^ϵ, a stronger convergence

than for the population process can be considered. The subspace of $D(\mathbb{N}^{2p})$ composed of $\mathbb{N}^{2p}$-valued functions, with components only increasing with unit jumps and with no common jumps, is denoted by $\mathcal{A}^{2p}$. The family $\mathbf{N}^{\mathrm{dem},\epsilon}$ is considered as a family of $\mathcal{A}^{2p}$-valued r.vs on the given space $(\Omega, \mathcal{G}, \mathsf{P})$.

The first result of the section concerns the tightness of $(\mathbf{N}^{\mathrm{dem},\epsilon})$:

Proposition 1.3.2 (Proposition 1.3.2)**.** *The family of $\mathcal{A}^{2p}$-valued random variables $(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ is tight. Equivalently, $(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ is stably relatively compact in $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$.*

The result is obtained rather straightforwardly as a consequence of the strong domination of $(\mathbf{N}^{\mathrm{dem},\epsilon})$ by $\mathbf{G}^{\mathrm{dem}}$.

Stable limit points of the demographic counting systems can be realized on the extended space $(\bar{\Omega}, (\bar{\mathcal{G}}_t)) = (\Omega \times \mathcal{A}^{2p}, (\mathcal{G}_t \otimes \mathcal{F}_t^{\mathcal{A}}))$, where $(\mathcal{F}_t^{\mathcal{A}})$ is the canonical filtration of $\mathcal{A}^{2p}$. The canonical variable on $\mathcal{A}^{2p}$ is denoted by $\bar{\mathbf{N}}^{\mathrm{dem}}(\omega, \alpha) = \alpha$.

By equivalence between tightness and stable relative compactness, there exists, for each sequence $(\mathbf{N}^{\mathrm{dem},\epsilon_k})$ converging in distribution, a subsequence (denoted with the same notations) converging stably to $\bar{\mathbf{N}}^{\mathrm{dem}}$ on $(\bar{\Omega}, (\bar{\mathcal{G}}_t)), \mathsf{R})$. $\mathsf{R} \in \mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$ is the limit rule, and the marginal of R on $\mathcal{A}^{2p}$ is the limit distribution of $(\mathbf{N}^{\mathrm{dem},\epsilon_k})$. In particular, the strong domination still holds at the limit and $\bar{\mathbf{N}}^{\mathrm{dem}} \prec \mathbf{G}^{\mathrm{dem}}$, R-a.s.. This representation of the limit on the extended space allows the structure of the initial problem to be maintained.

Joint stable convergence of $(\mathbf{N}^{\mathrm{dem}}, Z^{\epsilon})$ The main result of Section 3.4 is an identification result for stable limit points of $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^{\epsilon})$. The first step is to define both processes on the same probability space,since the population processes Z^{ϵ} are considered as $\mathbb{N}^p$-valued r.vs on the product space $\tilde{\Omega} = \Omega \times \mathsf{R}^+$, while the demographic counting systems $\mathbf{N}^{\mathrm{dem},\epsilon}$ are $\mathcal{A}^{2p}$-valued r.vs on Ω. However, the demographic counting systems can be extended without further difficulty to $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{\mathsf{P}})$ by setting $\mathbf{N}^{\mathrm{dem},\epsilon}(\omega, s) = \mathbf{N}^{\mathrm{dem},\epsilon}(\omega)$. Limit rules R in $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$ are extended to $\mathcal{R}(\tilde{\mathsf{P}}, \mathcal{A}^{2p})$ by $\tilde{\mathsf{R}}(\mathrm{d}(\omega, s), \mathrm{d}\alpha) = \mathsf{R}(\mathrm{d}\omega, \mathrm{d}\alpha)\lambda^e(\mathrm{d}s)$.

The family $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^{\epsilon})_{\epsilon}$ is stably relatively compact in $\mathcal{R}(\tilde{\mathsf{P}}, \mathcal{A}^{2p} \times \mathbb{N}^p)$. Let R^* be a limit rule of $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^{\epsilon})_{\epsilon}$. Due to a disintegration result on random kernels (see e.g. Theorem 1.25 in Kallenberg (2017)) and since $(\mathbf{N}^{\mathrm{dem},\epsilon}(\tilde{\omega}))$ does not depend on s, the limit rule can be rewritten as:

$$\mathsf{R}^*(\mathrm{d}\tilde{\omega}, \mathrm{d}(\alpha, z)) = \mathsf{R}(\mathrm{d}\omega, \mathrm{d}\alpha)\Gamma(\tilde{\omega}, \alpha, \mathrm{d}z).$$

In particular, if $(\mathbf{N}^{\mathrm{dem},\epsilon_k}, Z^{\epsilon_k})$ converges stably to R^*, then $(\mathbf{N}^{\mathrm{dem},\epsilon_k}(\omega))$ converges stably to R in $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$ and $(Z^{\epsilon_k}(\tilde{\omega}))$ converges stably to the R-expectation of Γ.

Thus, with $\bar{\mathbf{N}}^{\mathrm{dem}}(\omega, \alpha) = \alpha$ the canonical variable on $\Omega \times \mathcal{A}^{2p}$ as defined above, the stable

convergence of $(\mathbf{N}^{\mathrm{dem},\epsilon_k}, Z^{\epsilon_k})$ to R^* can be reinterpreted as follows:

$$\tilde{\mathsf{E}}[\mathbb{1}_{B\times[0,t]}h(\mathbf{N}^{\mathrm{dem},\epsilon})f(Z^\epsilon)] \underset{\epsilon\to 0}{\longrightarrow} \mathsf{R}[\mathbb{1}_B h(\bar{\mathbf{N}}^{\mathrm{dem}})\int_0^t \Gamma(\omega,s,\bar{\mathbf{N}}^{\mathrm{dem}},f)\lambda^e(\mathrm{d}s)], \tag{1.3.17}$$

for all $B\in\mathcal{G}$, $t\geq 0$, $h\in C_{cb}(\mathcal{A}^{2p})$ and $f\in C_b(\mathbb{N}^p)$[13]. Based on this representation, stable limit points of $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$ are denoted by (R,Γ).

Identification result We can now state the main result of the section, keeping the same notations as above:

Theorem 1.3.3 (Identification result)**.** *Let (R,Γ) be a limit point of the stably relatively compact family $(\boldsymbol{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$, where R is a rule in $\mathcal{R}(\mathsf{P},\mathcal{A}^{2p})$.*
Let us consider the process $(\Gamma(.,t,\bar{\boldsymbol{N}}^{\mathrm{dem}},\boldsymbol{\mu}^{\mathrm{dem}}))^o$ which is the $(\mathsf{R},\bar{\mathcal{G}})$-optional projection of process $t\mapsto\Gamma(.,t,\bar{\boldsymbol{N}}^{\mathrm{dem}},\boldsymbol{\mu}^{\mathrm{dem}})$.
Then, the canonical demographic process $\bar{\boldsymbol{N}}^{\mathrm{dem}}(\omega,\alpha)=\alpha$ is characterized by the following property:

$$\bar{\boldsymbol{N}}^{\mathrm{dem}} \text{ has the } (\mathsf{R},\bar{\mathcal{G}})\text{-compensator } \int_0^t (\Gamma(s,\bar{\boldsymbol{N}}^{\mathrm{dem}},\boldsymbol{\mu}^{\mathrm{dem}}))^o\mathrm{d}s. \tag{1.3.18}$$

Sketch of the proof: (1.3.18) is equivalent to the fact that $\bar{\mathbf{N}}^{\mathrm{dem}}_t - \int_0^t \Gamma(\omega,s,\bar{\mathbf{N}}^{\mathrm{dem}},\boldsymbol{\mu}^{\mathrm{dem}})\mathrm{d}s$ is a $\bar{\mathcal{G}}_t$-martingale. The proof of the theorem essentially relies on Equation (1.3.17) to prove that the $\mathcal{G}_t$-martingale property of $\mathbf{N}^{\mathrm{dem},\epsilon}_t - \int_0^t \boldsymbol{\mu}^{\mathrm{dem}}(s,Z^\epsilon)\mathrm{d}s$ still holds on the extended space by replacing $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$ by its stable limits.

Application to the case of Markov swaps

In last section of Chapter 3, Theorem 1.3.3 is applied to obtain the convergence in distribution of the demographic counting systems, in the particular case when the swap intensity functional is a *deterministic function* of the population process, $\boldsymbol{\mu}^s : z\in\mathbb{N}^p \to \boldsymbol{\mu}^s(z)\in\mathsf{R}^{p(p-1)}$. Thus, pure swap processes of intensity function $\boldsymbol{\mu}^s(z)$ are Markov processes. Under an ergodic assumption for pure swap processes associated with $\boldsymbol{\mu}^s$, we show in the main result of the section that the family of demographic counting systems converges in distribution towards an *averaged* jumps counting process of a pure *Birth Death system*. In particular, the dynamic of the aggregated population $Z^\natural$ is reduced to a Birth-Death process.
In the case when all intensities are deterministic and time homogeneous, an averaging

[13] Where $C_{cb}(\mathcal{A}^{2p})$ is the space of continuous bounded functions on $\mathcal{A}^{2p}$ and $C_b(\mathbb{N}^p)$ is the space of bounded functions on $\mathbb{N}^p$.

result for the demographic counting systems can be derived from Kurtz (1992), under an ergodic assumptions for the swap processes. In our framework, we rely on the results of the previous section and on the BDS decomposition algorithm to extend these results to the case where only swap intensities are deterministic.

Swap CTMC An example of realization of a Markov Swap is the pure Swap process X, solution of:

$$X_t = \xi_0 + \phi^s \odot \mathbf{N}^{\mathrm{sw}}_t, \quad \mathrm{d}N^{\mathrm{sw}}_t = \mathbf{Q}^s(\mathrm{d}t,]0, \boldsymbol{\mu}^s(X_{t^-})]).$$

X is a Continuous Time Markov Chain of intensity matrix:

$$L^{\mathrm{sw}} f(z) = \sum_{(i,j)\in\mathcal{J}^s} \mu^{ij}(z)(f(x+\phi(i,j)) - f(z)), \quad \forall x \in \mathbb{N}^p.$$

If the swap starts in state $\xi^n_0 \in \mathcal{U}_n$, where $\mathcal{U}_n$ is the state of populations of size n, then at all times $X_t \in \mathcal{U}_n$ (swap events don't change the size of the population). The following assumption is made on the restrictions of Markov swap processes to the spaces $\mathcal{U}_n$:

Ergodicity of the Swap process on $\mathcal{U}_n$(Assumption 1) $\forall n \geq 0$, The Swap CTMC restricted to $\mathcal{U}_n$ is assumed to be irreducible. Since $\mathcal{U}_n$ is finite, this means that the Swap CTMC restricted to $\mathcal{U}_n$ admits a unique stationary measure denoted by $(\pi(n, \mathrm{d}z))_{z\in\mathcal{U}_n}$.

Convergence of the demographic counting systems Due to Assumption 1, we show that at the limit, the intensity of the demographic counting systems $(\mathbf{N}^{\mathrm{dem},\epsilon})$ are *averaged* against the stationary measures of the Markov swaps. The demographic counting systems converge toward the jumps counting process of a true (non Markov) *Birth-Death system*, with intensities only depending on the size of the population:

Theorem 1.3.4 (Convergence of the demographic counting systems (Theorem (3.5.1))**.** *Let $(Z_0, \boldsymbol{N}^\epsilon, Z^\epsilon)$ be BDS system with deterministic swap intensity function. Under Assumption 1, the family of demographic counting systems $(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ converges in distribution in $\mathcal{A}^{2p}$.*

Furthermore, there exists an extension $(\bar{\Omega}, (\bar{\mathcal{G}}_t), \mathsf{R})$ of the given space $(\Omega, (\mathcal{G}_t), \mathsf{P})$and a multivariate counting process $\bar{\mathbf{N}}^{\mathrm{dem}} = (\bar{\mathbf{N}}^b, \bar{\mathbf{N}}^d) \in \mathcal{A}^{2p}$ defined on this space such that:

(i) $\bar{\mathbf{N}}^{\mathrm{dem}}$ *has the $(\bar{\mathcal{G}}_t)$-cumulative intensity:*

$$\mathcal{A}_t = \textstyle\int_0^t \pi(\bar{Z}^\natural_s, \boldsymbol{\mu}^{\mathrm{dem}}(\omega, s, \cdot))\mathrm{d}s, \quad \bar{Z}^\natural_s = Z_0 + \bar{N}^{b,\natural}_s - \bar{N}^{d,\natural}_s.$$

(ii) $(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ *converges to the distribution to* $\bar{\mathbf{N}}^{\mathrm{dem}}$.

It is important to note that at the limit, the aggregated process $\bar{Z}^\natural$ counting the number of individuals in the population is also a (one dimensional) Birth-Death process, which

was not the case before.
In the last part of Section 3.5, the result of the theorem is applied in the particular example of non linear swap intensities, coupled with linear (non deterministic) death intensities. More specifically, we show how non-linearities in the death intensity can emerge in the limit aggregated population, even when death intensities in each subgroup are linear. These non-linearities are the results of a non trivial aggregation of the subgroup specific death intensities, due to the swap events. Thanks to the approximation of the aggregated process by a simpler Birth Death process in the two timescale framework, we can see how swap events can modify the behavior of the population.

1.4 Part II: Overview and summary of the results

The results presented in Chapter 4 have been obtained as part of an ongoing collaboration with Héloïse Labit Hardy[14], Nicole El Karoui and Séverine Arnold (-Gaille)[15].

1.4.1 Introduction

In the second part of this thesis the impact of socioeconomic heterogeneity on aggregated mortality indicators is studied from an empirical point view. The widening of the socioeconomic gap in health and mortality has led a number of international organizations, governments, pension funds or insurance companies to rethink their models in order to tackle this heterogeneity issue and to understand the potential impact of socioeconomic inequalities. For instance the Life & Longevity Markets Association (LLMA) and the Institute and Faculty of Actuaries (IFoA) in the UK have recently commissioned a report on longevity basis risk[16]. The aim of this report was to assess the basis risk originating from using population level mortality to manage the longevity risk in pension benefits or annuitant liabilities (Haberman et al. (2014)). In the area of public health, Alai et al. (2017) have shown that the recommendations from the WHO targeting specific causes of death could actually increase life expectancy gaps in England, despite an increase of the national life expectancy.

A recent literature has taken an interest in the joint modeling and forecasting of the mortality of socioeconomic subgroups (see e.g Jarner and Kryger (2011); Villegas and Haberman (2014)), based on a relative approach. Individuals are grouped into so-called risk classes, assumed to have an acceptable level of homogeneity, and the mortality of risk classes is modeled with respect to the mortality of a large reference population, often the aggregated population. The impact of heterogeneity on cause-specific mortality has also been well-documented. However, there are still a lot of open questions on the consistency of sub-national and national estimations and forecasts. This subject was discussed very recently by Shang and Haberman (2017); Shang and Hyndman (2017), using methods based on recent developments in grouped functional time series methods.

However, the impact of composition changes on aggregated mortality indicators induced by the heterogeneous evolution of the population has been, to the best of our knowledge, less studied. Any changes in the composition of cohorts or age classes could lead to potential substantial changes in the age-specific mortality rates of the global population.

[14] ARC Centre of Excellence in Population Ageing Research (CEPAR), UNSW Australia, Sydney.

[15] Département de Sciences Actuarielles (DSA), Université de Lausanne.

[16] The basis risk is defined as the risk that policyholders experience a different longevity from a reference population, for instance the national population.

In particular, understanding the combined effect of changes in the composition of the population and of cause-specific reductions in socioeconomic subgroups cannot be done by the sole analysis of mortality rates time series.
Important recent releases of data allow a more detailed analysis of these composition effects. We argue that there is a real need and possibility to study the whole population and not only mortality rates to better understand the effect of heterogeneity on aggregate indicators. The population dynamics point of view allows us to represent the data differently than what is usually done when the focus is on mortality only, and to consider issues beyond those that are typically addressed.

Outline The aim of Chapter 4 is to illustrate possible impacts of changes in population composition, based on the English population and Cause-of-Death mortality data by deprivation level. In Section 4.2, we introduce the data used to carry out our study. Particular emphasis is being placed on the evolution of the age structure of the subpopulations grouped by level deprivation. In section 4.3, a simple deterministic of McKendrick-Von Foerster multi-population model is presented, for the modeling of age-structured heterogeneous population dynamics. Section 4.4 presents our numerical results, based on the model presented in the previous section. We first show how the heterogeneity of the initial age pyramid can impact the life expectancy and mortality improvement rates. In a second phase, we study how a cause of death reduction could be compensated by changes in population composition induced by heterogeneous birth rates.

1.4.2 How can a cause-of-death reduction be compensated in presence of heterogeneity? A population dynamics approach

In Section 4.2 of Chapter 4, we present the two databases used in this article. Particular emphasis is made on the evolution of the age structure of the subpopulations grouped by level deprivation. This is an important contribution of the study.

Databases

The data we use provide mid-year population estimates in England by age class and socio-economic circumstances for the years 1981-2015, combined with the number of deaths by age, cause and socioeconomic circumstances. The data have been obtained from two databases for the period 1981-2006 and 2001-2015. In particular, the last

database have been publicly released by the ONS in the UK very recently (April 2017).

In both databases, socioeconomic circumstances are measured by the Index of Multiple Deprivation (IMD). The IMD is a geographically based index, created in order to provide an official measure of multiple deprivation dimensions at the small area level of so-called LSOA[17]. The IMD is a measure of deprivation based on seven socioeconomic factors: income, employment, health, education, barriers to housing and services, living environment and criminality; and it is computed for each small living area in England. One advantage of the IMD is to provide information on the living environment (physical and social) of individuals, in addition to serve as substitute for individual socioeconomic characteristics.

Our two data sources are based on a relative measurement of deprivation. LSOAs are ranked by their IMD scores and grouped into five deprivation quintiles numbered from 1 to 5: IMD quintile 1 for the least deprived quintile to IMD quintile 5 for the most deprived quintile. Specific features of each database are summarized in Table 4.1 and 4.2 in Chapter 4.

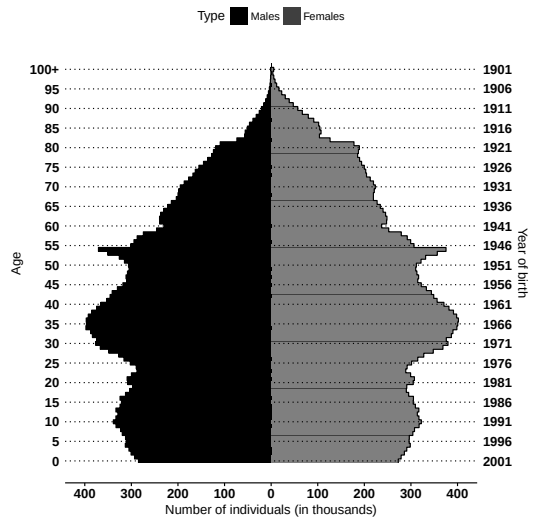
It is worth noting that in addition to updating the data for years 2007-2015, this second database provides disaggregated data by single year of age, as well as data for young ages below age 25. This constitutes an important improvement, in particular by facilitating significantly the interpretation of the graphical representations of data.

Evolution of the composition of the population

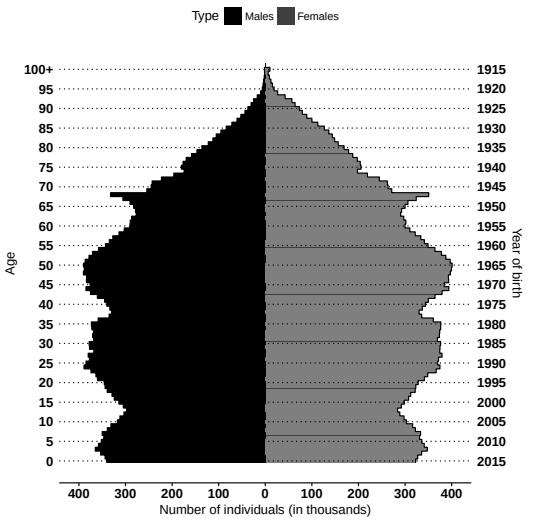
Following the presentation of the two databases, some insights are given on the important heterogeneity and evolution in time of the composition of the different age classes and cohorts. We summarize here some of the results, but we refer to Chapter 4 for a more complete description.

Age pyramids Figure 1.3 reproduces the age pyramids of the least and most deprived quintiles for years 2001 and 2015, along with England age-pyramid.

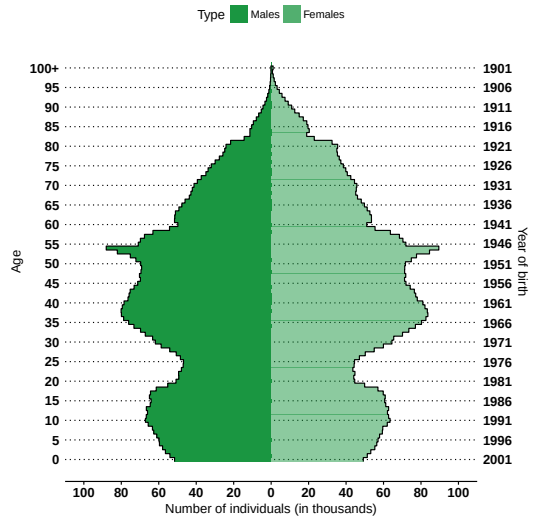
[17] In 2007 there were 32,482 LSOAs in England (34,753 in 2011), each composed of about 1500 individuals.



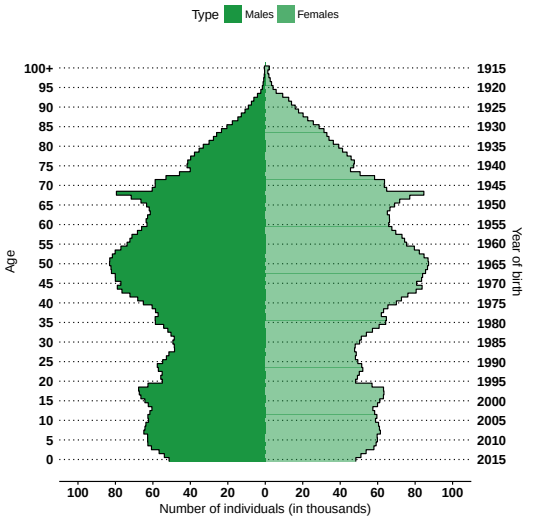
(a) England population, 2001



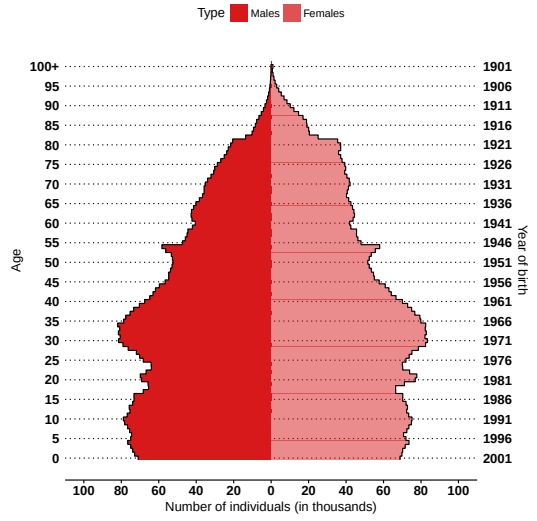
(b) England population, 2015



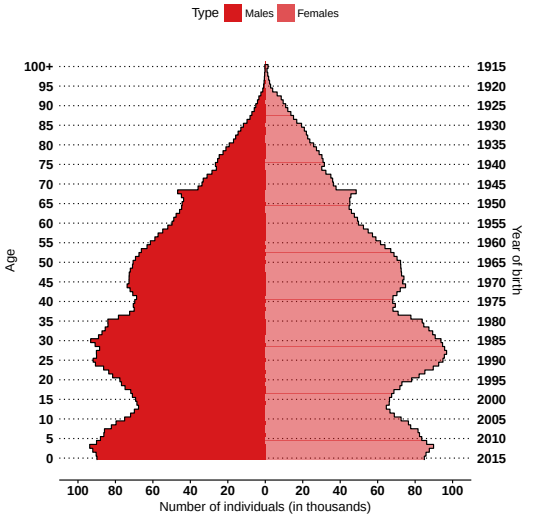
(c) Least deprived quintile ($$$$$), 2001



(d) Least deprived quintile ($$$$$), 2015



(e) Most deprived quintile ($), 2001



(f) Most deprived quintile ($), 2015

Fig. 1.3 Age pyramids in 2001 and 2015 (Figure 4.1)

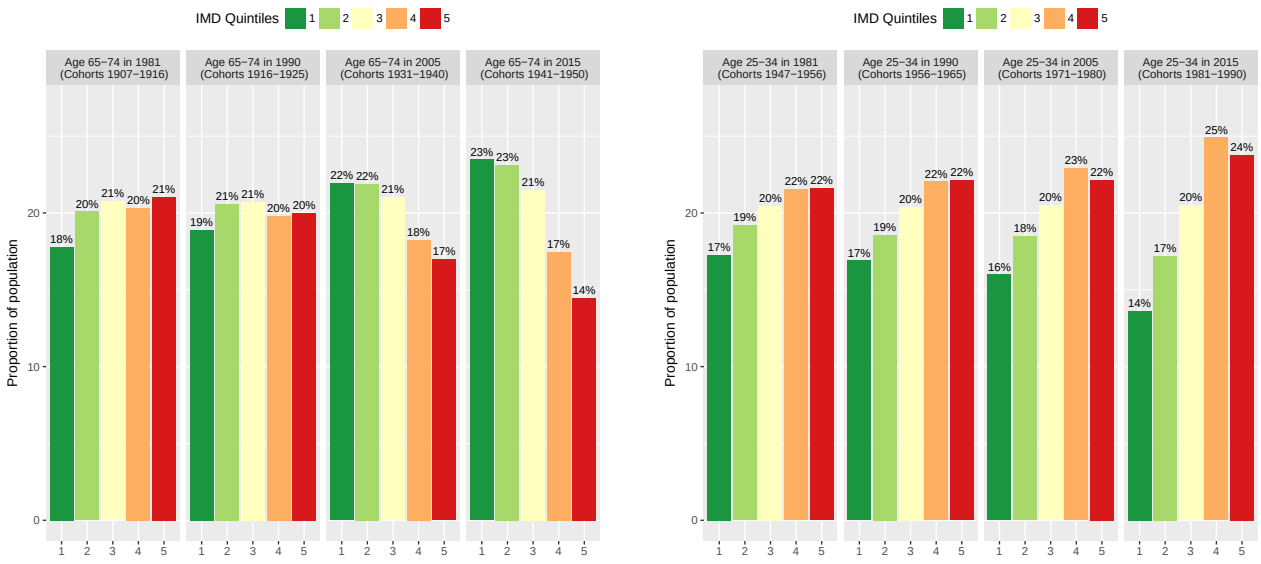
The vertical reading of Figure 1.3 shows for each point in time, the form of the age pyramids are very different among the deprivation quintiles and England population. The most deprived population (IMD quintile 5), (Figure 1.3e and 1.3f) is much younger on average than the least deprived population (IMD quintile 1), (Figure 1.3c and 1.3d), with a median age of 33 years, ten years younger than for the least deprived subpopulation. We also note that the baby-boom generation (born in the years after the World War II) and their children are more largely represented among the least deprived subpopulation. The horizontal reading Figure 1.3 shows that significant temporal changes have also occurred from 2001 to 2015, with different effects according to the deprivation level. As an example, the median age in the most deprived quintile has decreased of over 1%, from 33.4 to 33 years, while it has increased of almost 10% in the least deprived subpopulation, from 40 to 44. From 2001 to 2015, the most deprived population has thus rejuvenated, despite the general tendency of the ageing of the population.

There is an inherent difficulty in representing the time evolution of data structured in age. In order to better understand the evolution of the composition of the population, the evolution over the period 1981-2015 of the composition of specific *age classes* (the same age class is represented at different dates) and *cohorts* (the same cohort is represented at different dates) are also represented in Chapter 4.

Fixed age class Figure 1.4 represents the evolution of the distribution of males in each IMD quintile, for the years 1981, 1990, 2005 and 2015 and for two fixed ages classes: 65-74 (Figure 1.4a) and 25-34 (Figure 1.4b).

The composition of the age class 65-74 has significantly varied from 1981 to 2015, to the benefit of the least deprived populations. This could be explained by an improvement over time of living conditions for older individuals, but also, as noted above, by a baby-boom cohort effect. Indeed, the individuals born during the English baby-boom are less deprived than the immediately preceding and following cohorts, regardless of the global improving trend.

Younger age classes are more deprived than older ones, this being true for the whole period 1981-2015. However, the relative deprivation of younger age classes has increased with time.



(a) Age class 65-74 (b) Age class 25-34

Fig. 1.4 Proportion of males by age class and IMD quintile (Figure 4.2)

Fixed cohorts Looking at the data structured confirms the observations made on Figure 1.4 (see Figure 4.3 in Chapter 4, which is not reproduced in this introduction). By comparing the evolution of the cohorts born in 1956-1960 and in 1976-1980, we observe that not only the younger cohort (1976-1980) was more deprived at age 25-29 (in 2005) than the older one (age 25-29 in 1985), but also that the improvement over time of the deprivation of the older cohort is more important than the improvement of the deprivation of the younger one.

To summarize, the population data show an important heterogeneity in the composition of the different age classes, associated with important temporal changes in the age classes composition, such as the important evolution of the composition of the 65-74 age class between 1981 and 2015. Mortality rates are affected differently according to the age or time. Thus this varying heterogeneity generates additional complexity in the study of aggregated death rates. In particular, one might wonder how the increase of deprivation observed among younger cohorts will impact future mortality in England.

Mortality data

At the end of the second section of Chapter 4, mortality data by gender, deprivation quintile and cause of death is also represented, with central death rates (Figure 4.4), average annual mortality improvement rates (Figure 4.5 and 4.6), distribution of causes

of deaths (Figure 4.7) and life expectancy evolution (Figure 4.8). The results, which are more classical, are not reproduced here. Central death rates and mortality improvement rates for the period 1981-2015 have been higher in the most deprived quintiles, at all ages and for both males and females. It is worth noting that annual average mortality improvement rates are rather different for the period 1981-1995 and 2001-2015 (Figure 4.6). In particular males of the most deprived quintile have experienced over the period 2001-2015 a sharp deceleration of mortality improvement at ages above 60. All indicators support the thesis of a widening of socioeconomic gaps (see e.g. Figure 4.8).
The distribution of causes of death has also varied significantly during the period 1981-2015. In 1981, circulatory diseases were the first cause of death for the age class 65-85, whereas it was neoplasm in 2015. We also note the significant difference in the prevalence of respiratory diseases between the most and the least deprived quintile (Figure 4.7).

Population dynamics model

We begin in this first approach with simple assumptions on the evolution of the population dynamics. We show how mortality patterns can be impacted by heterogeneity even in a simple framework.

Evolution of the subpopulations The joint evolution of subpopulations inside a global population (IMD quintiles in our case) is modeled by an age-structured McKendrick-Von Foerster multi-population model, with no interactions. One of the advantages of this framework is to allow the derivation of closed formula and asymptotic results for a number of indicators, in order to better understand the complex interactions between heterogeneity and mortality.
The vital parameters of the model are the birth and mortality rates of each subpopulation and are assumed to be deterministic functions of age and time. The first assumption can actually be extended without difficulty to the broader scope of stochastic intensities, depending for instance on a random environment. The other assumptions are discussed at the beginning of the third section.
Each subpopulation $j = 1...p$ is described by the function $(g_j(\epsilon, a, t))$, where $g_j(\epsilon, a, t)$ is the number of individuals of gender $\epsilon = m$ of f, between age a and $a + \mathrm{d}a$ at time t and in subpopulation j. The evolution of the population is given by the transport partial

differential equation:

$$(\partial_t + \partial_a) g_j(\epsilon, a, t) = -\mu_j(\epsilon, a, t) g_j(\epsilon, a, t), \ \forall a, t > 0 \qquad \text{(balance law)} \quad (1.4.1)$$

$$g_j(\epsilon, 0, t) = p^\epsilon \int_0^{a_\dagger} b_j(a, t) g_j(f, a, t) \mathrm{d}a, \qquad \text{(birth law)} \quad (1.4.2)$$

$$g_j(\epsilon, a, 0) = g_0^j(\epsilon, a), \qquad \text{(initial population)}$$

where μ_j and b_j are the mortality and birth rates in subpopulation j.
The previous equation is usually solved along its characteristics curves, or equivalently by following the evolution of the population cohort by cohort. The resolution method can be interpreted as counting the number of survivors at time t in each cohort, with two distinct regimes: individuals who were present in the initial population and individuals born after the initial time.
Thus, at a small time t, the age pyramid is mostly shaped by the time translated initial age pyramid. On a longer term, the initial population is naturally erased and the shape of the future age pyramid is only characterized by the birth and survival functions.

Aggregated population We call aggregated population the global population composed of all subpopulations $g(\epsilon, a, t) = \sum_{j=1}^{p} g_j(\epsilon, a, t)$. The dynamics of the aggregated population is derived from the dynamics of the subpopulations,
$(\partial_t + \partial_a) g(\epsilon, a, t) = -\sum_1^p \mu_j(\epsilon, a, t) g_j(\epsilon, a, t), \quad g(\epsilon, 0, t) = p^\epsilon \int_0^{a_\dagger} (\sum_1^p b_j(a, t) g_j(f, a, t) \mathrm{d}a)$.
The mortality rate at age a and time t of the aggregated population is thus

$$d(\epsilon, a, t) = \sum_{j=1}^{p} \mu_j(\epsilon, a, t) w_j(\epsilon, a, t), \ \text{with} \quad \omega_j(\epsilon, a, t) = \frac{g_j(\epsilon, a, t)}{g(\epsilon, a, t)}. \qquad (1.4.3)$$

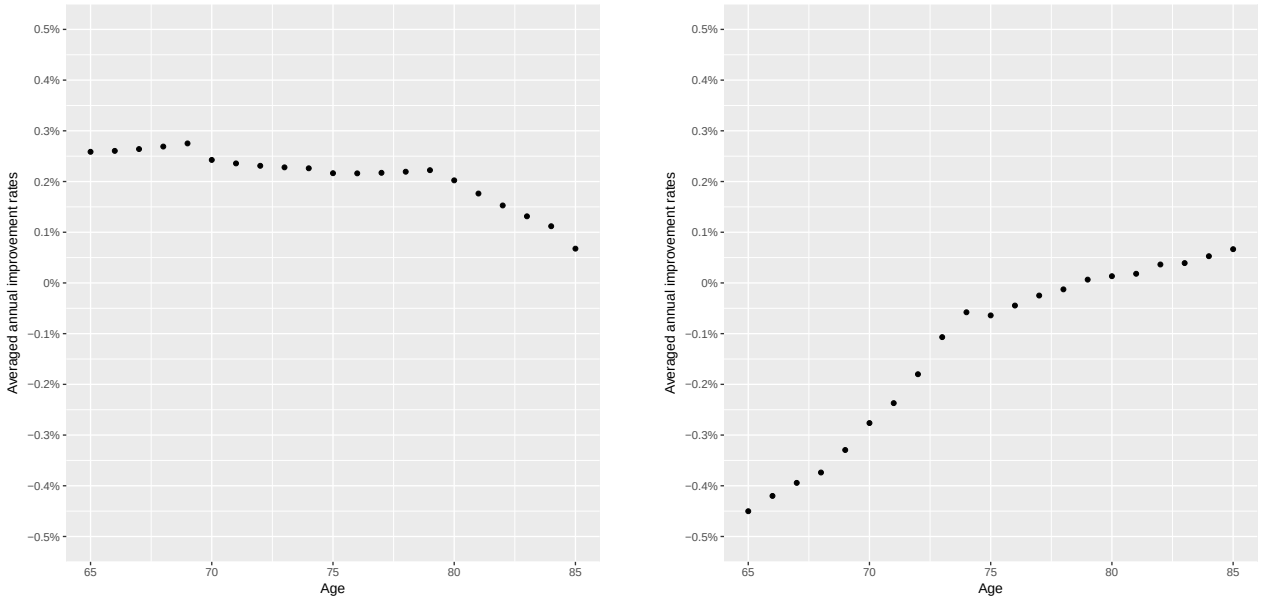
Actually, the mortality intensity in the aggregated population should be denoted by $d(\epsilon, a, t, (g_j)_{j=1..p})$, since it depends on the age pyramids of all subpopulations. By simple aggregation of linear dynamics, a non-linear dynamic is produced at the aggregated level, whose mortality rates are strongly dependent on the age pyramids of each subpopulation. At small times t, the weights $\omega_j(\epsilon, a, t)$ are mostly determined by the initial population and the cohort survival function. On a longer term, they are mainly governed by the birth patterns. Detailed formula for each regime are presented in Section 4.3.

Numerical results

In the last section of Chapter 4, the previous model is applied to show the impact of heterogeneity in two different cases: on the short term in Section 4.4.1 (influence of the

initial pyramid) and on a longer term (influence of birth patterns) in Section 4.4.2. For illustrative purposes, an heterogeneous population composed of only the most and least deprived IMD quintiles is considered.
In order to isolate the influence of changes in the population composition, we assume that mortality rates in each IMD quintile do not depend on time. The aggregated death rate is thus only modified through changes in the population composition.
In different applications, the initial age pyramids and mortality rates are estimated in each scenario for a given year from the data presented in Section 4.2.2. In the baseline scenario of Section 4.4.2, birth rates are assumed to be the same in each IMD quintile, and are estimated from England data available on the ONS website.

Influence of heterogeneity on the short-term In Section 4.4.1, the population is simulated on the short-term (30 years) in two scenarios: initial pyramids and mortality rates of year 1981 and initial pyramids and mortality rates of year 2015. The evolution of the composition of the aggregated population is shown in Figures 4.9 and 4.10, which are not reproduced in this introduction.
We observe the evolution of the period life expectancy at age 65 (Figure 4.11), along with average annual mortality improvement rates at ages above 65 which are reproduced in Figure 1.5 below. All computed indicators concern individuals who were of age 35 and above in the initial pyramids and only evolve due to changes in the age pyramids of the IMD quintiles. Birth rates have no influence here.
The results show that indicators move in opposite directions, according to the year chosen for the initial pyramids. For year 1981, the evolution of the age pyramids in the least and most deprived IMD quintiles contributes positively to the evolution of the period life expectancy at age 65, and average annual improvement rates are positive. On the contrary, the evolution of the age pyramids for year 2015 contributes negatively to the improvement of the aggregated mortality, supporting the observations made on the second section of more deprived younger cohorts.



(a) 1981 data

(b) 2015 data

Fig. 1.5 Males averaged annual improvement rates (Figure 4.12)

Influence of heterogeneity on the long term The effect of cause-specific reductions is often studied, especially when assessing or choosing public health measures. It is less common to study the impact of heterogeneity in birth rates. In the last part of Chapter 4, we study the influence on aggregated life expectancy of heterogeneity in birth patterns, in comparison with a cause reduction occurring in the most deprived IMD quintile, for instance as a consequence of a targeted public action. The effects of birth heterogeneity (a more important birth rate in the most deprived subpopulation) and of cause-specific reductions are first presented separately in Figure 4.14. Finally, we study the combined effect of both birth and mortality changes. In particular, we show how a cause-specific reduction can be compensated in presence of heterogeneity (Figure 4.15 and 4.16). Thus, if only the aggregated data are studied, a cause specific reduction could be missed in the presence of heterogeneity.

1.5 Part III: Overview and summary of the results

The survey presented in Chapter 5 has been written in collaboration with Nicole El Karoui and Kaouther Hadji[18].

1.5.1 Introduction

The last part of this thesis is a cross-disciplinary survey of selected topics on the evolution of human longevity and demographic models. Facing the considerable amount of literature, data and points of view concerning the evolution of human longevity and populations in the last two century, we came to the conclusion that it was necessary to bring out a number of key observations to avoid the pitfalls of an overly naive approach.
The goal of the survey presented in Chapter 5 is to help a modeler of human population dynamics to find a coherent way (for instance by taking into account the whole population dynamics and not only old ages) around this mass of multidisciplinary information. Based on numerous surveys from various academic disciplines and many contradictory readings, we propose a subjective selection of what we believe to be the most important ideas or facts, from a mathematical modeling perspective.
Whenever possible, the key ideas are illustrated with examples supporting the intuition about mentioned phenomena. However, if the discussion is greatly enriched by the multidisciplinary nature of the field, the presentation of ideas is also made more difficult, especially for matters of vocabulary. It should also be noted that issues related to medical advances and to the biology of human ageing are dealt with in a very cursory way in Chapter 5, as we focus mainly on economic and social issues.

Outline Chapter 5 is composed of three main parts which are briefly summarized in the next subsection. The first part deals with the historic demographic transition. The importance of public health is studied, with a specific focus on the cholera epidemic outbreaks that took place in France and in the UK during the nineteenth century. Other features of the historic demographic transition are also considered. In particular, we explore the relationship between the economic growth and mortality improvements experienced during the past century.
In the second part of Chapter 5, we examine the implications for population modeling and the key feature of this shift in paradigm that has been observed since the 2000s. We first give a brief overview of the so-called demographic transition, and the move toward the description of increased complexity and diverging trends that has recently

[18]LPMA, Université Pierre et Marie Curie, Paris

been observed, based principally on the experience of developed countries. A special attention is paid to socioeconomic differences in health and mortality.
The last part of Chapter 5 is dedicated to a short review of two types of demographic modeling exercises: microsimulation models and agent based models.

1.5.2 Inextricable complexity of two centuries of worldwide democratic transition, a fascinating modeling challenge

The historic demographic transition

In order to understand this "new modeling paradigm", it appeared necessary to understand first of all the main components of the historic demographic transition, which began in the nineteenth century in most European countries and countries with European roots, and was completed in over a century ($\sim (1850 - 1960)$). This historical process is mainly refered to as "the secular shift in fertility and mortality from high and sharply fluctuating levels to low and relatively stable ones" (Lee and Reher (2011)).
Infectious diseases constituted the vast bulk of the historic fall in mortality. The cholera epidemic of the nineteenth century is often viewed as a starting point of the demographic transition as well the real spark which lit the tinder of the global public health movement. Public health measures are considered as one a main determinants of the reduction of the mortality reduction during the first half of the twentieth century (Bloom and Luca (2016); Cutler et al. (2006)). The management of the crisis of the cholera, in France and England, but also at the international level, is very instructive on the contexts that enable the development of public measures. The nineteenth century was also a period of major upheavals brought about by the industrial revolution, which shaped many aspects of the contemporary society. There is a remarkable persistence of certain mechanisms, with for instance the important time delays (of about one or two generations) in public health response, which has not changed since the nineteenth century.
The cholera is also an iconic example where medicine was confronted to statistics, with the remarkable work W. Farr, "the architect of England's national system of vital statistics" (Eyler (1973)). The use of statistics made by Farr contributed to a better understanding of the disease. Yet the ambition to find causal factors by the sole analysis of data led to some important errors in interpretations, which is somehow reminiscent of the current challenges of this new " data science era".

The twentieth centurywas also the century of "the emergence for the first time in history of sustained increases in income per head" (Canning (2011)). The links between economic growth and mortality improvements have been extensively discussed by economists. In

his seminal article, Preston (1975) showed the existence of a curvilinear relation between life expectancy at birth and national income per head. But he also showed that economic growth was far from explaining of all the observed mortality decline. Furthermore, the relationship between demographic changes and level of income is often thought of as bidirectional, which makes interpretation very difficult. This issue still generates a lot of debate. Some points of view on these topics are presented at the end of the second section.

A new era of diverging trends

In the early 1970s, many demographers and population scientists supported the idea that populations would ultimately reach the last stage of the classical demographic transition, described as an "older stationary population corresponding with replacement fertility (i.e., just over two children on average), zero population growth, and life expectancies higher than 70years" (Lesthaeghe (2014)). Yet, it turned out that this state of equilibrium and homogeneity in populations was never realized. Actually, fertility rates remained too low to ensure the replacement of generations; mortality rates, especially at advanced ages, declined at a faster rate than ever envisaged; and contemporary societies seem to be defined by more and more heterogeneity and diverging trends. A the beginning of the third section, we present a brief overview of the main components of this second demographic transition, as formulated by Lesthaeghe and Van de Kaa.

The viewpoint of the second demographic transition shed an interesting light on recent observations of diverging trends in longevity, which is the main focus of the second part of the section. As the average life expectancy has been rising unprecedentedly, gaps have also been widening at several scales. What may be somehow surprising is that up until the 1980s, high income countries had roughly similar life expectancy levels. For example, the comparison of the female life expectancy at age 50 in ten high income countries showed that the gap was of less than one year in 1980. By 2007, the gap had risen to more than 5 years (National Research Council and Committee on Population (2011)).
On another scale, a great amount of evidence shows that socio-economic differentials have also widened within high income countries (Elo (2009)). If there is a general consensus on the pervasive effects of the socioeconomic status on health and longevity, underlying mechanisms are much less precisely understood, with several competing theories. Furthermore, the impact of this heterogeneity on the aggregated indicators or in term of public policy is even more complicated.
Thus, the aim of the last part of the section point out some aspects of the complexity in understanding current longevity trends, which cannot be disentangled from the evolution

of the whole population, requires a multiscale analysis of phenomena, with in the mean time keeping in mind that obtaining comparable and unbiased data is also a major challenge for explaining longevity.

Modeling complex population evolution

In the last section of Chapter 5, we present two types of "micro" population models, widely used in social sciences and in demographic applications: microsimulation models and agent based models.
Since the 60s, microsimulation models have been extensively used by governments and international (or regional) organizations, with the aim to test policy reforms or to evaluate pension costs for instance. Several examples are given in the Chapter 5. The main feature of microsimulation models is to study heterogeneous individual trajectories in order to obtain macro outcomes by aggregation, in the form of a data-driven complex model. As stated at the beginning of this introduction, these models often rely on vast amounts of data. Furthermore, there are mostly linear and cannot include many interactions.
In the last part of the section, we give a short review of agent based models. Agent based model allow the modeler to take into account individual interactions, in order to explain macroscopic regularities. In this class of models, agents are described by behavior rule, determining how they interact with the simulated environment. One of the major difficulty of agent based models is to defined these behavior rules.
After a discussion on microsimulation and agent based models, we present some very recent application combining the two types of models, and developed for instance by (Zinn (2017)).

Chapter 2

Pathwise representations of BDS systems

2.1 Introduction

The recent years have seen a renewed interest for the modeling of heterogeneous population dynamics, motivated in part by the study ecological systems. In these systems, the behavior of individuals is often influenced by a certain family of *characteristics*, also called *traits*. Examples of charactertistics are the the location of individuals or their level of agressiveness. For instance, Fournier and Méléard (2004) study a spatial ecological system in which the demographic behavior of individuals (birth and death rate) is determined by their location. Auger et al. (2000) considers a deterministic model population of fishes divided in two spatial patches. The first one contains food and resources but is also exposed to parasites attacks, while the second one is safe from parasitism but does not provide enough resources. Other examples are given in Auger and Pontier (1998); Marvá et al. (2013), in which individuals demographic rates are affected by the level of aggressiveness they deploy in order to have access to resources.
An important difference between Fournier and Méléard (2004) and Auger et al. (2000) is the possibility to move from one patch to another. In the former article, individuals cannot change of location (for instance in the case of plants). In the latter, fishes move *frequently* between the two patches, which adds significant complexity to the population dynamics. Furthermore, the behavior of individuals is also often influenced by interactions with other individuals, resulting in the introduction of *non-linearity* or *density dependence* in the population dynamics.
In human populations, it is also well-known that certain characteristics have a significant behavior on demographic rates. Gender is a key determinant of longevity, with differences

in males and females life expectancy at birth of up to six years. Socioeconomic differentials are also important. In particular, the neighborhood is considered to be an accurate predictor of socioeconomic status (see Chapter 4 for more details on this issue). This presents a difficulty since geographical characteristics cannot be taken into account in the same way as stable characteristics such as gender. Indeed, the main issue when modeling two-sex population is to maintain the stability of the composition of the population while changes in the composition of the population are determinant in the case of socioeconomic characteristics. In addition, human demographic intensities are characterized by a strong non stationarity and often depend on additional randomness expressing the variability of the global environment.

In this chapter, we study a general class of heterogeneous population dynamics structured by discrete characteristics, called *Birth Death Swap (BDS) systems*. Two types of events can modify the composition of the population: a *demographic events* describes a birth or death, and a *swap event* describes a change of characteristics of an individual. In particular, no specific assumptions are made to explain mechanisms responsible for swap or demographic events, so that the studied dynamics are very general.

Usually, the temporal evolution of such processes is assumed to be Markovian. When only demographic events happen, the process is known as the classical spatial (or multi-type) Birth-Death processes. When only swap events happen, the process is a Continuous Time Markov Chain (CTMC) describing the evolution of a population of constant size. These classes of Markov processes are very usual in many counting problems. A less common approach is to consider both dynamics together in a non Markovian framework, which is our setup. The age-structure of the population is however not taken into account in the modeling, due to the complexity already introduced by swaps. To the best of our knowledge, the terminology Birth Death Swap has been introduced in discrete time by Huber (2012) for the very different purpose of generating random variables as stationary distribution of pure jumps processes.

In Section 2.2, the setup of the model is presented. Our presentation gives an algebraic decomposition of the so-called population process, based on the multivariate process counting the number of events occurring in the population, called the *jumps counting process*, and a generalized initial condition, called *entry process*. In particular, some useful transformations on the population can be made without further assumptions on the distribution of the process.

In Section 2.3, we give a pathwise representation of jumps counting process, as the solution of a multivariate Stochastic Differential System (SDE) driven by an extended Poisson measure. The existence of BDS systems is derived from a more general result

obtained on the construction of multivariate counting processes by *strong domination* with a non-exploding process. The construction by strong domination offers many advantages, and for instance "free" tightness properties are obtained in Chapter 3 thanks to this construction.
In Section 2.4, an alternative construction of BDS systems is presented, called the *Birth Death Swap decomposition algorithm*. The decomposition algorithm allows us to disentangle swap events from demographic events, for instance when they are supposed to have their own timescale.

2.2 Birth Death Swap systems

2.2.1 Setup

We describe the evolution of a general class of heterogeneous population, in which individuals differ by a finite number of characteristics. The population is structured by discrete subgroups indexed by $1..p$, such that individuals in the same subgroup share the same set of characteristics and have similar demographic behaviors. The diversity of the population is measured by the size of the different subgroups. At a given time, the quantity of interest is thus the vector of the number of individuals in each subgroup, also called the *state of the population*.
Random changes in the population occur at random dates and are described by the move of *a single individual* at a time. Different types of events can modify the composition of the population. A *demographic event* is a birth (or entry) or death in a given subgroup. A *swap event* happens when a individual changes of characteristics, resulting in a move from one subgroup to another. No information is given on the "origin" of events. "Parents" for newborns are not specified, and interactions leading to a change of characteristics or death are not known. This description, only based on changes in the population composition allow us to include a large class of dynamics in the model.
The assumption that individuals in the population differ only by a finite number of characteristics is introduced for ease of exposition, and can easily be extended to a countable set of characteristics.

State space

Before introducing the model, let us introduce some algebraic notations in order to simplify the description of the state space of the population process.
NOTATIONS: The vector space $\mathbb{Z}^p$ is equipped with the canonical basis $(\mathbf{e}_i)_{1\leq i\leq p}$, with

$\mathbf{e}_i = {}^t(0, \cdots, 0, 1_i, 0)$. The usual scalar product is denoted by $<,>$,

$$z = {}^t(z_1, .., z_p), \quad z_i =< z, \mathbf{e}_i > \quad \text{and} \quad z = \sum_{i=1}^{p} < z, \mathbf{e}_i > \mathbf{e}_i.$$

The unit vector is denoted by $\mathbf{1} = {}^t(1, .., 1)$. Finally, the notation $z^\natural =< \mathbf{1}, z >= \sum_{i=1}^{p} z_i$ will often refer to the size of a "population" vector $z \in \mathbb{N}^p$ in the sequel.

Population state space: The population is divided in p subgroups of indistinguishable individuals sharing the same set of characteristics, for instance living in the same patch or neighborhood, or with the same level of income. Individuals in the same subgroup are assumed to have the same demographic or "swap" behavior, and the subgroups are indexed by the set $\mathcal{I}_p = \{1, 2, \cdots, p\}$. The state of the population is described by the number of individuals in the each subgroup, that is by a vector of $\mathbb{N}^p$, considered as a subset of $\mathbb{Z}^p$.

Elementary events representation.

We first described the events which can happen in the population. The analysis of moves (or events) in the population is inspired by the framework of Preston (1975) for spatial birth death process and Huber (2012) for birth death swap process in discrete time, but with continuous state space of characteristics. We give a more unified description, based on the observation that swap events and demographic events have a strong similarity.

a) A *swap event* from i to j ($j \neq i$) is a move of an individual from subgroup i to the subgroup j. Observe that since individuals in the same subgroup are indistinguishable, a swap move can also be interpreted as the *simultaneous* removing of an individual taken randomly in subgroup i and addition of an individual in subgroup j. This swap event linking at the same time the two subgroups i and j is indexed by the ordered pair $\kappa = (i, j)$.

A swap event from i to j can only occur if the subgroup i is not empty, $z_i \neq 0$. In this case, the change induced in a population $z = {}^t(z_1, .., z_p) \in \mathbb{N}^p$ affects the subgroup j whose size becomes $z_j + 1$ and the subgroup i whose size becomes $z_i - 1 \geq 0$. The new population composition is $\epsilon^{i,j}(z) = z + (\mathbf{e}_j - \mathbf{e}_i)$. The family of the *swap events* is indexed by the set $\mathcal{J}^s = \{\kappa = (i, j) |\, i, j \in \mathcal{I}_p, i \neq j\}$.

b) A *demographic event* is a *birth or death* in a subgroup. An event of type "birth in subgroup j" means that an individual is added to the subgroup j, inducing a change $z_j \mapsto z_j + 1$ in the jth subgroup. Then, after a birth event in subgroup j, the population is $\epsilon^{b,j}(z) = z + \mathbf{e}_j$. There is some ambiguity in using the word "birth" in this context since no information is given on the "parents" of the newborn, and a birth event is not a priori endogenous to the population.

An event of type "death in subgroup i" can only occur if the subgroup is not empty, and means that an individual is removed from subgroup i, inducing a change $z_i \mapsto z_i - 1$ in the ith population, if $z_i > 0$. Then, after a death event in subgroup i, the population is $\epsilon^{d,i}(z) = z - \mathbf{e}_i$. The families of birth events and death events are respectively by indexed by the sets $\mathcal{J}^b = \{(b,j)|\, j \in \mathcal{I}_p\}$ and $\mathcal{J}^d = \{(d,i)|\, i \in \mathcal{I}_p\}$. The set of all demographic events is denoted by $\mathcal{J}^{\text{dem}} = \mathcal{J}^b \cup \mathcal{J}^d$.

In order to have a unified description of swap and demographic events, it is convenient to add an "hypothetical" infinite subpopulation, denoted by $\{\infty\}$. All functions are assumed to be null on $\{\infty\}$, and the vector $\mathbf{e}_\infty$ is the null vector $\mathbf{e}_\infty = {}^t(0, \cdots, 0)$. Then, the event "birth in subgroup j" may be viewed as a swap event $(\infty, j) = (b, j)$ from ∞ to j, with $\epsilon^{b,j}(z) = \epsilon^{\infty,j}(z) = z + \mathbf{e}_j - \mathbf{e}_\infty$; and the event "death in the no empty subgroup i" may be viewed as a swap event $(i, \infty) = (d, i)$ from i to ∞, with $\epsilon^{d,i}(z) = \epsilon^{i,\infty}(z) = z + \mathbf{e}_\infty - \mathbf{e}_i$.

The indexing sets are modified as follows to include this additional subgroup:

- Characteristics set: $\mathcal{I} = \mathcal{I}_p \cup \{\infty\}$ with the additional notation $\mathcal{I}^{(i)} = \mathcal{I} \setminus \{i\}$.
- Set of all events: $\mathcal{J} = \{\gamma = (\alpha, \beta); \alpha \in \mathcal{I}, \beta \in \mathcal{I}^{(\alpha)}\}$.
- Set of swap events: $\mathcal{J}^s = \{\kappa = (i,j); i \in \mathcal{I}_p, i \in \mathcal{I}_p^{(i)}\}$.
- Set of demographic events: $\mathcal{J}^{\text{dem}} = \{(\infty, j);\, j \in \mathcal{I}_p\} \cup \{(i, \infty, j);\, i \in \mathcal{I}_p\}$.
- $\text{card}(\mathcal{I}) = p + 1$, $\text{card}(\mathcal{J}) = p(p+1)$, $\text{card}(\mathcal{J}^s) = p(p-1)$, $\text{card}(\mathcal{J}^{\text{dem}}) = 2p$.

General event notation and calculation

We close this section with some simple algebraic calculations. In particular, the focus is shifted from the population state space to the space of *counting vectors*, defined as integer-valued vectors indexed by the set of all events $\mathcal{J}$. These calculations will be particularly useful in the following sections, in which the population will be represented using a larger vector counting the number of events in the population.

Counting vectors are $\mathbb{N}^{p(p+1)}$ vectors indexed by $\mathcal{J}$, and denoted in the following by $\boldsymbol{\nu} = (\nu^\gamma)_{\gamma \in \mathcal{J}}$. For $\gamma \in \mathcal{J}$, ν^γ should be interpreted as the number of events of type γ which occurred in the population. It can be interesting to isolated the swap part and the demographic part. Thus, we also write $\boldsymbol{\nu} = (\boldsymbol{\nu}^s, \boldsymbol{\nu}^{\text{dem}})$, where $\boldsymbol{\nu}^s = (\nu^\kappa)_{\kappa \in \mathcal{J}^s}$ is the swap part of $\boldsymbol{\nu}$, and $\boldsymbol{\nu}^{\text{dem}} = (\boldsymbol{\nu}^b, \boldsymbol{\nu}^d)$ is the demographic part, with $\boldsymbol{\nu}^b = (\nu^{b,i})_{i \in \mathcal{I}_p}$ and $\boldsymbol{\nu}^d = (\nu^{d,i})_{i \in \mathcal{I}_p}$. Sometimes, we also write $\boldsymbol{\nu}^s + \boldsymbol{\nu}^{\text{dem}} = \boldsymbol{\nu}$ by abuse of notation.

Jumps For any $\gamma = (\alpha, \beta) \in \mathcal{J}$, $\phi(\gamma) = \mathbf{e}_\beta - \mathbf{e}_\alpha$ is the $\mathbb{Z}^p$-valued vector representing the *jump* of the population associated with the event of type $\gamma = (\alpha, \beta)$. As stated in the

previous paragraph, the state of the population after event γ is $\epsilon^{\alpha,\beta}(z) = z + \phi(\gamma) = z + \mathbf{e}_\beta - \mathbf{e}_\alpha$. In particular, $\phi(b,j) = \phi(\infty,j) = \mathbf{e}_j$, $\phi(d,i) = \phi(i,\infty) = -\mathbf{e}_i$.
ϕ can be seen as the $(p, p(p+1))$ *matrix* with columns $\phi(\gamma)$. With obvious notations, we write $\phi = \phi^s + \phi^{\text{dem}}$, as for counting vectors. Let us note that the matrix $\phi^{\text{dem}} = (\phi^b, \phi^d)$ is very simple since ϕ^b is the p identity matrix and $\phi^d = -\phi^b$.

FREQUENTLY USED CALCULATIONS For a counting vector $\boldsymbol{\nu}$ indexed by $\mathcal{J}$, the notation $\phi \odot \boldsymbol{\nu} \in \mathbb{Z}^p$ denotes the matrix product of ϕ and $\boldsymbol{\nu}$,

$$\phi \odot \boldsymbol{\nu} = \sum_{\gamma \in \mathcal{J}} \phi(\gamma) \boldsymbol{\nu}^\gamma = \phi^s \odot \boldsymbol{\nu}^s + \boldsymbol{\nu}^b - \boldsymbol{\nu}^d, \quad (\phi \odot \boldsymbol{\nu})_i = \sum_{\alpha \in \mathcal{I}^{(i)}} \nu^{\alpha,i} - \sum_{\beta \in \mathcal{I}^{(i)}} \nu^{i,\beta}. \quad (2.2.1)$$

The *population size* $(\phi \odot \boldsymbol{\nu})^\natural =< \mathbf{1}, \phi \odot \boldsymbol{\nu} >= \sum_{i \in \mathcal{I}_p} (\phi \odot \boldsymbol{\nu})_i$ is given by:

$$(\phi \odot \boldsymbol{\nu})^\natural = (\phi^{\text{dem}} \odot \nu^{\text{dem}})^\natural = \nu^{b,\natural} - \nu^{d,\natural}, \quad \text{since} (\phi^s \odot \boldsymbol{\nu}^s)^\natural = 0. \quad (2.2.2)$$

2.2.2 Population process and jumps counting process

From now on, all processes are assumed to be defined on a given probability space $(\Omega, \mathcal{G}, \mathsf{P})$, equipped with a filtration $(\mathcal{G}_t)$ verifying the usual assumptions of right-continuity and completeness. The predictable σ-field generated by adapted left-continuous processes is denoted $\mathcal{P}(\mathcal{G})$.
The evolution of the population is described by an adapted *càdlàg* $\mathbb{N}^p$-valued pure jump process Z, called the *population process*. Jumps of the population process are assumed to be swap or demographic jumps as described above. We also assume that two events cannot happen simultaneously. At time t the state of the population is $Z_t = (Z_t^1, .., Z_t^p)$, with Z_t^i the number of individuals in subgroup i.

Jumps counting process

JUMPS COUNTING PROCESS As a multivariate pure jump process, the population process Z has piecewise constant and càdlàg paths, and can be written as the sum of its jumps, denoted by $\Delta Z_s = Z_s - Z_{s-}$. With the unified notations introduced above, the jump size corresponding to the event $\gamma = (\alpha, \beta)$ is $\phi(\gamma)$, so that,

$$Z_t = Z_0 + \sum_{0<s\leq t} \Delta Z_s = Z_0 + \sum_{0<s\leq t} \sum_{\gamma \in \mathcal{J}} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}} \phi(\gamma) \quad (2.2.3)$$

For each type of event $\gamma \in \mathcal{J}$, the counting process N^γ defined by $N_t^\gamma = \sum_{0<s\leq t} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}}$ is the process which counts the number of events of type γ. By assumptions, the processes N^γ have no common jumps. Thus, the process $\mathbf{N} = (N^\gamma)_{\gamma\in\mathcal{J}}$ indexed by $\mathcal{J}$ is a well defined multivariate counting process, called the *jumps counting process* of Z.
Interchanging the sums in Equation (2.2.3) shows that the population process is an affine function of the jumps counting process $\mathbf{N}$,

$$\textit{Affine relation} \qquad Z_t = Z_0 + \sum_{\gamma\in\mathcal{J}} \phi(\gamma) N_t^\gamma = Z_0 + \phi \odot \mathbf{N}_t \tag{2.2.4}$$

Observe that Z_t is an affine function of $(Z_0, \mathbf{N}_t)$, but $\mathbf{N}_t$ is depending on all the history $Z_{[0,t]}$ of the population process Z on $[0,t]$, so that $\mathbf{N}_t = H(Z_{[0,t]})$. Using the notations introduced in 2.2.1, the affine relation can also be written as $Z_t = Z_0 + \phi^s \odot \mathbf{N}_t^s + \mathbf{N}_t^b - \mathbf{N}_t^d$.
An equivalent point of view is to consider the jump measure of the population process instead of the jumps counting process,

$$J(\mathrm{d}t, \mathrm{d}\gamma) = \sum_{s>0} \sum_{\gamma\in\mathcal{J}} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}} \delta_\gamma(\mathrm{d}\gamma)\delta_s(\mathrm{d}t) \tag{2.2.5}$$

Thus, this representation can be applied to any pure jump process generated by a finite or countable number of jumps types. In particular, a similar representation is used in Anderson and Kurtz (2015) in the case of Continuous Time Markov Chains for the modeling of chemical reaction network. To the best of our knowledge, this representation is less usual for population dynamics. Furthermore, the constraint $Z \in \mathbb{N}^p$ implies that the jumps counting process cannot be defined independently of the population, due to a *support condition*.

Support Condition An individual can be removed from subgroup i in several cases: a death, definitive exit, or a swap move to another subgroup j. These events are indexed by couples $(i,\beta), \beta \in \mathcal{I}^{(i)}$. Recall that such events cannot happen if the subpopulation i is empty. This means that for any such (i,β), the counting process $N^{i,\beta}$ satisfies the *support condition*

$$\textit{Support condition} \qquad \forall \beta \in \mathcal{I}^{(i)}, \quad \int_0^t \mathbb{1}_{\{Z_{s-}^i = 0\}} dN_s^{i,\beta} = 0, \quad \mathsf{P}\text{-a.s.} \tag{2.2.6}$$

The process $N^{i,\beta}$ cannot be defined independently of the population process in order to respect the constraint that the components of Z are non negative.
The situation is different for events leading to the addition of an individual in the

subgroup j. Since no reference is made on the origin of the new individual, such events include swap moves, but also birth (from parents in the population) or external entry (of immigrants). In a closed population (without immigration), it is assumed that no event can happen in an empty population, so that $\int_0^t \mathbb{1}_{\{Z_{s-}=0\}} d\mathbf{N}_s = 0$, or equivalently $\int_0^t \mathbb{1}_{\{Z_{s-}=0\}} dN_s^\natural = 0$, a.s..

Initialization procedure In the previous description of the population process, little attention has been paid to the role of the initial condition Z_0. In many problems, the date 0 for the initial condition is very arbitrary, and the state Z_0 gives a poor information on the past of the population dynamics. Massoulié (1998) suggests to define the population process on $\mathbb{R}$ in place of $\mathbb{R}^+$, and to model the initial condition as the state at time 0 of some $\mathbb{N}^p$-valued dynamic process $(\xi_t)_{t\leq 0}$. It is thus natural to consider a generalized initial condition starting at a random date $\tau \geq 0$ (generally a stopping time) from state $\zeta_\tau \in \mathbb{N}^p$. This information is summarized by the so-called *entry process* $\xi_t(\zeta) = \zeta_\tau \mathbb{1}_{\{\tau \leq t\}}$ (often denoted ξ_t or ξ_t^τ for simplicity), with in this generalized setting

$$Z_t = \xi_t + \phi \odot \mathbf{N}_t, \quad N_t^\gamma = \sum_{\tau < s \leq t} \mathbb{1}_{\{\Delta Z_s = \phi(\gamma)\}}.$$

In particular, the jumps counting process does not increase on the set $\{t;\ \xi_{t-} = 0\}$ and still satisfies the support condition (2.2.6).

Population system with entry process

The representation of the population using a larger multivariate counting process is very advantageous, due to the many tools available for the study of point processes. In the rest of this chapter and in Chapter 3, we strongly rely on the description of population processes using jumps counting processes.
We are thus interested in the inverse modeling approach, that is in the construction of a population process from an entry process and a jumps counting process. Given a couple $(\xi, \mathbf{N})$, the population is obviously characterized by the affine relation (2.2.4). However, such a process is not necessarily a well defined population since its components can take negative values. A necessary and sufficient condition for the population process to be well-defined is actually the support condition (2.2.6).

Definition 2.2.1 (Population system with random departure)**.**
a) *Let* $(\xi_t = \zeta_\tau \mathbb{1}_{\{\tau \leq t\}})$ *be an* $\mathbb{N}^p$*-valued entry process. A* $p(p+1)$*-multivariate counting*

process $\boldsymbol{N}$ *indexed by* $\mathcal{J}$ *is called a jumps counting process starting from* ξ *iff*

$$\begin{cases} \textit{Starting condition} & \mathbb{1}_{\{\xi_{t-}=0\}} d\boldsymbol{N}_t = 0 \\ \textit{Support condition} & \mathbb{1}_{\{\xi^i_{t-}+(\phi\odot\boldsymbol{N})^i_{t-}=0\}} dN_t^{i,\beta} = 0 \quad \forall i \in \mathcal{I}_p, \quad \forall \beta \in \mathcal{I}^{(i)} \end{cases} \tag{2.2.7}$$

b) *The companion population process of* $(\xi, \boldsymbol{N})$ *is defined by* $Z_t = \xi_t + \phi \odot \boldsymbol{N}_t$. *In paticular,* Z *is a well-defined population process with jumps counting process* $\boldsymbol{N}$ *and entry process* ξ. *The triplet* $(\xi, \boldsymbol{N}, Z)$ *is called a population system.*

A demographic event in the population corresponds to a jump of one of the processes $\mathbf{N}^b = (N^{\infty,j})$ or $\mathbf{N}^d = (N^{i,\infty})$. By application of the elementary calculations of Paragraph 2.2.1, the size of the companion population process (also called the aggregated process) is $Z^\natural = \xi^\natural + (\phi \odot \mathbf{N})^\natural = \xi^\natural + N^{b,\natural} - N^{d,\natural}$. This can be simply interpreted as the initial size of the population to which is added the total number of births minus the number of deaths. Since $N_t^{b,\natural}$ and $N_t^{d,\natural}$ have no common jumps, the size of the population process is constant between two demographic events. The process $N^{\text{dem},\natural} = N^{b,\natural} + N^{d,\natural}$ counts the total number of demographic events.

Swap processes Populations in which no demographic event occur are called *Swap processes*, and will play a very important role in the following. As for population systems, a swap system is defined by a triplet $(\xi, \mathbf{N}^{\text{sw}}, X)$, where the swap jumps counting process $\mathbf{N}^{\text{sw}}$ is now indexed by $\mathcal{J}^{\text{sw}}$, and verifies the starting and support conditions with ϕ^s in place of ϕ. The companion swap process is defined by $X_t = \xi_t + \phi^s \odot \mathbf{N}^{\text{sw}}_t$.
By Equation (2.2.2), The swap size $X^\natural$ after τ is equal to $\xi^\natural_\tau$. Thus, the size of the swap process is constant after the random start, and determined by the size of its initial condition. For each $n \in \mathbb{N}$, let $\mathcal{U}_n = \{z \in \mathbb{N}^p;\ z^\natural = n\}$ be the finite space of populations of size n. Then, the swap process lives in space $\mathcal{U}_{\xi^\natural_\tau}$ after the random start τ. We will come back to specific properties of swap processes in the following.

Temporal transformation of population system

Several transformations on population systems can be directly obtained from the previous algebraic definition. The two examples presented in this paragraph will be particularly useful in Section 2.4, in order to disentangle swap events from demographic events.

population decomposition Let $(Z_0, \mathbf{N}, Z)$ be a population system starting from 0 and τ a random time. The system stopped at time τ is denoted by $(Z_0, \mathbf{N}^\tau_t = \mathbf{N}_{t\wedge\tau}, Z^\tau_t = Z_{t\wedge\tau})$. Since $\mathbf{N}^\tau_t = \int_0^t \mathbf{1}_{]0,\tau]}(s)\mathrm{d}\mathbf{N}_s$, $\mathbf{N}^\tau$ verifies the support condition. Obviously $\mathbf{N}^\tau$ don't jump

after time τ.

A population system starting from τ in state Z_τ can also be defined, from the entry process $\xi_t^\tau(Z) = Z_\tau \mathbf{1}_{\{\tau \leq t\}}$ and the jumps counting process $\mathbf{N}^{\tau+}$:

$$\mathbf{N}_t^{\tau+} = \mathbb{1}_{]\tau,\infty)} * \mathbf{N}_t \quad \text{with } \mathbb{1}_{]\tau,\infty)} * \mathbf{N}_t = \int_0^t \mathbb{1}_{]\tau,\infty)}(s) \mathrm{d}\mathbf{N}_s = \mathbf{N}_t - \mathbf{N}_t^\tau.$$

Then, $\mathbf{N}^{\tau+}$ is a jumps counting process starting from $\xi^\tau(Z)$, associated with the companion population process $Z^{\tau+} = \xi^\tau(Z) + \phi \odot \mathbf{N}^{\tau+}$. By definition,

$$Z = Z_0 + \phi \odot (\mathbf{N} - \mathbf{N}^\tau + \mathbf{N}^\tau) = Z^\tau + \phi \odot \mathbf{N}^{\tau+}.$$

This means that the companion population $Z^{\tau+}$ coincides with Z on $[\tau, \infty)$. Furthermore, a decomposition of the initial population system is given by:

$$\mathbf{N}(Z_0) = \mathbf{N}^\tau(Z_0) + \mathbf{N}^{\tau+}(\xi^\tau(Z)), \quad Z(Z_0) = Z^\tau(Z_0) - \xi^\tau(Z) + Z^{\tau+}(\xi^\tau(Z)) \tag{2.2.8}$$

Population process with only **one** demographic event Let us now assume that only one demographic event (birth or death) occurs in the population, at time $T^{\text{dem}} = \tau$. τ is the first jump of the demographic counting process $\mathbf{N}^{\text{dem}}$, or equivalently of $N^{\natural,\text{dem}} = N^{\natural,b} + N^{\natural,d}$. On $[0, \tau[$, $\mathbf{N}^{\text{dem}}$ is the null process and $Z_t = Z_0 + \phi^s \odot \mathbf{N}_t^s$ is a "pure swap process". Since two events cannot occur at the same time, the process $Z_0 + \phi^s \odot \mathbf{N}_{t\wedge\tau}^s$ can be extended by continuity to $[0, \tau]$, and can be considered on this interval as a stopped pure Swap process X^τ. By definition, the population process has a jump at time τ equal to $\mathbf{N}_\tau^b - \mathbf{N}_\tau^d$, where $\mathbf{N}^{\text{dem}} = (\mathbf{N}^b, \mathbf{N}^d)$ has only one non-zero component. Then, the population process stopped at time τ can be written as $Z^\tau = X_t^\tau + \mathbf{N}_t^b - \mathbf{N}_t^d$. Finally, since only one demographic event occurs, the process $Z^{\tau+}$ starting at time τ is a pure Swap process.

Observe that this description actually defines a continuous pasting procedure of two Swap processes evolving on different state spaces. The demographic event is used as *switch process*, from state space $\mathcal{U}_{Z_0^\natural}$ to $\mathcal{U}_{Z_0^\natural + N_\tau^{b,\natural} - N_\tau^{d,\natural}}$, and also defines the starting state of the second Swap. This kind of decomposition will be the building block of the decomposition algorithm of Section 2.4.

2.2.3 Birth Death Swap Intensity

By describing the events changing the composition of the population rather than the behavior of individuals, we have obtained a very flexible algebraic description of the population, based on the jumps counting process. We can now define the so-called *Birth Death Swap (BDS) system*. Thanks to the theory on point processes and their pathwise representations, the BDS system is characterized by properties of the *multivariate*

intensity of the jumps counting process. Let us first give a brief recall on the intensity process, followed by the classical example of the Poisson measure.

Brief overview on intensity processes and Poisson measures

INTENSITY PROCESS A $\mathcal{G}_t$-adapted counting process N ($\mathcal{G}_t$ is not necessarily the canonical filtration of N) is said to have the $\mathcal{G}_t$-*intensity process* λ iff λ is a non-negative, $\mathcal{G}_t$-predictable process and

$$M_t^\lambda = N_t - \int_0^t \lambda_s \, ds \quad \text{is a } \mathcal{G}_t \text{ - local martingale} \tag{2.2.9}$$

Informally, $\lambda_t dt$ is the linear estimate of N between $]t, t+dt]$, conditionally to the strict information given by $\mathcal{G}_t$: $\mathsf{E}[N_{t+dt} - N_t | \mathcal{G}_{t-}] \simeq \lambda_t dt$. In particular, the predictable intensity process reflects any *predictable support condition* on the process N, since for any predictable subset Δ,

$$\int_0^t \mathbf{1}_\Delta(s) \, \mathrm{d}N_s = 0 \Longleftrightarrow \mathbf{1}_\Delta(s)\lambda_s = 0, \ \mathrm{d}s \times \mathrm{d}\mathsf{P} \ a.s.. \tag{2.2.10}$$

In the standard theory motivated by statistical estimation issues, the intensity is often defined in reference to the minimal filtration $\mathcal{F}_t = \sigma(N_s; s \leq t)$ generated by the past history of the counting process, also called *canonical filtration*. Then, the intensity may only be a function of the past of the counting process. Here, working with larger filtration actually facilitates the analysis of the jumps counting process, in particular by adding flexibility to the setting.

For a multivariate counting process $\mathbf{N} = (N^\gamma)_{\gamma \in \mathcal{J}}$ whose components have no common jumps, the $\mathcal{G}_t$-*multivariate intensity process* is the vector $\boldsymbol{\lambda} = (\lambda^\gamma)_{\gamma \in \Im}$, where λ^γ is the $\mathcal{G}_t$-intensity of process of N^γ. The concept of *spatial counting measure* $N(\mathrm{d}t, \mathrm{d}\gamma)$ is sometimes preferred to the vector representation. The correspondence is given by: $N([0,t] \times \{\gamma\}) = N_t^\gamma$. The associated intensity measure is $\lambda(\mathrm{d}t, \mathrm{d}\gamma) = \mathrm{d}t \ \lambda_t^\gamma \, \mathrm{d}\gamma$, where $\mathrm{d}\gamma$ is the counting measure on $\mathcal{J}$.

POISSON MEASURES Obviously, a $\mathcal{G}_t$-adapted spatial counting measure can be defined on a more general (complete, separable, metric) space $(E, \mathcal{E})$. When its $\mathcal{G}_t$-intensity measure is a deterministic product measure $\lambda(\mathrm{d}t, \mathrm{d}x) = \mathrm{d}t\mu(\mathrm{d}x)$, with μ is a σ-finite measure on $(E, \mathcal{E})$, the counting measure $Q(\mathrm{d}t, \mathrm{d}x)$ is known to have a Poisson distribution in the following sense: for disjoints sets $A_1, .., A_n \in \mathcal{E}$ such that $\mu(A_i) < \infty$, the processes $(Q_t(A_i) = Q([0,t] \times A_i))$, $i = 1..n$, are independent Poisson processes (they have no

jumps in common) of $\mathcal{G}_t$-intensity $\mu(A_i)$.
All properties of spatial Poisson measures are defined in reference to the large filtration $(\mathcal{G}_t)$. In particular, the conditional increments of the Poisson process $Q_{t+h}(A) - Q_t(A)$ are independent of $\mathcal{G}_t$. This property implies weak non causality in the sense of Florens and Fougere (1996) between the filtration $\mathcal{G}$ and the canonical filtration of the Poisson measure Q. In a probabilistic point of view, this notion is known as the (H)-assumption.
A constructive algorithm for the simulation of Q can be given when the measure $\mu(\mathrm{d}x)$ is finite, with mass $\mu(E)$. In this case, a sequence of random variables (T_i, χ_i) is defined, where $(\tau_i = T_i - T_{i-1}, \chi_i)$ is an iid sequence, with τ_i an exponential random variable with parameter $\mu(E)$ and χ_i an independent mark distributed as $\mu(\mathrm{d}x)/\mu(E)$. The Poisson measure Q is generated by using Dirac measures, $Q(\mathrm{d}t, \mathrm{d}x) = \sum_n \delta_{T_i}(\mathrm{d}t) \otimes \delta_{\chi_i}(\mathrm{d}x)$. In this case, Q is often called a *marked Poisson process*. In particular, an *ordered* enumeration of its jump times can be given. This construction, however, does not hold if μ is only σ-finite. In this case, a similar construction can be obtained by restricting the intensity to sets $\mathbb{R}^+ \times E_n$, with $E = \bigcup E_n$, $\mu(E_n) < \infty$. However, there is no ordered enumeration of the jump times of Q when μ is only σ-finite.

Birth Death Swap systems

The jumps counting process of a population system cannot have a deterministic intensity, due to the the support conditions (2.2.7) which are transferred onto the intensity process of the jumps counting process by (2.2.10). In order for the population system $(\xi, \mathbf{N}, Z)$ to become a *Birth Death Swap system*, additional assumptions are made on the multivariate intensity of $\mathbf{N}$. Only the companion population process Z is usually observed, and a natural assumption is that the intensity process depends on the population process rather than on the jumps counting process. This assumption is implicit in a Markov framework. To go further and take into account some additional time-dependent uncertainty, such as a random environment, the multivariate intensity process is assumed to depend in a predictable way on additional randomness, not explicitly modeled.

Definition 2.2.2 (BDS intensity functional and BDS system)**.**
a) *A* BDS intensity functional $\boldsymbol{\mu}(\omega, t, z) = \boldsymbol{\mu}(t, z) = (\mu^\gamma(t, z))_{\gamma \in \mathcal{J}}$ *is a multivariate* $\mathcal{G}_t$*-predictable non-negative functional depending on* $z \in \mathbb{N}^p$*, satisfying*

$$\boldsymbol{\mu}(t, \boldsymbol{0}) \equiv 0 \quad \textit{and} \quad \sum_{i \in \mathcal{I}_p} \sum_{\beta \in \mathcal{I}^{(i)}} \mu^{i,\beta}(t, z) \mathbf{1}_{\{z^i = 0\}} \equiv 0, \quad \mathrm{d}t \otimes \mathrm{d}\mathsf{P}\ a.s. \tag{2.2.11}$$

b) *A Birth Death Swap (BDS) system of intensity functional* $\boldsymbol{\mu}$ *is a population system* $(\xi, \boldsymbol{N}, Z)$ *such that the jumps counting process* $\boldsymbol{N}$ *is a multivariate counting process of* $\mathcal{G}_t$*-intensity* $\boldsymbol{\lambda}_t = \boldsymbol{\mu}(t, Z_{t-}) = \boldsymbol{\mu}(t, \xi_{t^-} + \phi \odot \boldsymbol{N}_{t^-})$.

The second part of (2.2.11) ensures that the BDS system $(\xi, \mathbf{N}, Z)$ verifies the support condition (2.2.7). The condition $\boldsymbol{\mu}(t, \mathbf{0}) \equiv 0$ ensures that the starting condition is also verified. Indeed, the starting condition means that the first jump time T_1 of $\mathbf{N}$ verifies $T_1 > \tau$, i.e $\int \mathbb{1}_{\{\xi_{t^-}=0\}\cap\{t\leq T_1\}} \mathrm{d}\mathbf{N}_t = 0$. By (2.2.10), this is equivalent to $\mathbb{1}_{\{\xi_{t^-}=0\}\cap\{t\leq T_1\}} \boldsymbol{\mu}(t, Z_{t^-}) = 0$, $\mathrm{d}t \times \mathrm{dP}$ p.s. By definition, $Z_{t^-} = 0$ on $\{\xi_{t^-} = 0\} \cap \{t \leq T_1\}$, and since $\boldsymbol{\mu}(t, \mathbf{0}) = 0$ by (2.2.11), this means that $\mathbf{N}$ verifies the starting condition. This condition is actually not necessary, and could have been replaced by more general condition $\boldsymbol{\mu}(t, z)\mathbb{1}_{\{\xi_{t^-}=0\}}$. However, we prefer the former condition which lead to simpler notations.

In the case of pure Swap processes, the swap intensity functional is defined by $(\mu^\kappa(t, x))_{\kappa\in\mathcal{J}^s}$ and satisfying a support condition equivalent to (2.2.11). Swap systems of intensity $\boldsymbol{\mu}^s$ are denoted by triplets $(\xi, \mathbf{N}^{\mathrm{sw}}, X)$.
The swap and demographic part of a population process can be isolated by writing $Z_t = \xi_t + \phi^s \odot \mathbf{N}^s_t + \mathbf{N}^b_t - \mathbf{N}^d_t$. However, disentangling the swap and the demographic part is not an easy task, since the support conditions introduce a non-linear dependence between them. In particular, N^s is not a swap jumps counting process. We come back to this problem in Section 2.4.

Examples of BDS intensities

Let us give some examples of BDS intensity functional.

a) Linear intensities and individual rates For an event $\gamma \in \mathcal{J}$, $\mu^\gamma(t, Z_{t^-})$ is the intensity corresponding to the occurrence of the event of type γ in *all* the population. This should not be confused with the *rate* at which the event γ can occur to *one individual*. The intensity μ^γ is called linear if the functional depends linearly on the number of individuals in each subgroup. Linear intensities allow a direct interpretation of intensities in term of individual rates:
-Death and Swap intensities: for an event (d, i) of type death in the subgroup i, a classical linear intensity functional is $\mu^{(d,i)}(\omega, t, z) = d^i(\omega, t)z^i$. The intensity can be interpreted as follow: all individuals in subgroup i die independently with a death rate $d^i(\omega, t)$. A similar interpretation can be given for linear swap events.
- Birth intensities: For an event (b, j) of type birth in the subgroup j, the equivalent is to take $\mu^{(b,j)}(\omega, t, z) = b^j(\omega, t)z^j$. This means that all individuals in subgroup j give

birth to an individual of same characteristics at rate $b^j(\omega, t)$. Mutations at birth can be included, $\mu^{(b,j)}(\omega, t, z) = \sum_{i=1}^{p} b^i(\omega, t) z^i m(\omega, t, i, j)$, where $m(\omega, t, i, j)$ is the random probability for a individual born at time t from a parent in subgroup i to be in subgroup j. A stochastic intensity λ_t can also be added to the birth intensity, in order to model the entry of immigrants at rate λ_t.

b) *Markov BDS system:* When the BDS intensity functional is an homogeneous deterministic function $\boldsymbol{\mu}(z)$, the BDS system is a Continuous Time Markov Chain (CTMC). In this case, the process can be described using classical tools of CTMC. If the population has the initial state Z_0, the first jump time is distributed as an exponential of parameter $\mu^\natural(Z_0)$, and the event of type γ is chosen as the first event independently with probability $\mu^\gamma(Z_0)/\mu^\natural(Z_0)$. The BDS system can be built by iterating the last two steps.

c) *Nonlinear Swap intensity:* Let us give an example of a BDS system with general birth and death intensities and deterministic nonlinear swap intensity functional. We consider here the case of two subgroups or patches ($p = 2$), where subgroup 2 is a favorable subgroup with lower death intensity: $\mu^{(d,1)}(t, z) \geq \mu^{(d,2)}(t, z)$, a.s. Two regimes can be distinguished for the swap events, depending on the size of the population: when the population is small, $z^\natural \leq M$, individuals can swap more easily to the favorable subgroup 2, at a rate $k_{12}(z^\natural)^\alpha$, with $\alpha > 0$. When the population size is large, $z^\natural > M$, access to the subgroup 2 is limited and individuals swap from 1 to 2 at a constant rate k_{12}^M. In both cases, individuals swap from the favorable subgroup 2 to subgroup 1 at constant rate k_{21}. To summarize, the swap intensity is defined as follow:

$$\mu^{(1,2)}(z) = k_{12}(z^\natural)^\alpha z^1 \mathbb{1}_{\{z^\natural \leq M\}} + k_{12}^M z^1 \mathbb{1}_{\{z^\natural > M\}}, \quad \mu^{(2,1)}(z) = k_{21} z^2. \tag{2.2.12}$$

In particular, pure Swap processes with intensity defined as above are CTMC. Since pure Swap processes have a constant size determined by their initial condition, the intensity regime is determined by the size n of the Swap process, and individuals swap independently from 1 to 2 at a constant rate equal to $k_{12}n^\alpha$ or k_{12}^M according to the size of the Swap. We come back to this example in the next chapter.

2.3 Birth Death Swap Differential Equations

The question of the existence of BDS systems can be complex considering the feedback effect induced by the intensity functional. Markovian assumptions are generally made for Birth Death processes or for Swap processes, yielding to a distributional point of view on the existence of such processes. Since the 1990s, a pathwise point of view, allowing

more flexibility on intensity processes, has been considered by many authors (see e.g Bansaye and Méléard (2015); Brémaud and Massoulié (1996); Fournier and Méléard (2004); Garcia and Kurtz (2008, 2006); Massoulié (1998); Nassar and Pardoux (2017)). This point of view is based on the pathwise realization of point processes as solutions of Stochastic Differential Equations (SDE) driven by Poisson measures. In particular, solutions are obtained from the thinning of an "augmented Poisson measure".
In this section, we rely on these representations in order to give a pathwise representation of the BDS system, based on the realization of the jumps counting process as the solution of a multivariate SDE. Focusing on the jumps counting process allows us to adopt the point of view of point processes, within a general framework similar to that of Massoulié (1998).
In their paper on the modeling of recurrent events (another terminology used in statistics for counting processes), Gjessing et al. (2010) highlight the importance of non-explosion properties, in order to avoid fitting errors due to the uncontrollable nature of exploding systems. Non-explosion is also central in the analysis of solutions of SDEs driven by Poisson measures. In the rest of this part, all solutions of stochastic differential equations driven by Poisson measures are considered to be well-defined if they are "honest" in the terminology of Cox and Miller (1977), also used in Gjessing et al. (2010), that is if they stay finite in finite time with probability one.
The existence of BDS systems is derived from a more general result obtained in 2.3.1, on the construction of multivariate counting processes by *strong domination* with a non-exploding process. In particular, this construction allows us to relax some of the usual assumptions on the intensity functional that are Lipschitz or sublinear growth conditions. Furthermore, the strong domination construction will be every useful in Chapter 3, by providing "free" tightness properties.

2.3.1 Thinning of Poisson measure and Markov birth process

Let us first introduce the basic principles of the thinning procedure. As a first example, a thinning construction of the classical Markov birth process is presented. By convention, all Poisson measures are assumed to be defined on the given probability space $(\Omega, (\mathcal{G}_t), \mathsf{P})$.

Thinning of Poisson measure

Let us first consider the simple case of a Poisson measure $Q(\mathrm{d}t, \mathrm{d}\theta)$, with marks θ defined on $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ and deterministic intensity measure $q(\mathrm{d}t, \mathrm{d}\theta) = \mathrm{d}t \times m(\mathrm{d}\theta)$, where m is the Lebesgue measure ($m(\mathrm{d}\theta) = \mathrm{d}\theta$). Let also λ be a non negative predictable process,

$\mathrm{d}t$-integrable over compact intervals, and Δ the the subset of $\mathcal{P}(\mathcal{G}) \times \mathcal{B}(\mathbb{R}^+)$ defined by $\Delta = \{(t,\theta);\, 0 < \theta \leq \lambda_t\}$.

A new random measure $Q^\Delta(\mathrm{d}t, \mathrm{d}\theta)$ can be defined as the restriction of the Poisson measure Q to the random set Δ,

$$Q^\Delta(\mathrm{d}t, \mathrm{d}\theta) = \mathbb{1}_\Delta(t,\theta) Q(\mathrm{d}t, \mathrm{d}\theta).$$

Q^Δ is a random measure of random intensity measure $q^\Delta(\mathrm{d}t, d\theta) = \mathbb{1}_\Delta(t,\theta)\mathrm{d}t \times m(d\theta)$. Furthermore, its projection on $\mathbb{R}^+$ is the *Cox process*

$$Q^\lambda_t = \int_0^t \int_{\mathbb{R}^+} \mathbb{1}_{\{0<\theta\leq\lambda_s\}} Q(\mathrm{d}s, \mathrm{d}\theta) = \int_0^t Q(\mathrm{d}s,]0, \lambda_s]),$$

of $\mathcal{G}_t$-compensator $\int_0^t \mathrm{d}s \int_{\mathbb{R}^+} \mathbb{1}_{\{0<\theta\leq\lambda_s\}} \mathrm{d}\theta = \int_0^t \lambda_s \,\mathrm{d}s$. The process Q^λ of $\mathcal{G}_t$-intensity λ is thus obtained by *thinning and projection* of the augmented Poisson measure Q. As stated in Paragraph 2.2.3, there is no monotone enumeration of the jump times of Q, and we cannot define an iterative construction of Q^λ from Q which would be based on an increasing sequence of jump times of Q, as it is sometimes incorrectly done.

However, since Q^λ is a well-defined counting process, its jump times can be enumerated increasingly, and Q^Δ can be characterized by a sequence (T_i, Θ_i), with (T_i) the increasing sequqence of jump times of Q^λ, and Θ_i the mark of Q associated with T_i. This description is particularly useful since it allows us to define a unique iterative construction for processes obtained by thinning of Q^Δ, using a similar construction to that presented in Lewis and Shedler (1979) for the construction of inhomogeneous Poisson processes.

The thinning procedure can be extended to more general Poisson measures $Q(\mathrm{d}t, \mathrm{d}x, \mathrm{d}\theta)$, with marks (x,θ) defined on $(E \times \mathbb{R}^+, \mathcal{E} \times \mathcal{B}(\mathbb{R}^+))$ and intensity measure $q(\mathrm{d}t, \mathrm{d}x, \mathrm{d}\theta) = \mathrm{d}t \times \nu(\mathrm{d}x) \times \mathrm{d}\theta$, where $\nu(\mathrm{d}x)$ is a σ-finite measure on E. A Cox measure $Q^\lambda(\mathrm{d}t, \mathrm{d}x) = Q(\mathrm{d}t, \mathrm{d}x,]0, \lambda(t,x)])$ can be defined similarly by thinning and projection of Q based on a $\mathcal{P}(\mathcal{G}) \otimes \mathcal{E}$ process $\lambda(t,x)$. The stochastic intensity of Q^λ is $q^\lambda(\mathrm{d}t, \mathrm{d}x) = \lambda(t,x)\,\mathrm{d}t \times \nu(\mathrm{d}x)$, and the measure $Q(\mathrm{d}t, \mathrm{d}x, \mathrm{d}\theta)$ is called the *driving Poisson measure* of Q^λ.

In many applications, the space E is a finite space of cardinal ρ, with elements denoted by $\imath$ and equipped with the uniform counting measure $\mathrm{d}\imath$. In the case of BDS systems for instance, E is the set $\mathcal{J}$ of all events types, with $\imath = \gamma$ ($\rho = p(p+1)$). It is thus interesting to use a vector version of the foregoing. The Poisson measure $Q(\mathrm{d}t, \mathrm{d}\imath, \mathrm{d}\theta)$ is reinterpreted as a family of independent Poisson measures $\bar{Q}(\mathrm{d}t, d\theta) = (Q^\imath(dt, d\theta); \imath \in E)$, and $\lambda(t,\imath)$ becomes a multivariate intensity $\bar{\lambda} = (\lambda^\imath)_{\imath\in\mathsf{E}}$. Finally, the Cox measure is rewritten as the multivariate Cox process $\bar{Q}^\lambda(dt,]0, \bar{\lambda}_t]) = \left(Q^\imath(dt,]0, \lambda(t,\imath)])\right)_{\imath\in E}$. A multivariate random measure $\bar{Q}^\Delta(\mathrm{d}t, \mathrm{d}\theta) = \mathbb{1}_{]0,\bar{\lambda}_t]}(\theta)\bar{Q}(\mathrm{d}t, \mathrm{d}\theta)$ can also be defined as in the one dimensional case.

Multivariate Markov Birth process

When the intensity $\bar{\lambda}$ of the multivariate counting process $\bar{Q}^\lambda$ is a predictable functional of $\bar{Q}^\lambda$ itself, the thinning equation becomes a stochastic differential equation, driven by the multivariate Poisson measure $\bar{Q}$. The most simple example of such a process is probably the one dimensional Markov pure Birth process, also called online Markov Birth process, and which has extensively been studied since the 60s. In particular, the famous Feller criterion (Feller (1968)) guarantees the non-explosion of the online Markov process. In this paragraph, we give a brief overview of the thinning construction of the Markov Birth process, followed by a generalization to multivariate Markov Birth processes.

Online Markov Birth process An online Markov Birth process B is characterized by its intensity functional, here denoted by $Kg(y)$ (with $g(0) = 0$). The most classical example is the *Yule process*, in the linear case when $g(y) = by$. The linear case is useful to prove the finitude of some moments as explained in Bansaye and Méléard (2015). A realization of B can be obtained as solution of the following SDE driven by a Poisson measure Q,

$$\begin{cases} B_t(y) = y + N_t^B \quad \text{and} \quad \mathrm{d}N_t^B = Q(\mathrm{d}t,]0, K\, g(B_{t-}(y))]). \\ \quad \text{Feller criterion} \qquad \sum_{j=1}^{\infty} \frac{1}{g(j)} = \infty \end{cases} \tag{2.3.1}$$

Let us give some elements of the proof of the (strong) existence and uniqueness of the previous equation.
Sketch of the proof: The first jump of the counting process N^B is the first jump time of the $\mathcal{G}_t$-Poisson process $Q_t(]0, K\, g(y)]) = Q(]0, t] \times]0, K\, g(y)])$, and is distributed as an exponential variable with parameter $Kg(y)$. By a recursive argument, the jumps of the counting process N^B occur at times $T_1, T_2, T_j..$, where T_j is the first jump time of a Poisson process starting at time T_{j-1} and of intensity $Kg(y + j - 1)$, obtained by thinning of Q. In particular, their increments are exponentially distributed with parameter $\frac{1}{K\, g(y+j-1)}$. The Feller condition for non-explosion is equivalent to the property that $\lim_n T_n = \infty$. Thus, B is a non-explosive pure Birth process, which is by construction the unique solution of Equation (2.3.1).

Multivariate Markov birth process The multivariate case is an easy extension under the following assumption: the process $\bar{B}$ is $\mathbb{N}^\rho$-valued and the multivariate intensity functional of $\bar{B}$ is a deterministic function of the size $\bar{y}^\natural = \sum_1^\rho y^\imath$ of the birth process, $\bar{g}(\bar{y}) = (g^\imath(\bar{y}^\natural))$. For each $\imath = 1..\rho$, the function $g^\imath$ is assumed to satisfy the Feller criterion,

and the multivariate birth process $\bar{B}$ is defined as solution of the differential system

$$\begin{cases} \bar{B}_t(\bar{y}) = \bar{y} + \bar{N}^B_t \quad \text{and} \quad \mathrm{d}\bar{N}^B_t = \bar{Q}(\mathrm{d}t,]0, K\,\bar{g}(\bar{B}^\natural_{t-}(\bar{y})]). \\ \quad \text{Multivariate Feller criterion} \qquad \forall\ \imath \le \rho, \qquad \sum_{n=1}^{\infty} \frac{1}{g^\imath(n)} = \infty \end{cases} \tag{2.3.2}$$

Sketch of the proof: The first jump time $T^{1,\natural}$ of $\bar{B}^\natural$ is also the first jump time of the processes $(\bar{N}^B)$ counting the number of births. As solution of Equation (2.3.2), the vector $(\bar{N}^B_t)$ is equal to the ρ-dimensional Poisson process $(\bar{Q}_t(]0, K\,\bar{g}(\bar{y}^\natural)]))$ before $T^{1,\natural}$. At time $T^{1,\natural}$, one and only one component I_1 of $\bar{Q}$ jumps and defines the component of $\bar{N}^B$ which jumps. In particular, $T^{1,\natural}$ is an exponential random variable with parameter $\sum_1^\rho K\,g^\imath(\bar{y}^\natural)$, independent of the random index I_1, which verifies $\mathsf{P}(I_1 = \imath) = g^\imath(\bar{y}^\natural)/\sum_1^\rho g^\imath(\bar{y}^\natural)$. This step can be iterated, by starting from $\bar{y}^{(I_1)} = \bar{y} + \mathbf{1}_{I_1}$ at time $T^{1,\natural}$.

Existence and strong comparison of multivariate SDEs driven by Poisson measures

The comparison of point processes with ordered (stochastic) intensities has been the subject of many papers (see e.g. Preston (1975); Rolski and Szekli (1991)). The thinning procedure is well-adapted to solve this problem, and for Cox processes, the answer is immediate. If two Cox processes $Q^{\lambda^i}_t = Q(]0,t]\times]0,\lambda^i_t])$, $i = 1, 2$ have ordered intensities $\lambda^1_t \le \lambda^2_t$, then the thinning construction using the same Poisson measure for both processes directly yields that $Q^{\lambda^1}_t \le Q^{\lambda^2}_t$, for all $t \ge 0$. Actually, a stronger property is verified. Since, $\lambda^1_t \le \lambda^2_t$, Q^{λ^1} can be rewritten as $Q^{\lambda^1}_t = Q(]0,t]\times]0,\lambda^1_t \wedge \lambda^2_t]) = Q^{\Delta_2}(]0,t]\times]0,\lambda^1_t])$. This means that Q^{λ^1} can be obtained by *thinning of* Q^{Δ_2} instead of Q. In particular, all jump times of Q^{λ^1} are jump times of Q^{λ^2}, and Q^{λ^1} is *strongly dominated* (or strongly majorized in the terminology of Jacod and Shiryaev (1987)) by Q^{λ^2}.

The direct application to general multivariate counting processes is not so easy, since the natural order of intensity functionals does not necessary imply a natural order on the stochasic intensities. Nevertheless, for online Markov Birth processes, Bhaskaran (1986) (see also Bezborodov (2015)) showed that for non exploding processes, an intensity functional inequality implied a strong domination result. The proof is based on a thinning procedure slightly different from ours and can be extended to Markov Birth Death processes.

The following theorem sets a general framework for the strong comparison of multivariate counting processes. A ρ-multivariate counting process $\bar{Y}^\alpha$ is said to be strongly dominated by $\bar{Y}^\beta$, $\bar{Y}^\alpha \prec \bar{Y}^\beta$ iff

$$\bar{Y}^\beta - \bar{Y}^\alpha \text{ is a multivariate counting process,}$$

or equivalently iff all jumps of $\bar{Y}^\alpha$ are jumps of $\bar{Y}^\beta$. The following result allows us to build the solution $\bar{Y}^\alpha$ of a multivariate SDE by strong comparison with a dominating process $\bar{Y}^\beta$, assuming that the multivariate intensity functional $\bar{\alpha}(t, y)$ of $\bar{Y}^\alpha$ is dominated by the intensity functional $\bar{\beta}(t, y)$ of $\bar{Y}^\beta$ in the following sense:

$$\alpha^\imath(t, \tilde{y}) \leq \beta^\imath(t, \bar{y}), \quad \forall 1 \leq \imath \leq \rho,\ \tilde{y} \leq \bar{y} \in \mathbb{N}^\rho.$$

Observe that $\bar{\beta}$ is dominated by itself iff $\bar{\beta}$ is non-decreasing in y.
The main argument of the result is similar to the case of Cox processes, and relies on replacing the driving Poisson measure $\bar{Q}$ by the random measure $\bar{Q}^{\Delta_\beta}$ associated with the dominating process.

Theorem 2.3.1 (Strong comparison of multivariate counting processes)**.** *Let $\bar{Q}(\mathrm{d}t, d\theta) = (Q^\imath(\mathrm{d}t, d\theta))_{\imath \in E}$ be a multivariate Poisson measure, and $\bar{\alpha}$ and $\bar{\beta}$ two predictable ρ-dimensional intensity functionals defined on $\mathcal{Y} = \mathbb{N}^\rho$, where $\bar{\alpha}$ is assumed to be dominated by $\bar{\beta}$ ($\bar{\alpha} \leq \bar{\beta}$).*
Assume the existence of a unique non-exploding solution $\bar{Y}^\beta \in \mathbb{N}^\rho$ of the multivariate SDE:

$$d\bar{Y}_t^\beta(\bar{y}) = \bar{Q}(dt,]0, \bar{\beta}(t, \bar{y} + \bar{Y}_{t-}^\beta(\bar{y}))]), \tag{2.3.3}$$

Then, for all $\tilde{y} \leq \bar{y}$, there exists a unique (non-exploding) solution to the equation,

$$d\bar{Y}_t^\alpha(\tilde{y}) = \bar{Q}\Big(dt,]0, \bar{\alpha}(t, \tilde{y} + \bar{Y}_{t-}^\alpha(\tilde{y}))]\Big) \tag{2.3.4}$$

Furthermore, $\bar{Y}^\alpha(\tilde{y})$ is strongly dominated by $\bar{Y}^\beta(\bar{y})$: $\bar{Y}^\alpha(\tilde{y}) \prec \bar{Y}^\beta(\bar{y})$.

Proof. a) The key of the proof is the introduction of a slightly different version of Equation (2.3.4), based on the multivariate random measure $\bar{Q}^{\Delta_\beta}$ defined by,

$$\bar{Q}^{\Delta_\beta}(\mathrm{d}t, \mathrm{d}\theta) = \mathbb{1}_{\Delta_\beta}(t, \theta) Q(\mathrm{d}t, \mathrm{d}\theta), \text{ with } \Delta_\beta = \{(t, \theta); 0 < \theta \leq \beta(t, \bar{y} + \bar{Y}_{t-}^\beta)\}.$$

By assumption, the solution $\bar{Y}^\beta$ of Equation (2.3.3) is a well-defined multivariate counting process. Thus, its jumps can be enumerated by a sequence $(T_j, \imath_j)_{j \geq 1}$, where (T_j) is the increasing sequence of the jump times of $\bar{Y}^\beta$, $\lim T_j = +\infty$, and $\imath_j$ is the index of the component of $\bar{Y}^\beta$ jumping at time T_j. As a consequence, $\bar{Q}^{\Delta_\beta}$ is a *marked* multivariate counting process, and can be characterized by the sequence $(T_j, \imath_j, \Theta_j)_{j \geq 1}$, where Θ_j is the mark of $\bar{Q}^{\imath_j}$ associated with T_j.
We can thus introduce the new equation, defined by:

$$d\tilde{Y}_t^\alpha = \bar{Q}^{\Delta_\beta}(\mathrm{d}t,]0, \bar{\alpha}(t, \tilde{y} + \tilde{Y}_{t-}^\alpha)]) = \bar{Q}(\mathrm{d}t,]0, \bar{\alpha}(t, \tilde{y} + \tilde{Y}_{t-}^\alpha) \wedge \bar{\beta}(t, \bar{y} + \bar{Y}_{t-}^\beta)]), \tag{2.3.5}$$

By following a similar constructive algorithm as in 2.2.3, the unique solution of the modified equation (2.3.5) is given in explicitly by:

$$\tilde{Y}_t^{\alpha,\imath} = \sum_{j=1}^{\infty} \mathbb{1}_{\{T_j \le t\}} \mathbb{1}_{\{\imath_j = \imath\}} \mathbb{1}_{\{\Theta_j \le \alpha(T_j, \tilde{y} + \tilde{Y}^{\alpha}_{T_{j-1}})\}}, \quad \forall \imath \in E.$$

b) *Existence for Equation* (2.3.4)*:* since $\tilde{Y}^{\alpha}$ is obtained by thinning of $\bar{Q}^{\Delta_\beta}$, the multivariate counting process is strongly dominated by $\bar{Y}^{\beta}$ by definition. In particular, $\tilde{y} + \tilde{Y}_t^{\alpha} \le \bar{y} + \bar{Y}_t^{\beta}$ for all $t \ge 0$ and $\tilde{y} \le \bar{y}$. Then, since $\bar{\alpha} \le \bar{\beta}$, $\bar{\alpha}(t, \tilde{y} + \tilde{Y}_t^{\alpha}) \le \bar{\beta}(t, \bar{y} + \bar{Y}_t^{\beta})$, and thus $\tilde{y} + \tilde{Y}_t^{\alpha}$ is solution of (2.3.4), which achieves to prove the existence part of the theorem.

c) *Uniqueness:* it remains to prove that any solution of (2.3.4) is solution of (2.3.5), in order to prove the uniqueness of the equation. This is easily obtained by using the previous iterative argument. Let $\bar{Y}^{\alpha}$ be a solution of (2.3.4) and T_1^{α} its first jump time, associated with the jumping component $\imath_1^{\alpha}$ and the mark Θ_1 of $\bar{Q}^{\imath_1^{\alpha}}$. By definition of the thinning procedure, $\Theta_1 \le \alpha(T_1, \tilde{y}) \le \beta(T_1, \bar{y})$. Thus, $(T_1^{\alpha}, \imath_1^{\alpha})$ is also a jump of $\bar{Y}^{\beta}$ and $\bar{Y}^{\alpha}_{T_1} \le \bar{Y}^{\beta}_{T_1}$. By iteration, we obtain that all jump times of $\bar{Y}^{\alpha}$ are jump times of $\bar{Y}^{\beta}$, or equivalently that $\bar{Y}^{\alpha} \prec \bar{Y}^{\beta}$. In particular, $\bar{Y}^{\alpha}$ is the unique solution of (2.3.5), which achieves the proof of the theorem. □

In order to apply Theorem 2.3.1, it is therefore very important to define a class of dominating non-exploding processes, large enough to enable us to obtain the existence of multivariate counting processes under reasonable assumptions. In particular, observe that since the existence or uniqueness is not obtain by controlling moments of the counting processes, the assumption of sublinear growth is not determinant.

Localization procedure and Cox Birth process A direct application of Theorem 2.3.1 is the existence and uniqueness of the solution of (2.3.4) for all intensity functionals $\bar{\alpha}(t, \bar{y})$ bounded by $\beta(t, \bar{y}) = K\bar{g}(\bar{y}^{\natural})$, with $\bar{g}$ is a deterministic function verifying the multivariate Feller criterion. In particular, such processes are honest. In the one dimensional case, this condition is sometimes known as the Jacobsen condition (Jacobsen (1982)). However, the domination by a Markov birth process is often not satisfactory. The assumption can be relaxed by using *Cox Birth processes* as dominating processes. Cox Birth processes are defined as multivariate counting processes with product intensity $\bar{\alpha}(t, \bar{y}) = k_t \bar{g}(\bar{y}^{\natural})$, where (k_t) is a locally bounded predictable process, i.e bounded by a sequence (K_p) along a nondecreasing sequence of stopping times (S_p) going to ∞. The existence of Cox Birth processes is a corollary of Theorem 2.3.1, since the solutions of (2.3.4) with intensity functional $\alpha^p(t, \bar{y}) = (k_t \wedge K_p)\bar{g}(\bar{y}^{\natural})$ do not depend on on p on the interval $[0, S_p]$.

2.3.2 BDS mulitvariate SDE controlled by a Cox Birth process

Let us now come back to the existence and pathwise realization of BDS systems of $\rho = p(p+1)$ BDS intensity functional $\boldsymbol{\mu}$, as in Definition 2.2.2. In the following, Theorem 2.3.1 is applied in order to realize jumps counting of processes of BDS systems, by strong domination with a multivariate Cox Birth process. We first present in the next paragraph the multivariate SDE associated with a BDS system, called the BDS multivariate SDE.

Birth Death Swap multivariate SDE

In reference to the jumps counting process, all multivariate quantities indexed by the set $\mathcal{J}$ of all events are now denoted by bold symbols instead of $\bar{}$. Recall that a BDS system $(\xi, \mathbf{N}, Z)$ of BDS intensity functional $\boldsymbol{\mu}$ is defined by an entry process ξ and a jumps counting process $\mathbf{N}$ of intensity $\boldsymbol{\lambda}_t = \boldsymbol{\mu}(t, Z_{t^-})$, with $Z_t = \xi_t + \phi \odot \mathbf{N}_t$ the companion population process. As a multivariate counting process, $\mathbf{N}$ can thus be represented as a solution of a multivariate SDE driven by a Poisson measure $\mathbf{Q} = (Q^\gamma)_{\gamma \in \mathcal{J}}$ indexed by $\mathcal{J}$.

Definition 2.3.1 (BDS multivariate SDE)**.**
Let $\boldsymbol{Q} = (Q^\gamma)_{\gamma \in \mathcal{J}}$ be a multivariate Poisson measure, $\boldsymbol{\mu}(t, z)$ a BDS intensity functional and (ξ_t) be an entry process. The Birth Death Swap multivariate SDE associated with the entry process ξ and intensity functional $\boldsymbol{\mu}$ is defined by

$$\mathrm{d}\boldsymbol{N}_t = \boldsymbol{Q}(\mathrm{d}t,]0, \boldsymbol{\mu}(t, \xi_{t-} + \phi \odot \boldsymbol{N}_{t-})]), \quad \textit{with } Z_t = \xi_t + \phi \odot \boldsymbol{N}_t. \tag{2.3.6}$$

If $\boldsymbol{N}$ is a solution of (2.3.6), then $(\xi, \boldsymbol{N}, Z)$ is a BDS system of entry process ξ and intensity functional $\boldsymbol{\mu}$.

Further assumptions on the BDS intensity functional are needed in order to obtain the existence and uniqueness of the solution (2.3.6). Given that the size of the population Z is dominated by the number of births, the idea is to control the birth intensity by a Cox Birth intensity, and then to apply Theorem 2.3.1 for an existence result. In particular, no assumptions are made on the swap and death intensities.

Existence and uniqueness of the BDS multivariate SDE

In order to apply Theorem 2.3.1, the BDS intensity functional has to be expressed in terms of jumps counting process, rather than in terms of population process. We use notations introduced in Subsection 2.2.1, where a counting vector $\boldsymbol{\nu}$ indexed by $\mathcal{J}$ can be decomposed into a demographic component $\boldsymbol{\nu}^{\mathrm{dem}} = (\boldsymbol{\nu}^b, \boldsymbol{\nu}^d)$ and a swap component $\boldsymbol{\nu}^s$.

$\mathbb{N}^p$-valued vectors can be written as $z = \xi + \phi \odot \boldsymbol{\nu}$. Recall that $(\phi \odot \boldsymbol{\nu})^\natural = (\phi^{\mathrm{dem}} \odot \boldsymbol{\nu}^{\mathrm{dem}})^\natural = \boldsymbol{\nu}^{b,\natural} - \boldsymbol{\nu}^{d,\natural} \leq \boldsymbol{\nu}^{b,\natural}$.

Cox Birth dominating assumption The BDS intensity functional $\boldsymbol{\mu}$ can be rewritten as a functional of $\boldsymbol{\nu}$ instead of z, $\boldsymbol{\lambda}(t, \boldsymbol{\nu}) = \boldsymbol{\mu}(t, \xi_t + \phi \odot \boldsymbol{\nu})$. Thanks to the result of Theorem 2.3.1, the problem of the existence and uniqueness of Equation (2.3.6) is reduced to the existence of a dominating honest multivariate counting process, whose intensity functional dominates the $\boldsymbol{\nu}$-intensity functional $\boldsymbol{\lambda}(t, \boldsymbol{\nu})$ of the BDS system.

As stated above, the main problem is the control of the size $Z^\natural$ of the population via the control of the size of the counting birth process $\mathbf{N}^{b,\natural}$. In the rest of this chapter and in Chapter 3, we now assume that the birth part $\boldsymbol{\lambda}^b$ of the intensity functional is dominated by a Cox Birth intensity functional:

$$\textit{Cox Birth Hyp} \quad \forall i \in \mathcal{I}_p,\ \lambda^{(b,i)}(t, \boldsymbol{\nu}) = \mu^{(b,i)}(t, \xi_t + \phi \odot \boldsymbol{\nu}) \leq k_t\, g^{(b,i)}(\xi_t^\natural + \boldsymbol{\nu}^{b,\natural}), \quad (2.3.7)$$

where the p components of the function $\mathbf{g}^b = (g^{(b,i)})$ are non-decreasing and satisfy the Feller criterion. As before, (k_t) is a predictable and locally bounded process.

The swap and death part of the intensity functional can be dominated by a multivariate functional depending only on the size of the population. When expressed in terms of population z, the swap and death intensity components are dominated for any $(i, \beta) \in \mathcal{J}^s \cup \mathcal{J}^d$ by

$$\mu^{(i,\beta)}(t, z) \leq \sup_{\{\zeta^\natural \leq z^\natural\}} \mu^{(i,\beta)}(t, \zeta) = \hat{\mu}^{(i,\beta)}(t, z^\natural)$$

By construction, the functional $\hat{\mu}^{(i,\beta)}(t, n)$ is non-decreasing in n. When $z = \xi_t + \phi \odot \boldsymbol{\nu}$, we have $z^\natural \leq \xi_t^\natural + \boldsymbol{\nu}^{b,\natural}$ and the previous inequality can be rewritten in terms of counting vector $\boldsymbol{\nu}$, for any $(i, \beta) \in \mathcal{J}^s \cup \mathcal{J}^d$:

$$\textit{Swap and death inequality} \quad \lambda^{(i,\beta)}(t, \boldsymbol{\nu}) = \mu^{(i,\beta)}(t, \xi_t + \boldsymbol{\nu}) \leq \hat{\mu}^{(i,\beta)}(t, \xi_t^\natural + \boldsymbol{\nu}^{b,\natural}) \quad (2.3.8)$$

Dominating multivariate counting process The previous paragraph shows that under Assumption (2.3.7), the $\boldsymbol{\nu}$- BDS intensity function $\boldsymbol{\lambda}$ is dominated by the the intensity functional $\hat{\boldsymbol{\lambda}}$ ($\boldsymbol{\lambda} \leq \hat{\boldsymbol{\lambda}}$), defined for all $\gamma \in \mathcal{J}$ by

$$\hat{\lambda}^\gamma(t, \boldsymbol{\nu}) = \begin{cases} k_t g^{(b,i)}(\xi_{t^-}^\natural + \boldsymbol{\nu}^{b,\natural}) & \text{if } \gamma = (b,i) \in \mathcal{J}^b \text{ is a birth event,} \\ \hat{\mu}^{(i,\beta)}(t, \xi_{t^-}^\natural + \boldsymbol{\nu}^{b,\natural}) & \text{if } \gamma = (i,\beta) \in \mathcal{J}^s \cup \mathcal{J}^d \text{ is a swap or death} \end{cases} \quad (2.3.9)$$

It thus remains to prove the existence of a non exploding $p(p+1)$-dominating process $\mathbf{G} = (\mathbf{G}^b, \mathbf{G}^d, \mathbf{G}^s)$ of intensity functional $\hat{\boldsymbol{\lambda}}$. The dominating multivariate counting

process is built in two steps:
(i) The first step is to introduce the p-Cox Birth process $\mathbf{G}^b_t(\xi)$, solution of the multivariate SDE:

$$\mathrm{d}\mathbf{G}^b_t = \mathbf{Q}^b(\mathrm{d}t,]0,\ k_t\, \mathbf{g}^b(\xi_{t^-} natural + G^{b,\natural}_{t^-})]). \tag{2.3.10}$$

By the localization procedure defined as a corollary of Theorem 2.3.1, $\mathbf{G}^b$ is a well-defined (non-exploding) p-Cox Birth process.

(ii) The second step is to add "swap and death coordinates" to $\mathbf{G}^b$, by defining the p and $p(p-1)$ multivariate Cox processes:

$$\mathrm{d}\mathbf{G}^d_t = \mathbf{Q}^d(\mathrm{d}t,]0, \hat{\boldsymbol{\mu}}(t, \xi^{\natural}_{t^-} + G^{b,\natural}_{t^-})]), \quad \mathrm{d}\mathbf{G}^s_t = \mathbf{Q}^s(\mathrm{d}t,]0, \hat{\boldsymbol{\mu}}(t, \xi^{\natural}_{t^-} + G^{b,\natural}_{t^-})]). \tag{2.3.11}$$

Observe that the previous thinning equation is not a differential equation, since the intensities of the processes $\mathbf{G}^d$ and $\mathbf{G}^s$ do not depend on the counting processes themselves. $\mathbf{G}^d$ and $\mathbf{G}^s$ are thus simple Cox processes, that is Poisson processes with stochastic intensity.
To summarize, the dominating processes is defined as follow:

Definition 2.3.2 (Dominating process)**.** *Let $\mathbf{G} = (\mathbf{G}^b, \mathbf{G}^d, \mathbf{G}^s)$ be the $p(p+1)$ multivariate counting process defined as the solution of (2.3.10)-(2.3.11). Then $\boldsymbol{G}$ is a multivariate counting process of intensity functional $\hat{\boldsymbol{\lambda}}$ defined in (2.3.9), called the* dominating process *of the BDS system.*

Observe that the dominating process $\mathbf{G}$ does not define a population process $\xi + \phi \odot \mathbf{G}$, since $\mathbf{G}$ does not verify the support conditions 2.2.7.

Existence and uniqueness of BDS SDE by strong domination Theorem 2.3.1 allows us to conclude with the main result of this section:

Theorem 2.3.2. *Assume that the Cox Birth assumption (2.3.7) is verified: $\boldsymbol{\mu}^b(t,z) \leq k_t\, \boldsymbol{g}^b(z^{\natural})$, where the components of $\boldsymbol{g}^b$ are non-decreasing and satisfy the Feller criterion. Moreover, assume that k_t, $\boldsymbol{\mu}^s(t,K)$ and $\boldsymbol{\mu}^d(t,K)$ are locally bounded in time for any K. Then, there exists a unique solution to Equation (2.3.6),*

$$d\boldsymbol{N}_t = \boldsymbol{Q}(dt,]0, \boldsymbol{\mu}(t, \xi_{t-} + \phi \odot \boldsymbol{N}_{t-}]), \quad \textit{with } Z_t = \xi_t + \phi \odot \boldsymbol{N}_t.$$

The triplet $(\xi, \boldsymbol{N}, Z)$ is a well-defined BDS system of BDS intensity functional $\boldsymbol{\mu}$ and entry process ξ. Furthermore, $\boldsymbol{N}$ is strongly dominated by $\boldsymbol{G}$, $\boldsymbol{N} \prec \boldsymbol{G}$.

Since $\mathbf{N}^b \prec \mathbf{G}^b$, all jumps of $\mathbf{N}^b$ are jumps of $\mathbf{G}^b$. This can interpreted as follow: "all individuals born in the Cox Birth population $\xi + \mathbf{G}^b$ are also born in the population Z".

There are many advantages in the construction by domination, in particular since the BDS jumps counting process $\mathbf{N}$ can be localized by a sequence of increasing stopping times which do not depend on the process itself. This property will be instrumental in the next chapter in order to obtain tightness properties. However, the same property can be a drawback when simulating the BDS system by strong domination. Indeed, the dominating process $\mathbf{G}$ can have much more jumps that $\mathbf{N}$, making the simulation inefficient.
In the last section of this chapter, an alternative construction of the solution of (2.3.6) is presented, called the *Birth Death Swap decomposition algorithm* and based on the disentanglement of swap and demographic events. The decomposition algorithm is better suited to the simulation of BDS systems when swap and demographic intensities are of a very different nature, for instance when they are supposed to have their own timescale. The disentanglement of swap and demographic events will also be instrumental in the proof of Theorem 3.5.1 in the next Chapter.

2.4 BDS decomposition algorithm

In the previous section, we have built BDS systems of initial condition ξ_t and intensity functional $\boldsymbol{\mu}$ as the solution of the multivariate SDE 2.3.6 driven by $\mathbf{Q}$.

$$Z_t = \xi_t + \phi \odot \mathbf{N}_t, \quad d\mathbf{N}_t = \mathbf{Q}(dt,]0, \boldsymbol{\mu}(t, Z_{t^-})]).$$

This construction of BDS systems, however, is not necessarily the most suitable when the goal is to disentangle the swap processes from demographic process, especially when they are supposed to have their own time scales. Recall that swap events can be distinguished from demographic events by writing the dynamic of the companion population process as:

$$Z_t = \xi_t + \phi^s \odot \mathbf{N}^s_t + \mathbf{N}^b_t - \mathbf{N}^d_t,$$

As stated in 2.2.3, $\mathbf{N}^s$ is not a swap jumps counting process associated with a well defined swap process, and $\mathbf{N}^{\text{dem}} = (\mathbf{N}^b, \mathbf{N}^d)$ can not be associated with a birth death process either, due to the occuring of swap events between two demographic events. In this section, we present an alternative construction of the BDS system multivariate intensity based on the study of the population process between two successive demographic events. This point of view will be instrumental for disentangling the dynamics of $\mathbf{N}^s$ and $\mathbf{N}^{\text{dem}}$ in the averaging result of the next Chapter.

2.4.1 Direct analysis of the BDS multivariate SDE

The main idea of the BDS decomposition algorithm is to generalize the decomposition of population with only demographic event given in 2.2.2 to BDS systems.

The sequence of demographic event times, defined as the jumps times of the counting process $N^{\mathrm{dem},\natural} = N^{b,\natural} + N^{d,\natural}$ is denoted by $(T_0 = 0, T_1..T_k, ..)$ (without the dem reference for simplicity). For all $k \geq 0$, if $T_k < \infty$ then:

$$T_{k+1} = \inf\{t > T_k, N_t^{\mathrm{dem},\natural} = k+1\}, \quad N_t^{\mathrm{dem},\natural} = k, \ \forall t \in [T_k, T_{k+1}[\tag{2.4.1}$$

Recall that $T_1 > \tau$ since the jumps counting process don't jump before the first jump time τ of the entry process. Furthermore, due to the construction by strong domination, $\lim_{k\to+\infty} T_k = +\infty$ and all demographic event times are also jump times of the demographic part $\mathbf{G}^{\mathrm{dem}}$ of the dominating process $\mathbf{G}$.

For all $k \geq 1$, the time interval $[T_{k-1}, T_k[$ is called the k-th demographic interval. By definition, $\mathbf{N}^{\mathrm{dem}}$ is constant on each demographic interval.

BDS system before the first demographic event The BDS system stopped at the first demographic event time T_1 is an example of population with only one demographic jump, as described in the previous section in Paragraph 2.2.2.

On the first demographic interval $[0, T_1[$, the demographic counting system is null and the population process behaves as a pure swap process since $Z_t = \xi_t^\tau + \phi^s \odot \mathbf{N}_t^s$ with $\mathrm{d}\mathbf{N}^s = \mathbf{Q}^s(\mathrm{d}t,]0, \boldsymbol{\mu}^s(t, \xi_t^\tau + \phi^s \odot \mathbf{N}_{t^-}^s)])$.

a) Since $\mathbf{Q}^s$ doesn't jump at T_1, the process defined by $X_t^{0,1} = \xi_t^\tau + \phi^s \odot \mathbf{N}_{t\wedge T_1}^s$ is continuous at time T_1 and behaves as a pure swap process X^0 stopped at T_1. The definition of the underlying swap process X^0 is not unique. The most straight forward construction might be to consider a swap of intensity functional $\boldsymbol{\mu}^s(t,z)\mathbb{1}_{\{t\leq T_1\}}$, null on $]T_1, \infty] = \{N_{t-}^{\mathrm{dem},\natural} > 0\}$. However, our goal is to disentangle Swap and demographic events, and we prefer to define X^0 as the solution of the following Swap multivariate SDE:

$$X_t^0 = \xi_t^\tau + \phi^s \odot \mathbf{N}_t^{\mathrm{sw},0}, \quad \mathrm{d}\mathbf{N}_t^{\mathrm{sw},0} = \mathbf{Q}^s(\mathrm{d}t,]0, \boldsymbol{\mu}^s(t, X_{t^-}^0)]). \tag{2.4.2}$$

By unicity of the previous multivariate SDE, $X_t^{0,1} = X_{t\wedge T_1}^0$ and the population process coincides with X^0 on the first demographic interval.

b) At the first demographic event time T_1, the population process jump of $\mathbf{N}_{T_1}^b - \mathbf{N}_{T_1}^d$. By definition, one and only one component of the vector $\mathbf{N}^{\mathrm{dem}}$ jumps at T_1 and an individual is either added (birth) or removed (death) to one of the subgroups.

The important point is that the first demographic event is completely characterized from the Poisson measure $\mathbf{Q}^{\text{dem}}$ and the Swap process X^0, since $\mathrm{d}\mathbf{N}^{\text{dem}}_t = \mathbf{Q}^{\text{dem}}(\mathrm{d}t, [0, \boldsymbol{\mu}^{\text{dem}}(t, Z_{t-})])$ and $Z_{t-} = X^0_{t-}$ on $]0, T_1]$. Thus,the first demographic event can be expressed as the first jump of a multivariate counting process $\mathbf{K}^1 = (\mathbf{K}^{1,+}, \mathbf{K}^{1,-})$, called *switch process* defined by the following thinning equation:

$$\mathrm{d}\mathbf{K}^1_t = \mathbf{Q}^{\text{dem}}(\mathrm{d}t, [0, \boldsymbol{\mu}^{\text{dem}}(t, X^0_{t-})]) \tag{2.4.3}$$

In particular, $]0, T_1] = \{t;\ \mathbf{K}^1_{t-} = 0\}$ and the demographic event *type* is characterized by the component of $\mathbf{K}^1$ which jumps at T_1.
The BDS system stopped at the demographic event can then be rewritten as:

$$Z_{t\wedge T_1} = X^{0,1}_t + \mathbf{K}^{1,+}_{t\wedge T_1} - \mathbf{K}^{1,-}_{t\wedge T_1} = X^{0,1}_t + \phi^{\text{dem}} \odot \mathbf{K}^1_{t\wedge T_1}. \tag{2.4.4}$$

c) From there, a new population process Z^{1+} can be defined, starting at time T_1 in state $Z_{T_1} = X^0_{T_1} + \mathbf{K}^{1,+}_{T_1} - \mathbf{K}^{1,-}_{T_1}$ and of jumps counting process $\mathbf{N}^1 = \mathbf{1}_{\{K^1_{t-}>0\}} * \mathbf{N}$. By definition of $\mathbf{K}^1$, $\mathbf{N}^1 = \mathbf{1}_{]T_1,\infty]} * \mathbf{N}$ and by the stopping procedure defined in 2.2.2, Z^{1+} is a population process coinciding with Z on $[T_1, \infty[$. Its first demographic event is T_2, and steps a) and b) can thus be applied to Z^{1+} in order to obtain a second swap process, coinciding with the population process on the second demographic interval $[T_1, T_2[$, and a switch process determining the second demographic event.

Observe that we can give a differential version of Equation (2.4.4), which avoids to enumerate jump times. By equation (2.4.2) and (2.4.4), $Z_{t\wedge T_1} = \xi^\tau_t + \phi \odot (\mathbf{N}^{\text{sw}}_{t\wedge T_1} + \mathbf{K}_{t\wedge T_1})$. Furthemore, $\mathrm{d}\mathbf{N}^{\text{sw}}_{t\wedge T_1} = \mathbb{1}_{]0,T_1]}\mathrm{d}\mathbf{N}^{\text{sw}}_t$ (idem for $\mathbf{K}_{t\wedge T_1}$),and since $]0, T_1] = \{t;\ \mathbf{K}^1_{t-} = 0\}$:

$$Z_{t\wedge T_1} = \xi^\tau_t + \phi \odot \mathbf{N}^{0,1}_t \quad \mathrm{d}\mathbf{N}^{0,1}_t = \mathbb{1}_{\{\mathbf{K}^1_{t-}=0\}}(\mathrm{d}\mathbf{N}^{\text{sw}}_t + d\mathbf{K}^1_t). \tag{2.4.5}$$

In particular, $\mathbf{N}^{0,1}$ is the jumps counting process of a population with only one demographic event.

Recursive decomposition on successive demographic intervals The population can be decomposed at any demographic event time T_k in:

$$Z_t = Z_{t\wedge T_k} + \phi \odot \mathbf{N}^k_t, \quad \mathbf{N}^k_t = \mathbf{1}_{]T_k,\infty]} * \mathbf{N}_t = \mathbf{N}_{t\vee T_k} - \mathbf{N}_{T_k}.$$

The analysis of the BDS before the the first demographic event can be applied to $Z^{k+} = \xi^k_t(Z) + \phi \odot \mathbf{N}^k$, where recall that $\xi^k_t(Z) = Z_{T_k}\mathbb{1}_{\{T_k\le t\}}$. The first demographic interval of Z^{k+} is $[T_k, T_{k+1}[$. A swap process with entry process $\xi^k_t(Z)$ can be defined by:

$$X^k_t = \xi^k_t(Z) + \phi^s \odot \mathbf{N}^{\text{sw},k}_t \quad \mathrm{d}\mathbf{N}^{\text{sw},k}_t = \mathbf{Q}^s(\mathrm{d}t, [0, \boldsymbol{\mu}^s(t, X^k_{t-})]),$$

associated with a switch process $\mathrm{d}\mathbf{K}_t^{k+1} = \mathbf{Q}^{\mathrm{dem}}(\mathrm{d}t, [0, \boldsymbol{\mu}^{\mathrm{dem}}(t, X_{t^-}^k)])$. Z^{k+} coincides with X^k on $[T_k, T_{k+1}[$ and T_{k+1} is the first jump of $\mathbf{K}^{k+1}$. Using the differential version 2.4.5,

$$Z_{t\wedge T_{k+1}} = Z_{t\wedge T_k} + \phi \odot \mathbf{N}_t^{k,k+1}, \quad \mathrm{d}\mathbf{N}_t^{k,k+1} = \mathbb{1}_{\{\mathbf{K}_{t^-}^{k+1}=0\}}(\mathrm{d}\mathbf{N}_t^{\mathrm{sw},k} + d\mathbf{K}_t^{k+1}), \tag{2.4.6}$$

and a new population can be generated, starting at time T_{k+1} in state $Z_{T_{k+1}}^{k+1} = X_{T_{k+1}}^k + K_{T_{k+1}}^{k+1,+} - K_{T_{k+1}}^{k+1,-}$.

2.4.2 BDS decomposition algorithm

The direct analysis of the BDS system serves as basis for defining the BDS decomposition algorithm. Our goal is to build a solution of (2.3.6) recursively on its demographic intervals, by continuous pasting of populations with one demographic event, defined recursively using Swap and Switch processes. Formally the BDS decomposition algorithm is defined as follow:

Initialization

Let $(X^0, \mathbf{K}^1)$ be the Swap-Switch process solution of (2.4.2)-(2.4.3). Observe that solutions of (2.4.2)-(2.4.3) are automatically well defined since swap processes have a constant size and $\mathbf{K}^1$ is a simple multivariate Cox process.

The population is defined on $]0, \tau_1] = \{t;\ K_{t^-}^1 = 0\}$ by:

$$Z_{t\wedge\tau_1} = \xi_t^\tau + \phi \odot \mathbf{N}_t^{0,1}, \quad \mathrm{d}\mathbf{N}_t^{0,1} = \mathbb{1}_{\{\mathbf{K}_{t^-}^1=0\}}(\mathrm{d}\mathbf{N}_t^{\mathrm{sw},0} + \mathrm{d}\mathbf{K}_t^1) = \mathbb{1}_{\{\mathbf{K}_{t^-}^1=0\}}\mathbf{Q}(\mathrm{d}t,]0, \boldsymbol{\mu}(t, X_{t^-}^0)]).$$

In particular, the first jump of $\mathbf{K}^1$ is the first demographic event in the population and $Z_{t^-} = X_{t^-}^0$ on $[0, \tau_1]$.

$k+1$th step

Assume now that a sequence $(X^l, \mathbf{K}^{l+1})_{1\le l\le k-1}$ of Swap-Switch processes have been defined, alongside with the population process defined on $[0, \tau_k]$.The $k+1$th Swap-Switch process is defined by a swap process X^k starting at time τ_k in state Z_{τ_k} and solution of the multivariate SDE:

$$X^k = \xi^{\tau_k}(Z) + \phi \odot \mathbf{N}_t^{\mathrm{sw},k}, \quad \mathrm{d}\mathbf{N}_t^{\mathrm{sw},k} = \mathbf{Q}^s(\mathrm{d}t,]0, \boldsymbol{\mu}^s(t, X_{t^-}^k)]),$$

associated with the Switch process $d\mathbf{K}_t^{k+1} = \mathbf{Q}^{\mathrm{dem}}(0,]0, \boldsymbol{\mu}^{\mathrm{dem}}(t, X_{t^-}^k)])$. By construction, the first jump time τ_{k+1} of $\mathbf{K}^{k+1}$ is strictly greater than τ_k, and the population is defined on $[0, \tau_{k+1}]$ by:

$$\begin{aligned} &Z_{t\wedge\tau_{k+1}} = Z_{t\wedge\tau_k} + \phi \odot \mathbf{N}_t^{k,k+1}, \\ &d\mathbf{N}_t^{k,k+1} = \mathbb{1}_{\{\mathbf{K}_{t^-}^{k+1}=0\}}(d\mathbf{N}_t^{\mathrm{sw},k} + d\mathbf{K}_t^{k+1}) = \mathbb{1}_{\{\mathbf{K}_{t^-}^{k+1}=0\}}\mathbf{Q}(\mathrm{d}t,]0, \boldsymbol{\mu}(t, X_{t^-}^k)]). \end{aligned}$$

By construction, τ_{k+1} is the $k+1$th demographic event time and $Z_{t^-} = X^k_{t^-}$ on $]\tau_k, \tau_{k+1}]$.

By iterating the last step, we obtain a BDS system defined by:

$$Z_t = \xi^\tau_t + \phi \odot \mathbf{N}_t, \quad \text{with } \mathrm{d}\mathbf{N}_t = \sum_{k\geq 0} \mathbb{1}_{\{\mathbf{K}^{k+1}_{t^-}=0\}} \mathbf{Q}(\mathrm{d}t,]0, \boldsymbol{\mu}(t, X^k_{t^-})]). \tag{2.4.7}$$

Firstly, for all $k \geq 0$, $X^k_{t^-} = 0$ on $[0, \tau_k]$, so that $\{\mathbf{K}^{k+1}_{t^-} = 0\}$ can be replaced by $]\tau_k, \tau_{k+1}]$ in the r.h.s of the previous equation. Furthermore, since $Z_{t^-} = X^k_{t^-}$ on $]\tau_k, \tau_{k+1}]$, the previous equation can be rewritten as:

$$Z_t = \xi^\tau_t + \phi \odot \mathbf{N}_t, \quad \text{with } \mathrm{d}\mathbf{N}_t = \mathbf{Q}(\mathrm{d}t, [0, \boldsymbol{\mu}(t, Z_{t^-})]) \textstyle\sum_{k\geq 0} \mathbb{1}_{]\tau_k, \tau_{k+1}]}$$

In order to conclude, it remains to show that $\lim \tau_k = \infty$. By the Cox Birth domination assumption (2.3.7), we can actually show using the same reasoning as in the previous section that the multivariate counting process defined above is strongly dominated by the non-exploding process $\mathbf{G}$. Thus, $\lim \tau_k = \infty$ a.s. and this achieves to prove that the solution of the BDS decomposition algorithm is a solution of the BDS thinning equation (2.3.6).

Chapter 3

Averaging of BDS systems in the presence of two timescales

3.1 Introduction

Aggregation methods are important in several fields such as in economics, ecology, biology or operations research. Indeed, they provide a better understanding of the link between finer-grained dynamics and aggregated variables. Another common feature of complex systems is that one can often observe a separation of time scale between different types of phenomena.

In many populations, the rates at which demographic events and swap events happen appear to be very different. This is the case for human populations where some changes in the social structure of the population can sometimes be fast in comparison with the demographic timescale. In the ecological systems described in Auger et al. (2000); Marvá et al. (2013), migrations between different patches or changes of strategies occur at a much faster timescale than demographic events. This separation of timescale often leads to averaging approximations for the aggregated process. In this chapter, we consider that swap events happen at a much faster timescale than demographic events, and study the Birth Death Swap (BDS) system in presence of these two timescales.

The literature on the evolution of populations in presence of several time scales is wide and interdisciplinary. Motivated by the study of ecological systems, study of deterministic populations with two timescales can be found in Auger et al. (2000, 2012); Sánchez et al. (2000), where most of the results are obtain using singular perturbation theory. In the field of operation research, Yin and Zhang have obtained approximation results for continuous time Markov chains with two timescales, using asymptotic expansions also based on singular perturbation theory (Yin and Zhang (2004),Yin and Zhang (2012)).

In Taylor and Véber (2009), the genealogy of individuals from a subdivided population that experiences sporadic mass extinction events is studied, where the "fast" process converge to an absorbing state between two extinction events. However, in other works, the population is often renormalized in presence of two timescales, which is not the case here. Futhermore, due to our non Markov framework, we cannot apply classical averaging results such as Kurtz (1992).
Before going further, we first introduce in the next section the two timescales BDS sytem, and then describe the scope of the Chapter.

3.2 BDS system in the presence of two timescales

The framework introduced in the previous chapter is the same. We consider BDS systems as defined in 2.2.2, whose jumps counting processes are built by thinning of a multivariate Poisson measure $\mathbf{Q}$, as presented in Definition 2.3.6. In particular, all BDS intensity functionals are assumed to verify the Cox Birth dominating assumption 2.3.7.
In this chapter, we now consider that swap events occur at a much faster timescale than demographic events. We us first define the BDS in this two timescale framework, and the present the scope of this chapter.

3.2.1 Two timescales BDS system

Swap events are assumed to occur at a much faster timescale than demographic events. Intuitively, this means that demographic events happen with low intensities of order "$O(1)$" in comparison with swap events occurring with greater intensities of order "$O(\frac{1}{\epsilon})$", depending on a small parameter ϵ. The following figure gives an example of the distribution of the different types of events in this two timescale framework.

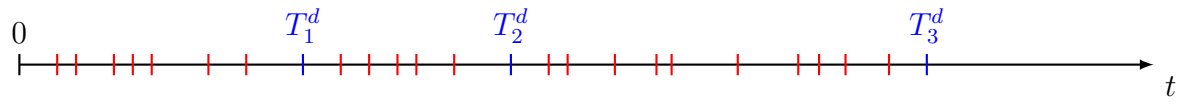


Fig. 3.1 Example of distribution of swap events and demographic events

Definition of the two timescales BDS system More formally, the BDS system associated with the given multivariate Poisson measure $\mathbf{Q}$ and the BDS intensity functional $(\boldsymbol{\mu}^\epsilon = \boldsymbol{\mu}^{\text{dem}} + \frac{1}{\epsilon}\boldsymbol{\mu}^s)$ is now depending on a small parameter ϵ. The BDS system is now

denoted by $(\xi_t, \mathbf{N}^\epsilon, Z^\epsilon)$. The BDS multivariate SDE 2.3.6 becomes:

$$\begin{aligned} Z_t^\epsilon &= \xi_t + \phi^s \odot \mathbf{N}_t^{s,\epsilon} + \mathbf{N}_t^{b,\epsilon} - \mathbf{N}_t^{d,\epsilon}, \quad \mathbf{N}_t^{\mathrm{dem},\epsilon} = (\mathbf{N}_t^{b,\epsilon}, \mathbf{N}_t^{b,\epsilon}) \\ \mathrm{d}\mathbf{N}_t^{s,\epsilon} &= \mathbf{Q}^s(\mathrm{d}t,]0, \frac{1}{\epsilon}\boldsymbol{\mu}^s(t, Z_{t^-}^\epsilon)]), \quad \mathrm{d}\mathbf{N}_t^{\mathrm{dem},\epsilon} = \mathbf{Q}^{\mathrm{dem}}(\mathrm{d}t,]0, \boldsymbol{\mu}^{\mathrm{dem}}(t, Z_{t^-}^\epsilon)]). \end{aligned} \tag{3.2.1}$$

The swap part $\mathbf{N}^{s,\epsilon}$ of the jumps counting process is called the swap counting system, and the demographic part $\mathbf{N}^{\mathrm{dem},\epsilon}$ is called the demographic counting system. As stated in the previous chapter, the swap and the demographic counting systems are entangled due to their dependence on $Z_t^\epsilon = F(\mathbf{N}_t^{\mathrm{dem},\epsilon}, \mathbf{N}_t^{s,\epsilon})$. However, the behavior of the two multivariate counting processes is very different in this separation of timescale framework:

- On one hand, the swap counting system $\mathbf{N}^{s,\epsilon}$ evolves on a fast timescale. The process depends on the small parameter ϵ through its intensity functional $\frac{1}{\epsilon}\boldsymbol{\mu}^s(t,z)$ *and* through the population process Z^ϵ. Thus, $\mathbf{N}^{s,\epsilon}$ will explode when $\epsilon \to 0$.
- On the other hand, the demographic counting system $\mathbf{N}^{\mathrm{dem},\epsilon}$ only depends on ϵ through Z^ϵ.

The construction of the two timescale BDS differs significantly from usual constructions proposed in the literature. Often, the fast evolution is modeled by a change of time and adapted to a filtration indexed by ϵ. Here, the thinning construction (3.2.1) of the two timescale BDS system allows all processes to be adapted to the same filtration $(\mathcal{G}_t)$, for all $\epsilon > 0$. This property simplifies the interpretations.

Dominating counting process As before, the solution of (3.2.1) can be obtained by strong domination with a dominating counting process. The dominating process of Definition 2.3.2 is also index by ϵ and denoted by $\mathbf{G}^\epsilon$. The critical property of the dominating process in the two timescale framework is that its demographic part $\mathbf{G}^{\mathrm{dem}} = (\mathbf{G}^b, \mathbf{G}^d)$ *does not depend on* ϵ. Indeed, the Cox Birth process $\mathbf{G}^b$ and the Cox process $\mathbf{G}^d$ are still defined by Equations 2.3.10 and 2.3.11, which are not modified:

$$\mathrm{d}\mathbf{G}_t^b = \mathbf{Q}^b(\mathrm{d}t,]0, k_t\, \mathbf{g}^b(\xi_{t^-}^\natural + G_{t^-}^{b,\natural})]), \quad \mathrm{d}\mathbf{G}_t^d = \mathbf{Q}^d(\mathrm{d}t,]0, \hat{\boldsymbol{\mu}}^d(t, \xi_{t^-}^\natural + G_{t^-}^{b,\natural})]), \tag{3.2.2}$$

On the other hand, the swap part $\mathbf{G}^{s,\epsilon}$ is now depending on ϵ since the swap intensity functional is $\frac{1}{\epsilon}\boldsymbol{\mu}^s$. $\mathbf{G}^{s,\epsilon}$ is the multivariate Cox process defined by the modified version of 2.3.11,

$$\mathrm{d}\mathbf{G}_t^{s,\epsilon} = \mathbf{Q}^s(\mathrm{d}t,]0, \frac{1}{\epsilon}\hat{\boldsymbol{\mu}}^s(t, \xi_{t^-}^\natural + G_{t^-}^{b,\natural})]) \tag{3.2.3}$$

Since $\mathbf{G}^{s,\epsilon}$ strongly dominates $\mathbf{N}^{s,\epsilon}$, the process also explodes when ϵ goes to 0.

These properties have a major impact on the behavior of the two timescale BDS system.

Since the demographic part $\mathbf{G}^{\text{dem}}$ of the dominating process does not depend on ϵ, the demographic counting systems $\mathbf{N}^{\text{dem},\epsilon}$ are uniformly strongly dominated by the multivariate counting process $\mathbf{G}^{\text{dem}}$. Furthermore, the size of the population $Z_t^{\epsilon,\natural} = N_t^{b,\natural,\epsilon} - N_t^{d,\natural,\epsilon}$ is also uniformly bounded by $G_t^{b,\natural}$ at each time t. These domination properties will be instrumental in obtaining tightness properties for the family of multivariate counting processes $(\mathbf{N}^{\text{dem},\epsilon})_{\epsilon>0}$, as well as in showing the tightness of (Z^ϵ), when seen as a family a random variables defined *on a well chosen space*.

3.2.2 Scope of the chapter

The aim of this chapter is to study the convergence of the BDS system (3.2.1) when swap events become instantaneous with respect to demographic events ($\epsilon \to 0$). We focus on convergence properties of the family of demographic counting systems $(\mathbf{N}^{\text{dem},\epsilon})$ and of the family of population processes (Z^ϵ).

As stated in the previous paragraph, the family of demographic counting systems $(\mathbf{N}^{\text{dem},\epsilon})$ is uniformly strongly dominated by the process $\mathbf{G}^{\text{dem}}$. This allows us to consider the *convergence in distribution* of the demographic counting systems *as dynamic processes*, i.e viewed as random variables taking values in the space in the space of $2p$-multivariate counting processes.

The situation is different for the family of population processes $(Z^\epsilon)_{\epsilon>0}$, which is not tight in the space of càdlàg pure jump processes on $\mathbb{N}^p$, due to the explosion of swap events. However, the tightness of (Z^ϵ) can be obtained in a weaker framework, by considering the population process as an $\mathbb{N}^p$-valued random variable $Z^\epsilon(\omega, s)$ defined on the product space $\Omega \times \mathbb{R}^+$ (equipped with the product measure $\mathsf{P} \otimes \text{Leb}$), rather than a dynamic process.

In the non-Markov framework of BDS systems, intensities are random functionals, and classical averaging results based on the weak convergence, such as in Kurtz (1992), cannot be applied here. Nevertheless, the notion of *stable convergence* (Jacod and Mémin (1981)) allows us to overcome these difficulties. For ease of reading, we first give an overview of this mode of convergence in Section 3.3, whose advantage is to maintain martingale properties as well as the "thinning" structure of the problem. At the end of the section, *stable limits* of the population processes $Z^\epsilon(\omega, s)$ are defined on $\Omega \times \mathbb{R}^+$.

The study of the demographic counting systems is particularly interesting, since the *aggregated population process* $Z^{\natural,\epsilon}$ counting the number of individuals in the population is a simple function of $\mathbf{N}^{\text{dem},\epsilon}$. In general, $Z^{\natural,\epsilon}$ is not an "autonomous" Birth Death process, since its birth and death intensities functionals $\mu^{b,\natural}$ and $\mu^{d,\natural}$ depend on the

whole structure of the population and not just on $Z^{\natural,\epsilon}$. In addition, these intensities are not constant between two demographic events due to the occurrence of swap events. However, in the presence of two timescales, swap events can have an *averaging effect* on demographic intensities, thus reducing the complexity of the aggregated process.

In Section 3.4, a general identification result is proven, characterizing the intensity of stable limits of the demographic counting systems. By considering the joint stable convergence of $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^{\epsilon})$ on the right space, stable limits of $(\mathbf{N}^{\mathrm{dem},\epsilon})$ are identified as "autonomous" demographic counting systems with an *averaged intensity*, corresponding to the demographic intensity functional $\boldsymbol{\mu}^{\mathrm{dem}}$ averaged against stable limits of the population process $Z^{\epsilon}(\omega, s)$.

In Section 3.5, a convergence result for the family of demographic counting systems is obtained in the case of deterministic swap intensities. We show that in this particular case, the demographic counting systems converge in distribution to the jumps counting system of a "true" (Non-Markov) multi-type Birth Death process. Due to the averaging effect of swap events, the limit birth and death intensity functionals only depend on the size of the process. In the particular, the limit aggregated process is also a (one dimensional) Birth-Death process. The proof of this result relies heavily on the BDS decomposition algorithm presented in Section 2.4 of Chapter 2.
Finally, we show on an example how non-linearities in the aggregated birth and death intensities can emerged, resulting from a non-trivial aggregation of subgroup specific birth and death intensities.

3.3 Overview on the stable convergence and application to the population process

In order to prove the identification result of this section, we shall need the notion of stable convergence. In the sequel, the stable convergence is applied on one hand to identify the intensity of stable limit points of the demographic counting systems, and on the other hand to realize stable limits of the demographic counting systems, with the aim to preserve the initial structure (martingale properties, thinning ...).

3.3.1 Stable Convergence

Originated by Alfred Rényi, the notion of stable convergence, which is stronger than the classical weak convergence of probability measures, is used in many limit theorems in probability and statistics in random environment. Some useful characterizations and

properties of the stable convergence may be found in Jacod and Mémin (1981), Jacod and Shiryaev (1987) or Aldous et al. (1978). A very detailled presentation is also given in the recent book of Häusler and Luschgy (2015).
The introduction of this mode of convergence can be motivated in different ways. The point of view best suited to our purpose defines the stable convergence as a convergence of probability measures on an extended space, in which the randomness of the initial structure is taken into account without change.

Enlarged space of Rules The initial probabilistic structure is a given probability space $(\Omega, \mathcal{G}, \mathsf{P})$. We add to this given structure a "canonical" Polish space $\mathcal{X}$, equipped with its Borel σ-field $\mathcal{B}(\mathcal{X})$. $\mathcal{X}$ is the state space of the random variables of interest.
A natural extension preserving the initial structure is to consider the enlarged measurable product space $(\bar{\Omega}, \bar{\mathcal{G}}) = (\Omega \times \mathcal{X}, \mathcal{G} \otimes \mathcal{B}(\mathcal{X}))$, where the "canonical" (identity) variable on $\mathcal{X}$ is denoted by $\Upsilon(\omega, \chi) = \chi$. The admissible probability measures on $(\bar{\Omega}, \bar{\mathcal{G}})$, called **rules**, are characterized by the constraint to have their marginal on $(\Omega, \mathcal{G})$ equal to the given probability measure P. The set of rules on $(\bar{\Omega}, \bar{\mathcal{G}})$ is denoted by $\mathcal{R}(\mathsf{P}, \mathcal{X})$ (the notation $\mathcal{R}(\mathsf{P})$ is sometimes used when there is no ambiguity on the space $\mathcal{X}$).
The rule associated with an $\mathcal{X}$-valued random variable $Y(\omega)$ is defined by $\mathsf{R}^Y(\mathrm{d}\omega, \mathrm{d}\chi) = \mathsf{P}(\mathrm{d}\omega)\delta_{Y(\omega)}(\mathrm{d}\chi)$. The integral of any $\bar{\mathcal{G}}$-measurable bounded r.v. $H(\omega, x)$ is given by

$$\mathsf{R}^Y(H) = \int_{\Omega\times\mathcal{X}} \mathsf{R}^Y(\mathrm{d}\omega, \mathrm{d}\chi)\, H(\omega, \chi) = \int_\Omega \mathsf{P}(\mathrm{d}\omega) H(\omega, Y(\omega)) = \mathsf{E}[H(., Y)]. \tag{3.3.1}$$

In particular, the restriction to $\mathcal{X}$ of the rule R^Y is the probability distribution μ^Y of Y. The R^Y-distribution of the canonical variable Υ is also μ^Y, and $\delta_{Y(\omega)}(\mathrm{d}\chi)$ is the R^Y-conditional distribution of Υ given $\mathcal{G}$.

The Dirac measure can be replaced by a random probability kernel $\Gamma(\omega, \mathrm{d}\chi)$ from $(\Omega, \mathcal{G})$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, and generated a rule defined by

$$\mathsf{R}^\Gamma(\mathrm{d}\omega, \mathrm{d}\chi) = \mathsf{P}(\mathrm{d}\omega)\Gamma(\omega, \mathrm{d}\chi).$$

Conversely, any general rule $\mathsf{R}(\mathrm{d}\omega, \mathrm{d}\chi)$ can be disintegrated in the R-conditional distribution of Υ given $\mathcal{G}$, denoted by $\Gamma^\Upsilon(\omega, \mathrm{d}\chi)$, such that

$$\mathsf{R}(h(\Upsilon)|\mathcal{G}) = \Gamma^\Upsilon(\cdot, h), \quad \mathsf{P}.a.s., \quad \mathsf{R}(\mathrm{d}\omega, \mathrm{d}\chi) = \mathsf{P}(\mathrm{d}\omega)\Gamma^\Upsilon(\omega, \mathrm{d}\chi).$$

Weak and Stable convergence Different notions of convergence can be defined on the space of probability measures defined on a polish space $\mathcal{X}$, based on different classes of test functions. For instance, the usual *"weak" convergence* is defined from the class

$\mathcal{C}_b(\mathcal{X})$ of bounded and continuous functions on $\mathcal{X}$ by

$$\mu_n \text{ converges weakly to } \mu \text{ iff } \mu_n(f) \text{ converges to } \mu(f), \quad \forall f \in \mathcal{C}_b(\mathcal{X}). \tag{3.3.2}$$

By abuse of language, a sequence of $\mathcal{X}$-valued random variables (Y_n) is said to converge in distribution to μ iff the sequence (μ_n) of the distributions of (Y_n) converges weakly to μ, or equivalently

$$(Y_n) \text{ converges weakly to } \mu \text{ iff for any } f \in \mathcal{C}_b(\mathcal{X}),\ \mathsf{E}(f(Y_n)) \to \textstyle\int_{\mathcal{X}} f(y)\mu(\mathrm{d}y).$$

The weak convergence is a convergence of probability measures on $\mathcal{X}$: a limit random variable does not have to be specified, and the random variables Y_n need not to be realized on the same probability space.

The *stable convergence* can be interpreted as an extension of the weak convergence to the probability rules R defined on the extended space $\bar{\Omega}$. The class of test functions is extended to the family $\mathcal{C}_{bmc}(\Omega \times \mathcal{X})$, of bounded functionals $H(\omega, \chi)$, $\bar{\mathcal{G}}$-measurable, continuous in χ for any ω but without any regularity in ω. In particular, no topological structure is required on the space Ω.

Definition 3.3.1 (Stable convergence)**.** *Let $(\bar{\Omega}, \bar{\mathcal{G}}) = (\Omega \times \mathcal{X}, \mathcal{G} \otimes \mathcal{B}(\mathcal{X}))$ be the enlarged measurable space defined above.*

A sequence of probability rules (R^n) is said to converge stably to $\mathsf{R} \in \mathcal{R}(\mathsf{P}, \mathcal{X})$ iff

$$\forall\, H \in \mathcal{C}_{bmc}(\Omega \times \mathcal{X}),\ \mathsf{R}^n(H) \text{ converges to } \mathsf{R}(H).$$

This property is equivalent to the convergence of the random kernels (Γ^n) associated with (R^n) to the random kernel Γ associated with R, in the sense that $\mathsf{E}[\Gamma^n(H)]$ converges to $\mathsf{E}[\Gamma(H)]$ for all $H \in \mathcal{C}_{bmc}(\Omega \times \mathcal{X})$.

It is not necessary to test the convergence on all functionals in $\mathcal{C}_{bmc}(\Omega \times \mathcal{X})$. For instance, it is sufficient to consider only the bounded functionals $H(\omega, \chi) = K(\omega)f(\chi)$ where $K(\omega)$ is $\mathcal{G}$-measurable and f is continuous on $\mathcal{X}$.

Moreover, many properties true for the weak convergence are still valid for the stable convergence (Jacod and Mémin (1981)). For instance, the "porte-manteau" theorem can be extended to the stable convergence: let $F \in \bar{\mathcal{G}}$ such that $\forall \omega \in \Omega$, $F_\omega = \{y;\ (\omega, y) \in F\}$ is closed. Then, if (R^n) converges to R in $\mathcal{R}(\mathsf{P})$,

$$\limsup_n \mathsf{R}^n(F) \leq \mathsf{R}(F). \tag{3.3.3}$$

In particular, if the rules R^n have supports in the same space F ($\mathsf{R}^n(F) = 1$, $\forall n \geq 0$) verifying the previous condition, then the same is true for the limit rule ($\mathsf{R}(F) = 1$).

3.3.2 Stable convergence of random variables

When the rules R^n are associated with random variables (Y_n) defined on the given probability space, then $\mathsf{R}^n(H) = \mathsf{R}^{Y_n}(H) = \mathsf{E}[H(.,Y_n)]$. In this case, R^n converges stably to R, iff for any bounded $\mathcal{G}$-random variable K and continuous f on $\mathcal{X}$, $\mathsf{E}[K\, f(Y_n)]$ converges to $\mathsf{E}[K\, \Gamma(f)]$. In particular, (Y_n) converges weakly to the marginal distribution $\mathsf{R}_{\mathcal{X}}$ of R on $\mathcal{X}$.
The advantage of this representation is that the stable limit of (Y_n) can be *realized* on the extended space $(\bar{\Omega}, \bar{\mathcal{G}}, \mathsf{R})$. The random variables Y_n can be naturally extended on $(\bar{\Omega}, \bar{\mathcal{G}}, \mathsf{R})$ by setting $Y_n(\omega, x) = Y_n(\omega)$, and the stable convergence of (R^n) to R can be interpreted as the stable convergence of (Y_n) to the canonical variable Υ on the extended space $(\bar{\Omega}, \bar{\mathcal{G}}, \mathsf{R})$.

Some properties of stable convergence of variables In particular, for every $\mathcal{G}$-measurable r.v U, $((Y_n, U))$ converges in distribution to (Υ, U) (take $K(\omega) = h(U(\omega))$). Conversely, if $((Y_n, U))$ converges in distribution for every $\mathcal{G}$-measurable random variable U, then (Y_n) converges stably to a rule R. Proving the existence of a well-defined limit rule is actually not trivial and derive from results on bi-measures (see e.g. Jacod and Mémin (1981)).
When the limit random kernel Γ is a Dirac mass, $\delta_{Y(\omega)}(\mathrm{d}\chi)$ (i.e. $\mathsf{R} = \mathsf{R}^Y$), the stable convergence of (Y_n) to R means that (Y_n) converges to Y *in probability* on the given space. Thus, as stated in the title of Jacod and Mémin (1981), the stable convergence is a mode of convergence stronger than convergence in distribution but weaker than convergence in probability.

Stable and weak-$\mathbb{L}^1$ convergence: It is interesting to compare the stable convergence with another mode of convergence which preserves the initial structure $(\Omega, \mathcal{G}, \mathsf{P})$, *the weak-$\mathbb{L}^1$ convergence*, defined as follows:
A sequence of random variables (ξ_n) converges to ξ *weakly* in $\mathbb{L}^1(\Omega, \mathcal{G}, \mathsf{P})$ iff for any bounded random variable $H \in b\mathcal{G}$, $\mathsf{E}[H\, \xi_n] \to \mathsf{E}[H\, \xi]$.
If (Y_n) converges stably to R, $(f(Y_n))$ converges weakly in $\mathbb{L}^1(\mathsf{P})$ to $\Gamma(f)$ for all $f \in \mathcal{C}_b(\mathcal{X})$. Thus , the stable convergence can be reinterpreted as a mode of convergence for the sequence random kernels $(\Gamma^{Y_n}) = (\delta_{Y_n})$:

$$\forall f \in C_b(\mathcal{X}), \quad \Gamma^{Y_n}(f) = f(Y_n) \text{ converges to } \Gamma(f) \text{ weakly in } \mathbb{L}^1(\Omega, \mathcal{G}, \mathsf{P}).$$

This does not mean that (Γ^{Y_n}) converges in distribution (as $\mathcal{M}^e(\mathcal{X})$-valued random variables) to Γ, since the weak-$\mathbb{L}^1$ convergence of $(\Gamma^{Y_n}(f))$ does not imply that $(\Gamma^{Y_n}(f))$ converges in distribution to $\Gamma(f)$. Thus, the interpretation of the stable convergence as a

convergence of random measures is quit different from the usual weak convergence of random measures, in the sense of Kallenberg (1975).

STABLE RELATIVE COMPACTNESS One of the most interesting property of the stable convergence is that relative compactness for the stable convergence is "free". This is due to a result of Jacod and Mémin (1981), stating that a sequence (R^n) of rules is relatively compact in $\mathcal{R}(\mathsf{P}, \mathcal{X})$ iff the sequence $(\mathsf{R}^n_{\mathcal{X}})$ of their marginals on $\mathcal{X}$ is tight. Here, we recall that a sequence (μ^n) of probability measures on $\mathcal{X}$ is tight iff:

$$\forall \eta > 0, \text{ there exists a compact } K \text{ of } \mathcal{X} \text{ such that } \mu^n(K^c) \leq \eta, \quad \forall n \geq 0.$$

When the (μ^n) are the probability measures associated with a sequence (Y_n) of random variables, we also say by abuse of language that (Y_n) is tight in $\mathcal{X}$.
In terms of random variables, the property means that from any sequence (Y_n) converging weakly, there exists a subsequence of (Y_n) which converges stably to a rule R, such that $R_{\mathcal{X}}$ is the limit distribution. Equivalently,

$$(\mathsf{R}^{Y_n}) \text{ is stably relatively compact in } \mathcal{R}(\mathsf{P}, \mathcal{X}) \text{ iff } (Y_n) \text{ is tight in } \mathcal{X}.$$

For simplicity, we also say "(Y_n) or (Γ^{Y_n}) is stably relatively compact."
Finally, the boundedness requirement of test functions H can by replaced by a uniform integrability condition. If Y_n converges stably and

$$\lim_{a \to \infty} \sup_n \mathsf{E}[|H(\omega, Y_n)| \mathbb{1}_{\{|H(\omega, Y_n)| > a\}}] = 0, \tag{3.3.4}$$

then the stable convergence can also be applied to H.

3.3.3 Application to the population process Z^ϵ

As stated at the beginning of this section, the family of population processes (Z^ϵ) is not tight in the space of $\mathbb{N}^p$-valued càdlàg processes, due to the explosion of swap events. However, by construction, the demographic counting systems $\mathbf{N}^{\mathrm{dem},\epsilon}$ are all strongly dominated by the multivariate counting process $\mathbf{G}^{\mathrm{dem}}$. As a consequence, the size of the population $(Z^{\epsilon,\natural}) = (N^{b,\natural,\epsilon} - N^{d,\natural,\epsilon})$ is controlled by $G^{b,\natural}$. Our aim is to define the right probability space on which this property can be used to obtain the stable relative compactness of the family of population processes.

THE NEW "GIVEN" PROBABILITY SPACE The idea is to see the population processes not as dynamic processes anymore, but rather as $\mathbb{N}^p$-valued random variables, defined on the product-space:

$$(\tilde{\Omega}, \tilde{\mathcal{G}}) = (\Omega \times \mathbb{R}^+, \mathcal{G} \otimes \mathcal{B}(\mathbb{R}^+)), \text{ by } Z^\epsilon(\tilde{\omega}) = Z^\epsilon(\omega, s) = Z^\epsilon_s(\omega).$$

The space $(\tilde{\Omega}, \tilde{\mathcal{G}})$ should a priori be equipped with the product measure $\mathsf{P} \otimes \text{Leb}$, which is only a σ-finite measure. To overcome this difficulty, we replace the Lebesgue measure by the exponential distribution $\lambda^e(ds) = e^{-s}\mathrm{d}s$ and use the measure $\tilde{\mathsf{P}} = \mathsf{P} \otimes \lambda^e$ as probability measure on $(\tilde{\Omega}, \tilde{\mathcal{G}})$. In this new space, for any $\tilde{\mathcal{G}}$-measurable bounded functional $H(\tilde{\omega}, z)$,

$$\tilde{\mathsf{E}}[H(\tilde{\omega}, Z^\epsilon(\tilde{\omega}))] = \mathsf{E}[\textstyle\int_0^\infty H(\omega, s, Z^\epsilon_s(\omega))\lambda^e(\mathrm{d}s)] = \mathsf{E}[\int_0^\infty e^{-s} H(\omega, s, Z^\epsilon_s(\omega))\mathrm{d}s],$$

and the $\tilde{\mathsf{P}}$-probability distribution of Z^ϵ is the probability measure on $\mathbb{N}^p$, defined by

$$\tilde{\mu}^\epsilon(f) = \mathsf{E}[\textstyle\int_0^\infty e^{-s} f(Z^\epsilon_s)\mathrm{d}s].$$

Tightness and stable relative compactness Thanks to the uniform domination by the random variable $G^{b,\natural}(\tilde{\omega})$ of the **size** $Z^{\epsilon,\natural}(\tilde{\omega})$ of the population $Z^\epsilon(\tilde{\omega})$, it directly follows that the family $(\tilde{\mu}^\epsilon)$ is tight in $\mathbb{N}^p$, by a simple application of the definition of tightness in $\mathbb{N}^p$. From Subsection 3.3.2, it follows that (Z^ϵ) is stably relatively compact in $\mathcal{R}(\tilde{\mathsf{P}}, \mathbb{N}^p)$. Up to a subsequence, the family $(Z^\epsilon(\tilde{\omega}))$ converges stably to a random kernel $\tilde{\Gamma}(\tilde{\omega}, \mathrm{d}z)$ from $\tilde{\Omega}$ to $\mathbb{N}^p$. This means that up to a subsequence, the following convergence holds for all $H(\omega, s, z) \in \mathcal{C}_{bmc}(\tilde{\Omega} \times \mathcal{X})$:

$$\begin{aligned} \tilde{\mathsf{E}}[H(\omega, s, Z^\epsilon)] &\underset{\epsilon \to 0}{\longrightarrow} \tilde{\mathsf{E}}[\textstyle\int_{\mathbb{N}^p} H(\omega, s, z)\tilde{\Gamma}(\omega, s, \mathrm{d}z)] \\ &= \mathsf{E}[\textstyle\int_0^\infty \Big(\int_{\mathbb{N}^p} H(\omega, s, z)\tilde{\Gamma}(\omega, s, \mathrm{d}z)\Big)\lambda^e(\mathrm{d}s)] = \mathsf{E}[\int_0^\infty \tilde{\Gamma}(\omega, s, H)\lambda^e(\mathrm{d}s)] \end{aligned} \tag{3.3.5}$$

Finally, the kernel $\tilde{\Gamma}(\omega, t, dz)$ considered as a dynamic random measure $(\tilde{\Gamma}(\omega, t, dz))$ inherits the $(\mathcal{G}_t)$-*adaptation* of the population processes (Z^ϵ_t). Indeed, let us consider a test function $H(\omega, s, z) = M_s(\omega)\, f(z)$, where M_s is a bounded $(\mathsf{P}, \mathcal{G}_t)$-martingale on $[0, T]$, and f a bounded continuous function.

By the chain rule, $\mathsf{E}[\textstyle\int_0^T M_s f(Z^\epsilon_s)\mathrm{d}s] = \mathsf{E}[\int_0^T M_T\, f(Z^\epsilon_s)\mathrm{d}s]$.

The same properties hold at the limit by definition of the stable convergence, so that

$$\mathsf{E}[\textstyle\int_0^T M_s\tilde{\Gamma}(., s, f)\mathrm{d}s] = \mathsf{E}[\int_0^T M_T\tilde{\Gamma}(., s, f)\mathrm{d}s].$$

In the first term, the process $s \mapsto \tilde{\Gamma}(., s, f)$ can be replaced by its $\mathcal{G}-$optional projection $s \mapsto \tilde{\Gamma}(., s, f)^o$, and the two variables $\int_0^T \tilde{\Gamma}(., s, f)\mathrm{d}s$ and $\int_0^T \tilde{\Gamma}(., s, f)^o\mathrm{d}s$ are equal $\mathsf{P}.a.s$ as defining the same measure on $\mathcal{G}_T$. Since this property is true for any T, the processes $(\tilde{\Gamma}(., t, f))$ can be chosen $\mathcal{G}$-optional, and since $\mathcal{X}$ is separable, the property is also true for the kernel $(\tilde{\Gamma}(., t, \mathrm{d}z))$.

Application to the demographic intensity The example of the application of the stable convergence to the demographic intensity functional is a good illustration of all these

properties. On $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{\mathsf{P}})$, the demographic intensity processes $(\boldsymbol{\mu}^{\text{dem}}(\omega, t, Z_t^\epsilon))$ can be considered as a functional of the r.v $Z^\epsilon(\tilde{\omega})$. $\boldsymbol{\mu}^{\text{dem}}(\tilde{\omega}, \cdot)$ is not bounded, but $(\boldsymbol{\mu}^{\text{dem}}(\tilde{\omega}, Z^\epsilon))$ verifies the uniform integrability condition (3.3.4) (by strong domination), and the stable convergence can be still be applied. Then, by application of the convergence 3.3.5, for all $\mathcal{G}$-stopping time τ and $B \in \mathcal{G}$, up to a subsequence,

$$\begin{aligned}\mathsf{E}\left[\mathbb{1}_B \int_0^{t\wedge\tau} \boldsymbol{\mu}^{\text{dem}}(s, Z_s^\epsilon)\mathrm{d}s\right] &= \tilde{\mathsf{E}}[\mathbb{1}_B \mathbb{1}_{[0,t\wedge\tau]}(s) e^s \boldsymbol{\mu}^{\text{dem}}(s, \tilde{Z}^\epsilon)] \\ &\underset{\epsilon\to 0}{\longrightarrow} \mathsf{E}\left[\mathbb{1}_B \int_0^{t\wedge\tau} \tilde{\Gamma}(s, \boldsymbol{\mu}^{\text{dem}}(s, .))\mathrm{d}s\right] \end{aligned} \tag{3.3.6}$$

where $\tilde{\Gamma}$ is a $\mathcal{G}_t$-optional kernel. This is interesting since the multivariate process $\mathbf{A}_t^\epsilon = \int_0^t \boldsymbol{\mu}^{\text{dem}}(s, Z_s^\epsilon)\mathrm{d}s$ is the predictable compensator (or cumulative intensity process) of the demographic counting system $\mathbf{N}^{\text{dem},\epsilon}$. We summarize these results in a proposition:

Proposition 3.3.1. *Under the dominating assumption and up to a subsequence, the predictable compensator $\boldsymbol{A}_{t\wedge\tau}^\epsilon$ (or cumulative intensity process) of the demographic counting system $\boldsymbol{N}_{t\wedge\tau}^{\text{dem},\epsilon}$ converges weakly in $\mathbb{L}^1(\Omega, \mathcal{G}, \mathsf{P})$ to $\boldsymbol{A}_{t\wedge\tau}$ for all $t \geq 0$; where $\boldsymbol{A}$ is a continuous non decreasing process whose density is $\boldsymbol{\mu}^{\text{dem}}$ averaged again the optional kernel $(\tilde{\Gamma}(., t, \mathrm{d}z))$, that is:*

$$\boldsymbol{A}_{t\wedge\tau}^\epsilon = \int_0^t \boldsymbol{\mu}^{\text{dem}}(s, Z_s^\epsilon)\mathrm{d}s \; cv \; weakly - \mathbb{L}^1 \; to \quad \boldsymbol{A}_{t\wedge\tau} = \int_0^{t\wedge\tau} \tilde{\Gamma}(s, \boldsymbol{\mu}^{\text{dem}}(s, .))\mathrm{d}s \tag{3.3.7}$$

Remark 3.3.1. The advantage of the weak-$\mathbb{L}^1$ convergence is to preserve martingale properties. However, the stable convergence of Z^ϵ in $\mathcal{R}(\tilde{\mathsf{P}}, \mathbb{N}^p)$ does not imply the convergence in distribution of $\mathbf{A}_t^\epsilon$ to the the averaged process, which is a key point of the averaging result of Kurtz (1992). Indeed, the weak-$\mathbb{L}^1$ convergence does not imply the convergence in distribution, and the reverse is not true either. Moreover, if both modes of convergence are true, the weak-$\mathbb{L}^1$ limit does not necessarily have the same distribution than the limit distribution. In fact, this is true only if convergence in probability holds (Jajte and Paszkiewicz (1999)).

3.4 Convergence of the demographic counting systems

The previous property will be instrumental in proving the main result of this section, that is the characterization of the limit points of the demographic counting systems, $(\mathbf{N}^{\text{dem},\epsilon})$, which only depends on ϵ only through the population process Z^ϵ. The proof

relies on the identification of their intensity processes with the demographic intensity functional $\boldsymbol{\mu}^{\text{dem}}$, averaged against stable limits in $\mathcal{R}(\tilde{\mathsf{P}}, \mathbb{N}^p)$ of the population processes. Contrary to the above, we are also interested in realizing limit distributions of the demographic counting system as stable limits on the extended space, as defined in 3.3.2.

3.4.1 Tightness of demographic counting systems

Let us first study the tightness of the family of demographic counting systems $(\mathbf{N}^{\text{dem},\epsilon})$ as dynamic processes. The demographic counting systems are considered as random variables on $(\Omega, \mathcal{G}, \mathsf{P})$, taking values in a subspace $\mathcal{A}^{2p}$ of the Skorohod space $\mathrm{D}(\mathbb{N}^{2p})$. As stated before, the key to the tightness of the demographic counting systems is the strong domination by $\mathbf{G}^{\text{dem}}$. The second ingredient is to prove the tightness of the family in the space of *point measures* rather than in $D(\mathbb{N}^{2p})$, which simplifies the problem. But let us first define properly the topology of the state space and recall some properties of the convergence of counting processes.

Convergence of multivariate counting processes A multivariate counting process ξ with values in $\mathbb{N}^k$ can either be seen as a dynamic process or as random measure:
As a stochastic process, ξ is considered as an $\mathcal{A}^k$-valued random variable, where $\mathcal{A}^k$ is the subspace of the Skorohod space $D(\mathbb{N}^k)$, composed of $\mathbb{N}^k$-valued functions whose components only have unit jump and no common jump (see Jacod and Shiryaev (1987) for more details on the space $\mathcal{A}^k$). In general, it is not so easy to give tightness conditions in the Skorohod space.
This problem is simplified by considering ξ as a random measure of the space $\mathcal{N}_v^k = \mathcal{N}_v(\mathbb{R}^+ \times \mathcal{I}_k)$ of σ-finite *simple* point measures on $\mathbb{R}^+ \times \mathcal{I}_k$, endowed with the *vague topology*. In $\mathcal{N}_v^k$, a sequence of random measures (ξ^n) converges *vaguely* to ξ iff for all $(t_1, .., t_l)$ and $(k_1, .., k_l)$ such that $\xi(\{t_i\} \times \{k_i\}) = 0$ a.s, the random vector $(\xi^n([0, t_1] \times \{k_1\}), .., \xi^n([0, t_l] \times \{k_l\})) = (\xi^n_{k_1}(t_1), .., \xi^n_{k_l}(t_l))$ converges in distribution to $(\xi_{k_1}(t_1), .., \xi_{k_l}(t_l))$ (see Kallenberg (2017) for more details on the vague convergence of random measures).
Obviously, convergence for the Skorohod topology implies convergence in the vague topology, but the reverse is usually not true. However, in the case of multivariate counting processes, the convergence in distribution for the vague and Skorohod topology are equivalent (Theorem 4.20, Kallenberg (2017)). From this equivalence, tightness conditions in $\mathcal{A}^k$ can be derived from simpler conditions obtained from the random measure point of view. In particular, the following tightness criterion in $\mathcal{N}_v^k$ (Lemma 14.15, Kallenberg (2017)) can be easily verified:

A family (ξ^n) is tight iff for all $i \in \mathcal{I}_k$ and $t \geq 0$ the family $(\xi^n_i(t))$ is tight in $\mathbb{R}^+$.

TIGHTNESS OF DEMOGRAPHIC COUNTING SYSTEMS For each $\epsilon > 0$ the demographic counting system $\mathbf{N}^{\mathrm{dem},\epsilon}$ is an $\mathcal{A}^{2p}$ (or $\mathcal{N}_v^{2p}$) random variable. Furthermore, for all $t \geq 0$, $(\mathbf{N}_t^{\mathrm{dem},\epsilon})$ is tight in $(\mathbb{R}^+)^{2p}$ since $\mathbf{N}_t^{\mathrm{dem},\epsilon} \leq \mathbf{G}_t^{\mathrm{dem}}$, $\forall \epsilon > 0$. Hence, $(\mathbf{N}^{\mathrm{dem},\epsilon})$ is tight in $\mathcal{N}_v^{2p}$ by the previous tightness condition. By equivalence between the the convergence in distribution for the vague topology and the Skorohod topology in $\mathcal{A}^{2p}$, we obtain that:

Proposition 3.4.1 (Tightness of the demographic counting systems)**.**
The family of demographic counting systems $(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ *is tight in* $\mathcal{A}^{2p}$ *(endowed with the induced Skorohod topology). Equivalently,* $(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ *is also stably relatively compact in* $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$.

Remark 3.4.1. Note that we could also have proven Proposition 3.4.1 without using the equivalence with the random measure point of view, by applying more general results on the tightness of increasing processes (see e.g. Jacod and Shiryaev (1987)).

3.4.2 Realization of stable limits of the demographic counting systems

SET OF RULES FOR COUNTING PROCESSES As stated before, the advantage of the realizing stable limits on the extended space is to preserve the given structure of the problem. Let us detail in this subsection how stable limits of the the demographic counting systems are realized on the extended space, as well as properties verified by these limits.
Using notations of 3.3.1, the given space is $(\Omega, \mathcal{G}, \mathsf{P})$, and $\mathcal{X}$ is the space of processes $\mathcal{X} = \mathcal{A}^{2p}$, equipped with the Skorohod topology. The canonical filtration of $\mathcal{A}^{2p}$ is denoted by $(\mathcal{F}_t^{\mathcal{A}}) = (\sigma(\alpha(\cdot \wedge t)\ ;\ \alpha \in \mathcal{A}^{2p}))$. The canonical variable on $\mathcal{A}^{2p}$ is denoted by $\bar{\mathbf{N}}^{\mathrm{dem}}(\omega, \alpha) = \alpha$.
On the extended filtered space $(\bar{\Omega}, (\bar{\mathcal{G}}_t)) = (\Omega \times \mathcal{A}^{2p}, (\mathcal{G}_t \otimes \mathcal{F}_t^{\mathcal{A}}))$, the set of admissible probability measures are the rules R in $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$. The space $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$ is closed for the stable convergence, whose tests functions are the product functions $\mathbb{1}_B\, h(\bar{\mathbf{N}}^{\mathrm{dem}})$ with $B \in \mathcal{G}$ and h continuous and bounded on $\mathcal{A}^{2p}$.
By the support property deduced from the "portemanteau" inequality (3.3.3), the subset $\mathcal{R}^G(\mathsf{P}, \mathcal{A}^{2p})$ of $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$ of probability measures with support included in the domain $\bar{F} = \{(\omega, \alpha) \in \bar{\Omega};\ \alpha \prec \mathbf{G}^{\mathrm{dem}}(\omega)\}$ (whose sections in ω are closed in of $\mathcal{A}^{2p}$) is closed for the stable convergence, and for all rule $\mathsf{R} \in \mathcal{R}^G(\mathsf{P}, \mathcal{A}^{2p})$, the canonical variable $\bar{\mathbf{N}}^{\mathrm{dem}}$ is R a.s. strongly dominated by $\mathbf{G}^{\mathrm{dem}}$.

STABLE LIMITS OF $\mathbf{N}^{\mathrm{dem},\epsilon}$ We are now interested in *realizing* stable limits of the demographic counting systems $(\mathbf{N}^{\mathrm{dem},\epsilon})$, strongly dominated by $\mathbf{G}^{\mathrm{dem}}$, on the extended filtered space $(\bar{\Omega}, (\bar{\mathcal{G}}_t))$ associated with limit rules in $\mathcal{R}^G(\mathsf{P}, \mathcal{A}^{2p})$.
By equivalence between tightness and stable relative compactness, there exists for each sequence $(\mathbf{N}^{\mathrm{dem},\epsilon_k})$ converging in distribution a subsequence (still denoted with the same notation) converging stably to a rule $\mathsf{R} \in \mathcal{R}^G(\mathsf{P}, \mathcal{A}^{2p})$, whose marginal on $\mathcal{A}^{2p}$ is the limit distribution of $(\mathbf{N}^{\mathrm{dem},\epsilon_k})$ and for which the canonical variable $\bar{\mathbf{N}}^{\mathrm{dem}}$ is R-a.s. strongly dominated by $\mathbf{G}^{\mathrm{dem}}$, by the extension of the porte-manteau inequality.
This representation of limit distributions is particularly interesting since it allows the structure of the primary problem to be maintained. In particular, jump times of $\bar{\mathbf{N}}^{\mathrm{dem}}$ are also jump times of $\mathbf{G}^{\mathrm{dem}}, \mathsf{R}, a.s.$.

3.4.3 Characterization of limit points

Let us now state the main result of this section, which is the characterization of the limit points of the demographic counting systems. In this result, the intensity of the stable limits are identified with the optional projection of the demographic intensity functional $\boldsymbol{\mu}^{\mathrm{dem}}$ averaged against the conditional kernel of stable limits in $\mathcal{R}(\tilde{\mathsf{P}}, \mathbb{N}^p)$ of the population processes. To that matter, we shall need to consider the joint stable convergence of $(\mathbf{N}^{\mathrm{dem},\epsilon})$ and (Z^ϵ).

EXTENSION OF $(\mathbf{N}^{\mathrm{dem},\epsilon})$ TO $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{\mathsf{P}})$ The first step is to define $(\mathbf{N}^{\mathrm{dem},\epsilon})$ and (Z^ϵ) on the same probability space. In Subsection 3.3.3, the population processes Z^ϵ were considered as $\mathbb{N}^p$-valued random variables on the product space $\tilde{\Omega} = \Omega \times \mathbb{R}^+$ equipped with the product probability $\tilde{\mathsf{P}} = \mathsf{P} \otimes \lambda^e$.
In Subsection 3.4.1, the demographic counting systems $\mathbf{N}^{\mathrm{dem},\epsilon}$ have been seen as $\mathcal{A}^{2p}$-valued random variables on Ω. However, as all random variables defined on Ω, these processes $\mathbf{N}^{\mathrm{dem},\epsilon}$ can also be considered as random variables depending only on the first coordinate of the product space $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{\mathsf{P}})$.
It is direct to show that if $(\mathbf{N}^{\mathrm{dem},\epsilon})$ converges stably to $\mathsf{R} \in \mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$, then $(\mathbf{N}^{\mathrm{dem},\epsilon})$ converges stably on the new "reference space" $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{\mathsf{P}})$ to $\mathsf{R} \otimes \lambda^e \in \mathcal{R}(\tilde{\mathsf{P}}, \mathcal{A}^{2p})$. In particular, the new conditional kernel is not changed and given by the initial P-kernel, Γ^{dem}.
Conversely, if $(\mathbf{N}^{\mathrm{dem},\epsilon})$ converges stably to the rule $\tilde{\mathsf{R}}(\mathrm{d}\tilde{\omega}, \mathrm{d}\alpha) = \tilde{\Gamma}(\tilde{\omega}, \mathrm{d}\alpha)\tilde{\mathsf{P}}(\mathrm{d}\tilde{\omega}) \in \mathcal{R}(\tilde{\mathsf{P}}, \mathcal{A}^{2p})$, then $\tilde{\mathsf{E}}[\mathbb{1}_{[0,t]}(.)\tilde{\Gamma}(., h)] = \lambda^e([0,t])\tilde{\mathsf{E}}[\tilde{\Gamma}(., h)]$, for all $t \geq 0$ and $h \in C_{bc}(\mathcal{A}^{2p})$, since this property holds for the processes $\mathbf{N}^{\mathrm{dem},\epsilon}$. Hence, $\tilde{\Gamma}$ does not depend on s, i.e $\tilde{\Gamma}(\omega, s, \mathrm{d}\alpha) := \Gamma^{\mathrm{dem}}(\omega, \mathrm{d}\alpha)$ $\tilde{\mathsf{R}}$-a.s.

JOINT STABLE CONVERGENCE OF $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$ By tightness of $(\mathbf{N}^{\mathrm{dem},\epsilon})$ and (Z^ϵ), the family of random vectors $((\mathbf{N}^{\mathrm{dem},\epsilon}, Z^\epsilon))$ is thus stably relatively compact in $\mathcal{R}(\tilde{\mathsf{P}}, \mathcal{A}^{2p} \times \mathbb{N}^p)$. Our goal is to define stable joint limits in such way that the point of view of 3.3.3 and 3.4.2 are preserved.

For any limit rule $\mathsf{R}^*(\mathrm{d}\tilde{\omega}, \mathrm{d}(\alpha, z))$, there exists a kernel $\tilde{\Gamma}^*(\tilde{\omega}, \mathrm{d}(\alpha, z))$ from $\tilde{\Omega}$ to $\mathcal{A}^{2p} \times \mathbb{N}^p$ such that $\tilde{\mathsf{R}}^*(\mathrm{d}\tilde{\omega}, \mathrm{d}(\alpha, z)) = \tilde{\Gamma}^*(\tilde{\omega}, \mathrm{d}(\alpha, z))\tilde{\mathsf{P}}(d\tilde{\omega})$. By Theorem 1.25 in Kallenberg (2017), the kernel $\tilde{\Gamma}^*$ can be disintegrated into: $\tilde{\Gamma}^*(\tilde{\omega}, \mathrm{d}(\alpha, z)) = \tilde{\Gamma}^{\mathrm{dem}}(\tilde{\omega}, \mathrm{d}\alpha)\Gamma(\tilde{\omega}, \alpha, \mathrm{d}z)$.

In this decomposition, $\tilde{\Gamma}^{\mathrm{dem}}$ is a kernel from $\tilde{\Omega}$ to $\mathcal{A}^{2p}$ and Γ is a kernel from $\tilde{\Omega} \times \mathcal{A}^{2p}$ to $\mathbb{N}^p$. In particular, the stable limit of the sequence $(\mathbf{N}^{\mathrm{dem},\epsilon})$ (defined on Ω) in $\mathcal{R}(\tilde{\mathsf{P}}, \mathcal{A}^{2p})$ is $\tilde{\mathsf{R}}(\mathrm{d}\tilde{\omega}, \mathrm{d}\alpha) = \tilde{\Gamma}^{\mathrm{dem}}(\tilde{\omega}, \mathrm{d}\alpha)\tilde{\mathsf{P}}(\mathrm{d}\tilde{\omega})$, which means from the paragraph above that the kernel $\tilde{\Gamma}^{\mathrm{dem}}$ does not depend on s, $\tilde{\mathsf{R}}$ $.a.s.$ and can be denoted Γ^{dem}. This kernel is associated with the rule in $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$ defined by $\mathsf{R}(\mathrm{d}\omega, \mathrm{d}\alpha) = \Gamma^{\mathrm{dem}}(\omega, \mathrm{d}\alpha)\mathsf{P}(\mathrm{d}\omega)$.

The limit rule $\tilde{\mathsf{R}}^*$ can be rewritten as $\tilde{\mathsf{R}}^*(\mathrm{d}\tilde{\omega}, \mathrm{d}\alpha) = \mathsf{R}(\mathrm{d}\omega, \mathrm{d}\alpha)\lambda^e(ds)\Gamma(\omega, s, \alpha, \mathrm{d}z)$.

The stable convergence (up to a subsequence) of $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$ to $\tilde{\mathsf{R}}^*$ can be thus reinterpreted as follows: for all $B \in \mathcal{G}$, $t \geq 0$, $h \in C_{cb}(\mathcal{A}^{2p})$ and $f \in \mathcal{C}_b(\mathbb{N}^p)$,

$$\begin{aligned}\tilde{\mathsf{E}}\Big[\mathbb{1}_{B\times[0,t]}h(\mathbf{N}^{\mathrm{dem},\epsilon})f(Z^\epsilon)\Big] &= \mathsf{E}[\mathbb{1}_B h(\mathbf{N}^{\mathrm{dem},\epsilon})\int_0^t f(Z^\epsilon_s)\lambda^e(\mathrm{d}s)\Big] \\ &\underset{\epsilon\to 0}{\longrightarrow} \mathsf{R}\Big[\mathbb{1}_B h(\bar{\mathbf{N}}^{\mathrm{dem}})\int_0^t \Gamma(.,s,\bar{\mathbf{N}}^{\mathrm{dem}}, f)\lambda^e(\mathrm{d}s)\Big], \qquad (3.4.1)\end{aligned}$$

In particular, $(\mathbf{N}^{\mathrm{dem},\epsilon})$ converges stably to R and $(Z^\epsilon(\omega, s))$ converges stably to the R-expectation of Γ, whose kernel is $\tilde{\Gamma}(\omega, t, \mathrm{d}z) = \int_{\mathcal{A}^{2p}} \Gamma^{\mathrm{dem}}(\omega, \mathrm{d}\alpha)\Gamma(\omega, t, \alpha, \mathrm{d}z)$, as studied in Proposition 3.3.1. Recall that this marginal kernel, when considered as measure-valued process, is $\mathcal{G}$-optional.

Based on this representation, limit points for the joint distribution are now denoted by (R, Γ).

Theorem 3.4.1 (Identification result)**.** *Let (R, Γ) be a limit point of the stably relatively compact family $(\boldsymbol{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$, where R is a rule in $\mathcal{R}(\mathsf{P}, \mathcal{A}^{2p})$.*

Let us consider the process $(\Gamma(., t, \bar{\boldsymbol{N}}^{\mathrm{dem}}, \boldsymbol{\mu}^{\mathrm{dem}}))^o$ which is the $(\mathsf{R}, \bar{\mathcal{G}})$-optional projection of process $t \mapsto \Gamma(., t, \bar{\boldsymbol{N}}^{\mathrm{dem}}, \boldsymbol{\mu}^{\mathrm{dem}})$.

Then, the canonical demographic process $\bar{\boldsymbol{N}}^{\mathrm{dem}}(\omega, \alpha) = \alpha$ is characterized by the following property:

$$\bar{\boldsymbol{N}}^{\mathrm{dem}} \text{ has the } (\mathsf{R}, \bar{\mathcal{G}})\text{-compensator } \int_0^t (\Gamma(s, \bar{\boldsymbol{N}}^{\mathrm{dem}}, \boldsymbol{\mu}^{\mathrm{dem}}))^o \mathrm{d}s. \qquad (3.4.2)$$

Proof. Let $\epsilon_k \to 0$ be a subsequence along which $\{(\mathbf{N}^{\mathrm{dem},\epsilon_k}, \Gamma^{\epsilon_k})\}$ converges stably to (R, Γ). Let also $0 \leq u \leq t$, $A \in \mathcal{G}_u$ and $h_u \in C_{bc}(\mathcal{A}^{2p})$, an $\mathcal{F}^{\mathcal{A}}_u$ measurable function. The

martingale property on $(\mathbf{N}^{\mathrm{dem},\epsilon_k} - \mathbf{A}^{\epsilon_k})$ gives

$$\mathsf{E}[\mathbb{1}_A h_u(\mathbf{N}^{\mathrm{dem},\epsilon_k})(\mathbf{N}_t^{\mathrm{dem},\epsilon_k} - \mathbf{N}_u^{\mathrm{dem},\epsilon_k})] = \mathsf{E}[\mathbb{1}_A h_u(\mathbf{N}^{\mathrm{dem},\epsilon_k}) \int_u^t \boldsymbol{\mu}^{\mathrm{dem}}(s, Z_s^{\epsilon_k})\mathrm{d}s]$$

By the stable convergence of $(\mathbf{N}^{\mathrm{dem},\epsilon_k})$ to R, the right hand side of the previous equation converges to $\mathsf{R}[\mathbb{1}_A h_u(\bar{\mathbf{N}}^{\mathrm{dem}})(\bar{\mathbf{N}}_t^{\mathrm{dem}} - \bar{\mathbf{N}}_u^{\mathrm{dem}})]$.
Furthermore, let $H(\omega, s, \alpha, z) = \mathbb{1}_{A\times[u,t]}(\omega, s)e^s h_u(\alpha)\boldsymbol{\mu}^{\mathrm{dem}}(\omega, s, z)$. The left hand side of the previous equation is equal to $\tilde{\mathsf{E}}[H(\omega, s, \mathbf{N}^{\mathrm{dem},\epsilon}, Z^{\epsilon})]$, and by applying the joint stable convergence 3.4.1 to H, we obtain that

$$\mathsf{R}[\mathbb{1}_A h_u(\bar{\mathbf{N}}^{\mathrm{dem}})(\bar{\mathbf{N}}_t^{\mathrm{dem}} - \bar{\mathbf{N}}_u^{\mathrm{dem}})] = \mathsf{R}[\mathbb{1}_A h_u(\bar{\mathbf{N}}^{\mathrm{dem}}) \int_u^t \int_{\mathbb{N}^p} \boldsymbol{\mu}^{\mathrm{dem}}(u, z)\Gamma(\omega, s, \bar{\mathbf{N}}^{\mathrm{dem}}, \mathrm{d}z)\mathrm{d}s],$$

which gives a non-adapted version of the theorem. To achieve the proof, we have just to introduce an optional version of the density kernel tested again $\boldsymbol{\mu}^{\mathrm{dem}}$. □

3.5 Application to case of Markov swaps

In the previous section, we characterized the limit points of the distribution of the demographic counting systems in the presence of a fast evolution of the composition of the population, due to *fast swaps*. Without more information on the model and especially on the BDS intensity functionals, we cannot hope to have a unique limit distribution.
Most of the papers concerned with similar questions of different times scales are set in a Markov framework, with often ergodic requirements (see e.g. Kurtz (1992); Yin and Zhang (2012)). As a consequence, the intensity functionals are deterministic functions of the state of the population. Under these assumptions, the BDS process belongs to the largely documented family of Continuous Time Markov Chain. Actually, in the case where all intensities are deterministic and time homogeneous, an averaging result for the demographic counting systems can be derived for Kurtz (1992) or Yin and Zhang (2012), under ergodic assumptions for the swap processes.
In our setting, we want to restrict only the behavior of the swap component of the process, by assuming that the swap intensity functionals are deterministic functions of the population, and time homogeneous, in the perspective of using ergodicity results to characterize the limit distribution. On the other hand, no assumptions will be made on the dependence of the demographic intensity functional $\boldsymbol{\mu}^{\mathrm{dem}}(\omega, t, z)$ in (ω, t).
In order to prove the result, we will rely on the decomposition algorithm of Section 2.4 in order to characterize the intensity of the limit distribution. Thus, we assume for the

remainder of this section that the swap intensity functional $\boldsymbol{\mu}^s$ is a deterministic and time homogeneous function, $\boldsymbol{\mu}^s : z \in \mathbb{N}^p \to \boldsymbol{\mu}^s(z) \in \mathbb{R}^{p(p-1)}$.
The two timescale BDS $(Z_0, \mathbf{N}^\epsilon, Z^\epsilon)$, solution of (3.2.1), is now solution of:

$$\begin{aligned} Z_t^\epsilon &= Z_0 + \phi^s \odot \mathbf{N}_t^{s,\epsilon} + \mathbf{N}_t^{b,\epsilon} - \mathbf{N}_t^{d,\epsilon}, \\ d\mathbf{N}_t^{s,\epsilon} &= \mathbf{Q}^s(dt,]0, \frac{1}{\epsilon}\boldsymbol{\mu}^s(Z_{t^-}^\epsilon)]), \quad d\mathbf{N}_t^{\mathrm{dem},\epsilon} = \mathbf{Q}^{\mathrm{dem}}(dt,]0, \boldsymbol{\mu}^{\mathrm{dem}}(\omega, t, Z_{t^-}^\epsilon)]). \end{aligned} \tag{3.5.1}$$

As for the BDS decomposition algorithm, Swap processes will play an central role in the study in the following, even if we are eventually interested in the demographic counting systems. Recall that the swap counting system $\mathbf{N}^{s,\epsilon}$ is *not* a jumps counting system associated with a pure Swap process, due to its entanglement with the demographic counting system $\mathbf{N}^{\mathrm{dem},\epsilon}$. However, the BDS decomposition algorithm allows the disentanglement of these two processes, by building the population as a continuous *pasting of pure Swap processes*, stopped at demographic events times and reborn on a new space in which an individual has been either added or retrieved to the population.
These Swap processes of intensity functional $\boldsymbol{\mu}^s$ are obtained by thinning of the $p(p-1)$ multivariate Poisson measure $\mathbf{Q}^s$.
In the next subsection, we thus begin by studying these Swap processes in the two timescale setting, under the assumption of deterministic swap intensity functional.

3.5.1 Fast Markov Swap processes

Markov Swap

Swap CTMC When there is are no differences in the events timescale, a realization of the pure (Markov) Swap of intensity $\boldsymbol{\mu}^s$ starting in state ξ_0 can be given by the solution of the following Swap multivariate SDE:

$$X_t = \xi_0 + \phi^s \odot \mathbf{N}_t^{\mathrm{sw}}, \quad \mathrm{d}\mathbf{N}_t^{\mathrm{sw}} = \mathbf{Q}^s(\mathrm{d}t,]0, \boldsymbol{\mu}^s(X_{t^-})]). \tag{3.5.2}$$

Recall that the size of the swap is constant equal to $\xi_0^\natural$ (swap events don't change the size of the population). Thus, conditionally to the initial population ξ_0, X has a finite state space and the existence and uniqueness of (3.5.2) is trivial. Since $\boldsymbol{\mu}^s$ is a deterministic function, the Swap X solution of (3.5.2) is a $\mathcal{G}_t$-*CTMC*, of intensity matrix L^{sw} defined by:

$$L^{\mathrm{sw}} f(z) = \sum_{(i,j)\in\mathcal{J}^s} \mu^{ij}(z)(f(z + \phi(i,j)) - f(z)), \quad \forall x \in \mathbb{N}^p.$$

In the following, Continuous Time Markov Chains of intensity matrix L^{sw} are called Swap CTMC

Swap CTMC stationary measures Swap CTMC, when viewed as Markov processes on the whole state space $\mathbb{N}^p$, cannot be ergodic processes since if the Swap starts in the space $\mathcal{U}_n$ of population of size n, it will stay in $\mathcal{U}_n$ and not visit the entire state space. Nevertheless, restrictions of Swap processes to the spaces $\mathcal{U}_n$ are from now on assumed to be ergodic:

Hypothesis 1 (Ergodicity of the Swap process on $\mathcal{U}_n$)**.** $\forall n \geq 0$, The Swap CTMC restricted to $\mathcal{U}_n$ is assumed to be irreducible. Since $\mathcal{U}_n$ is finite, this means that the Swap CTMC restricted to $\mathcal{U}_n$ admits a unique stationary measure denoted by $(\pi(n, \mathrm{d}z))_{z\in\mathcal{U}_n}$.

Thus, if $X(\xi_0^n)$ is a swap process starting with an initial population $\xi_0^n \in \mathcal{U}_n$, the swap has a deterministic constant size $\xi_0^{d,\natural} = n$ ($X \in \mathcal{U}_n$), and $\forall f : \mathcal{U}_n \to \mathbb{R}^+$,

$$\frac{1}{t}\int_0^t f(X_s(\xi_0^n))\mathrm{d}s \underset{t\to+\infty}{\longrightarrow} \pi(n, f) = \int_{\mathcal{U}_n} f(z)\pi(n, \mathrm{d}z) \quad \mathsf{P}\text{-a.s}, \tag{3.5.3}$$

$$\mathsf{E}[f(X_t(\xi_0^n))] \underset{t\to+\infty}{\longrightarrow} \pi(n, f). \tag{3.5.4}$$

The previous equation shows the importance of the initial size of the population in the asymptotics of the swap system. To emphasize this dependence, the swap process is sometimes denoted in what follows by $X_t(\xi_0)$, or $X_t(\tau, \xi_\tau)$ when the starting time is random. In order to simplify equations when the starting time is random, we also assume without loss of generality that $f(0) = 0$.

Accelerated and Fast Swap CTMC

In the two-time scale framework, we are now interested in the realization of Swap CTMC of large intensity $(\frac{1}{\epsilon}\boldsymbol{\mu}^s(z))$, or equivalently of intensity matrix $\frac{1}{\epsilon}L^{\mathrm{sw}}$. Let us present two realizations of such processes in order to make the link between swap stationary distributions $\pi(n; \cdot)$ and the two-timescale process.

Accelerated Markov Swap Usually, realizations of Swap CTMC $\hat{X}^\epsilon$ of intensity matrix $\frac{1}{\epsilon}L^{\mathrm{sw}}$, which we call *accelerated Markov Swap*, are obtained by change of time. For instance, if X is a Swap CTMC defined by the thinning equation (3.5.2), then the process $\hat{X}^\epsilon$ defined by change of time $t \mapsto \frac{t}{\epsilon}$,

$$\hat{X}_t^\epsilon = X_{\frac{t}{\epsilon}} = \xi_0 + \textstyle\int_0^{\frac{t}{\epsilon}} \phi^s \odot \mathbf{Q}^s(\mathrm{d}s,]0, \boldsymbol{\mu}^s(X_{s^-})])$$

is a $(\mathcal{G}_{\frac{t}{\epsilon}})$ CTMC of intensity matrix $\frac{1}{\epsilon}L^{\mathrm{sw}}$.

FAST MARKOV SWAP The second realization, called the *fast Markov Swap*, is based on the thinning representation, and is defined by:

$$X_t^\epsilon = \xi_0 + \phi^s \odot \mathbf{N}_t^{\mathrm{sw},\epsilon}, \qquad d\mathbf{N}_t^{\mathrm{sw},\epsilon} = \mathbf{Q}^s(\mathrm{d}t,]0, \frac{1}{\epsilon}\boldsymbol{\mu}^s(X_{t^-}^\epsilon)]). \tag{3.5.5}$$

X^ϵ and $\hat{X}^\epsilon$ have the same distribution, but contrary to $\hat{X}^\epsilon$, the fast Markov Swap X^ϵ is $(\mathcal{G}_t)$-adapted for all $\epsilon > 0$. These two processes illustrate two different constructions of the Swap CTMC of intensity matrix $\frac{1}{\epsilon}L^{\mathrm{sw}}$: the accelerated process rely on the time component of the Poisson measure $\mathbf{Q}^s$ in order to increase the intensity, while the fast process uses the space component θ of the Poisson measure. In the sequel, the second construction is used since it coincides with swap processes obtained from the BDS decomposition algorithm.

The trade-off of the second construction lies in the convergence properties of the fast Markov Swap. On the one hand, by the ergodic Assumption 3.5.3 the accelerated Markov Swap $\hat{X}^\epsilon$ verifies the following a.s convergence property:

$$\int_0^t f(\hat{X}_s^\epsilon)\mathrm{d}s = \epsilon \int_0^{\frac{t}{\epsilon}} f(X_u)\mathrm{d}u \underset{\epsilon\to 0}{\longrightarrow} t\,\pi(\xi_0^\natural, f), \quad \mathsf{P}\text{-a.s.} \quad \forall f \in C_b(\mathbb{N}^p), \tag{3.5.6}$$

On the other hand, the above convergence does not holds almost surely for the fast Markov Swap X^ϵ, but *a priori* only in distribution. However, since the limit distribution $t\,\pi(\xi_0^\natural, f)$ is only random through its dependence on the initial condition ξ_0, the convergence (3.5.6) actually holds in probability for the fast Markov Swap X^ϵ.

This result can easily be extended to fast Swap processes starting at a random time τ in state ξ_τ, solution of:

$$X_t^\epsilon(\tau, \xi_\tau) = \xi_\tau \mathbb{1}_{\{t\geq\tau\}} + \phi^s \odot \mathbf{N}_t^{\mathrm{sw},\epsilon}, \qquad d\mathbf{N}_t^{\mathrm{sw},\epsilon} = \mathbf{Q}^s(\mathrm{d}t,]0, \frac{1}{\epsilon}\boldsymbol{\mu}^s(X_{t^-}^\epsilon)]). \tag{3.5.7}$$

On $\{t \geq \tau\}$, $X_t^\epsilon(\tau, \xi_\tau)$ is a fast Markov swap, and by the foregoing:

$$\forall f \in C_b(\mathbb{N}^p) \text{ s.t. } f(0) = 0, \quad \int_0^t f(X_s^\epsilon(\tau, \xi^\tau))\mathrm{d}s \underset{\epsilon\to 0}{\longrightarrow} (t-\tau)^+\pi(\xi_\tau^\natural, f), \quad \text{in probability.} \tag{3.5.8}$$

Fast Swap with random start depending on ϵ

In the sequel, we wish to study fast Swap processes with random start $(\tau^\epsilon, \xi^\epsilon)$ depending on ϵ. The following lemma extends the convergence in probability of Equation (3.5.8).

Lemma 3.5.0.1. *Let $(\tau^\epsilon, \xi^\epsilon)_{\epsilon>0}$ be a family of random variables in $\mathbb{R}^+ \times \mathbb{N}^p$, such that $(\tau^\epsilon, \xi^{\natural,\epsilon})$ converges in distribution. For each $\epsilon \geq 0$, let $X^\epsilon(\tau^\epsilon, \xi^\epsilon)$ be the fast Swap process solution of the Swap multivariate SDE (3.5.7), starting at time τ^ϵ in state ξ^ϵ.*
Then, for all bounded function f on $\mathbb{N}^p$ and $t \geq 0$,

$$\int_0^t f(X_s^\epsilon(\tau^\epsilon, \xi^\epsilon))\mathrm{d}s - (t-\tau^\epsilon)^+\pi(\xi^{\epsilon,\natural}, f) \text{ converges in probability to } 0.$$

Proof. In all the proof, we denote by $X^\epsilon(u,z)$ the fast Swap process solution of the Swap multivariate SDE (3.5.7), starting in state (u,z). Let $M > 0$.
The first step of the proof is to reduce the problem to the study of a functional only depending on the initial state ξ^ϵ through it size $\xi^{\natural,\epsilon}$. Thus, let:

$$Y^\epsilon(\omega, t, u, n) = \max_{z\in\mathcal{U}_n} |\int_0^t f(X_s^\epsilon(u,z))\mathrm{d}s - (t-u)^+\pi(n,f)|.$$

Then, by simple domination,

$$\mathsf{P}(|\int_0^t f(X_s^\epsilon(\tau^\epsilon,\xi^\epsilon)\mathrm{d}s - (t-\tau^\epsilon)^+\pi(\xi^{\epsilon,\natural},f)| \geq \frac{1}{M}) \leq \mathsf{P}(Y^\epsilon(t,\tau^\epsilon,\xi^{\epsilon,\natural}) \geq \frac{1}{M})$$

By denoting $g^\epsilon(u,n) = \mathsf{P}(Y^\epsilon(t,u,n) \geq \frac{1}{M})$, we have:

$$\mathsf{P}(Y^\epsilon(t,\tau^\epsilon,\xi^{\epsilon,\natural}) \geq \frac{1}{M}) = \mathsf{E}[g^\epsilon(\tau^\epsilon,\xi^{\epsilon,\natural})].$$

By Theorem 3.27 in Kallenberg (2006), if $(\tau^\epsilon, \xi^{\epsilon,\natural})$ converges in distribution and $g^\epsilon(u_\epsilon, n) \to 0$ for all $u_\epsilon \underset{\epsilon\to 0}{\longrightarrow} u$, then $(g^\epsilon(\tau^\epsilon,\xi^{\epsilon,\natural}))$ converges in distribution to 0, and thus converges in expectation since $(g^\epsilon(\tau^\epsilon,\xi^{\epsilon,\natural}))$ is uniformly bounded by 1. By assumption, $(\tau^\epsilon,\xi^{\epsilon,\natural})$ converges in distribution, is thus remains to prove the second condition. Let $u_\epsilon \underset{\epsilon\to 0}{\longrightarrow} u$. Firstly, observe that

$$g^\epsilon(u_\epsilon, n) \leq \sum_{z\in\mathcal{U}_n} \mathsf{P}(|\int_0^t f(X_s^\epsilon(u_\epsilon,z))\mathrm{d}s - (t-u_\epsilon)^+\pi(n,f)| \geq \frac{1}{2M})$$

For all $z \in \mathcal{U}_n$, $(\int_0^t f(X_s^\epsilon(u,z))\mathrm{d}s)$ converges in probability to $(t-u)^+\pi(n,f)$ by (3.5.8), and $(t-u_\epsilon)^+\pi(n,f) \underset{\epsilon\to 0}{\longrightarrow} (t-u)^+\pi(n,f)$. Thus, $\int_0^t f(X_s^\epsilon(u_\epsilon,z))\mathrm{d}s - (t-u_\epsilon)^+\pi(n,f)$ converges to 0,for all $z \in \mathcal{U}_n$. Since $\mathcal{U}_n$ is finite, this achieves to prove the Lemma. □

3.5.2 Birth Death limit of demographic counting systems

We can now state the main result of this section. As stated many times since the beginning of Chapter 2, the demographic counting system $\mathbf{N}^{\mathrm{dem},\epsilon}$ is not the jumps counting system of a Birth Death system, due to the presence of Swap events. Furthermore, the *aggregated process* $Z^{\natural,\epsilon}$ is also not an "autonomous" one dimensional Birth Death process. This is

due to swap events, but also to the fact that the aggregated birth and death intensities functional $\mu^{b,\natural}$ and $\mu^{d,\natural}$ depend on the whole structure of the population and not just on $Z^{\natural,\epsilon}$.

In the setting of deterministic swap intensities, along with the ergodic Assumption 1, the following theorem shows that the family of demographic counting systems converges in distribution to the jumps counting system of a "true" (Non-Markov) p Birth-Death process. Furthermore, due to the averaging effect of the Swap CTMC defined between two successive demographic events, the aggregated birth and death intensity functionals only depend on the size of the population at the limit. Thus, the limit aggregated is a much simpler Birth-Death process.

Recall that the two timescale BDS system $(Z_0, \mathbf{N}^\epsilon, Z^\epsilon)$, with deterministic swap intensity functions, has been defined a the beginning of this section as the solution of the following BDS multivariated SDE:

$$
\begin{aligned}
&Z_t^\epsilon = Z_0 + \phi^s \odot \mathbf{N}_t^{s,\epsilon} + \mathbf{N}_t^{b,\epsilon} - \mathbf{N}_t^{d,\epsilon},\\
&d\mathbf{N}_t^{s,\epsilon} = \mathbf{Q}^s(dt, [0, \frac{1}{\epsilon}\boldsymbol{\mu}^s(Z_{t^-}^\epsilon)]), \quad d\mathbf{N}_t^{\mathrm{dem},\epsilon} = \mathbf{Q}^{\mathrm{dem}}(dt, [0, \boldsymbol{\mu}^{\mathrm{dem}}(\omega, t, Z_{t^-}^\epsilon)]).
\end{aligned}
$$

The convergence in distribution of the family of demographic counting systems is stated as follows.

Theorem 3.5.1 (Convergence of the demographic counting systems)**.** *Let $(Z_0, \boldsymbol{N}^\epsilon, Z^\epsilon)$ be the two timescale BDS system with deterministic swap intensity functions, defined above. Under the ergodic Assumption 1 for Swap CTMC, the family of demographic counting systems $(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ converges in distribution in $\mathcal{A}^{2p}$.*

Furthermore, there exists an extension $(\bar{\Omega}, (\bar{\mathcal{G}}_t), \mathsf{R})$ of the given space $(\Omega, (\mathcal{G}_t), \mathsf{P})$ and a multivariate counting process $\bar{\mathbf{N}}^{\mathrm{dem}} = (\bar{\mathbf{N}}^b, \bar{\mathbf{N}}^d) \in \mathcal{A}^{2p}$ defined on this space, such that:

(i) *$\bar{\mathbf{N}}^{\mathrm{dem}}$ has the $(\bar{\mathcal{G}}_t)$-compensator:*

$$\bar{\boldsymbol{A}}_t = \textstyle\int_0^t \pi(\bar{Z}_s^\natural, \boldsymbol{\mu}^{\mathrm{dem}}(s, \cdot))\mathrm{d}s, \quad \bar{Z}_s^\natural = \bar{\mathbf{N}}_s^{b,\natural} - \bar{\mathbf{N}}_s^{d,\natural}.$$

(ii) *$(\boldsymbol{N}^{\mathrm{dem},\epsilon})$ converges in distribution to $\bar{\mathbf{N}}^{\mathrm{dem}}$.*

Proof. By Theorem 3.4.1, $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$ is stably relatively compact in $\mathsf{R}(\tilde{\mathsf{P}}, \mathcal{A}^{2p} \times \mathbb{N}^p)$. Let (R, Γ) be a limit point as defined in 3.4.1, and $\tilde{\mathsf{R}} = \mathsf{R} \otimes \lambda^e$. In order to prove Theorem 3.5.1, we need to prove that:

$$\Gamma(\omega, s, \bar{\mathbf{N}}^{\mathrm{dem}}, \mathrm{d}z) = \pi(\bar{Z}_s^\natural, \mathrm{d}z) \ \tilde{\mathsf{R}}\text{-a.s.} \tag{3.5.9}$$

Since we only want to identify Γ, we may assume that $(\mathbf{N}^{\mathrm{dem},\epsilon}, Z^\epsilon)$ converge stably to (R, Γ). Let $A \in \mathcal{G}$, $h \in C_{bc}(\mathcal{A}^{2p})$ and $f \in C_b(\mathbb{N}^p)$. Then,

$$
\begin{aligned}
&|\mathsf{E}[\mathbb{1}_A h(\mathbf{N}^{\mathrm{dem},\epsilon}) \int_0^t f(Z_s^\epsilon)\mathrm{d}s] - \mathsf{R}[\mathbb{1}_A h(\bar{\mathbf{N}}^{\mathrm{dem}}) \int_0^t \pi(\bar{Z}_s^\natural, f)\mathrm{d}s]| \\
&\leq \underbrace{|\mathsf{E}[\mathbb{1}_A h(\mathbf{N}^{\mathrm{dem},\epsilon}) \int_0^t f(Z_s^\epsilon)\mathrm{d}s] - \mathsf{E}[\mathbb{1}_A h(\mathbf{N}^{\mathrm{dem},\epsilon}) \int_0^t \pi(Z_s^{\epsilon,\natural}, f)\mathrm{d}s]|}_{I} \\
&\qquad + \underbrace{|\mathsf{E}[\mathbb{1}_A h(\mathbf{N}^{\mathrm{dem},\epsilon}) \int_0^t \pi(Z_s^{\epsilon,\natural}, f)\mathrm{d}s] - \mathsf{R}[\mathbb{1}_A h(\bar{\mathbf{N}}^{\mathrm{dem}}) \int_0^t \pi(\bar{Z}_s^\natural, f)\mathrm{d}s]|}_{II}.
\end{aligned}
$$

II: The function $\alpha \to \int_0^t \pi(\alpha_s^{b,\natural} - \alpha_s^{d,\natural}, f)\mathrm{d}s$ is a continuous bounded function from $\mathcal{A}^{2p}$ to R, and thus the stable convergence of $(\mathbf{N}^{\mathrm{dem},\epsilon})$ to R imply that the second term II of the r.h.s of the previous equation converges to 0.

I: In order to prove that I converges to 0, we rely on the BDS decomposition algorithm. For each $\epsilon > 0$, let (T_k^ϵ) be the sequence of demographic event times (i.e the jump times of $(\mathbf{N}^{\mathrm{dem},\epsilon})$), and take $\eta > 0$.
By Lemma 3.2 of Jacod (1987), the tightness of $(\mathbf{N}^{\mathrm{dem},\epsilon})$ imply that there exists an integer K such that $\forall k > K$,

$$\limsup_\epsilon \mathsf{P}(T_k^\epsilon \leq t) \leq \frac{\eta}{2t\|f\|_\infty \|g\|_\infty}.$$

By decomposing I onto the Kth first demographic intervals, we obtain that:

$$
\begin{aligned}
I &\leq \|g\|_\infty \left(\sum_{k=0}^{K} \mathsf{E}[| \int_{T_k^\epsilon \wedge t}^{T_{k+1}^\epsilon \wedge t} (f(Z_s^\epsilon) - \pi(Z_s^\epsilon, f))\, \mathrm{d}s|] + \mathsf{E}[| \int_{T_{K+1}^\epsilon \wedge t}^{t} (f(Z_s^\epsilon) - \pi(Z_s^\epsilon, f))\mathrm{d}s|] \right) \\
&\leq \|g\|_\infty \sum_{k=0}^{K} \mathsf{E}[| \int_{T_k^\epsilon \wedge t}^{T_{k+1}^\epsilon \wedge t} (f(Z_s^\epsilon) - \pi(Z_s^\epsilon, f))\, \mathrm{d}s|] + \frac{\eta}{2}
\end{aligned}
$$

For each $0 \leq k \leq K$, the population process Z^ϵ coincides on the $k+1$th demographic interval $[T_k^\epsilon, T_{k+1}^\epsilon[$ with the kth fast Swap process $X^{k,\epsilon}$, starting at the kth demographic event time T_k^ϵ in state $Z_{T_k^\epsilon}^\epsilon$ and solution of:

$$X_t^{k,\epsilon} = X_t^\epsilon(Z_{T_k^\epsilon}^\epsilon, T_k^\epsilon) = Z_{T_k^\epsilon}^\epsilon \mathbb{1}_{\{t \geq T_k^\epsilon\}} + \phi^s \odot \mathbf{N}_t^{\mathrm{sw},k,\epsilon}, \quad d\mathbf{N}_t^{\mathrm{sw},k,\epsilon} = \mathbf{Q}^{\mathrm{sw}}(\mathrm{d}t,]0, \tfrac{1}{\epsilon}\boldsymbol{\mu}^s(X_{t^-}^{k,\epsilon})]).$$

By definition, the size of the population $Z^{\epsilon,\natural}$ is constant on demographic intervals, hence:

$$\int_{T_k^\epsilon \wedge t}^{T_{k+1}^\epsilon \wedge t} (f(Z_s^\epsilon) - \pi(Z_s^\epsilon, f))\, \mathrm{d}s = \int_0^{T_{k+1}^\epsilon \wedge t} f(X_s^{k,\epsilon})\mathrm{d}s - (T_{k+1}^\epsilon \wedge t - T_k^\epsilon)^+ \pi(Z_{T_k^\epsilon}^\epsilon, f), \ \mathsf{P} \text{ a.s.}$$

As in the proof of Lemma 3.5.0.1, the previous expression can be dominated using the random variable $Y^\epsilon(\omega, r, u, n) = \max_{z \in \mathcal{U}_n} |\int_0^r f(X_s^\epsilon(u, z))\mathrm{d}s - (r-u)^+ \pi(n, f)|$ which depends on the size of the Swap process rather than on its initial state:

$$|\int_0^{T^\epsilon_{k+1}\wedge t} f(X^{k,\epsilon}_s)\mathrm{d}s - (T^\epsilon_{k+1}\wedge t - T^\epsilon_k)^+\pi(Z^\epsilon_{T^\epsilon_k}, f)| \leq \sup_{r\leq t} Y^\epsilon(\omega, r, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k}).$$

By Lemma 3.5.0.1, $(Y^\epsilon(\omega, r, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k}))$ converges in probability to 0 for all $r \leq t$, and by using a diagonal argument, there exists a subsequence along which:

$$Y^\epsilon(\omega, q, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k}) \underset{\epsilon\to 0}{\longrightarrow} 0 \text{ for all rationals } q \leq t, \mathsf{P}\text{-a.s.}$$

The convergence can be extended by continuity to $[0, t]$. In addition, the family $(Y^\epsilon(\omega, r, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k}))$ is uniformly bounded by $2t\|f\|_\infty$ and for all $r, s \in [0, t]$,

$$|Y^\epsilon(\omega, s, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k}) - Y^\epsilon(\omega, r, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k})| \leq 2\|f\|_\infty|s - r|.$$

We can thus apply of Arzela-Ascoli theorem, and there exists a subsequence along which $(Y^\epsilon(\omega, r, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k}))$ converges uniformly on $[0, t]$ P-a.s, so that:

$$\sup_{r\leq t} Y^\epsilon(\omega, r, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k}) \underset{\epsilon\to 0}{\longrightarrow} 0 \text{ in probability.}$$

Since the family is also uniformly bounded,

$$\mathsf{E}[|\int_0^{T^\epsilon_{k+1}\wedge t} f(X^{k,\epsilon}_s)\mathrm{d}s - (T^\epsilon_{k+1}\wedge t - T^\epsilon_k)^+\pi(Z^\epsilon_{T^\epsilon_k}, f)|] \leq \mathsf{E}[\sup_{r\leq t} Y^\epsilon(\omega, r, T^\epsilon_k, Z^\epsilon_{T^\epsilon_k})] \underset{\epsilon\to 0}{\longrightarrow} 0.$$

This is true for all $k = 1..K$, and thus for ϵ small enough, $I \leq \eta$, i.e $I \to 0$.

This achieves to prove that $(\mathbf{N}^{\mathrm{dem},\epsilon}(\omega), Z^\epsilon(\omega, s))$ converges stably to $(\mathsf{R}, \pi(\bar{Z}^\natural_s, \mathrm{d}z))$. But $(\mathbf{N}^{\mathrm{dem},\epsilon}(\omega), Z^\epsilon(\omega, s))$ also converges to (R, Γ), hence

$$\tilde{\Gamma}(\omega, s, \bar{\mathbf{N}}^{\mathrm{dem}}, \mathrm{d}z) = \pi(\bar{Z}^\natural_s, \mathrm{d}z) \quad \tilde{\mathsf{R}}\text{-a.s.}$$

□

3.5.3 Application

In this last subsection, we apply Theorem 3.5.1 in the example of non-linear swap intensities presented in (2.2.12). Let us first recall the model.

A population composed of two subgroups is considered. The second subgroup (subgroup 2) has a lower death intensity, $\mu^{(d,1)}(t, z) \geq \mu^{(d,2)}(t, z)$, a.s. When the population is smaller than a given size M, individuals swap to the favorable subgroup 2, at rate $k_{12}(z^\natural)^\alpha$, $\alpha > 0$. When the population is larger than M access to the subgroup 2 is restricted and individuals swap from 1 to 2 at a lower constant rate k^M_{12}. Individuals swap from the favorable subgroup 2 to subgroup 1 at constant rate k_{21}. The swap intensity is thus defined by:

$$\mu^{(1,2)}(z) = k_{12}(z^\natural)^\alpha z^1 \mathbb{1}_{\{z^\natural \leq M\}} + k^M_{12} z^1 \mathbb{1}_{\{z^\natural > M\}}, \quad \mu^{(2,1)}(z) = k_{21} z^2.$$

SWAP CTMC STATIONARY MEASURE The intensity matrix of the Swap CTMC restricted to the space $\mathcal{U}_n$ (of population of size n) is defined for all $f : \mathcal{U}_n \to \mathbb{R}$ by,

$$L_n^{\mathrm{sw}} f(z) = \begin{cases} k_{12} n^\alpha z^1 (f(z + \mathbf{e}_2 - \mathbf{e}_1) - f(z)) + k_{21} z^2 (f(z + \mathbf{e}_1 - \mathbf{e}_2) - f(z)), & \text{if } n \leq M \\ k_{12}^M z^1 (f(z + \mathbf{e}_2 - \mathbf{e}_1) - f(z)) + k_{21} z^2 (f(z + \mathbf{e}_1 - \mathbf{e}_2) - f(z)), & \text{if } n > M \end{cases}$$

Actually, the Swap CTMC can be reinterpreted as follow: all individuals evolve as "independent CTMC" on the state space $\{1, 2\}$, with constant transition rates depending on the initial number of individuals. In particular, this means that the stationary measure of the Swap CTMC restricted to $\mathcal{U}_n$, $\pi(n, \cdot)$, is defined as the distribution of the sum of n i.i.d random variables of distribution ν defined by:

$$\nu_2(n) = \begin{cases} \dfrac{k_{12} n^\alpha}{k_{12} n^\alpha + k_{21}} & \text{if } n \leq M \\ \dfrac{k_{12}^M}{k_{12}^M + k_{21}} & \text{if } n > M \end{cases},$$

and with $\nu_1(n) = 1 - \nu_2(n)$. In particular, $\pi(n, z^2) = n\nu_2(n)$.
Thus, there can be two situations at the limit. If the size of the population is small enough individuals are able to move more easily to the favorable subgroup 2 which is more populated on average. On the other hand, when the population becomes crowded, the access to subgroup 2 is restricted, the proportion of individual in subgroup 2 becomes smaller.

LINEAR DEATH INTENSITIES Let us now assume that the death intensity functionals $\mu^{(d,i)}(\omega, t, z)$, $i = 1, 2$, are linear, equal to:

$$\mu^{(d,i)}(\omega, t, z) = d^i(\omega, t) z^i.$$

By Theorem 3.5.1, the aggregated death intensity of the limit process is:

$$\lambda^d(\omega, t, n) = \pi(n, \mu^{d,\natural}) \quad \text{, where } \mu^{d,\natural} = \mu^{(d,1)} + \mu^{(d,2)}.$$

In the case of linear death intensities, the previous equation can be rewritten as:

$$\lambda^d(\omega, t, n) = \sum_{i=1,2} d^i(\omega, t) \pi(n, z^i) = n(d^1(\omega, t)(1 - \nu_2(n)) + d^2(\omega, t)\nu_2(n)). \tag{3.5.10}$$

This shows how non-linearities in the death intensity can emerge in the limit aggregated population, even when death intensities in each subgroup are linear. This results from a non trivial aggregation of the subgroup specific death intensities, due to the swap events. In this particular example, the *death rate of individuals* in the limit aggregated

population - which approximates the behavior of the aggregated population when swaps occur on a faster timescale - *depends non trivially on the size of the population.* Due to the two regimes of swap events, the individual death rate in the limit aggregated population is lower when the population is small.
Thus, thanks to the approximation of the aggregated process by a simpler Birth Death process in the two timescale framework, we can better understand how swap events modify the behavior of the population ,by creating non-linearities at the aggregated level.

Chapter 4

How can a cause-of-death reduction be compensated in presence of heterogeneity? A population dynamics approach

4.1 Introduction

When studying a large population such as a national population, it is well known that the population usually shows some heterogeneity, in the sense that individuals with different characteristics (e.g. gender, social characteristics, neighborhood...) can exhibit different demographic behaviors (different mortality and fertility rates). Whenever possible, to take into account these individual characteristics for modeling mortality and population evolution gives additional information.

Research on the relationship between socioeconomic status and mortality can be traced back as far as the nineteenth century. Since then, an important body of work have investigated the links between socioeconomic status (SES) and health, and there is now a broad consensus on the strong correlation between SES and mortality. This correlation holds as well for causes of death. Indeed, many studies have shown that individuals with different socioeconomic status are affected differently by diseases. For example, diseases linked to smoking habits or obesity are more likely to impact individuals with lower socioeconomics status (see e.g. National Research Council and Committee on Population (2011)).

More recently, a growing number of studies, in different countries, have highlighted the

increasing gap between different socioeconomic subgroups (see e.g. the review of Elo (2009)). Those divergences generate a high heterogeneity in populations, and socioeconomic subgroups can experience mortality rates rather different from the mortality rates estimated from aggregated data such as national data.

The widening of these socioeconomic gaps has led a number of pension funds and insurance companies to rethink their models in order to tackle this heterogeneity issue and to understand the potential impact of socioeconomic inequalities. Indeed, the non-consideration of socioeconomic differences can have substantial impacts for insurance companies or governments, by leading for instance to errors in funding of annuity and pension obligation (see e.g. Meyricke and Sherris (2013); Villegas and Haberman (2014)) or to increase the basis risk (variation between sample and population mortality).

For instance, the Life & Longevity Markets Association (LLMA) and the Institute and Faculty of Actuaries (IFoA) in the UK have recently commissioned a report on longevity basis risk (Haberman et al. (2014)). By studying recommendations from the World Health Organization targeting specific causes of death, Alai et al. (2017) have shown that these recommendations could increase life expectancy gaps in England, despite an increase of the national life expectancy. Indeed, a public policy could as well increase inequalities when differences between SES are disregarded.

Consequently, a growing literature has recently taken an interest in the joint modeling and forecasting of the mortality of socioeconomic subgroups (see e.g Jarner and Kryger (2011); Villegas and Haberman (2014)). This new class of models is based on a relative approach. Individuals are grouped into so-called risk classes, assumed to have an acceptable level of homogeneity, and the mortality of risk classes is modeled with respect to the mortality of a large reference population, often the aggregated population. The impact of heterogeneity on cause-specific mortality has also been well-documented. However, there are still a lot of open questions on the consistency of sub-national and national estimations and forecasts. An approach to this issue has been proposed very recently by Shang and Haberman (2017); Shang and Hyndman (2017), based on recent developments in grouped functional time series methods.

However, the impact of composition changes on aggregated mortality indicators induced by the heterogeneous evolution of the population has been, to the best of our knowledge, less studied. Any changes in the composition of cohorts or age classes could lead to substantial changes in the age-specific mortality rates of the global population. In particular, understanding the combined effect of changes in the composition of the population and of cause-specific reductions in socioeconomic subgroups cannot be done by the sole analysis of mortality rates time series.

Important recent releases of data allow a more detailed analysis of these composition effects. We argue that there is a real need and possibility to study the whole population and not only mortality rates to better understand the effect of heterogeneity on aggregate indicators. The population dynamics point of view allows us to represent the data differently than what is usually done when the focus is on mortality only, and to consider issues beyond those that are typically addressed.

The aim of this chapter is to illustrate possible impacts of changes in population composition, and is organized as follows. In Section 4.2, we introduce the data used to carry out our study, which is a unique database obtained from the UK Office for National Statistics (ONS), containing information on cause-specific death rates by age, gender and deprivation level for the period 1981-2015. Particular emphasis is being placed on presenting the main features of the age structures of the subpopulations grouped by level deprivation, and their evolution over time. In Section 4.3, we present the population dynamics framework which is used in Section 4.4. One of the advantages of this framework is to allow the derivation of closed formula and asymptotic results for a number of indicators, in order to better understand the complex interactions between heterogeneity and mortality. Section 4.4 presents our numerical results. We first show how different socioeconomic composition of the age classes can impact the life expectancy and mortality improvement rates. In a second phase, we show how a cause of death reduction can be compensated by changes in the composition of the population induced by heterogeneity in fertility rates, and thus could be misinterpreted.

4.2 What can be learned from the data

More and more data are available from various statistical institutions, allowing us to take into account more complex information in population modeling, in order to better understand the past and future evolution. When assessing demographic changes, those new data bring to light complex interactions between longevity and the population dynamics, especially due to the social heterogeneity of the population. In this section, we present the two databases used in this article. Particular emphasis is made on the evolution of the age structure of the subpopulations grouped by level deprivation, which we consider to be an important contribution of the study.

4.2.1 Databases

The data we use provide mid-year population estimates in England by age class and socio-economic circumstances for the years 1981-2015[1], combined with the number of deaths by age, cause and socioeconomic circumstance. Our study is based two data sources:

(i) The first database was provided to us by the Department of Applied Health Research (DAHR) of the University College London in UK, and is based on the Index of Multiple Deprivation 2007 (IMD 2007) for the 1981-2006 period[2].

(ii) The second database was released in 2017 by the Office for National Statistics in UK (ONS)[3], and is based on the Index of Multiple Deprivation 2015 (IMD 2015) for the 2001-2015 period.

Deprivation Criterion In both databases, socio-economic circumstances are measured by the Index of Multiple Deprivation (IMD). The concept of deprivation has been defined by Townsend as the lack of *"types of diet, clothing, housing, household facilities and fuel and environmental, educational, working and social conditions, activities and facilities which are customary"* (Townsend (1979), cited in Noble et al. (2007)). The IMD is a geographically based index, created by the British Department for Communities and Local Government in order to provide an official measure of these multiple deprivation dimensions at the level of small areas called LSOAs[4] (see Noble et al. (2007) and

[1]Population estimations are based on decade census data, see Office for National Statistics (2012) for more details.

[2]The authors thank Madhavi Bejekal for her great help in obtaining the data and her explanations.

[3]Publicly available on the ONS website (www.ons.gov.uk) under the reference number 006925.

[4] In 2007, there were 32,482 Lower Layer Super Outupt areas (LSOAs) in England (34,753 in 2011), each composed of about 1500 individuals (Office for National Statistics (2012)).

Department for Communities and Local Government (2015) for more details). More precisely, the IMD is based on the measure of seven broad socioeconomic factors: income, employment, health, education, barriers to housing and services, living environment and crime; and it is computed for each small living area in England. The IMD score of a small area is used as SES proxy for individuals living in the LSOA. Furthemore, the index also includes information on the physical and social environment of individuals (by including for instance the road distance to a GP surgery and supermarket, or crime statistsics), which can have a significant influence on health outcomes (Diez Roux and Mair (2010); Nandi and Kawachi (2011)).

One limitation of using area based measurements is to apply the same deprivation to all individuals living in the same area. However, LSOAs are rather small areas[4] and geographical data are often more available on a large scale than multiple individual socioeconomic measurements. We also note that the IMD is computed at fixed dates, while being applied to a longer time period (e.g. IMD 2007 applied to the period 1981-2007). The implications of using a fixed IMD quintile allocation have been discussed comprehensively in Bajekal et al. (2013) and in Appendix D of Lu et al. (2014), based on the period 1981-2001. Our comparison of data computed with the IMD 2007 and IMD 2015 for the overlapping period 2001-2006 also gives very similar results. See Appendix A for a more detailed discussion on this issue.

Structure of data sources Our two data sources are based on a *relative* measurement of deprivation. Small living areas are ranked by their IMD scores and grouped into five deprivation quintiles numbered from 1 to 5: IMD quintile 1 for the least deprived quintile, to IMD quintile 5 for the most deprived quintile. It is worth noting that for each year, the five deprivation quintiles have approximately the same number of individuals, see Section 4.2.2.

Population data are structured by age-class and deprivation quintile, while deaths data are structured by age-class, deprivation quintile and cause of deaths. Specific features of each database are summarized in Table 4.1 and 4.2.

	Database 1	Database 2
Deprivation index	IMD 2007	IMD 2015
Time period	1981-2006	2001-2015
Ages	25-85+	0-90+
Populations age group	5 years	1 year
Deaths age group	5 years	1 year 5 years per cause

Table 4.1 Databases

Causes of death also differ in both databases. The following table summarizes the different causes of death available in both databases[5].

	Database 1	Database 2
Circulatory diseases	×	×
Neoplasms	×	×
Respiratory diseases	×	×
External causes	×	×
Diabetes		×
Digestives diseases	×	
Mental diseases	×	
Neonatal deaths		×

Table 4.2 Causes of death in each database

The compilation method of Database 1 is described in Lu et al. (2014). See also Labit Hardy (2016) for a detailed description of the dataset. Those data have also been used in papers focusing on the study of mortality improvements and healthy life expectancy by deprivation (e.g. Bajekal (2005), Lu et al. (2014)) or in mortality modeling (e.g. Villegas and Haberman (2014)).

Database 2 has been only very recently released by ONS. In addition to updating the data for years 2007-2015, this second database provides disaggregated data by single year of age, as well as data for young ages below age 25. This constitutes an important

[5]In Database 2, circulatory diseases are divided in ischemic heart diseases, strokes and other circulatory diseases; external causes in intentional and unintentional injuries.

contribution to our paper, by allowing a more precise analysis of the population by deprivation.

4.2.2 About the data

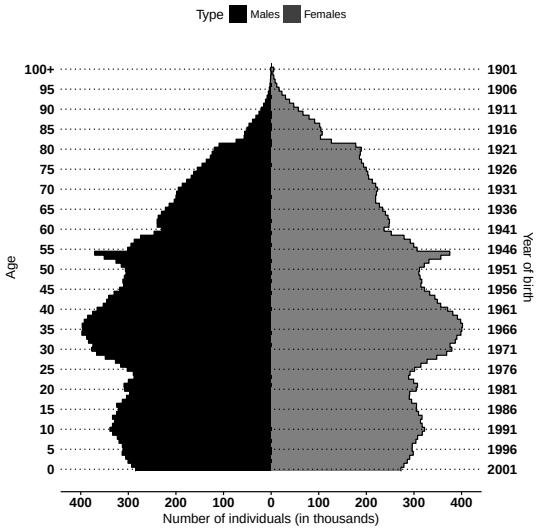
In the rest of this section, selected representations of the data are presented, with a particular focus on the population evolution. Our goal is to use the data to give the reader some insights on the complex interactions between the aggregated mortality and the evolution of the composition of the population.

Population composition

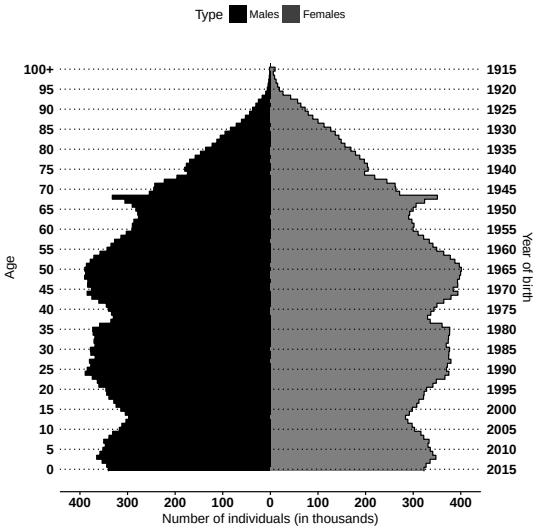
Age Pyramids In order to illustrate differences in the composition of populations per deprivation quintile, age pyramids of the least and most deprived quintiles are represented on Figure 4.1 for years 2001 and 2015, along with the age pyramid of England.
By reading Figure 4.1 vertically, we can see that for each year, the form of the age pyramids are very different between the different deprivation quintiles and England population. The most deprived population, IMD quintile 5, represented on Figure 4.1e and 4.1f, is much younger on average than the least deprived population, IMD quintile 1, which is represented on Figure 4.1c and 4.1d. For instance, the median age in 2015 was of 33 years (35.5 for the mean age) in the most deprived population, while the median in the least deprived population was of 44.2 (42.6) and of 39 years (39.7) in England. Some cohorts are also more represented among a particular subpopulation. For instance, the English baby-boom generation (born in the years after the World War II) and their children are more represented among the least deprived subpopulation.
The horizontal reading of Figure 4.1 shows that in addition to this heterogeneity in age, significant temporal changes in the age pyramids occurred from 2001 to 2015, due to the population ageing but also to internal and external migrations[6]. During this period, the largest flows of populations can be observed around ages 20-35. They are characterized by an inflow of individuals in the most deprived quintiles and in the general population, and an outflow of individuals around age 20 in the least deprived quintiles, follow by an inflow of individuals at ages 25-35. Furthermore, these changes in the age pyramids are quite different according to the level of deprivation. For instance, the median age in the most deprived quintile has dropped of over 1%, from 33.4 to 33 years, while it has increased of more than 9% in the least deprived subpopulation, from 40 to 44, and of

[6]Here, internal migrations correspond to the migration of individuals in between IMD quintiles, whereas external migrations correspond to the migration of individuals from/to places outside of England.

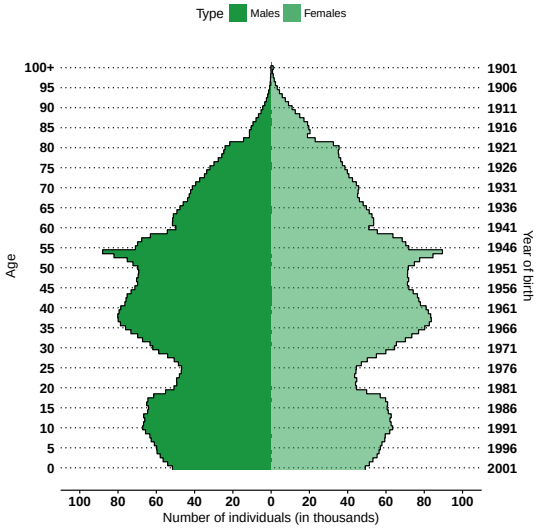
about 5% in the general population, from 37.1 to 39. Thus, the most deprived population has become more youthful from 2001 the 2015, despite the general tendency of the ageing of the population.



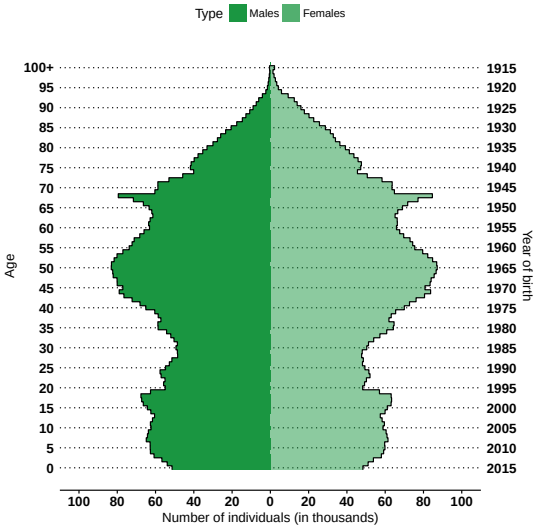
(a) England population, 2001



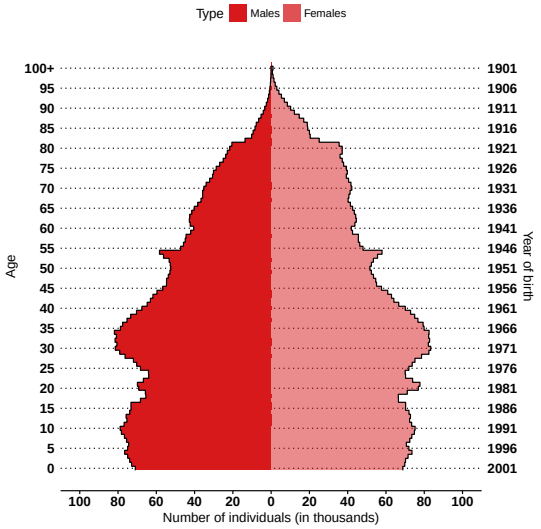
(b) England population, 2015



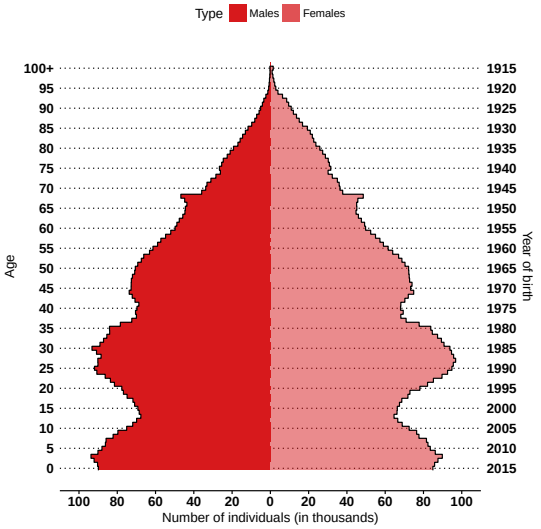
(c) Least deprived quintile ($$$$$), 2001



(d) Least deprived quintile ($$$$$), 2015



(e) Most deprived quintile ($), 2001



(f) Most deprived quintile ($), 2015

Fig. 4.1 Age pyramids in 2001 and 2015

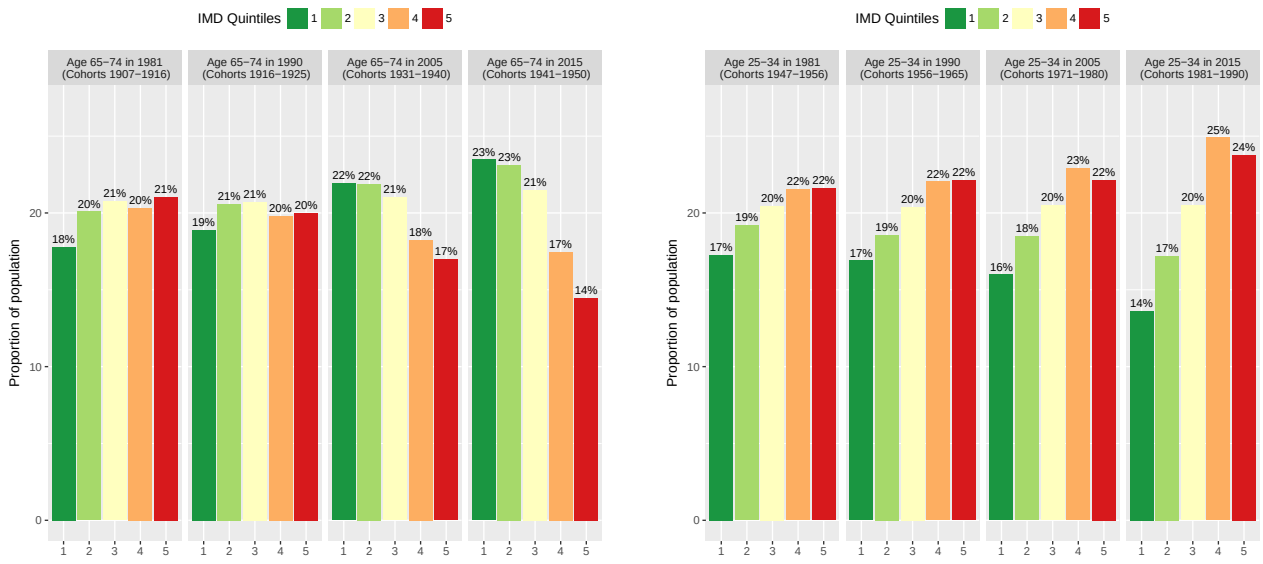
To compare the age distribution of the subpopulations more continuously through time, the evolution of the ages distribution in each of the IMD quintiles is represented in Appendix B.1, Figure B.1.

There is an inherent difficulty in representing the time evolution of data structured in age. In order to better understand the evolution of the composition of the population, we also represent the evolution of the composition of specific *age classes* (the same age class is represented at different dates), and of specific *cohorts* (the same cohort is represented at different dates), over the period 1981-2015.

Fixed age classes Figure 4.2 represents the evolution of the distribution of males in each IMD quintile, for the years 1981, 1990, 2005 and 2015 and for two fixed ages classes: 65-74 (Figure 4.2a) and 25-34 years (Figure 4.2b); plots for females are available in Appendix B.2. All results for males also hold for females.

The composition of the age class 65-74 significantly varied from 1981 to 2015, to the benefit of the least deprived populations. Thus, the proportion of males in the two least deprived quintiles (IMD quintiles 1 and 2) increased from 38% in 1981 to 46% percent in 2015; on the contrary the proportion of males in the two most deprived quintiles (IMD quintiles 4 and 5) decreased from 41% to 32%. This could be explained by an improvement over time of living conditions for older individuals, but also, as noted above, by a baby-boom cohort effect. Indeed, individuals born during the English baby-boom are less deprived than the immediately preceding and following cohorts, regardless of the global improving trend.

As already observed in Figure 4.1, the level of deprivation is more important in younger age classes, this being true for the whole period 1981-2015. However, as shown in Figure 4.2b, the composition of the age class 25-34 also varied from 1981 to 2015, and the relative deprivation of this age class increased over time. For instance, the proportion of males in the age class 25-34 for IMD quintiles 1 and 2 has decreased from 36% to 31%, while the proportion of males in the most deprived quintiles 4 and 5 has increased from 43% to 49%.



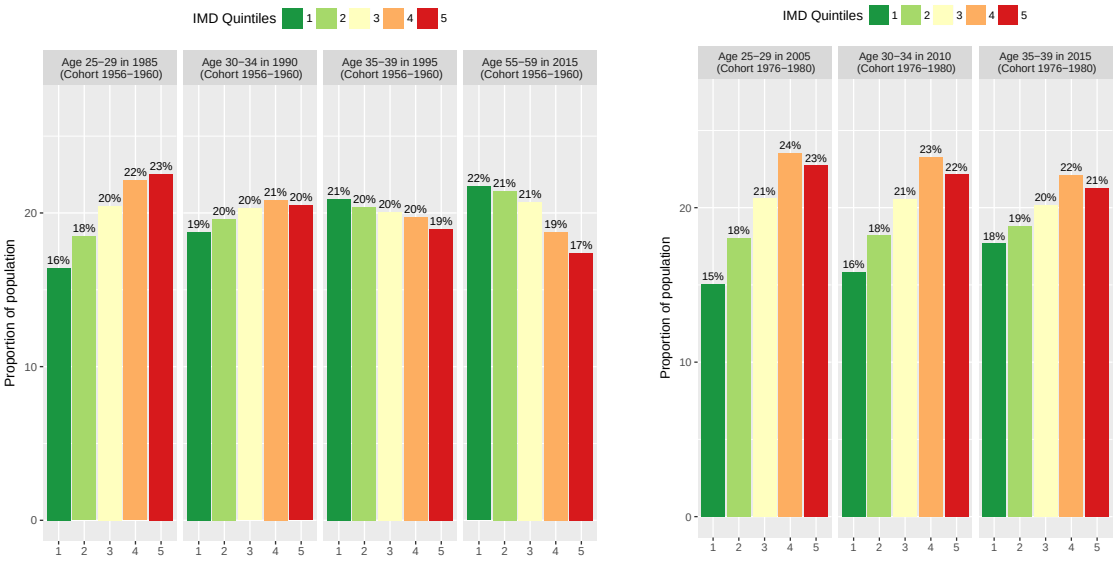
(a) Age class 65-74 (b) Age class 25-34

Fig. 4.2 Proportion of males by age class and IMD quintile

Fixed cohorts Data can also be represented in the cohort dimension. Figures 4.3a and 4.3b represent the evolution of the proportion of males in each IMD quintile for the cohorts with ages 25-34 in 1985 (born in 1956-1960) and in 2005 (born in 1976-1980); plots for females are available in Appendix B.2. The 1976-1980 cohort could only be represented up to ages 35-39.

The average deprivation of both cohorts improved over time, due to internal and external migrations (internal migrations correspond to individuals moving) and of different level of mortality rates. However, the improvement of the deprivation for the older cohort, Figure 4.3a, is much more important than that of the younger cohort, Figure 4.3b, which confirms observations of Figure 4.2b. Thus, the proportion of males in the 1956-1960 cohort at age 35-39, in year 1995, for the two most deprived quintiles was of 39%, against 43% in the 1976-1980 cohort at the same age (in 2015). Similarly, the proportion in the two least quintiles was of about 41% in the oldest cohort (1956-1960) against 37% in the youngest cohort (1976-1980).

At old ages, a classical selection effect can be observed, as cohorts become less and less deprived due to differences in the mortality rates, see also Figure B.2 in Appendix B.2.



(a) Cohort born in 1956-1960

(b) Cohort born in 1976-1980

Fig. 4.3 Proportion of males by cohort and IMD quintile

The population data show an important heterogeneity in the composition of the different age classes, combined with significant temporal changes in the age classes composition, such as the striking evolution of the composition of the 65-74 age class between 1981 and 2015.
This varying heterogeneity impacts mortality rates differently according to the age or time, and generates additional complexity in the study of aggregated death rates. In particular, one might wonder how the increase of deprivation observed among younger cohorts will impact future mortality in England.

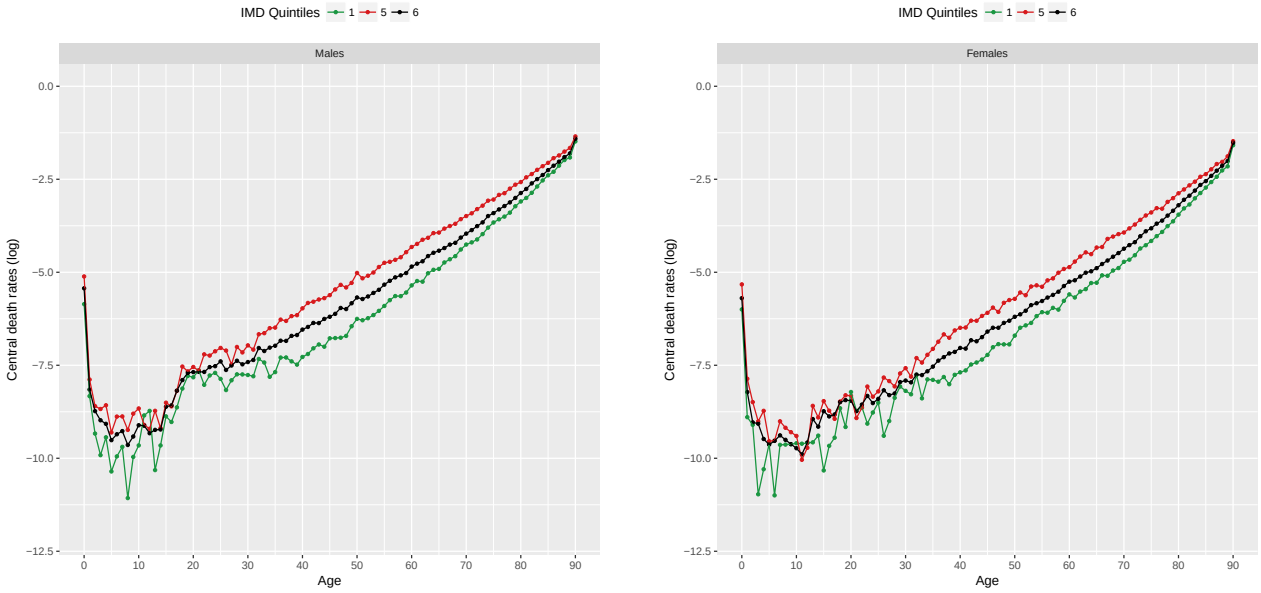
Mortality

Let us now give some insights on the mortality per deprivation in England, over the period 1981-2015. As mortality data are more commonly studied, we only give here a brief overview of the main stylized facts, with a particular focus on database 2. For more details on the mortality data of database 1, we refer to Bajekal (2005), Lu et al. (2014) or Villegas (2015).

Mortality rates Central death rates can be computed by age, gender, deprivation quintiles and cause of death. They are estimated from our data by taking the number of deaths over the mid-year population, which is assumed to be an accurate estimate

of the exposure to risk. For more details on the estimation of central death rates, see e.g. Delwarde and Denuit (2006). Central death rates by single year of age in 2015 are represented on a log-scale in Figure 4.4 for the English population, the least and the most deprived IMD quintiles. Central death rates for all IMD quintiles are represented in Figure B.3 of Appendix B.3. Both levels and shapes of central death rates vary with the level of deprivation, with a mortality higher at all ages for the most deprived IMD quintile 5.

Before age 35-40 (with the exception of age 0), differences are less pronounced and central death rates in all IMD quintiles are lower than 1‰. Let us give an example to illustrate differences in shape and level after age 40. After age 40, central death rates first attain the level of 10‰ at age 58 for males in the most deprived quintile 5, while this value is only attained at age 68 for males in the least deprived quintile 1. Similar differences can be observed for females. In the most deprived quintile, central death rates attain 10‰ at age 62, in comparison with 72 in least deprived quintile.



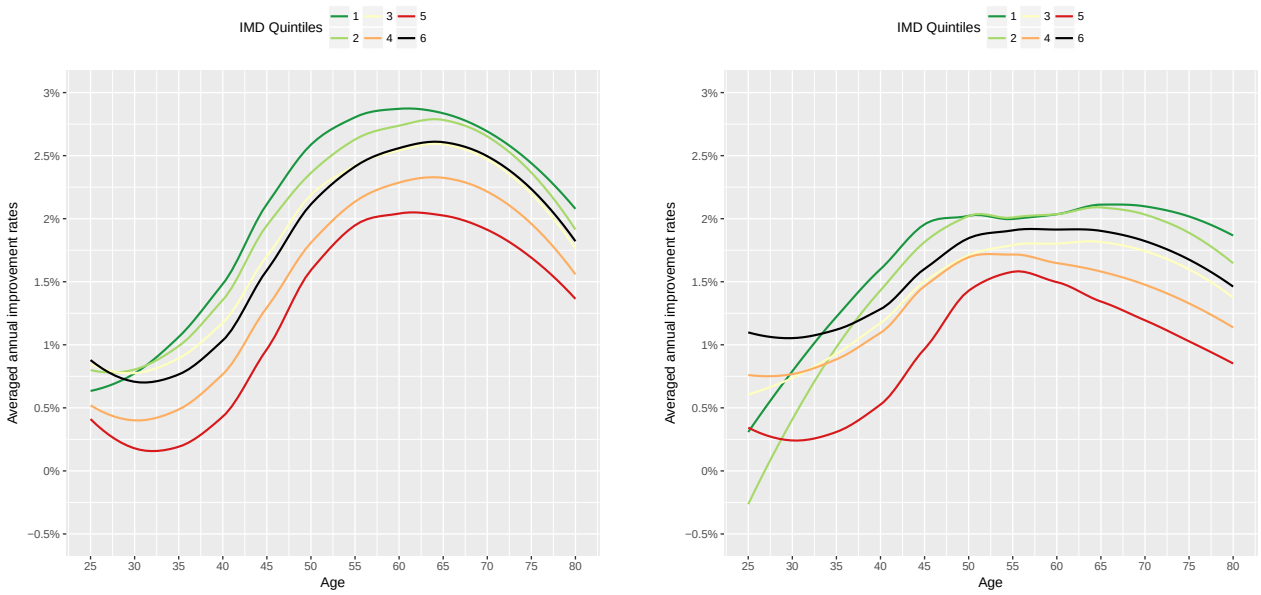
(a) Males (b) Females

Fig. 4.4 Central death rates per single year of age and IMD quintile in 2015

Improvement rates Average annual rates of improvement in mortality over the 1981-2015 period are represented in Figures 4.5a and 4.5b[7], for ages classes 25 and older. In this paragraph, improvement rates are smoothed for visualization purpose only. Figure

[7]Improvement rates are computed as the yearly improvement rates of central death rates over five year age classes.

4.5 shows that at all ages, the most deprived quintiles have experienced higher rates of improvement in mortality than the least deprived quintiles (with the exception of ages 25-34 for females). Males experienced overall higher improvement in mortality than females, with the highest differentials in deprivation being at ages 40-44 and 45-50. Females experienced the highest differentials in the rate of mortality improvement at ages 35-39 and 40-44.



(a) Males (b) Females

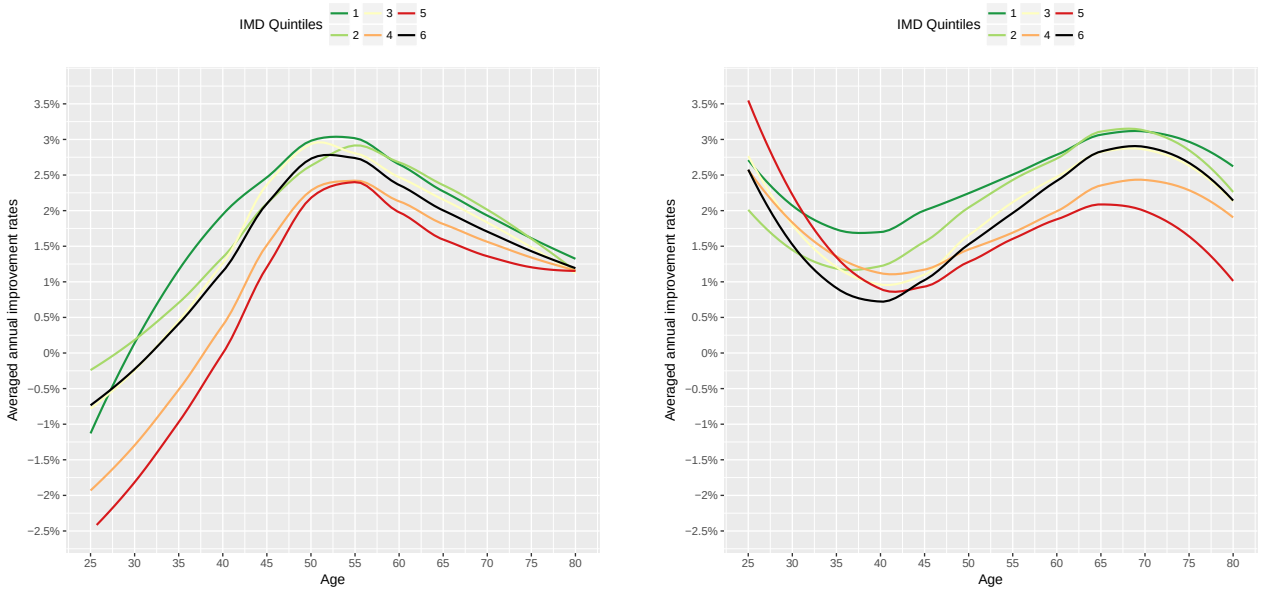
Fig. 4.5 Average annual rates of improvement in mortality, 1981-2015

Figures 4.6a and 4.6b represent the average annual mortality improvement rates for males over two distinct periods: 1981-1995 and 2001-2015. The two figures illustrate a clear widening of the gap in annual improvement in mortality at older ages, which is consistent with the observations of Lu et al. (2014) and Villegas and Haberman (2014) over the period 1981-2007.

Over the period 1981-1995, the two most deprived quintiles (4 and 5) actually experienced a deterioration of mortality at ages under 40, while improvements in mortality under 40 have been more important in IMD quintiles 4 and 5 than in the two least deprived quintiles over the period 2001-2015. Over the period 2001-2015, at ages above 60, the gap in mortality improvement rates increased significantly, with the highest differentials being at ages 75-79 and 80-84 (in comparison with 30-34 and 35-39 over the period 1981-1995). The improvements in mortality at younger ages have also changed significantly.

Plots for females are presented in Appendix B.3, Figure B.4. It is worth noting that over

the period 2001-2015 and for females of age above 65, the gap in mortality improvement rates between the most and least deprived IMD quintile has been higher than for males. This can be explained by a sharp deceleration of mortality improvement for females of the most deprived quintile, already reported by Villegas and Haberman (2014) for the period 1981-2007.
A discussion on the potential drivers of these widening of socioeconomic gaps can be found in Lu et al. (2014).



(a) 1981-1995 (b) 2001-2015

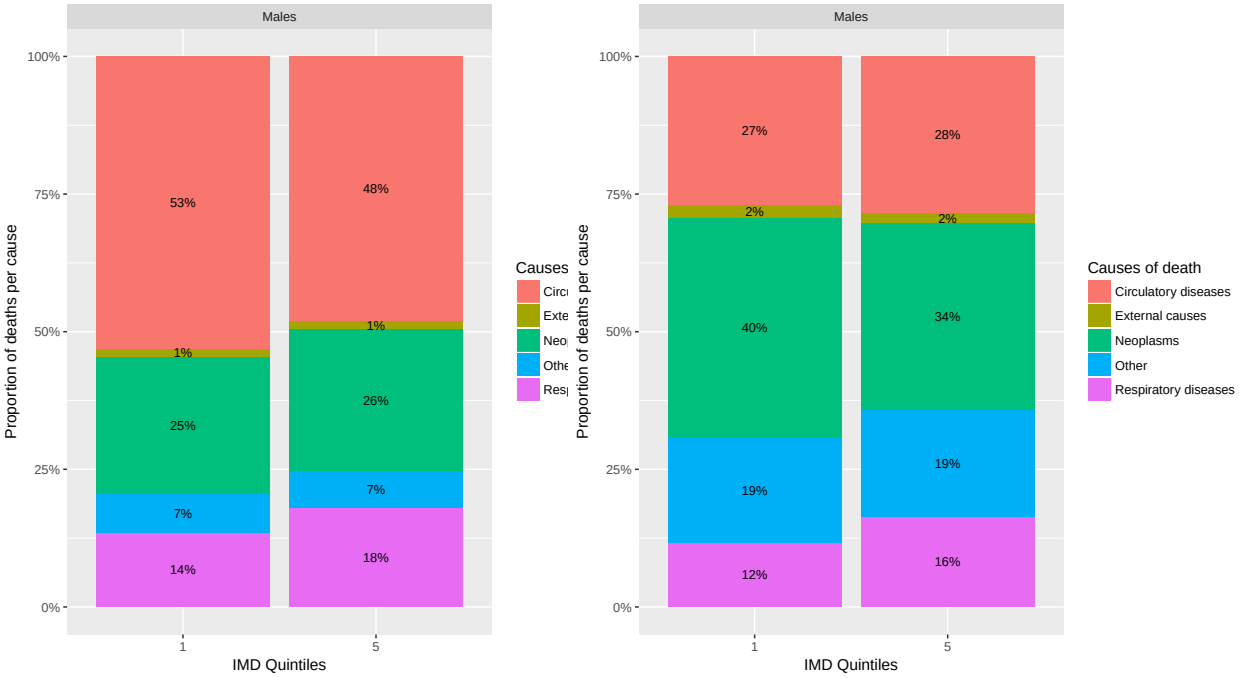
Fig. 4.6 Average annual rates of improvement in mortality, males

Causes of death Over the whole period 1981-2015, circulatory diseases constituted the first cause of death in England for ages above 25, followed by cancers (neoplasms) and respiratory diseases. However, recent changes in cause of death trend have been observed since the early 2000s, with neoplasms becoming a more prominent cause of death, larger than circulatory diseases. These evolutions occurred at different speed following the deprivation degree, the gender and the age class. For example, for males of age 25-85 in the least deprived IMD quintile, neoplasms became the first cause of death in 2005, while it only became the first cause of death in 2010 for males in the same age class in the most deprived IMD quintile.
Figures 4.7a and 4.7b represent the proportion of deaths per causes of death for males at ages 65-85, in the most and least deprived IMD quintiles and for years 1981 and 2015. Plots for females are available in Appendix B.4. For this age class, circulatory diseases

were the first cause of death in 1981 for both quintiles, whereas the first cause of death was neoplasms in 2015. Differences in cause-of-death mortality per deprivation can be observed for neoplasms, circulatory and respiratory diseases. In 1981, differences between the least and the most deprived quintiles were mainly on circulatory diseases (53% of all deaths for IMD quintiles 1 and 48% for IMD quintile 5), while differences were mostly on neoplasms in 2015 (40% of all deaths for IMD quintiles 1 and 34% for IMD quintile 5). During the whole period, differences in the proportion of death from respiratory diseases remained rather stable.

It is interesting to note as well that at young ages, the most deprived quintiles are more affected by neonatal deaths and accidents (see e.g. Guildea et al. (2001); Oakley et al. (2009) for more details on mortality at younger ages). We aslo refer to Villegas (2015) for more details on trends in cause-of-death mortality per deprivation over the period 1981-2007.

Plots for males and females in the age class 45-65 are also presented in Appendix B.4, Figures B.5 and B.5.

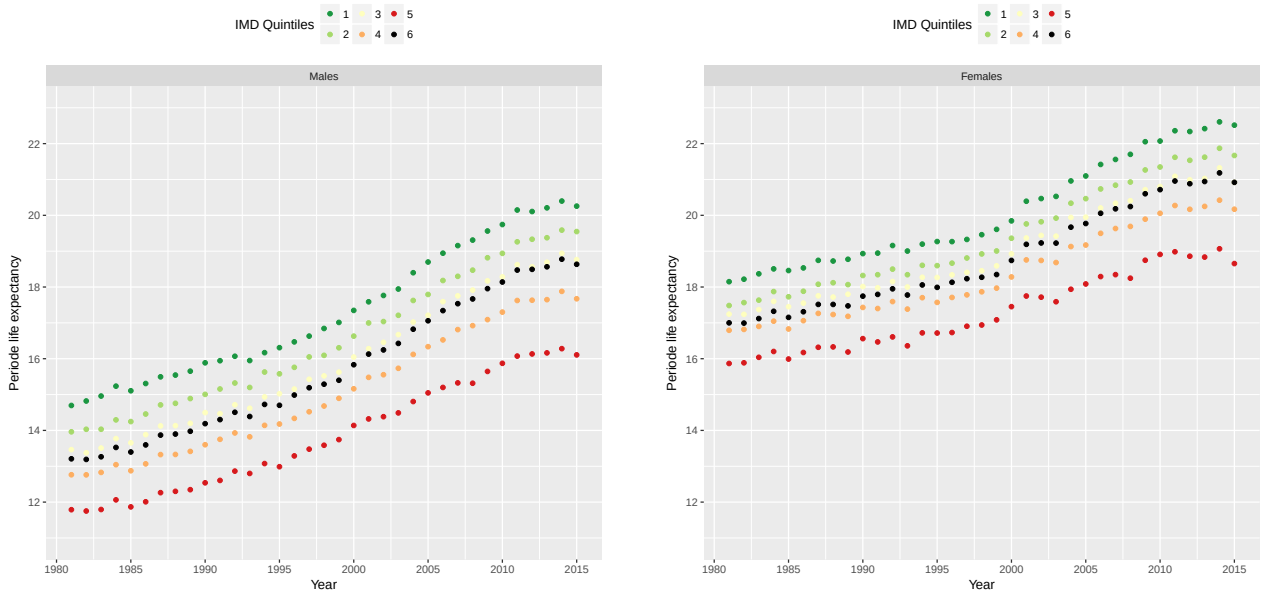


(a) 1981 (b) 2015

Fig. 4.7 Males deaths per cause and IMD quintile for ages 65-85

Period life expectancy The period life expectancy at age 65 per IMD quintile is represented over the period 1981-2015 in Figures 4.8a and 4.8b, for males and females.

As for death rates, we note a difference in level between the subpopulations for both genders, especially for males. The period life expectancy at age 65 increased over the years for all IMD quintiles. But despite a common improvement in mortality, the gap in life expectancy between IMD quintiles appears to have widened over time. For instance, the gap in life expectancy between the least and the most deprived quintiles has grown from 2.2 years for females and 2.9 years for males in 1981, to respectively 4.2 and 3.9 years in 2015. The evolution of the period life expectancy at age 25 for males and females is represented in Appendix B.5, Figure B.7.



(a) Males (b) Females

Fig. 4.8 Life expectancy at age 65 over 1981-2015

To summarize, the analysis of the data show that at the same time, gaps in mortality improvements by deprivation seem to have widened, while significant changes of composition have taken place in the population. Thus, understanding the evolution of the population dynamics appears to be instrumental in order to better understand the impact of this heterogeneity on the aggregated mortality. In the next section, the population dynamics is modeled under simple assumptions, in order to illustrate how mortality patterns can be impacted by compositional changes, even in a simple framework.

4.3 Population dynamics model

The joint evolution of subpopulations inside a global population (IMD quintiles in our case) is modeled by a linear and deterministic McKendrick-Von Foerster multi-population model, with no interactions. One of the advantages of this framework is to allow the derivation of closed formula and asymptotic results for a number of indicators, in order to better understand the complex interactions between heterogeneity and mortality.
The vital parameters of the model, i.e the birth and mortality rates, are assumed to be deterministic and only depending on age and time. The first assumption can actually be extended without difficulty to the broader scope of stochastic rates, depending for instance on a random environment. The second assumption means that vital rates do not depend on the population dynamics itself (the model is linear), which means that heterogeneity impacts the aggregated mortality only by "composition" effects. The assumption of no-interactions means that we assume that subpopulations evolve independently, in the sense that there are no internal (or external) migrations between the subpopulations. Even if this assumption can be unrealistic for young adults, it allows us to isolate the impact of heterogeneity in birth patterns, which can also to a certain extent reproduce certain patterns of internal migrations by changing cohort compositions. Finally, subpopulations dynamics are assumed to be deterministic: the evolution of the age pyramid is known, given the vital rates. The deterministic model is linked to stochastic population models recently developed by Tran (2008) in the field of mathematical ecology, and extended by Bensusan (2010) and Boumezoued (2016) to the framework of heterogeneous human populations, possibly in a random environment. Indeed, the deterministic model can be seen as a limit model when the size of the population becomes very large. However, the notion of limit model should be applied cautiously, since it does not allow the computation of "average" indicators when the studied quantities are non linear (see Boumezoued (2016), Arnold et al. (2016), Ferriere and Tran (2009) for further discussions on these matters).
In this section, we first recall briefly the McKendrick-Von Foerster model for a two-sex population with time dependent vital rates, and then move on to the description of the joint evolution of the subpopulations and of the aggregated population.

4.3.1 McKendrick-Von Foerster population dynamic model

The so-called McKendrik-Von Foerster model is a classical age-structured deterministic population model, first introduced by McKendrick (1926) and Von Foerster (1959), and which can be easily adapted into a two-sex model. The model is continuous in age and

time, in which the population is described at time t by the function of gender and age $(g(\epsilon, a, t))$, for $a \in [0, a^\dagger[$ and $\epsilon = f$ or m respectively for females and males. $g(\epsilon, a, t)$ should be understood as the number of individuals of gender ϵ between age a and $a + \mathrm{d}a$ at time t. It follows that $\int_0^{a^\dagger} g(\epsilon, a, t)\mathrm{d}a$ is the total number of individuals of gender ϵ in the population at time t. The model is determined by two vital rates functions: mortality and birth rates.

Vital rates functions

(i) *Mortality rate:* For each gender ϵ, the mortality rate (or force of mortality) at age a and time t is denoted by $\mu(\epsilon, a, t)$. Two types of mortality indicators are usually computed, cohort indicators and period indicators:

- The classical period survival function is denoted $S(\epsilon, a, t) = e^{-\int_0^a \mu(\epsilon, x, t)\mathrm{d}x}$ and represents the probability to survive to age a, *in the mortality conditions of time* t. The period life expectancy at age a and time t is given by:

$$e(\epsilon, a, t) = \int_a^{a^\dagger} e^{-\int_a^x \mu(\epsilon, s, t)\mathrm{d}s}\mathrm{d}x = \frac{1}{S(\epsilon, a, t)} \int_a^{a^\dagger} S(\epsilon, x, t)\mathrm{d}x.$$

- The cohort survival function is denoted by $\mathcal{S}^c(\epsilon, a, t) = e^{-\int_0^a \mu_j(\epsilon, s, t-a+s)\mathrm{d}s}$, and represents the probability for *an individual born at time* $t - a$ to survive until age a. The cohort life expectancy, which is the averaged time that individuals born at time $t - a$ will live after age a, conditional to surviving to this age, is denoted by:

$$\mathcal{E}^c(\epsilon, a, t) = \int_a^{a^\dagger} e^{-\int_a^x \mu(\epsilon, s, t-a+s)\mathrm{d}s}\mathrm{d}x = \frac{1}{\mathcal{S}^c(\epsilon, a, t)} \int_a^{a^\dagger} \mathcal{S}^c(\epsilon, x, t)\mathrm{d}x.$$

We also denote by $S(\epsilon, a - x, a, t)$ and $\mathcal{S}^c(\epsilon, a - x, a, t)$ the respective period and cohort survival probabilities from age $a - x$ to age a at time t, with $S(\epsilon, 0, a, t) = S(\epsilon, a, t)$ (the same holds for $\mathcal{S}^c$).

Observe that the period and cohort indicators do not provide the same information at all. The cohort life expectancy is "real", in the sense that it gives information on particular individuals living in the population. On the contrary, the period life expectancy is only an indicator which aggregates information on all individuals living in a given population at a specific date t. The period life expectancy can be interpreted as "the average duration of life of a representative individual living in the mortality conditions of time t".

(ii) *Birth rate* The rate of birth for an individual with gender ϵ and age a at time t is be denoted by $b(\epsilon, a, t)$. In a two-sex population, modelling births can be quite complex (Iannelli et al. (2005), Arnold et al. (2016)). In this article, we adopt the usual assumption that only women give births (the number of births does not depend on the number of

males in the population), so that $b(m,a,t)=0$. Females are assumed to give birth to a female with a probability $p^f=p$ and to a male with probability $p^m=1-p$. For sake of simplicty, the female birth rate $b(f,a,t)$ is now denoted by $b(a,t)$.

McKendrick-Von Foerster transport equation

The evolution of the population is given by the solution of the following transport partial differential equation:

$$(\partial_t+\partial_a)g(\epsilon,a,t)=-\mu(\epsilon,a,t)g(\epsilon,a,t),\ \forall a,t>0 \qquad \text{(balance law)} \qquad (4.3.1)$$

$$g(\epsilon,0,t)=p^\epsilon\int_0^{a^\dagger}b(a,t)g(f,a,t)\mathrm{d}a \qquad \text{(birth law)} \qquad (4.3.2)$$

$$g(\epsilon,a,0)=g_0(\epsilon,a) \qquad \text{(initial population)}$$

Due to the balance law, a proportion $(1-\mu(\epsilon,a,t))$ of individuals of age a and gender ϵ ages to age $a+\mathrm{d}a$ between time t and $t+\mathrm{d}t$. The birth law states that at each time t women of age a give birth to $b(a,t)g(f,a,t)$ individuals of age 0.
Equations (4.3.1)-(4.3.2) are usually solved along its characteristics curves, or equivalently by following the evolution of the population cohort by cohort. The resolution method can be interpreted as counting the number of survivors at time t in each cohort of individuals, either taken from the initial population or born after the initial time. Two regimes can thus be distinguished in the evolution of the population:
(i) *Individuals present in the initial population* $(a\geq t)$: At time t, individuals who were already present in the initial population are individuals of age $a\geq t$. Their number at time t is the number of individuals of age $a-t$ in the initial population, and who survived until time t:

$$g(\epsilon,a,t)=g_0(\epsilon,a-t)\mathcal{S}^c(\epsilon,a-t,a,t),\quad a\geq t. \qquad (4.3.3)$$

where recall that $\mathcal{S}^c(\epsilon,a-t,a,t)=e^{-\int_{a-t}^{a}\mu(\epsilon,s,t-a+s)\mathrm{d}s}$ is the cohort survival function from age $a-t$ to a, for individuals of age a at time t or equivalently born at time $t-a$.
(ii) *Individuals born after the initial time* $(a<t)$: At time t, individuals born after $t=0$ are individuals of age $a<t$. Their number at time t is thus the number of individuals born at time $t-a$ and who survived until time t:

$$g(\epsilon,a,t)=p^\epsilon B(t-a)\mathcal{S}^c(\epsilon,a,t),\quad a<t, \qquad (4.3.4)$$

where $B(t)$ is the birth function at time t which corresponds to the number of individuals (males and females) born at time t:

$$B(t) = \int_0^{a^\dagger} b(a,t)g(f,a,t)\mathrm{d}a. \tag{4.3.5}$$

Thus, if we look at the population at a small time t, the age pyramid will be mostly shaped by the time translated initial age pyramid, and will follow the idea that "today's youths give us most of the information on tomorrow's seniors". On a longer term, the initial population is naturally erased and the shape of the future age pyramid is only characterized by the birth and survival functions.

Stable age profile

The Stable theory provides useful indicators of the vital rates of a population, see Keyfitz and Caswell (2005) for more details and application of the stable theory and see Webb (1985) for a more general framework of multi-populations. The stable theory defines a *stable age profile*, given a fixed regime of time-independent age-specific vital rates. As for the period life expectancy, the stable age profile gives information on a fictive population, living *in the mortality and fertility conditions of a given time*. For instance its comparison with the real age profile of the population allows us to observe if strong changes in birth or mortality rates have occurred in the past.
For now, the vital rates are assumed to depend only on age and are denoted by $\mu(\epsilon, a)$ and $b(a)$. Since only women give birth, only the female stable age-profile is defined in a first step, and for simplicity of notation the variable ϵ will be omitted when considering the female population.

Stable solution of the McKendrick-Von Foerster equation A stable solution of the McKendrick-Von Foerster evolution Equation (4.3.1)-(4.3.2) (for females) is a solution which can be expressed as $\phi(a)T(t)$, with ϕ a function of age and T a function of time. In particular, a solution of this type is called stable because its age profile, or age distribution, remains constant over time. The size of the population is not constant but the proportion of individuals in each age class $[a_1, a_2[$ is constant and equal to:

$$\frac{\int_{a_1}^{a_2} \phi(a)\mathrm{d}a}{\int_0^{a^\dagger} \phi(a)\mathrm{d}a}.$$

By replacing $g(a,t)$ by $\phi(a)T(t)$ in (4.3.1)-(4.3.2), it follows that:

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

where λ^* is the unique solution of:

$$1 = p^f \int_0^\infty b(a) e^{-\lambda^* a - \int_0^a \mu(x)\mathrm{d}x} \mathrm{d}a. \tag{4.3.6}$$

The previous equation is called the characteristic equation of the model and λ is called the intrinsic growth rate of the population, which we justify now in the next paragraph.

Asymptotic exponential growth Actually, a much deeper property is that the solution of (4.3.1-4.3.2) with initial population g_0 and time independent vital rates $\mu(\epsilon, a)$ and $b(\epsilon, a)$ behaves asymptotically as a stable solution. After a long period of time, the population increases or decreases exponentially at rate λ^*. This means that the birth function B, which determines the shape of the age pyramid on the long term, has the following asymptotic exponential growth:

$$B(t) \underset{t\to+\infty}{\sim} C(\lambda^*, g_0) e^{\lambda^* t}.$$

Recalling that the population evolution is defined on the long term by equation (4.3.4), the previous equation gives an asymptotic stable equivalent for the age pyramid:

$$g(\epsilon, a, t) \underset{t\to\infty}{\sim} C(\lambda^*, g_0) e^{\lambda^*(t-a)} S(a). \tag{4.3.7}$$

The constant $C(\lambda^*, g_0)$ can be rewritten as $C(\lambda^*, g_0) = \dfrac{V(\lambda^*, g_0)}{M(\lambda^*)}$, where $V(\lambda^*, g_0)$ is called the total reproductive value of the initial population and $M(\lambda^*)$ the mean age at childbirth for the stable distribution.

The total reproductive value tells us how the initial age pyramid of women will weigh on the future number of births. $V(\lambda^*, g_0)$ is an increasing function of the total size of the initial population of women and decreasing in its mean age.

The mean age at childbirth under the stable distribution is a decreasing function of λ^*; Intuitively, the greater the intrinsic growth rate is, the younger females will have children.

To come back to the solution of the McKendrick-Von Foerster equation with time dependent rates, Equation (4.3.7) can be interpreted as follows: at a given time t_0, the right hand side of (4.3.7) is the shape that the age pyramid at t_0 would have if the vital rates had been constant in the past, equal to $\mu(a, t_0)$ and $b(a, t_0)$. Thus, the constant λ^* summarizes the combined effects on the population of birth and death rates at time t_0.

4.3.2 Joint evolution of the subpopulations

Subpopulations evolution

In the sequel, we consider the evolution of p socioeconomic subpopulations (for instance IMD quintiles). For each $j = 1...p$, the population j is described by the solution of the McKendrick-Von Foerster Equations (4.3.1)-(4.3.2)): $g_j(\epsilon, a, t)$ with initial population g_0^j and vital rates $\mu_j(a,t)$ and $b_j(a,t)$. Let us recall that $g_j(\epsilon, a, t)$ is the number of individuals at time t in population j, gender ϵ and between age a and $a + \mathrm{d}a$.

Aggregated population

We call aggregated population the global population composed of all subpopulations, denoted by $g(\epsilon, a, t)$ with:

$$g(\epsilon, a, t) = \sum_{j=1}^{p} g_j(\epsilon, a, t).$$

The dynamics of the global population is derived from the dynamics of the subpopulations,

$$(\partial_t + \partial_a)g(\epsilon, a, t) = -\sum_1^p \mu_j(\epsilon, a, t)g_j(\epsilon, a, t), \quad g(\epsilon, 0, t) = p^\epsilon \int_0^\infty (\sum_1^p b_j(a,t)g_j(f, a, t)\mathrm{d}a).$$

Aggregated mortality The mortality rate at age a in the aggregated population corresponds to the proportion of individuals of age a (for each gender) dying between a short interval of time $\mathrm{d}t$. Here, the previous partial differential equation can be rewritten as:

$$(\partial_t + \partial_a)g(\epsilon, a, t) = -\left(\sum_1^p \mu_j(\epsilon, a, t)\frac{g_j(\epsilon, a, t)}{g(\epsilon, a, t)}\right) g(\epsilon, a, t).$$

The mortality rate of the aggregated population is thus:

$$d(\epsilon, a, t) = \sum_{j=1}^{p} \mu_j(\epsilon, a, t)w_j(\epsilon, a, t), \text{ with } \quad \omega_j(\epsilon, a, t) = \frac{g_j(\epsilon, a, t)}{g(\epsilon, a, t)}. \tag{4.3.8}$$

Actually, the mortality rate in the aggregated population should be denoted by $d(\epsilon, a, t, (g_j)_{j=1..p})$, since it depends on the age pyramids of all subpopulations: the dynamics is non-linear. The time dependence of the aggregate mortality rate is caused not only by the time-dependence of the specific mortality rates in each subpopulation, but also by the evolution of the proportions $w_j(\epsilon, a, t)$ of individuals in each subpopulation and age class.
In order to better understand how these weights could impact the aggregated population, let us again make the distinction between the two cases $a \geq t$ and $a < t$. In order to simplify the notations, we consider for the remainder of this section that the aggregated

population is composed of two subpopulations ($p = 2$). When there is no ambiguity, we also omit the gender variable ϵ.

Aggregated mortality on the short term For individuals present in the initial population (of age $a \geq t$ at time t), the age pyramid of the subpopulations are mostly shaped by the initial subpopulations and Equation ((4.3.3)) yields for $a \geq t$:

$$d(a,t) = \frac{g_0^1(a-t)\mathcal{S}_1^c(a-t,a,t)\mu_1(a,t) + g_0^2(a-t)\mathcal{S}_2^c(a-t,a,t)\mu_2(a,t)}{g_0^1(a-t)\mathcal{S}_1^c(a-t,a,t) + g_0^2(a-t)\mathcal{S}_2^c(a-t,a,t).} \tag{4.3.9}$$

For small times t, the previous equation holds for most ages and the aggregated mortality depends on three ingredients:
- The subpopulations mortality rates μ_1 and μ_2.
- The initial subpopulations g_0^1 and g_0^2: the aggregated mortality rate at age a depends on the initial composition of the age class $a - t$, since individuals are assumed to stay in the same subpopulation. In particular, if the initial age pyramid is very heterogeneous in age, i.e if the age classes are composed very differently, aggregate death rates could experience significant changes (for instance if younger individuals are more deprived than older ones, this could lead to an increase of aggregated mortality rate on the short term).
- Cohorts survival: if the initial age pyramids in each subpopulation are equal ($g_0^1 = g_0^2$) Equation ((4.3.9)) becomes:

$$d(a,t) = \frac{\mathcal{S}_1^c(a-t,a,t)\mu_1(a,t) + \mathcal{S}_1^c(a-t,a,t)\mu_2(a,t)}{\mathcal{S}_1^c(a-t,a,t) + \mathcal{S}_2^c(a-t,a,t)} \quad a \geq t.$$

This illustrates a well known "selection" effect which is that if a subpopulation, say subpopulation 2, experiences a higher overall mortality, population 1 will have more and more weight at older ages and the aggregated mortality will tend to the mortality rate of subpopulation 1.

Aggregated mortality on the long-term On a longer term when t is large, the subpopulations evolution is mainly governed by the birth functions B_1 and B_2 and for $a < t$

$$d(a,t) = p^{\epsilon}\frac{B_1(t-a)\mathcal{S}_1^c(a,t)\mu_1(a,t) + B_2(t-a)\mathcal{S}_2^c(a,t)\mu_2(a,t)}{B_1(t-a)\mathcal{S}_1^c(a,t) + B_2(t-a)\mathcal{S}_2^c(a,t)}. \tag{4.3.10}$$

Thus, if the subpopulations experience heterogeneity in birth patterns for a certain period of time, this can induce a temporary variation in the aggregated mortality rates and generate a so-called *cohort effect* (see for instance Boumezoued (2016) for an example in a stochastic framework).

In the case of time-independent vital rates, we can obtain a stable equivalent for the aggregated death rate by replacing the birth functions by their asymptotic expression as in equation (4.3.7):

$$d(a,t) \sim p^{\epsilon} \frac{C(\lambda_1^*, g_0^1) e^{\lambda_1^*(t-a)} S_1(a)\mu_1(a) + C(\lambda_2^*, g_0^2) e^{\lambda_2^*(t-a)} S_2(a)\mu_2(a)}{C(\lambda_1^*, g_0^1) e^{\lambda_1^*(t-a)} S_1(a) + C(\lambda_2^*, g_0^2) e^{\lambda_2^*(t-a)} S_2(a)} \tag{4.3.11}$$

4.4 Numerical Results

This section presents presents the preliminary results of different simulations using the model presented above.
The model presented in Section 4.3 is applied in this section in order to illustrate different potential impacts of heterogeneity on the mortality experienced by the aggregated population, in two different cases:
We shall first start by showing in Section 4.4.1 how heterogeneity in age in the initial pyramid can affect the aggregated life expectancy and improvement rates on the short term. Then, in 4.4.2, we study the impact of heterogeneity in birth patterns on a longer term. In particular, we show how a cause specific reduction could be compensated in the presence of unfavorable birth patterns.
In order to isolate the influence of changes in the population composition, we assume in the following applications that the death rates in each subgroup do not depend on time. Thus, the death rate in each subgroup j is now denoted by $\mu_j(\epsilon, a)$. The aggregated mortality rate defined in Equation (4.3.8) is now equal to $d(\epsilon, a, t) = \sum_{j=1}^{p} \mu_j(\epsilon, a) w_j(\epsilon, a, t)$, and is only modified through changes in the composition of the population.
For illustrative purposes, we consider in the remainder of this section the evolution of a synthetic heterogeneous population composed of two subgroups: the most and least deprived IMD quintiles (IMD quintile 1 and 5).

Parameters of the model A common denominator of all population models is that vital rates are considered as input functions of the model. Coupled to an initial age pyramid, they are the determinants of the future shape of the age pyramid. Any changes in these inputs will thus impact the population evolution on the short and long term, which will, therefore, impact the aggregate longevity indicators of the global population. In our case, the parameters of the model, i.e. the birth rates, mortality rates, and initial age pyramids for the two IMD quintiles have to be specified.
In the different applications presented in this section, initial age pyramids and mortality rates are estimated for each scenario based on a given year, from the data presented in

Section 4.2.2. Since only central death rates per five-year age class are given in both databases, death rates are estimated based on the fitting procedure described in Appendix C.1. Furthermore, individuals in each IMD quintile of age above 85 are grouped in the age class "85 and older" (90 and older for database 2). To overcome this difficulty, we assume that individuals are distibuted in age classes until age 110 as in the UK population[8]. This assumption is consistent with the observation that deaths rates of all IMD quintles converge at old ages (see e.g. Figure 4.4).
In the baseline scenario, birth rates are assumed to be the same in each population, and estimated from Office for National Statistics (2015) (see Arnold et al. (2016) for more details on the estimation). The English females birth rates are given for 5-year age classes for ages 15 to 44. Over the reporting period 1981-2015, the probability p^f to give birth to a female is estimated to be between 0.4843 and 0.4886 (depending on the year).

Numerical implementation The general model is implemented by discretization of the transport partial differential equation 4.3.1, using a first order implicit Euler scheme. More details on the convergence and stability of the scheme, as well as other numerical methods can be found in the review of Pelovska and Iannelli (2006). The implementation of the model is further discussed in Appendix C.2.

4.4.1 Impact of heterogeneity in the initial age pyramid

As seen in Section 4.2.2, the English population shows a strong *heterogeneity in age*, meaning that the composition of the population can vary a lot according to the age class. Furthermore, the composition of the population has also varied significantly over time. In this first subsection, we illustrate how the heterogeneity in age of the age-pyramid can impact two different indicators of the evolution of mortality in the aggregated population: the period life expectancy at age 65 and the average annual mortality improvement rates at ages above 65.
The population is simulated on the short term (30 and 40 years) according to two different scenarios. In the first scenario, the initial age pyramids and mortality rates of the synthetic population (composed on the most and least deprived quintiles) are based on the population and mortality data of year 1981. In the second scenario, the initial age pyramids and mortality rates of the IMD quintiles are based on the population and mortality data of year 2015.

[8]The English age pyramid after age 85 is provided by the Human Mortality Database (The Human Mortality Database (2016)).

Evolution of the aggregated population The evolution of the aggregated population is shown in Figure 4.9 for the population based on the 1981 data, and in Figure 4.10 for the population based on the 2015 data. The initial age pyramids are displayed in Figure 4.9a and Figure 4.10a. Each age class is represented by the addition of individuals in the most deprived IMD quintile (in red) and in the least deprived IMD quintile (in green). The green line in each graph represents the age-pyramid of the least deprived IMD quintile.

The age-pyramids are structured differently, due to the two different databases, database 1 - structured by 5 year age class - and database 2 - structured by single year of age. It makes the comparison of the two pyramids more difficult. However, we can see that there are more individuals from the baby boom cohort (of age around 30-35 in 1981 and 65-70 in 2015) in the least deprived quintile than in the most deprived quintile. We can also see that older individuals are more deprived in the 1981 synthetic age pyramid than in the 2015 age pyramid, as already noted in section 4.2.2.

Moreover, there are no data for the age class 0-25 in database 1 (differences between the two databases are summarized in Table 4.1). In Figure 4.9a, individuals of age in 0-25 are distributed uniformly in each IMD quintile, based on the English population. However, this hypothesis has no influence here since only indicators at ages above 65 are considered over a period of 40 years. Thus, only individuals who were initially more than 25 years old are taken into account in the computation of the aggregated indicators.

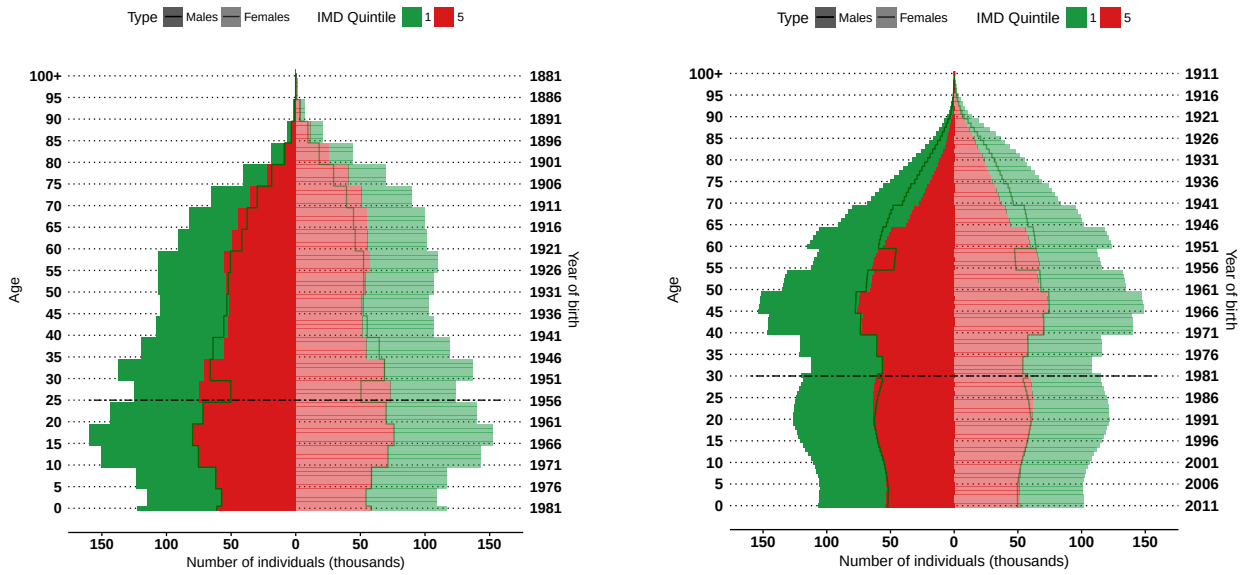
The age pyramids after 30 years are represented in Figures 4.9b and 4.10b. The two regimes in the population evolution described in (4.3.1) are distinguished in each figure by the black dashed line. Individuals of age over 30 were initially present in the population, and the age pyramid is defined by Equation (4.3.3). Individuals of age under 30 were born after the initial time, and the age pyramid is defined by Equation (4.3.4). For both populations, birth rates are assumed to be the same in each IMD quintile. However, birth rates have no influence in this first part since as stated above, we only consider indicators based on individuals who were initially more than 25 years old. This justifies our terminology “short term”.

For the population based on the 1981 data (Figure 4.9b), a decrease in deprivation can be observed for individuals over 60. This can be explained by the fact that younger cohorts were initially less deprived than older ones. Despite the important differences in mortality rates between the most and the least deprived quintiles, there was initially more individuals of age over 60 in the most deprived IMD quintile.

The population based on the 2015 data (Figure 4.10b) presents important differences in cohorts sizes. We can observe the evolution of the larger cohort of individuals born

during the the 60s, who were initially of age 45-55, and of age 75-85 after 30 years. The size of the age class 75-85 increases significantly over the 30 year period. However, the proportion of individuals in each IMD quintile appears to be rather stable, with around 60% of individuals in the least deprived IMD quintile.

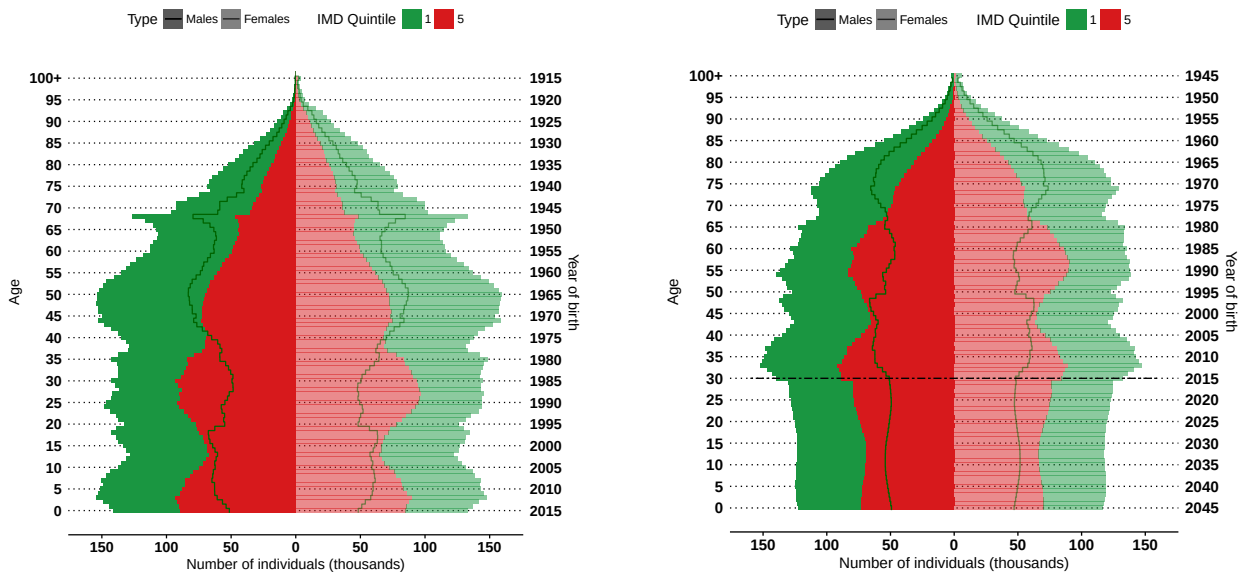
The composition of the age class 55-70 has changed significantly. Initially in this age class, individuals were much less deprived, with more individuals in the least deprived quintile. However, the situation is reversed 30 years later with an increase in the number of individuals in the most deprived quintile, as shown Figure 4.10b.



(a) Initial age pyramid

(b) Age pyramid at t=30

Fig. 4.9 Aggregated age pyramid with initial population based on the 1981 data



(a) Initial age pyramid

(b) Age pyramid at t=30

Fig. 4.10 Aggregated age pyramid with initial population based on the 2015 data

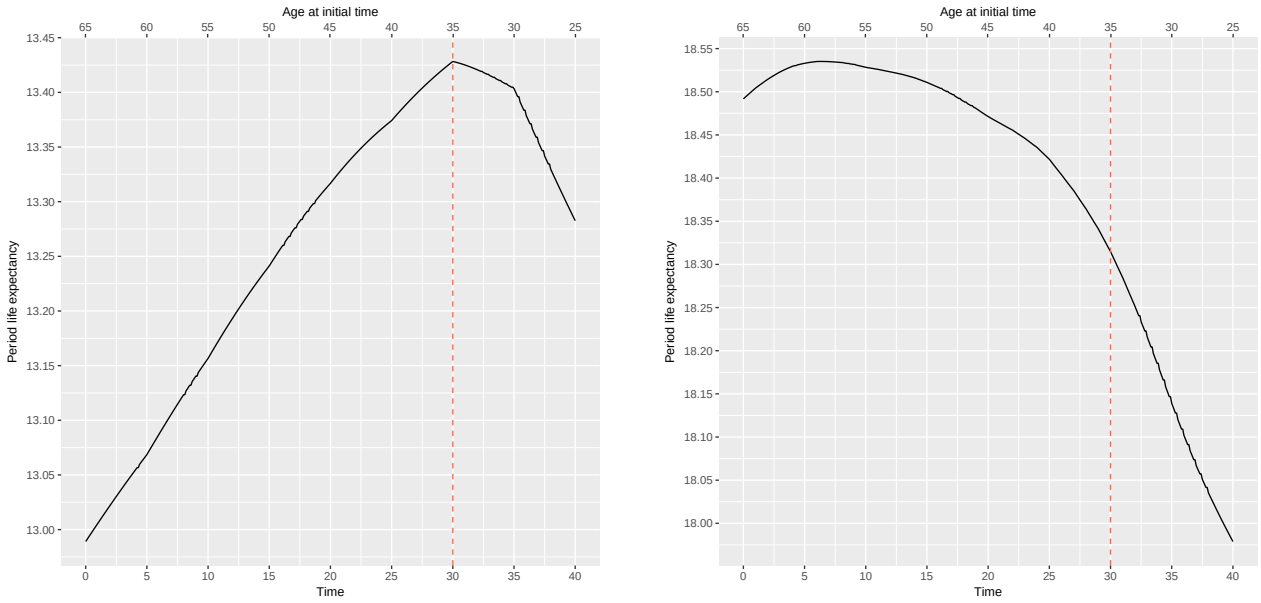
Period life expectancy Let us now study the impact on the period life expectancy at age 65 of the changes in the age pyramids observed in the previous paragraph. The evolution of the male period life expectancy at age 65 is represented for each scenario in Figure 4.11. Graphs for the female period life expectancy are presented in Appendix D.1, Figure D.1. As explained at the beginning of this subsection, only individuals who were initially of age over 25 are taken into account. The period life expectancy is thus computed based on the aggregated death rate defined in Equation (4.3.9). Note that since mortality rates in each IMD quintile are assumed to be not dependent on time, the cohort and period survival functions are equal in each IMD quintile, and do not depend on time. Thus, the period life expectancy only evolves due to changes in the age pyramids of the IMD quintiles.

The results show that the indicator moves in *opposite directions*, according to the year chosen for the initial pyramid, Figure 4.11a for initial year 1981 and Figure 4.11b for initial year 2015. The axe on top of each graph represents the age at initial time ($t = 0$) of individuals who are 65 years old at time t.

The evolution of the life expectancy at age 65 for the population based on the 1981 data is represented on Figure 4.11a. The evolution of life expectancy between $t = 30$ and $t = 40$ should be interpreted with caution. Indeed, we observe significant internal and external migrations in the data at ages 25-35, due to individuals moving (see e.g. Figure

4.3). Thus, the decrease of the period life expectancy between $t = 30$ and $t = 40$ in Figure 4.11a can be caused by the fact that the model does not take these changes into account.

The evolution of life expectancy at age 65 for the population based on the 2015 data is represented on Figure 4.11b. Contrary to Figure 4.11a, the life expectancy *decreases* of about 6 months in this case, after a short initial increase of the life expectancy. Thus, the evolution of the age pyramids of 2015 in the least and most deprived quintiles contributed *negatively* to the evolution of the life expectancy.



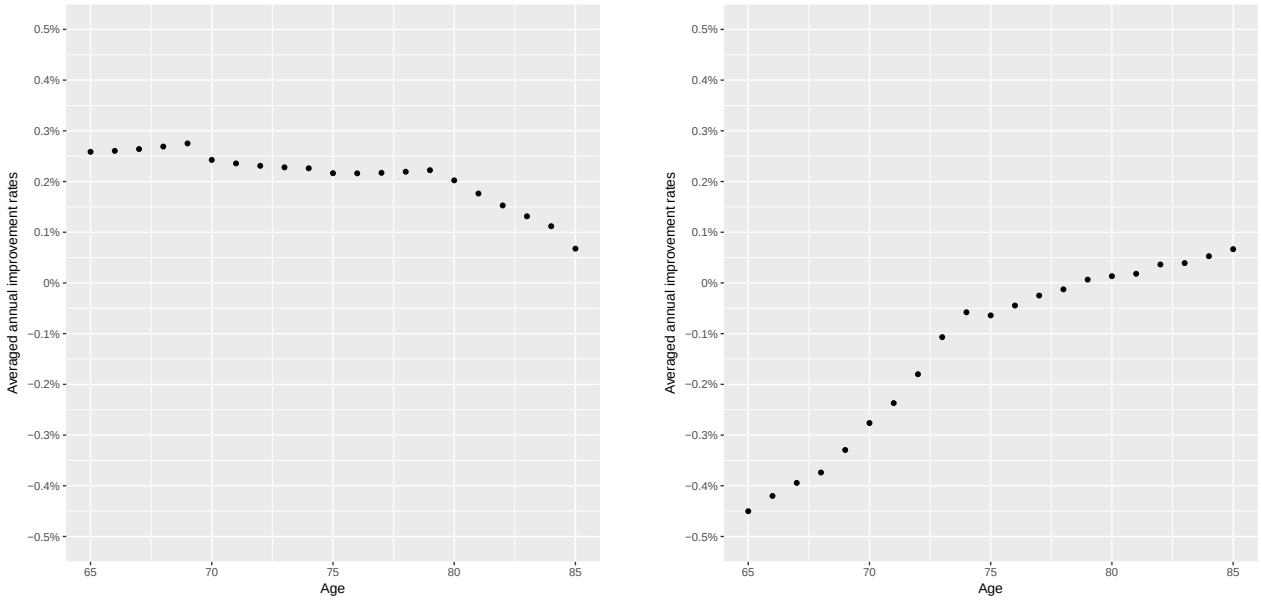
(a) 1981 inputs

(b) 2015 inputs

Fig. 4.11 Evolution of males life expectancy at age 65

Average annual mortality improvement rates To supplement the analysis, we compute the average annual mortality improvement rates over the first 30 years of simulation, for ages above 65. Average improvement rates for males are presented in Figure 4.12. Graphs for females are also presented in Figure D.2 of Appendix D.1.

The study of average annual mortality improvement rates confirms the observations made on the evolution of the life expectancy. In the conditions of 1981, averaged mortality improvement rates are *positive* at all ages, meaning that the evolution of the composition of the population contributed positively to the reduction of mortality rates. On the contrary, in the conditions of 2015, average improvement rates are *negative* at all ages below 80.



(a) 1981 data

(b) 2015 data

Fig. 4.12 Males average annual mortality improvement rates

To summarize, the two examples presented above demonstrate the importance of the heterogeneity of the age pyramid. The results show that the favorable changes in the composition of the population since 1981 probably contributed to the increase in the aggregated life expectancy. On the other hand, the model shows that based on the population of 2015, changes in the composition of the population might contribute negatively to the evolution of the aggregated life expectancy, and might offset improvements in the IMD quintiles mortality rates.

4.4.2 Cause of death reduction and compensation effect

In this subsection, we now focus on a particular mechanism which can create changes in the composition of the population: birth patterns. In order to understand the effect of changes in birth patterns, the population is thus studied on a longer term.
Many studies have shown that fertility differs also according to the socioeconomic status (The fertility rates of the most deprived quintile were around 1.18 times higher than the English general population fertility rates in 2015). Thus, in presence of heterogeneity, different birth patterns can induce changes in the composition of the population, and in turn influence the aggregated mortality of the global population. For example, in his

study of the so-called golden cohort in England[9], Willets (2004), noted that the important mortality improvements experienced by the cohort might have been a consequence of specific fertility rates in the 1930s, inducing a particular socioeconomic composition of the cohort (see Boumezoued (2016) for a more detailed discussion on the golden cohort). On the other hand of the spectrum, the consequence of causes of death reductions due to medical breakthroughs or public policies are more commonly investigated, and more and more studies take into account the heterogeneity of the population (see e.g. Alai et al. (2017); Labit Hardy (2016); National Research Council and Committee on Population (2011)). More and more recommendations on public health policy focus on the reduction of socioeconomic inequalities, for instance by advocating for policies targeting more deprived areas and/or causes of deaths which are more important among the most deprived individuals.

At the same time, in presence if changes in population composition and/or in cause-specific mortality in a given socioeconomic subgroup, the analysis of aggregated data can be quite complex. Based on these remarks, we illustrate in this subsection how a cause of death reduction can be compensated in presence of heterogeneity, and could be misinterpreted due to changes in birth patterns.

We consider the same synthetic population as in subsection 4.4.1, composed of the most deprived and the least deprived IMD quintile. In order to minimize the impact of the initial pyramid and focus on changes induced by birth rates, we now use the same initial age-pyramid for each IMD quintile, based on the age-pyramid of the English population. The initial pyramid is then composed of 50% of IMD quintile 1 and 50% of IMD quintile 5 at each age. In all scenarios, the demographic rates are based on the data of year 2015. The evolution of the aggregated population is simulated under different scenarios. The indicator used is the aggregated life expectancy at 25. In a first step, changes in birth rates and in mortality rates (due to a cause specific reduction) are studied separately. More precisely, we study the evolution of the aggregated life expectancy when a cause is reduced or when birth rates are increased in the most deprived quintile (IMD quintile 5). Since the IMD is based on geographic areas, the cause reduction could for instance be interpreted as an effect of a target policy. In a second step, the evolution of the population is studied when both demographic rates change at the same time.

In order to compare the results, we define a baseline scenario (numbered 0) in which birth rates and initial pyramids of each subpopulation are identical: only mortality rates

[9]The golden cohort is composed of individuals in the UK born around 1930. The terminology golden cohort is used because the cohort appears to have experienced particularly high mortality improvement rates.

differ among the subpopulations. In scenario 1(a), the impact on the aggregated period life expectancy of birth rate changes in IMD quintile 5 (most deprived) is studied. The impacts of cause of death reduction in IMD quintle 5 are then analyzed in scenario 1(b), under the assumption of independence between causes of death. We refer to Chiang (1968) or Arnold et al. (2016) for a discussion on this assumption. Finally, both changes are coupled in scenario 2.

To summarize, the aggregate life expectancy at age 25 is studied under the following scenarios: in scenario 0, we consider the same initial age pyramids for both subpopulations by using the age pyramid of the global English population in 2015, both populations have the same birth rate equal to the English birth rate in 2015. In scenarios 1(a) and 1(b), the population dynamics is modified by changing demographic rates separately (births 1(a) and deaths 1(b)). Finally, scenario 1(a) and (b) are coupled in scenario 2. The different scenarios are summarized in Table 4.3.

#	Initial pyramids: $[g_1(a,t_0); g_5(a,t_0)]$	Birth rates: $[b_1(a); b_5(a)]$	Death rates: $[\mu_1(a,t_0); \mu_5(a,t_0)]$
0	$[g_{Eng}(a,t_0); g_{Eng}(a,t_0)]$	$b_{Eng}(a,t_0)$	$[\mu_1(a,t_0); \mu_5(a,t_0)]$
1(a)	$[g_{Eng}(a,t_0); g_{Eng}(a,t_0)]$	$[b_{Eng}(a,t_0); b^{(*)}_{Eng}(a,t_0)]$	$[\mu_1(a,t_0); \mu_5(a,t_0)]$
1(b)	$[g_{Eng}(a,t_0); g_{Eng}(a,t_0)]$	$b_{Eng}(a,t_0)$	$[\mu_1(a,t_0); \mu^{(*)}_5(a,t_0)]$
2	$[g_{Eng}(a,t_0); g_{Eng}(a,t_0)]$	$[b_{Eng}(a,t_0); b^{(*)}_{Eng}(a,t_0)]$	$[\mu_1(a,t_0); \mu^{(*)}_5(a,t_0)]$

Table 4.3 Scenarios tested, $t_0 = 2015$

Baseline scenario (0) The evolution of the male (period and cohort) life expectancy at age 25 over time is represented in Figure 4.13. The plot for females is available in Appendix D.2. The initial (aggregated) period life expectancy at age 25 under scenario 0 is of 54.2, while the initial cohort life expectancy is of 54.5. Those values are between the life expectancy of the least deprived quintile 1, which is around 58, and the most deprived quintile 5, around 51. Let us note that since mortality rates are assumed to be independent of time in each subpopulation, the period and cohort life expectancy in each subpopulation are equal and constant over time.

The difference between the initial aggregated period and cohort life expectancy can be explained by the choice of the initial age pyramid. On the one hand, the aggregated period life expectancy is computed based on the initial age pyramid, composed of 50% of the two subpopulations at each age. On the other hand, the cohort life expectancy is computed by considering the evolution of the cohort composed of individuals who were

initially 25 years old. Even if the mortality rates in each subpopulation are considered to be time independent, the composition of the cohort is modified over time, due to the difference in mortality rates between the two subopulations. Since individuals in the most deprived quintile have a higher mortality, the cohort composition will change in favor of the least deprived quintile, as individuals grow older. This explains why the period life expectancy is lower than the initial cohort life expectancy.
Another consequence of the difference in mortality rates is that the intrinsic growth rate (as defined in (4.3.6)) of the least deprived subpopulation is higher than in the most deprived subpopulation. As interactions between the subpopulations are not modeled here, this means that the aggregated life expectancy will tend over a long run to the life expectancy of the least deprived subpopulation.

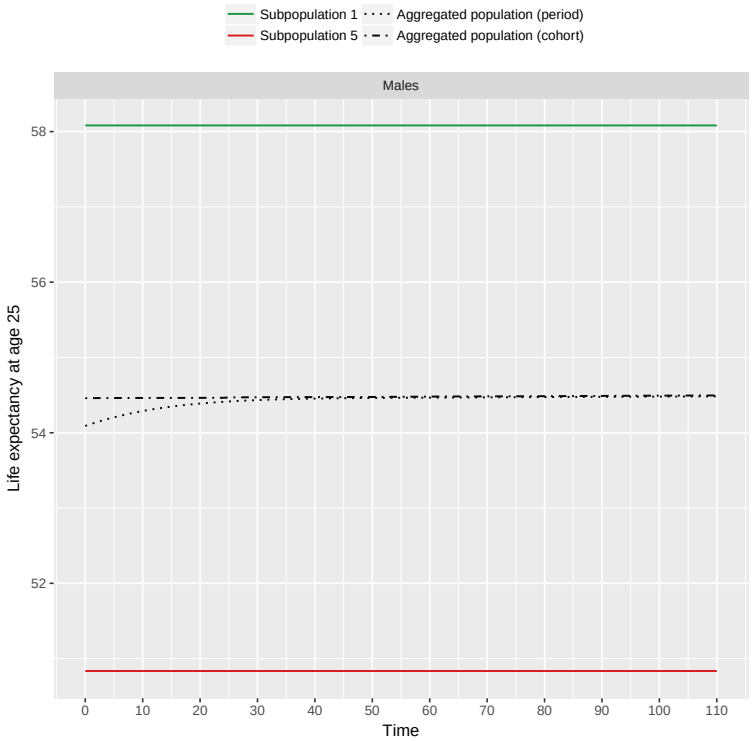


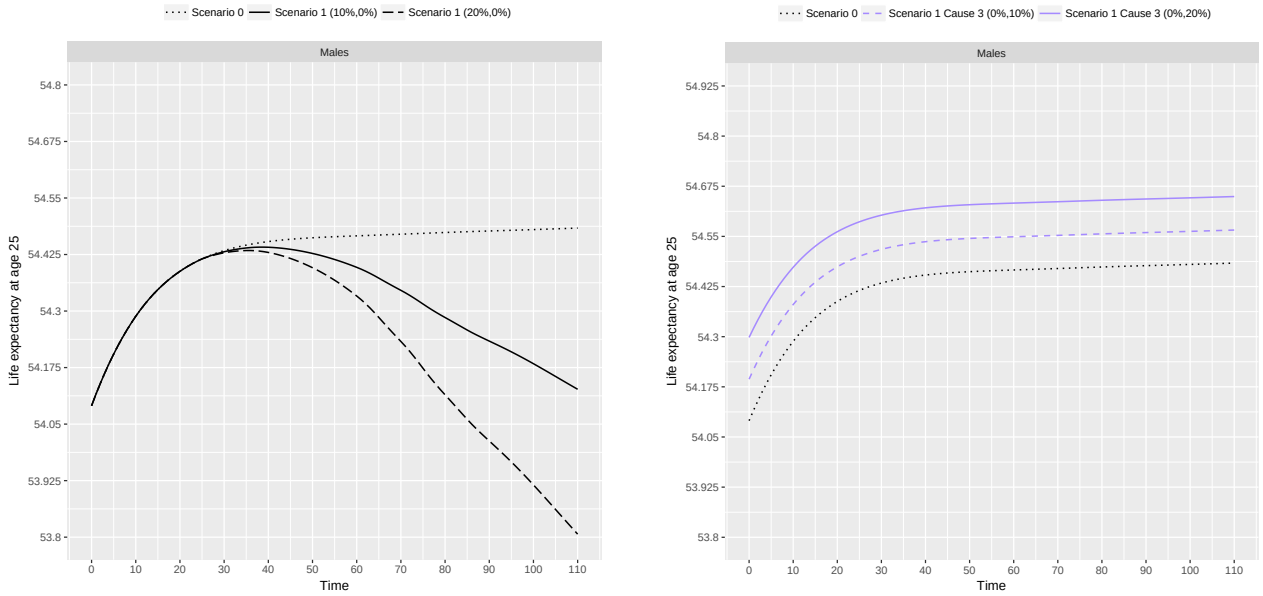
Fig. 4.13 Life expectancy over time

Scenario 1(a) and 1(b) Let us now compare effects of two different natures: changes in births rates on the one hand (scenario 1(a)), and on the other hand cause specific reductions (scenario 1(b)), which are more commonly studied. The evolution of the period life expectancy under those scenarios is represented in Figures 4.14a and 4.14b.
In scenario 1(a), the birth rate of the most deprived subpopulation is slightly increased (+10% and +20% from the baseline scenario), and is thus higher than in the least deprived

subpopulation. When increasing the birth rate in the most deprived subpopulation, the composition of the population changes and the weight of the most deprived IMD quintile becomes more important in the aggregated population/mortality. Before time $t = 25$, the period life expectancy at age 25 is based on individuals present in the initial population, and thus the three scenarios represented in Figure 4.14a do not differ up until this time. From time $t = 25$, the computation of the period life expectancy at age 25 includes individuals born after the initial time. Due to the increase in the birth rate of the most deprived subpopulation, new cohorts are more deprived. Compared to the baseline scenarios, the most deprived subpopulation has more weight in scenario 1(a), which implies a degradation of the aggregated mortality.

On the other hand, we slightly reduce the mortality of the most deprived subpopulation after age 25 and observe the effects on the aggregated life expectancy (scenario 1(b)). The population dynamics is modified by reducing ischemic heart diseases (Cause 3) for the most deprived subpopulation (-10% and -20%), under the independence assumption between causes of death. Reducing death rates in the most deprived subpopulation increases the initial aggregated life expectancy, which then increases again due to changes in the composition of the population, as described for the baseline scenario. It is interesting to note that whichever cause is *eliminated* in the most deprived quintile, the corresponding life expectancy is still smaller than the life expectancy of the least deprived quintile.

By changing the birth rate in the most deprived subpopulation in scenario 1(a), the intrinsic growth rate of the population increases and becomes higher than the growth rate of the least deprived subpopulation, hence the decrease in the aggregated life expectancy. In scenario 1(b), the growth rate of the most deprived subpopulation also increases and is higher than in the baseline scenario. However, the intrinsic growth rate in the most deprived subpopulation is still lower than in the least deprived population.



(a) Birth changes

(b) Mortality changes

Fig. 4.14 Aggregated period life expectancy over time

Scenario 2 Finally we couple both changes in births and deaths rates in the subpopulation 5, in order to study the combined effects. We demonstrate how a cause of death reduction can be compensated by births and thus could misinterpreted.
In this scenario, we assume that birth rates are higher in most deprived subpopulation (+10% and +20% from the baseline scenario), while the subpopulation also experiences a decrease in ischemic heart diseases (−20% from the baseline scenario). The evolution of the aggregated life expectancy under these scenarios, in comparison with the baseline scenario, is presented in Figure 4.15. The aggregated life expectancy is initially higher due to the cause reduction. As in scenario 1(a), birth rates impact the life expectancy from $t = 25$. Due to the changes in the cohort composition engendered by the heterogeneous birth rates, the life expectancy decreases in both versions of scenario 2, and eventually become lower than in the baseline scenario. For instance, when birth rates of the most deprived subpopulation are 20% higher than in the baseline scenario, the increase in life expectancy is compensated when the first individuals born in the population attain age 65.

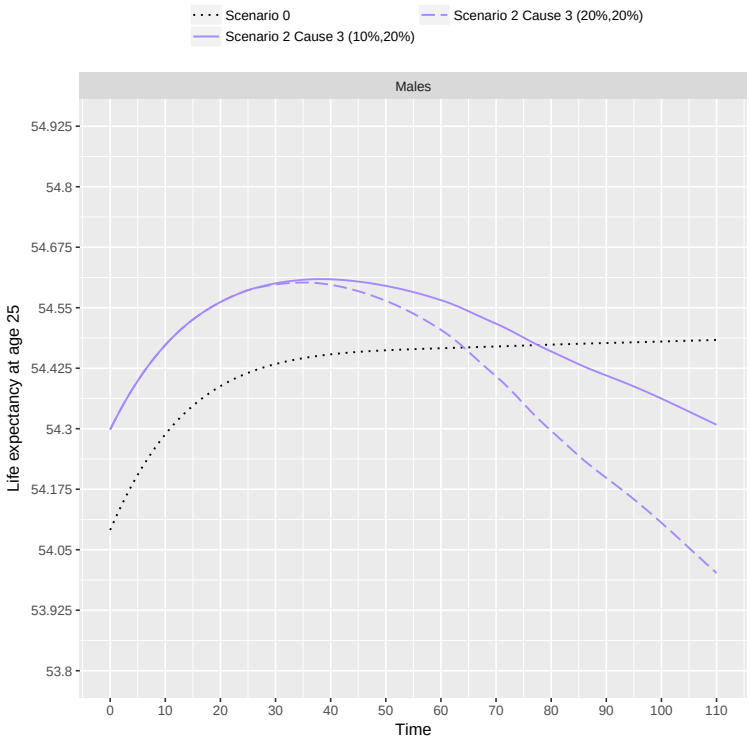


Fig. 4.15 Aggregated period life expectancy over time: Ischemic heart diseases reductions

Figure 4.16 presents the aggregate life expectancy at age 25 following the reduction (-20%) for 5 other causes of death, and with a higher birth rate ($+10\%$) for the most deprived population. The causes considered are numbered from 1 to 5: neoplasms (1), respiratory diseases (2), ischemic heart diseases (3), strokes (4), and external causes (5). The life expectancy evolution under this third scenario is similar for each cause reduction. Naturally, the direction and magnitude of the impact of the cause reductions depend on the cause of death that is reduced, since different causes do not impacting all age groups and socioeconomic categories in the same way. Per order of importance, the causes are the following: neoplasms, ischemic heart diseases, respiratory diseases, external causes and strokes.

For instance, we see on Figure 4.16 that a reduction of -20% of strokes in the most deprived quintile induces variation of life expectancy very similar in magnitude to those induced by changes in birth rates.

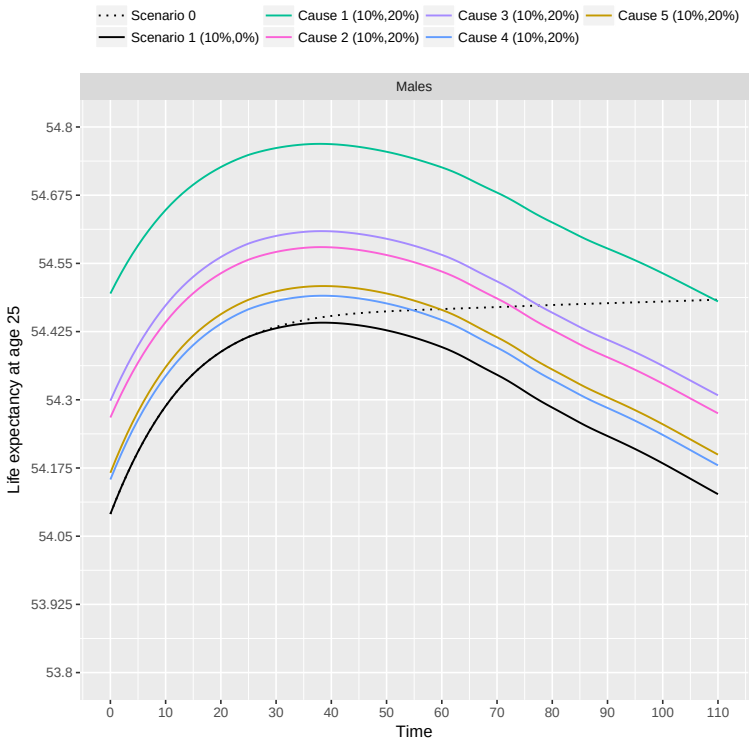


Fig. 4.16 Aggregated period life expectancy over time

4.5 Concluding Remarks

To conclude, reducing or eliminating a cause of death may not necessarily result in an improvement of aggregated mortality rates or life expectancy, if the composition of the population changes at the same time.

The data analysis and the first tests presented in the previous section reveals interesting interactions between the population dynamics and mortality indicators. They also stress the importance of studying the whole population dynamics in presence of heterogeneity, even when only interested in longevity indicators. Indeed, based on our data, we see that the future evolution of the composition of population could have a very different impact on mortality than 30 years ago. Furthermore, effects of birth or death rates modifications in an heterogeneous population are not straightforward. More particularly, a "reverse" cohort effect could easily compensate a cause-of-death reduction. In this case, the effect of a public health policy could be missed if only aggregated data are studied.

As part of an ongoing work, we are currently extending these results in order to include time-dependent demographic rates and interactions between subgroups. However, these

first tests already highlight interesting insights on the analysis of the aggregated mortality for a population composed of different socioeconomic categories.

Appendix A

IMD over time

The IMD is computed at fixed times, year 2007 and year 2015, and applied to larger time periods, see Figure A.1. Therefore, the socio-economic evolution of living areas over the reporting periods is not taken into account. However, the aggregation of small living areas into quintiles deprivation might dim changes over reasonable periods of time. For the 1981-2006 period, previous studies have show that the majority of small areas, especially areas in the extreme quintiles, have stayed in the same deprivation quintile (se e.g. Bajekal et al. (2013), Appendix A of Lu et al. (2014) or Norman and Darlington-Pollock (2016) for the 1970-2011 period).

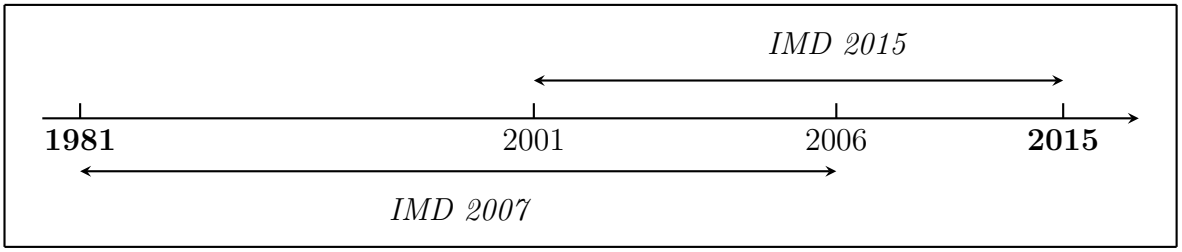


Fig. A.1 IMD2007-IMD2015

We have also compared data over the 2001-2006 period, computed with the IMD 2007 in the one hand (database 1) and with the IMD 2015 in the other hand (database 2). For the overlapping period 2001-2006, we obtained differences in life expectancy of less than one percent.

Appendix B

Miscellaneous information on data

B.1 Age profiles

Figure B.1[1] represents the evolution of the age profiles of each IMD quintiles from 2001 to 2015. The age profile of a population corresponds to the distribution of age inside the population.

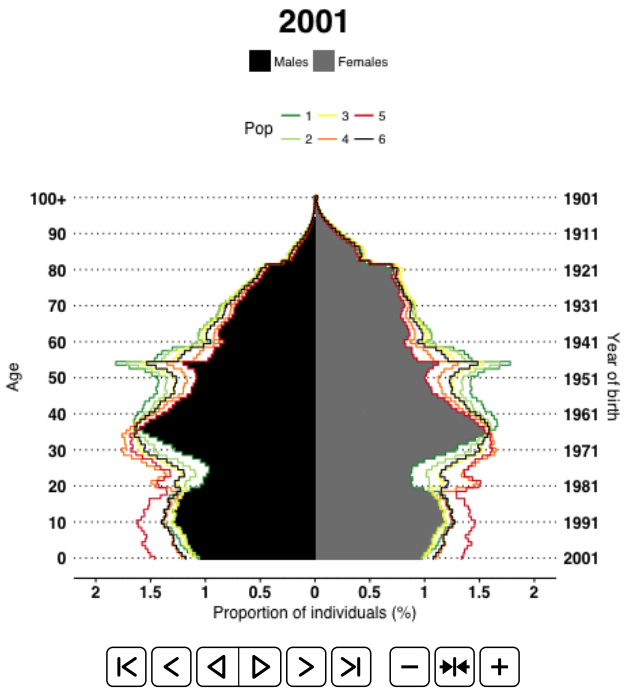
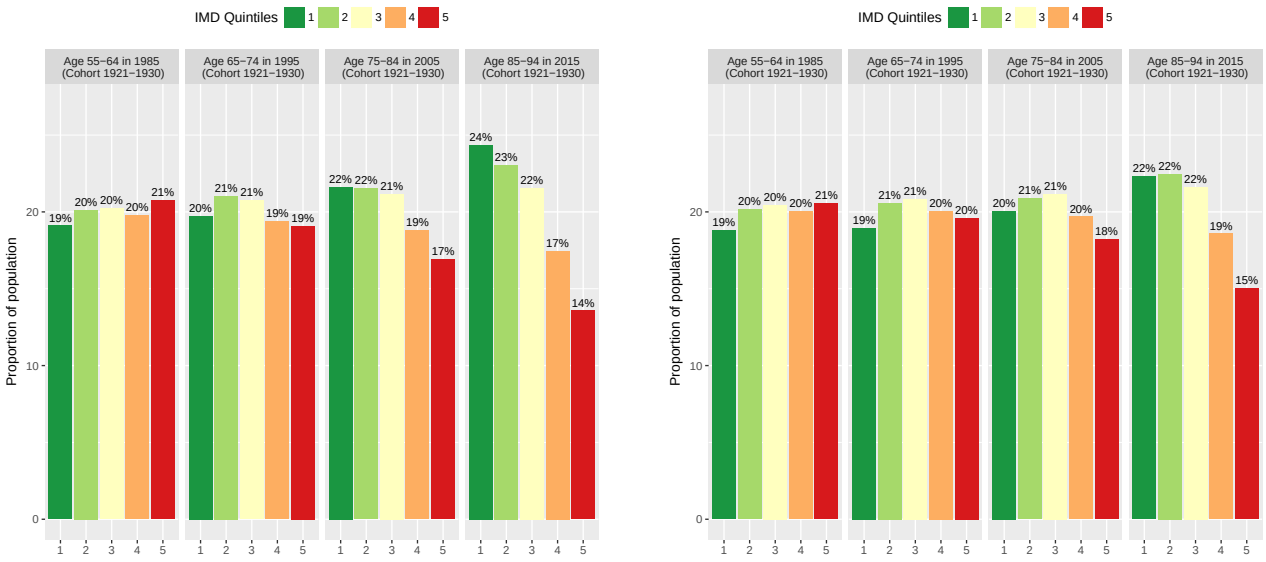


Fig. B.1 Age profile per IMD 2015 quintile from 2001 to 2015 (GIF)

[1]The figure is in GIF format. In order to see the animation, please open the PDF file with a PDF reader supporting this format.

B.2 Ages classes and cohorts



(a) Males proportions (b) Females proportions

Fig. B.2 IMD 2015 quintile proportions for the Cohort 1921-1930

B.3 Central death rates

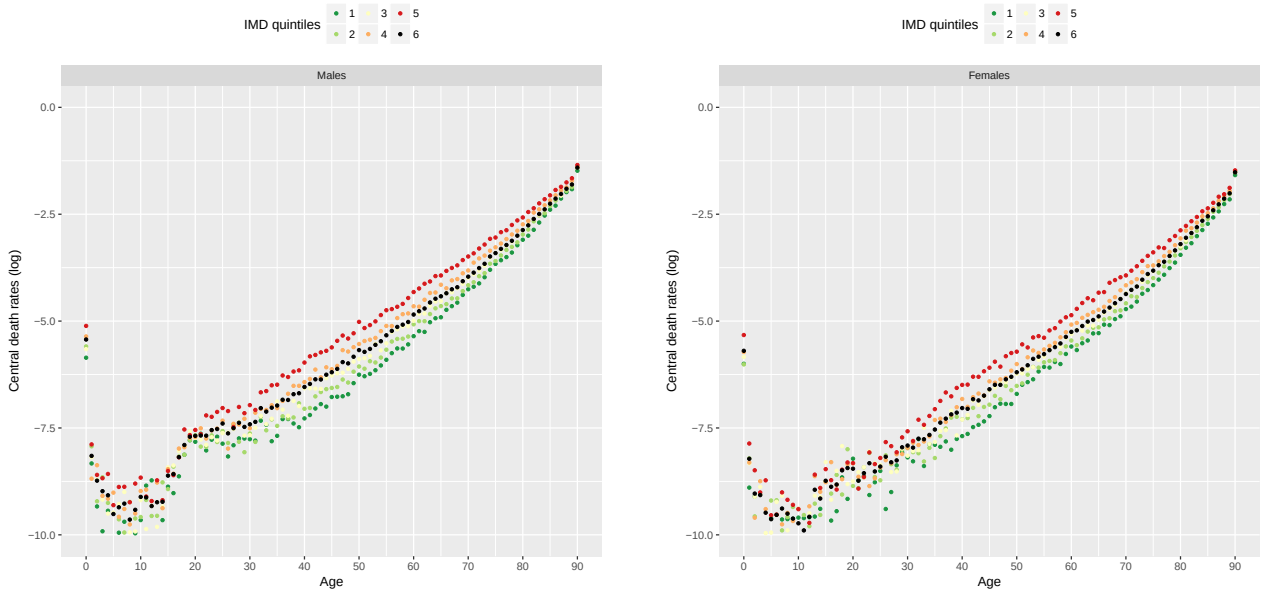
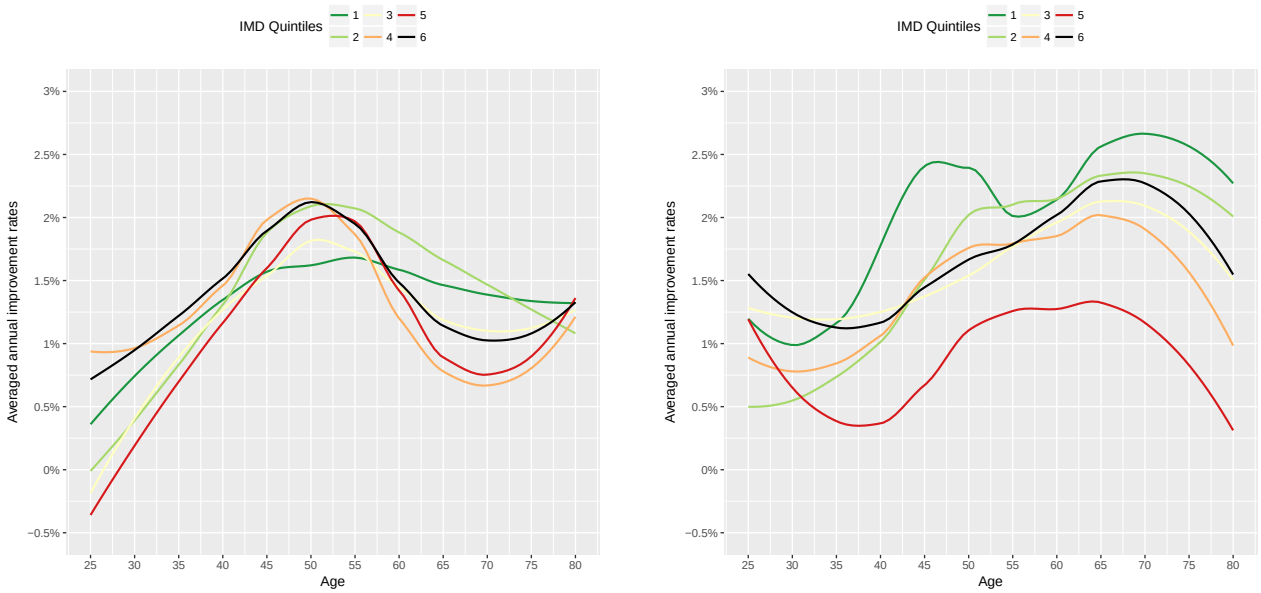


Fig. B.3 Central death rates per IMD 1y 1981, 2015

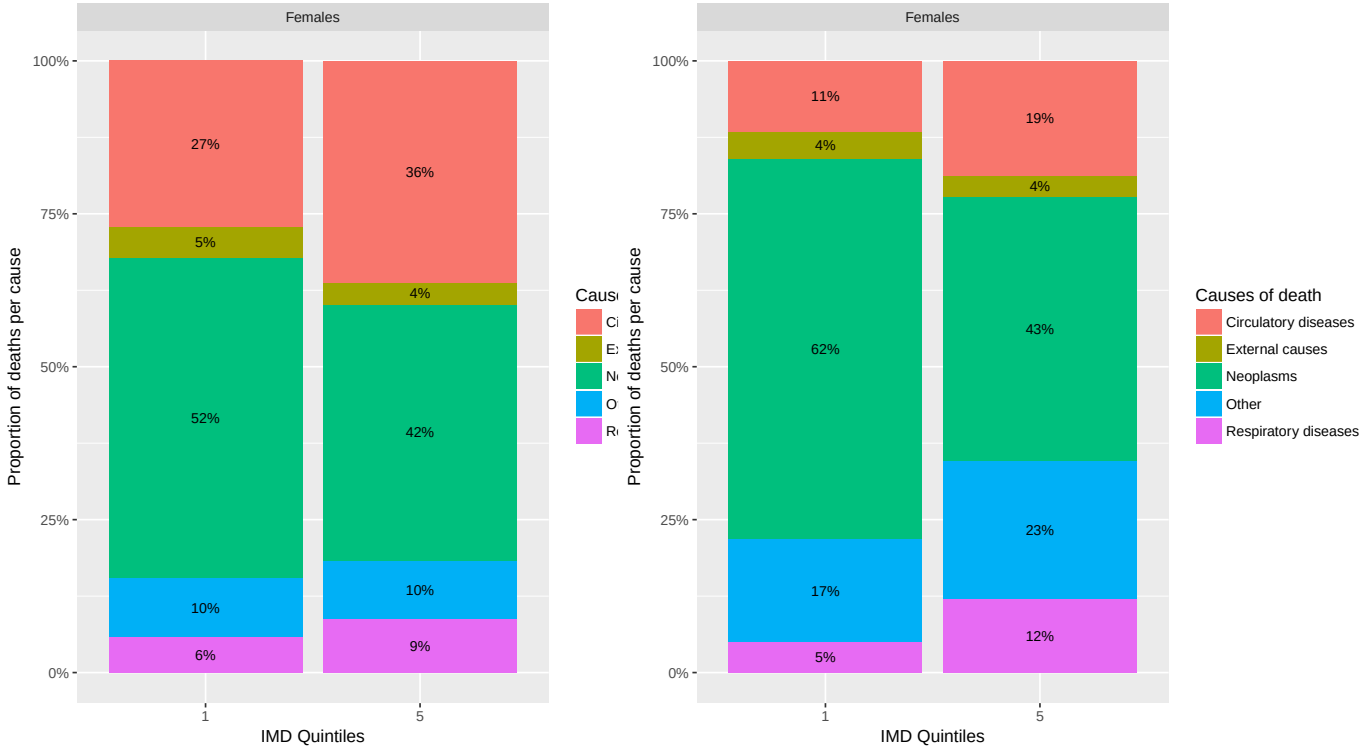


(a) 1981-1995

(b) 2001-2015

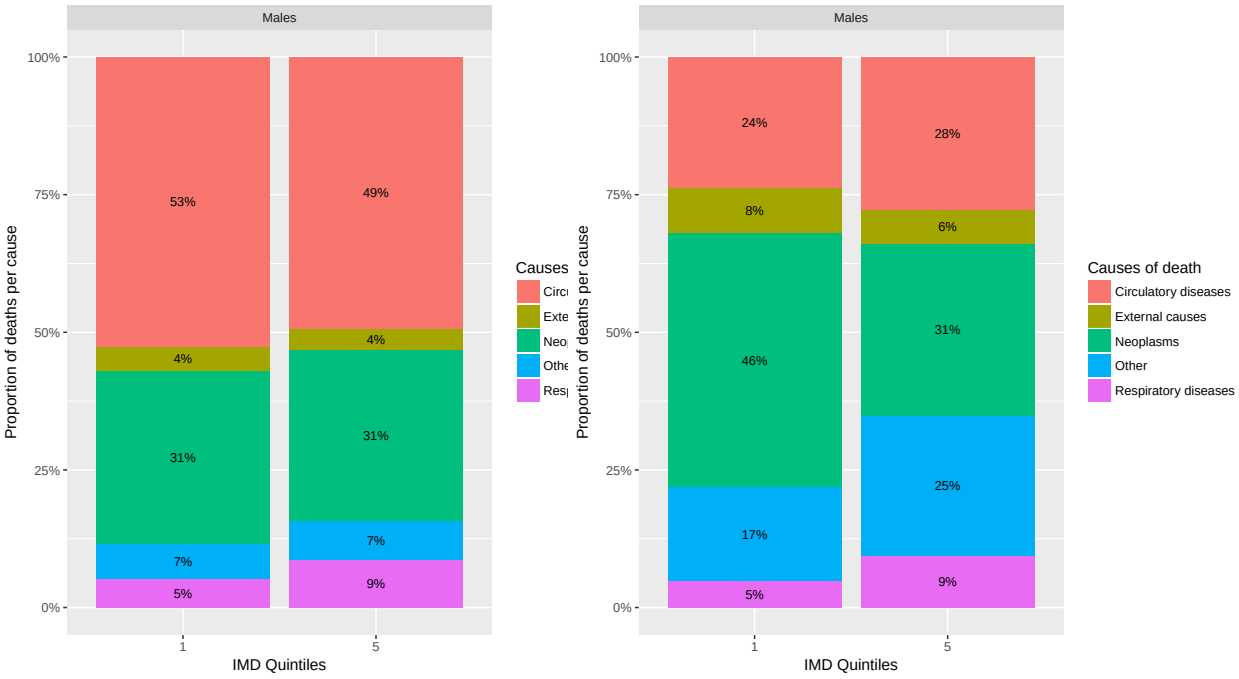
Fig. B.4 Females improvement rates

B.4 Causes of death



(a) 1981 (b) 2015

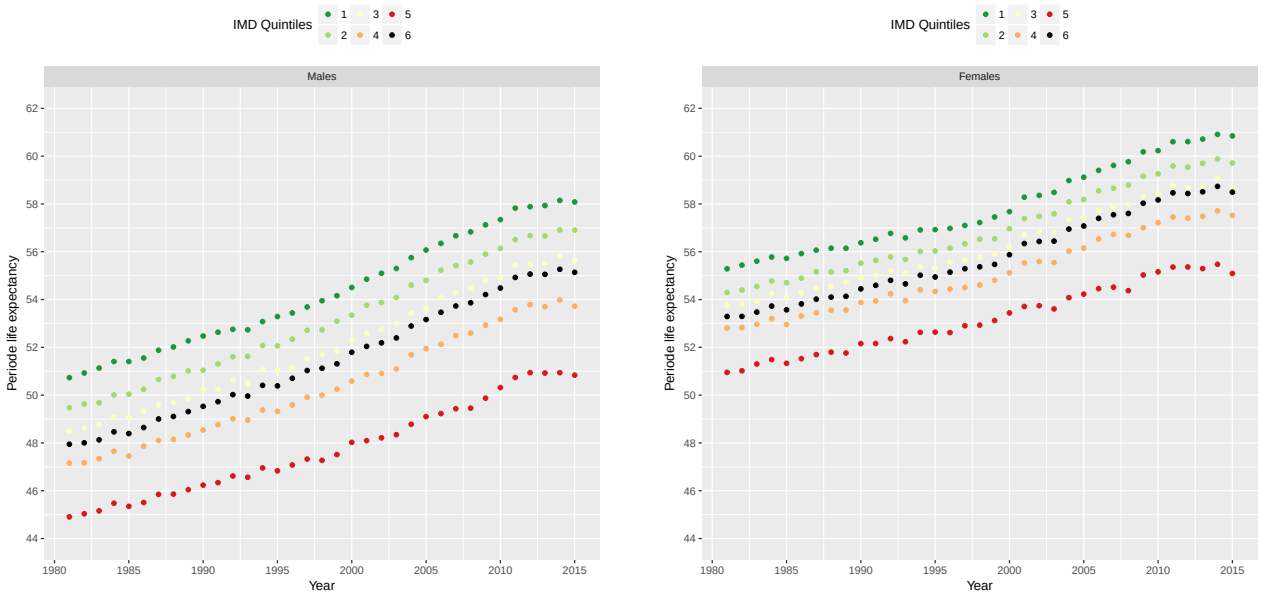
Fig. B.5 Deaths per cause and IMD quintile for females of age in 45-65



(a) 1981 (b) 2015

Fig. B.6 Deaths per cause and IMD quintile for males of age in 45-65

B.5 Life Expectancy



(a) Males (b) Females

Fig. B.7 Life expectancy at age 25 over 1981-2015

Appendix C

Numerical implementation and inputs

C.1 Fitting of death rates

In this paragraph, we consider the mortality of individuals of gender ϵ in a given IMD quintile j. For simplicity of notations, we omit these variables when there is no ambiguity. In both databases, deaths are given for each calendar year by 5 years age-classes (with the exception of the age class 0-5 year divided in two classes, 0-1 and 1-5). As explained in Section 4.2.2, central death rates at age a and year t for 5 five years age classes can been estimated by ${}_5\hat{m}(a,t) = \frac{{}_5D(a,t)}{{}_5\hat{E}(a,t)}$; where ${}_5D(a,t)$ is the number of individuals who died during year t (in $[t, t+1[$) at an age in $[a, a+5[$, and where ${}_5\hat{E}(a,t)$ is the estimated exposure. Recall that the real exposure ${}_5E(a,t)$ is the cumulative time lived during year t by individuals in the age class[1].

In our model, we need to estimate the force of mortality $\mu(a,t)$ which is linked to the theoretical death rate by the following formula:

$$ {}_5m(a,t) = \int_t^{t+1} \int_a^{a+5} \mu(x,s) \frac{g(x,s)}{\int_t^{t+1} \int_a^{a+5} g(u,h) \mathrm{d}u \mathrm{d}h} \mathrm{d}x \mathrm{d}s \tag{C.1.1} $$

Equation (C.1.1) can be interpreted as follow: the central death rate is the average force of mortality on $[t, t+1[\times[x, x+5[$, weighted against the population distribution on this interval.

In the sequel, we make assumption that the force of mortality is constant over 1 year

[1] An individual attaining age a at time $t+s$ and who died at time $t+h+s < t+1$ will weight as h in the exposure.

periods, so that for all calendar years t, $\mu(a, t+s) = \mu(a,t) \quad \forall s \in [0,1[$. In this case, Equation C.1.1 can be rewritten as:

$$ {}_5m(a,t) = \int_a^{a+5} \mu(x,t) \frac{\int_t^{t+1} g(x,s)ds}{\int_a^{a+5}(\int_t^{t+1} g(u,s)du)ds} \mathrm{d}x $$

When data is structured by single year of age, the force of mortality is usually also assumed to stay constant over age class, so that $\mu(x,t) = {}_1m(a,t)$ for $a \leq x < a+1$. However, this assumption seems quite unrealistic when data is aggregated over 5 years classes. The next simpler parametric assumption is to assume that the force of mortality is piecewise linear in age over the age classes:

$$ \mu(x,s) = \alpha(a,t)x + \beta(a,t) \quad \forall (x,s) \in [a, a+5[\times[t, t+1[. \tag{C.1.2} $$

When information on the population by single years of age is available, the distribution of the population on the age class can be approximated by a discrete distribution defined for $0 \leq k \leq 4$ by:

$$ \frac{\int_t^{t+1} g(x,s)ds}{\int_a^{a+5} \left(\int_t^{t+1} g(u,s)du\right) ds} = \hat{E}_k, \quad \forall x \in [a+k, a+k+1[, $$

with $\hat{E}_k$ is the estimated proportion of individuals of age in $[a+k, a+k+1[$ in the the age class. By replacing $\mu(a,t)$ and $g(x,s)$ in C.1.1 with the new assumptions, we obtain that Equation C.1.2 should be a line passing through ${}_5m(x,t)$ at the mean age of individuals in the age class, that is $\bar{x} = \sum_{k=0}^{4}(x+k+0.5)\hat{w}_k$.

From these calculation, an inductive method can thus be defined in order to fit the force of mortality of year t:

(i) *Initialization:* choose $\mu(0,t)$.

(ii) *Induction:* Assume that the mortality rate has been fitted until the ith age class $[a_i, a_{i+1}[$. The mortality rate on the next age class $[a_{i+1}, a_{i+2}[$ is the line passing through $\mu(a_{i+1}, t)^-$ at a_{i+1} and ${}_5m(x_{i+1}, t)$ at $\bar{x}_{i+1}$.

(iii) Reiterate the second step on the next age class...

The main advantage of the piecewise linear approximation is to be consistent with the theoretical definition of the aggregated and specific central death rates in our heterogeneous population model. However, the degree of liberty in the choice of the initial point $\mu(0,t)$ is a drawback of the method, and the fitting does not guarantee to obtain positive death rates. In the numerical applications the initial point $\mu(0,t)$ by an optimization. A similar approach have been proposed by Hautphenne and Latouche

(2012), with possible discontinuities in the death rates. See also Villegas and Haberman (2014) for an alternative method.

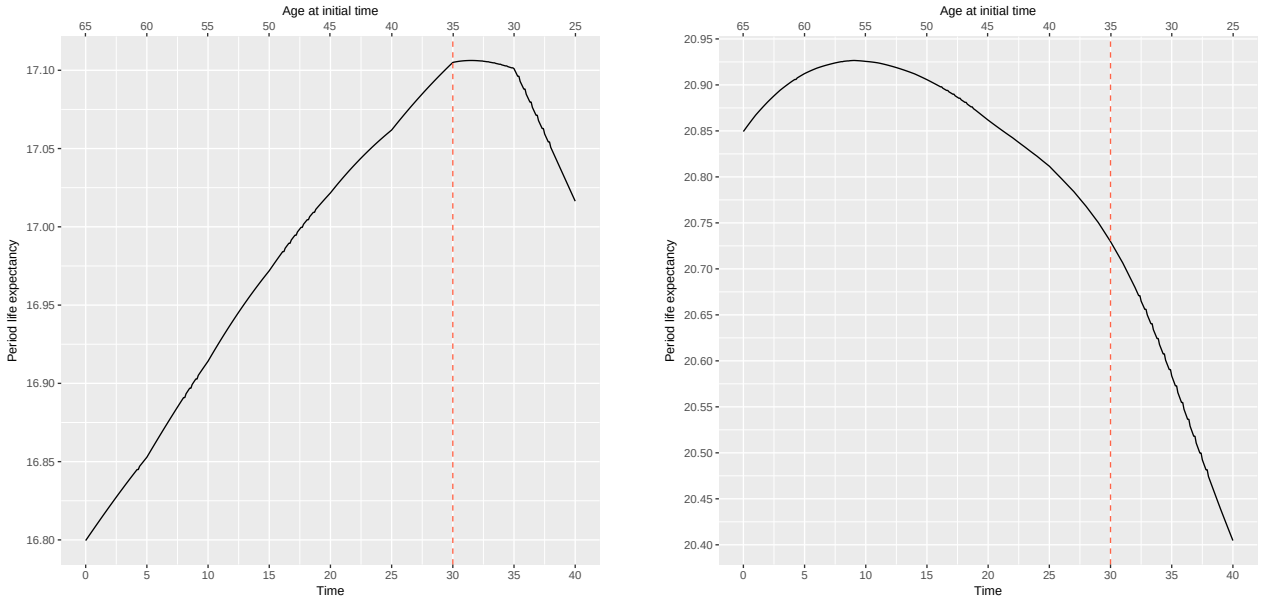
C.2 Implementation of the model

The numerical scheme corresponding to the heterogeneous population dynamics presented in Section 4.3 is implement in C++.
The inputs (demographic rates and initial age pyramids) and outputs of the model are processes using R, and interfaced with the C++ code using the package Rcpp. All graphs have been made using the R package ggplot2.

Appendix D

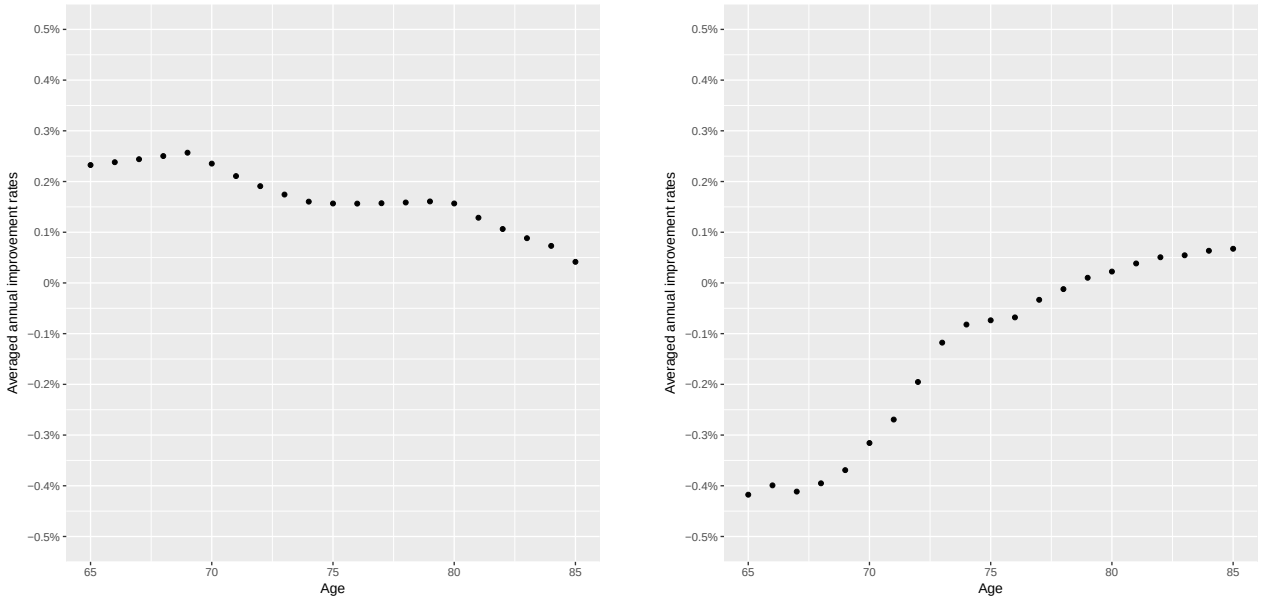
Additional results

D.1 Short term population dynamics (females)



(a) 1981 data (b) 2015 data

Fig. D.1 Evolution of female life expectancy at age 65



(a) 1981 data (b) 2015 data

Fig. D.2 Females averaged annual improvement rates

D.2 Long-Term population dynamics (females)

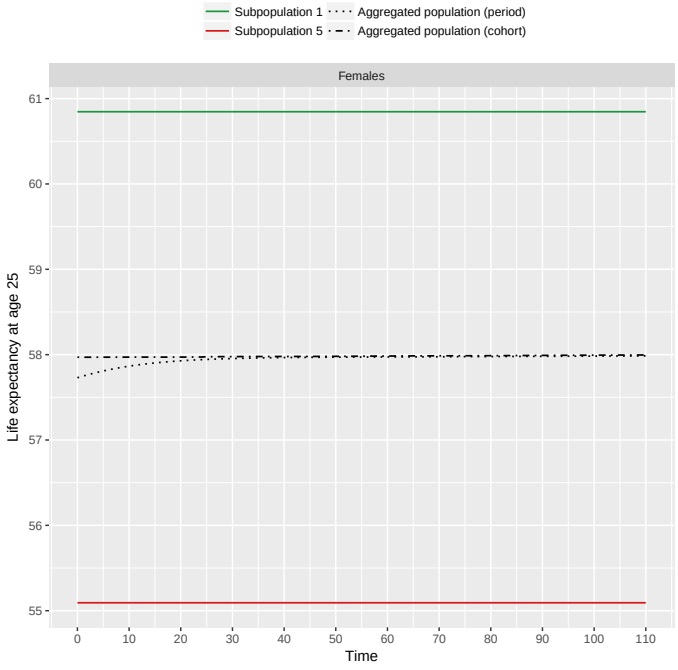
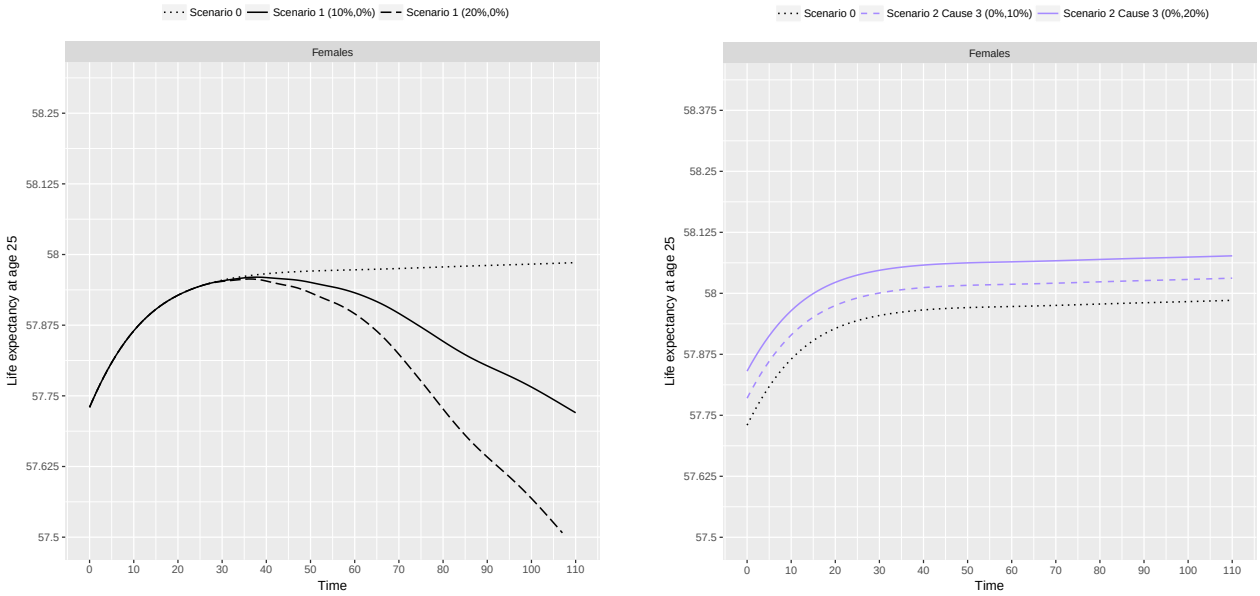
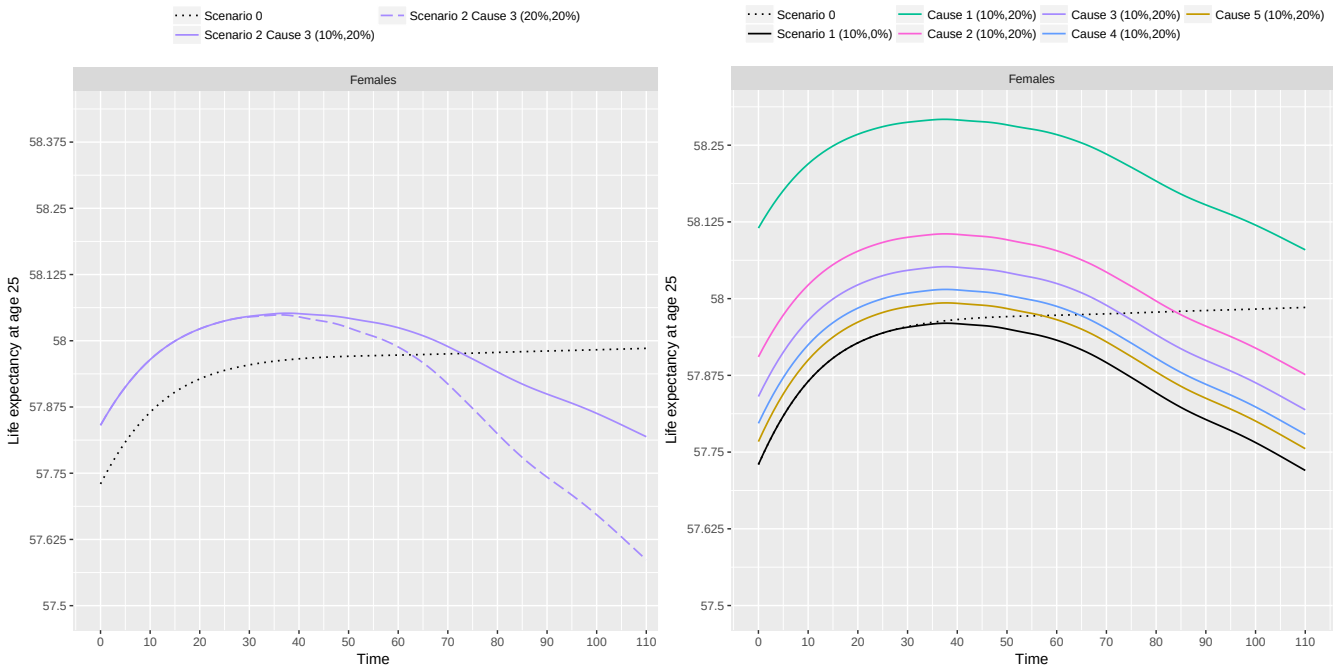


Fig. D.3 Aggregated period life expectancy over time



(a) Birth changes (b) Mortality changes

Fig. D.4 Aggregated life expectancy over time



(a) Cause 3 (b) All causes

Fig. D.5 Aggregated period life expectancy over time with birth and mortality changes

Chapter 5

Inextricable complexity of two centuries of worldwide demographic transition: a fascinating modeling challenge

5.1 Introduction

The world population has been transformed dramatically over the past two centuries. These transformations, or rather revolutions, have occurred on multiple levels: demographic, economic, technological, medical, epidemiological, political or social to name but a few. The ageing of the developed world has brought on an unprecedented situation, and the complexity of involved phenomena makes the projection of future developments very difficult.

In the recent years, a considerable amount of data have been collected at different levels. For instance, a number of international organizations and national statistical institutes have their own open databases. Numerous empirical studies and reports are published each year by academics in various domains, but also by governments and international organizations. The private sector is also very active on these issues, especially pension funds and insurance companies which are strongly exposed to the increase in life expectancy at older ages.

The past few years have also been marked by a renewed demand for new demographic models, motivated by recent observations which seem to be in contradiction with some firmly established ideas. New available data seem to indicate a paradigm shift - or second

demographic transition - toward a more complex and individualized world. Countries which had similar mortality experiences until the 1980s now diverge, and a widening of health and mortality gaps inside countries has been reported by a large number of studies. These new trends have been declared as key public issues by several organizations, including the WHO in its latest World report on ageing and health (World Health Organization (2015)), and the National Institute on Aging in the United States which created in 2008 a panel on Understanding Divergent Trends in Longevity in High-Income Countries, leading to the publication of a comprehensive report in National Research Council and Committee on Population and others (2011).

This cross-disciplinary survey aims at helping a modeler of human population dynamics to find a coherent way (for instance by taking into account the whole population dynamics and not only old ages) around this mass of multidisciplinary information. In front of the considerable body of literature, data and contradictory points of views on the evolution of human longevity, an cross-disciplinary approach appear to be necessary in order to avoid the pitfalls of an overly naive approach. Our goal is to offer a subjective selection, based on numerous surveys from various academic disciplines, of what we believe to be the most important ideas or facts from a mathematical modeling perspective.

As we will not be able to devote the necessary time to each point, we try to illustrate some of our points with examples that will support the intuition about mentioned phenomena. It should be emphasized from the start that if the discussion is greatly enriched by the multidisciplinary nature of the field, the presentation of ideas is also made more difficult, especially for matters of vocabulary. It should also be noted that issues related to medical advances and to the biology of human ageing are dealt with in a very cursory way, as we focus mainly on economic and social issues.

The survey is composed of three main parts which are summarized in the next subsection. The first part deals with the historic demographic transition. The importance of public health is dealt with, with a specific focus on the cholera epidemic outbreaks that took place in France and in the UK during the nineteenth century. Other features of the historic demographic transition are also considered. In particular, we explore the relationship between the economic growth and mortality improvements experienced during the past century.

In the second part of the survey, we examine the implications for population modeling and the key features of this shift in paradigm that have been observed since the 2000s. We first give a brief overview of the so-called demographic transition, and the move toward the description of increased complexity and diverging trends that have been recently observed, based principally on the experience of developed countries. A special

attention is paid to socioeconomic differences in health and mortality.
In the last part of the survey, we give a short review of microsimulation models and agent based models widely used in social sciences, and in particular in demographic applications. We first describe the main components of a dynamic microsimulation exercise to study heterogeneous individual trajectories in order to obtain macro outcomes by aggregation in the form of a data-driven complex model. Then, we present the agent based models which take into account individual interactions for explaining macroscopic regularities.

5.2 The historic demographic transition

Since the nineteenth century, most countries have experienced a remarkable evolution of their populations, referenced by demographers and economists as the *demographic transition* (Bongaarts (2009)). The historic demographic transition of the developed world[1] began in the nineteenth century and was completed in over a century ($\sim (1850 - 1960)$). This historical process is mainly referred to as "the secular shift in fertility and mortality from high and sharply fluctuating levels to low and relatively stable ones" (Lee and Reher (2011)). These substantial demographic changes caused life expectancy at birth to grow by more than 40 years over the last 150 ans (for instance, life expectancy at birth rose in the United Kingdom and France from about 40 years in the 1870s [2], to respectively 81 and over 82 years in 2013[3]), and the world population to grow from around 1 billion in 1800 to 2.5 billions by 1950 (Bloom and Luca (2016)). The treatment of infectious diseases constituted the vast bulk of the causes that explain the historic fall in mortality. For example, infectious diseases had virtually disappeared by 1971 in England and Wales while they were responsible for 60 percent of deaths in England and Wales in 1848 (Cutler et al. (2006)). The causes of this reduction have been extensively debated. Among the main causes that have been put forward are economic growth, improvement in living standards, education and most importantly social and public health measures (Bloom and Luca (2016); Cutler et al. (2006)). For instance, Cutler and Miller (2005) estimated that the purification of water explained half of the mortality reduction in the US in the first third of the twentieth century.

[1] The historic demographic transition affected most of European countries and countries with European roots (Argentina, Uruguay, the United States, Canada, New Zealand (Reher (2011)).
[2] Cambois et al. (2009); Cutler et al. (2006).
[3] Source: World Bank, (https://data.worldbank.org/indicator/SP.DYN.LE00.IN).

5.2.1 The cholera pandemic: a starting point of the demographic transition

In order to understand the unprecedented rise of life expectancy during the first half of the twentieth century, one has to go back 50 years which foreshadow the demographic transition. At the beginning of the nineteenth century, the Industrial Revolution led to a total upheaval of society, associated with unbridled urban sprawl and unsanitary living conditions. In Paris, the population doubled from 1800 to 1850 to attain over one million inhabitants (Jardin and Tudesq (1983)), while London grew by 2.5 fold during those 50 years, to attain more than 2 million inhabitants (Chalklin (2001)). In this context, epidemics were frequent and deadly.

The cholera pandemic, which struck fear and left indelible marks of blue-black dying faces due to cyanose' on the collective imagination (hence the nickname "blue death"), had the most important social and economic consequences. It is often refered to as an iconic example where medicine was confronted to statistics (Dupaquier and Lewes (1989)) and was regarded as "the real spark which lit the tinder of the budding philanthropic movement, culminating in the social reforms and the foundation of the official public health movement seventeen years later" (Underwood (1948)). The cholera pandemics originated in India and spread to Europe in the 1830s. Four subsequent outbreaks (1831, $1848-1854$, $1866-1867$ and $1888-1889$) mainly affected France and England, causing 102.000 deaths in France in 1832 and 143.000 in the $1850s$ over a population of 36 million (Haupt and Laroche (1993)). In London, 6536 deaths were reported in 1831 and 14137 deaths during the 1848-1849 cholera outbreak (Underwood (1948)).

In the following paragraphs, we will focus on the cholera outbreaks in France and England, in order to illustrate the profound changes which occured at different levels (city, state and international), and which still give valuable insight on contemporary challenges.

Cholera in England

The intensity of the first cholera outbreak in London in 1831, combined with the growing influence of advocates of public health, brought to light the need for public measures to improve sanitation. At that time, a lot of reformers considered that statistics were a prerequisite for any intervention, and the enthusiasm in the field expanded very quickly, which is somehow reminiscent of the current craze for data science.

In this context, the General Register Office (G.R.O) was created in 1836, with the aim of centralizing vital statistics. England and Wales were divided in 2193 registration sub districts, administered by qualified registrars (often doctors). In charge of compiling

data from registration districts, W. Farr served as statistical superintendent from 1839 to 1880 and became "the architect of England's national system of vital statistics" (Eyler (1973)). The precise mortality data collected by the G.R.O during cholera outbreaks turned out to be instrumental in the analysis of the disease.
In his pioneering *Report on the Mortality of Cholera in England, 1848-49* (Farr (1852)), Farr and the G.R.O produced almost four hundred pages of statistics. His main finding, based on the collected data of the 1848-49 outbreak, was the existence of an inverse relationship between cholera mortality rates and the elevation of registration districts above the Thames. Farr was particularly pleased with this statistical law, since it validated his beliefs in the prevailing miasmatic theories, which predicted that the passing on of the disease was airborne. It was actually J. Snow who first claimed that cholera communication was waterborne, with his famous experiment of the Broad street pump (Brody et al. (2000)). However, Farr's statistics were decisive in supporting and validating his theory.
Although Snow's theory was not widely accepted, he contributed to raising the issue of water quality. Under the impulsion of the General Board of Health created in 1848, the Metropolis Water Act of 1852 introduced for the first time regulations for (private) water supply companies, to take effect by 1855. At the time of the 1853-54 cholera outbreak, Farr found out that only one company had complied with the new regulations, and that in a number of districts, it was competing with another company drawing water from a highly polluted area. The perfect conditions for a full-scale experiment were brought together, and Farr and Snow joined their investigations to conclude that without doubt, water played an important role in the communication of the disease. In 1866, a smaller outbreak hit London. More specifically, the reintroduction of sewage contaminated water by the *East London Water Company* caused in just one week 908 over 5596 deaths in London (Dupaquier and Lewes (1989),Underwood (1948)) . Despite the overwhelming amount of evidence, the Medical Officer himself tried to exonerate the company, causing the wrath of Farr. This event, however, was a wake-up call for the English political class to guarantee the supply of clean water. Several public health measures were taken from the 1850s in order to improve public health and water quality. A new administrative network was established in London in 1855, which undertook the development of the city's main drainage system which was completed in 1875. Among other measures were the Rivers Pollution act in 1876 and the carrying out of monthly water reports from the 1860s (Hardy (1984)).

Cholera in France

France's experience with cholera varied from England's, due to its different scientific environment and unstable political situation. The first epidemic reached Paris by the spring of 1832 causing, in four months, the death of almost 2.1% of the 774.338 Parisians (Paillar (1832)).

In his remarkable report addressed to the Higher Council of Health (Moreau de Jones (1831) [4]), the former military A. Moreau de Jonnès (1778 - 1870) gave considerable details on the international spread along trade routes of the pandemic that started in India in 1817, including the treatments and precautions taken against the disease. He clearly attested that cholera was "incontestably" contagious. In 1833, he became the first chief of the Statistique Generale de la France (SGF), the nearest equivalent to the G.R.O in England. Like Farr, Moreau de Jonnès published many reports (13 volumes) and contributed to the development of Statistics and its applications in France. Unfortunately, little attention was paid to his findings by the French health care community.

In the end of 1831, France was anticipating a cholera epidemic. Health commissions were established in Paris and in other departments in order to control the disease with the help of health councils (conseils de salubrité); but the organization was less systematic and data collection was less reliable than in England. During the first epidemic, social unrest among the lower classes, who saw the disease "as a massive assassination plot by doctors in the service of the state", were the worst fears of the government (Kudlick (1996)). The government was supported by the Faculty of medicine in its efforts to reduce fear and avoid a population uprising, and the latter stated in 1832 that the disease was not communicable (Fabre (1993)). In 1848, a public health advisory committee was created and attached to the Ministry of Agriculture and Commerce, in charge of sanitary issues (housing, water and protection of workers) and prophylactic measures to prevent the epidemic from spreading Le Mée (1998). As in 1832, this committee stated that cholera was not contagious (Dupaquier and Lewes (1989)).

In 1849, the second epidemic broke out after the 1848 revolution. Contrary to the first epidemic characterized by riots and tensions, the reaction to the outbreak was more peaceful, with more efficient collaboration between scientists and the administration. At the same time, the perception of the lower classes also changed with the idea of struggling against destitution in order to prevent revolt (Kudlick (1996)). As a consequence, the response to the second epidemic was better organized and social laws were passed in $1850 - 51$. In the following years, hygiene problems and unsanitary living conditions caused by the rapid growth of Paris's population were addressed to by important public

[4] The report is available on the website of BNF.

health measures. In particular, the massive public work projects led by Baron Haussman [5] in less than two decades from 1852 to 1870 remains as a symbol of the modernization of Paris at the end of the nineteenth century (Raux (2014)).

Cholera Pandemic and International Health Organization

The international dimension of the problem raised by cholera, reported in France by Moreau de Jonnès in 1824-31, was widely publicized by The Lancet, which published in 1831 a map on the international progress of cholera [6] (Koch (2014)). This map suggested a relation between human travel and the communication of the disease, accelerated by the industrial revolution in transport, in particular with steamships and railways. Cholera was regarded as an issue transcending national boundaries, which needed international cooperation to control it (Huber (2006)). Europe had succeeded in setting up an efficient protective system against the plague, based on ideas such as quarantine and "cordon sanitaire". But those measures were very restrictive and seemed inefficient against cholera. Moreover, in the second half of the nineteenth century, Western European countries were involved in competitive colonial expansion, and were rather hostile to travel restrictions, even if increased global circulation was a threat to populations. The opening of the Suez Canal in 1869 was an emblematic example of those changes.

Under the influence of French hygienist doctors, the first International Sanitary Conference opened in Paris in 1851 (Huber (2006)) gathering European states and Turkey. It was the first international cooperation on the control of global risk to human health, and so the beginning of international health diplomacy. It took more than ten international conferences over a period of over 50 years to produce tangible results. During the first five conferences [7], the absence of clear scientific explanation on the origin of cholera prevented any agreement. It was only with the formal identification of the *V. cholerae* bacterium by R. Koch in 1883[8] and the work of L. Pasteur that infectious diseases were clearly identified and efficiently fought against. Indeed, technological progress as evinced through disinfection machines could allow the technological implementation of new measures (Huber (2006)). Furthermore, advances on germ theory “allowed diplomats to shape better informed policies and rules” (Fidler (2001)).

At the Seventh Conference (1892), the first maritime regulation treaty was adopted for ship traveling via the Suez Canal. During the ninth conference (1894), sanitary

[5] Napoléon III appointed Baron Haussmann as Préfet de la Seine

[6] The map was completed in 1832 by Brigham to include Canada and the USA

[7] (1851,1859,1866, 1874 and 1881)

[8] The bacterium had been isolated before by other scientists such as F.Pacini in 1854, but his work did not had a wide diffusion.

precautions were taken for pilgrims traveling to Mecca. Participants finally agreed that cholera was a waterborne disease in 1903 during the eleventh conference.
The International Sanitary Conferences provided a forum for medical administrators and researchers to discuss not only on cholera but also on other communicable diseases, and brought about the first treaties and rules for international health governance. Ultimately, this spirit of international cooperation gave birth in 1948 to the World Health Organization, an agency of the United Nations, conceived to direct and coordinate intergovernmental health activities.

Discussion

In England, Farr's discoveries could not have been made without the cutting edge organization and the power of the G.R.O. It is worth noting that only a governmental organization such as the G.R.O was able to collect the data fast enough for the 1854 experiment of Farr and Snow (Dupaquier and Lewes (1989)) to be possible. The modern organization of the G.R.O undoubtedly contributed to the remarkable quality of today's England vital databases. Across the Channel, France did not manage to create the same kind of centralized authority. On the grounds of their hostility to the communicable disease theory, French doctors did not rely on statistics.
On the other hand, the use of statistics made by Farr contributed to a better understanding of the disease. It was only more than a century and a half later that a major breakthrough was made in the understanding of the origins of the disease, with the work of R. Colwell showing that the *V. cholerae* bacterium appears naturally in the environment. Yet the ambition to find causal factors by the sole analysis of data is not devoid of risks, and thus constitutes a major challenge for the data science era. Farr's elevation law is a textbook case of an unexpected correlation that turns out to have a great influence. Despite claimed impartiality, his choice to highlight the elevation law among all the findings mentioned in his report on the 1848-49 outbreak was clearly biased by his beliefs in the predominant (though false) miasma theory. While he later accepted that epidemics could be waterborne, Farr continued to believe in the prevailing role of elevation, even when deaths due to cholera during the 1854 and 1866 outbreak were not consistent with the elevation law. Rather than allowing the discovery of the causes of cholera, Farr's statistics were actually more useful for testing and validating the relationships predicted by Snow's theory.
Another point is that the conditions that made the 1854 experiment possible were quite extraordinary. Testing theories regarding the complex events of health and mortality in human communities is often nearly impossible. Only a handful of studies can take advan-

tage of natural experiments. More often than not, as stated in National Research Council and Committee on Population and others (2011), “they are limited ethical opportunities to use randomized controlled trials to study the question at issue". Furthermore, governments failed to come to an agreement during the first international conferences because of the lack of scientific explanations on the origin of cholera. The need of theoretical arguments for public decisions to be made is still an important issue, especially when considering human health and longevity, for which no biological or medical consensus has emerged. As will be developed further in this survey, the use of a mathematical model and simulations can operate as a proxy to real life experiments and help decision making. Even when theories are publicized, there are often important delays (one or two generations) before action is taken. For instance, even if Snow's theory was better known in 1866, and despite the development of germ theory in the early 1880s, political divergences prevented any action before 1892. The example of asbestos, which took 50 years to be banned after the exhibition of its link with cancer, shows us that these delays in public response did not diminish over time (Cicolella (2010)). More generally speaking, around 30 years elapsed between the first epidemic and the real development of public health policies in England and in France.

The example of cholera illustrates the complexity of studying mortality evolution, inseparable from societal and political changes. Although cholera outbreaks occurred at about the same time in France and in England, they were experienced very differently owing to the different political and scientific climates in both countries. This shows that the sole study of mortality data could not be sufficient to understand the future trends of mortality. In particular, the explosion of the London population, whose size was twice as large as that of Paris, brought about social problems on a much greater scale, which played a determining role as a catalyst of public health changes.

The cholera outbreaks contributed to the development of important public health measures, which played a major role in the reduction of infectious diseases. For instance, Cutler and Miller (2005) estimated that water purification explained half of the mortality decline in the United States between 1900 and 1930. In comparison, the discoveries of new vaccines for a number of diseases at the beginning of the twentieth century seem to have had little impact on the reduction of mortality from those diseases. For instance, the reduction in mortality due to those diseases (except tuberculosis) following the introduction of those vaccines is estimated to have contributed to the emergence of only 3 percent of total mortality reduction (Cutler et al. (2006)). The second half of the twentieth century was marked by the rise of more intensive medical interventions, and by an epidemiological transition from infectious diseases to chronic diseases. But public

health should not be underestimated in this new age of "degenerative and man-made diseases" (Bongaarts (2014)). Public health campaigns against tobacco have played an important role in reducing cardiovascular disease mortality caused by smoking (Cutler et al. (2006)), although with varying degrees of success depending on countries, gender or social classes. The increase of environmental risks constitutes one of the major challenges faced by contemporary societies, and public action will play a central role in preventing and successfully reducing those risks (Cicolella (2010)).

5.2.2 A century of economic growth

The twentieth century was the century of "the emergence for the first time in history of sustained increases in income per head" (Canning (2011)), and the association of economic growth and mortality improvements have been extensively discussed by economists. During the nineteenth century, individuals in rich and poor countries experienced similar health conditions. The 1870s were a turning point with the improvement of health in rich countries (Bloom and Canning (2007)). In his seminal article, Preston (1975) was one of the first economists to examine the relationship between life expectancy at birth and national income per head in different countries[9], for three different decades: the 1900s, 1930s and 1960s (see Figure 5.1). In each decade, Preston brought to light a strong positive association between life expectancy and national income. He also stated that the relationship was curvilinear. For instance, the so-called Preston curve of 1960 appeared "to be steeper at incomes under 400$ and flatter at incomes over 600$" Preston (1975). Preston also noted an upward shift of the curve characterized by a rise of life expectancy over time at all income levels. These empirical results showed that economic growth alone did not explain the remarkable mortality decline. For instance, the income level corresponding to a life expectancy of 60 was about three time higher in 1930 than in 1960. Another example is China which had in 2000 the same income level as the USA in the 1880s, but the life expectancy level of the USA in 1970. Preston (1975) estimated that national income accounted for only 10 to 25 percent of the growth of life expectancy between 1930s and 1960s. Bloom and Canning (2007) also estimated that increases in income between 1938 and 1963 were responsible for about 20% of the increase in the global life expectancy.

[9]The national income per head was converted in 1963 U.S dollars.

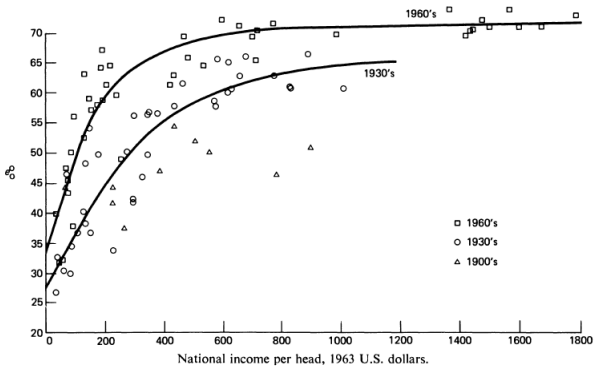


Fig. 5.1 Preston curves, 1900, 1930, 1960, reproduced from Preston (1975)

Deaton (2003) represented the Preston curve in 2000 (see Figure 5.2), in which countries are represented by circles that are proportional to the size of population. The Preston curve in 2000 (see Figure 5.2) shows that correlation of income with life expectancy is more tenuous for high income countries.

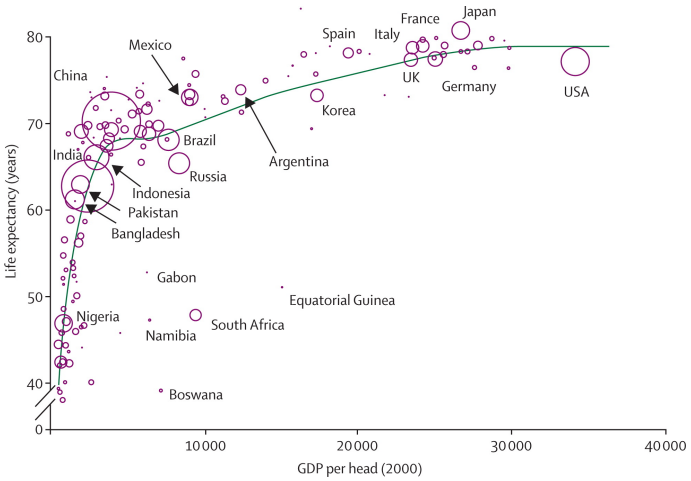


Fig. 5.2 Preston curve in 2000, reproduced from Deaton (2003)

Furthermore, the relationship between mortality decline and level of income is often thought as bidirectional. This issue still generates a lot of debate (see e.g. Acemoglu and Johnson (2007, 2014); Bloom and Canning (2007); Bloom et al. (2014); Cutler et al. (2006)). On the one hand, some studies on the causal link between health and wealth suggested that "health can be a powerful instrument of economic development" (Bloom and Canning (2007)). On the opposite side, Acemoglu and Johnson (2007) argue that improvements in population health, especially the reduction of children mortality, might have negative impacts on economic growth, due to the increase in the population size.

They argue that a positive effect of economic growth on health may be counterbalanced by the negative effect of population growth on health. However, Reher (2011) describes the increase in the proportion of working age poeple in the population that occurred in developed countries during the twentieth century as a situation which had "profound economic implications for society, as long as the economy was able to generate enough jobs to accommodate the growing population of working age".

For a more complete picture, it is thus interesting to go beyond "macro" environmental indicators such as public health and economic growth, and to look at mortality experiences on different scales, by exploring differences between countries, and within countries.

5.3 A new era of diverging trends

5.3.1 A second demographic transition?

In the early 1970s, many demographers and population scientists had supported for the idea that populations would ultimately reach the last stage of the classical demographic transition, described as an "older stationary population corresponding with replacement fertility (i.e., just over two children on average), zero population growth, and life expectancies higher than 70y" (Lesthaeghe (2014)). More generally speaking, populations were supposed to attain an equilibrium state, characterized by a significant level of homogeneity. For instance, the nuclear family composed of a married couple and their children was expected to become the predominant family model.
Yet, in most countries which experienced the historic transition, the baby boom of the 1960s[10] was characterized by higher fertility rates, followed by a decline in fertility in the 1970s (baby bust). In response to these fluctuations, attempts were made to modify the original theory. For instance, Easterlin (1980) developed a cyclical fertility theory, linking fertility rates to labor-market conditions. Smaller cohorts would benefit from better living conditions when entering the labor market, leading to earlier marriage and higher fertility rates. On the contrary, larger cohorts would experience worse living conditions, leading to later marriage and lower fertility rates.
However, it turned out that this state of equilibrium and homogeneity in populations was never realized. Actually, fertility rates remained too low to ensure the replacement of generations; mortality rates, especially at advanced ages, declined at a faster rate than ever envisaged; and contemporary societies seem to be defined by more and more

[10]The baby boom affected several countries such as France, the United Kingdom or the United States, although with different timings from the early 1950s to 1970s.

heterogeneity and diverging trends. The idea of a renewed or second demographic transition, distinct from the classical demographic transition, was originally formulated by Lesthaeghe and Van de Kaa (1986) in an article in Dutch, followed by a series of articles (Lesthaeghe (2010, 2014); Van de Kaa (2010)).

In the early 1980s, a number of researchers had already observed that a shift of paradigm (Van de Kaa (2010)) had occured. In particular, the French historian P. Ariès suggested that motivations explaining the decline birth rate in the West had changed (Ariès (1980)). During the historic demographic transition, the decline in fertility rates was assumed to originate from an increased parental investment in the child. P. Ariès explained more recent declines in fertility by an increasing interest of individuals in self-realization in which parenthood is only one particular life course choice among many others. More specifically, Lesthaeghe (2014) characterized the second demographic transition by multiple lifestyle choices and a more flexible life course organization. A striking example can be found in the emerging of multiple types of family arrangements.

Lesthaeghe and Van de Kaa also define the second demographic transition as a shift in the value system. The first phase of the demographic transition was a period of economic growth and aspirations to better material living conditions. In contrast, the past few years have seen a rise of "higher order" needs and individualization. In this new paradigm, individuals are overwhelmingly preoccupied by individual autonomy, self-realization and personal freedom of choice, resulting in the creation of a more heterogeneous world.

Even if the framework of the second demographic transition has been criticized, this viewpoint shed an interesting light on recent longevity trends. Indeed, divergences in mortality levels and improvements between and within high income countries are at the heart of numerous debates and research works. As the average life expectancy has been rising unprecedentedly, gaps have also been widening at several scales. What may be somehow surprising is that up until the 1980s, high income countries had roughly similar life expectancy levels. For example, the comparison of the female life expectancy at age 50 in ten high income countries[11] shows that the gap was of less than one year in 1980. By 2007, the gap had risen to more than 5 years, with the United States at the bottom of the panel with Denmark, more than 2 years behind Australia, France, Italy and Japan (National Research Council and Committee on Population and others (2011)). On another scale, a great amount of evidence shows that socioeconomic differentials have also widened within high income countries. For instance, the gap in male life expectancy at age 65 between higher managerial and professional occupations and routine occupations

[11] Australia, Canada, Denmark, England and Wales, France, Italy, Japan, Netherlands, Sweden and the United States.

in England and Wales was of 2.4 years in 1982 to 1986, and rose to 3.9 years in 2007 to 2011[12].

The following part focuses on two angles of analysis on these diverging trends: the impact of smoking behaviors and socioeconomic inequalities. The goal of the following discussion is not to detail further the impact of these risk factors, but rather to show the complexity of understanding current longevity trends, which cannot be disentangled from the evolution of the whole population, and which require a multiscale analysis of phenomena while keeping in mind that obtaining comparable and unbiased data is also a challenge in order to explain longevity.

5.3.2 Diverging trends between high-income countries: the impact of smoking behaviors.

In the comprehensive report of the National Research Council on explaining divergent levels of longevity in high income countries (National Research Council and Committee on Population and others (2011), a panel of experts have debated on the role of different risk factors for explaining the slower increase of life expectancy in the United States over the last 30 years, in comparison with other high income countries. From 1980 to 2015, the world ranking for life expectancy of the United States kept falling significantly. Furthermore, the gap between the United States and other high income countries widened, due to the slower increase of life expectancy at all ages in the United States. The ranking of the United States for male life expectancy at age 50 fell from 17th in 1980-85 to 28th in 2010-2015, with an increase of life expectancy of 4.58 years, smaller in absolute and relative terms than the average of high income countries[13]. The evolution is even more striking for women: the ranking for the female life expectancy at age 50 fell from the 13th to the 31th position, with an increase of only about 60 percent of the average increase of high income countries. In addition, the gap with higher achieving countries such as France or Japan grew from less than one year in 1980-85 to more than 3 years in 2010-15[14]. Netherlands and Denmark also show similar patterns of underachievement in life expectancy increases.

Although many methodological problems may arise when using cause-of-death statistics, a cause-of-death analysis can provide a powerful tool for understanding divergences in mortality trends. In a commissioned background article for the report, Glei et al. (2010)

[12]Source: Office for National Statistics (ONS).

[13]High income country classification based on 2014 GNI per capita from the World Bank.

[14] United Nations, Department of Economic and Social Affairs, Population Division (2015). World Population Prospects: The 2015 Revision, custom data acquired via website.

have studied cause-of-death patterns for 10 different countries in order to identify the main causes of death possibly responsible for diverging trends. The case of lung cancer or respiratory diseases, which are relevant indicators concerning smoking is particularly interesting. Age-standardized mortality rates from lung cancer among men aged 50 and older in the U.S decreased from 1980 to 2005 while they increased for women, although they remain higher for men than for women. In addition, the increase of age-standardized mortality rates due to lung cancer for women was much faster in the U.S, Denmark and Netherlands than the average increase of the studied countries and especially than Japan where age-standardized mortality rates remained flat.

These findings of Glei et al. (2010) clearly point out to smoking as the main underlying factor explaining those divergences. Over the past 30 years, evolution of mortality due to lung cancer and respiratory diseases has had a positive effect on gains in life expectancy for males, while the effect was negative for females. These gender differences can be linked to the fact that women began to smoke later than men, and have been quitting at a slower pace (Cutler et al. (2006)). In addition, fifty years ago, people smoked more intensively in the United States, Denmark and the Netherlands than in other European countries or in Japan.

These differences can give precious information as for future mortality patterns. Because of its delayed effects on mortality, the impact of smoking behaviors on future trends is somehow predictable. Just as the causes of death of individuals aged 50 and older give some insight on what happened in the past, current behaviors among younger individuals can be a useful indicator of future trends. Thus, life expectancy for males in the United States is likely to increase rather rapidly following reductions in the prevalence of smoking over the past twenty years, while slower life expectancy improvements can be expected for women in the coming years (National Research Council and Committee on Population and others (2011)). According to a panel of experts, life expectancy in Japan is also expected to increase at a slower pace in the future due to an increase in the prevalence of smoking. Differences in the timing of evolutions in smoking behaviors across gender and countries might also give additional information. The impact of smoking on male life expectancy in the past could help predict future trends for women, and the experience of the United States could shed light on the future impact of smoking in Japan.

But smoking is certainly not a sufficient explanation, and other risk factors may have contributed to the underachievement of the United States. In particular, the obesity epidemic may partly account for the slower increase in life expectancy experienced by the United States. Quantifying the impact of the obesity epidemic is much more complicated, since no clear markers are available such as lung cancer and respiratory

diseases concerning smoking. According to some researchers, the obesity epidemic in the United States might even offset gains in life expectancy due to the decline of smoking (Olshansky et al. (2005); Stewart et al. (2009)).

Discussion The "predictable" effects of smoking could be integrated in a population dynamic framework taking into account the whole age structure of the population. Countries experiencing similar phenomena but with different timings could also be compared in a theoretical framework of population dynamics. Furthermore, a finer-grained model could help to better understand the future impacts of emerging issues such as the obesity epidemic, as well as the potential compensating effect of a decrease in smoking prevalence.

5.3.3 Differences within countries: the impact of social inequalities

Research on the relationship between socioeconomic status and mortality and health can be traced back as far as the nineteenth century. In France, Villermé (1830) compared mortality rates in Paris' boroughs with the rates of non-taxable households in each borough (Mireaux (1962)). In England and Wales, a systematic documentation of mortality by occupational class was made by the G.R.O starting from 1851 (Elo (2009)). Since then, studies have consistently exhibited a pervasive effect of socioeconomic inequalities on longevity, regardless of the period or country. A recent study based on the French longitudinal survey has found out that males with managerial and higher professional occupations have a life expectancy 6.3 years higher than working-class males (in the mortality conditions of 2000-2008, Blanpain (2011)). Numerous other examples can be found in the review of Elo (2009).

Moreover, despite unprecedented rise in life expectancy during the 20th century, evidence shows that socioeconomic inequalities have widened in many developed countries in recent decades (Elo (2009)), or have remained identical at best (Blanpain (2011)). For instance, Meara et al. (2008) (cited in National Research Council and Committee on Population and others (2011)) argue that the educational gradient in life expectancy at age 25 rose from the eighties to the nineties of about 30 percent. Similarly in England, socioeconomic status measurements using the geographically based Index of Multiple Deprivation (IMD) (see next Chapter for more details) have shown that the average mortality improvement rates at age 65 and older have been about one percent higher in the least deprived quintile that in the most deprived quintile during the period 1982-2006

(Haberman et al. (2014), Lu et al. (2014)).
The persistence and widening of socioeconomic inequalities in longevity has created a new paradigm, in which the increased heterogeneity has brought out even more complexity in understanding longevity evolution, and which has now to be taken into account in mortality predictions. New interlinked problems have arisen on multiple scales. On an individual level, underlying factors linking individuals' health to their socioeconomic status are still debated. Another subject of no little interest to us is the critical challenge of understanding the impact of this rising heterogeneity on aggregated variables.
In the following part, we will focus on some selected topics which have been discussed by sociologists, demographers, social epidemiologists and other scientists, with the aim of highlighting modeling challenges and solutions hidden beyond these reflections.

Measuring the socioeconomic status

The concept of socioeconomic status (SES) is broad and can encompass numerous characteristics, observable or not. Translating socioeconomic status into empirical measurements in order to better understand the links between SES and health and mortality, is in itself a challenge. Proxy variables such as educational attainment, occupation, income or wealth usually serve as SES measures, with different practices and habits in different countries (Elo (2009)). However, their ability to model the complexity of the social hierarchy and to produce comparable data through different times and places are often quite limited. Educational systems, even in groups of similar countries, can differ substantially from one country to another and make cross-national comparisons difficult (Elo (2009)). Furthermore, there is a real difficulty in comparing certain groups at different periods in time. Important changes can occur in group sizes and composition. For instance, the proportion of women in France with higher managerial and professional occupations increased from about 2 percent in 1975 to 6 percent in 1999 (10 to 14 percent for males). The evolution in the number of women long term unemployed or not in the labor force is even more striking. Their proportion decreased from 45 percent in 1975 to only 21 percent in 1999. Besides, Blanpain (2011) observed an important widening of mortality inequalities between this subgroup and other occupational subgroups over the period. The widening of these gaps is actually a typical consequence of important changes in the composition of the long term unemployed or not in the labor force subgroup. The major decrease in the size of the subgroup can be explained by the important decrease of the number of housewives over time, leaving only the most precarious in the subgroup. Proxies for the SES can be measured at different periods in the life course of an individual, and can have different causal relations with health or mortality. Education is rather

consistent across the lifespan (which allows for an easier dynamic modeling), and permits to assess the stock of human capital accumulated early in life and available throughout the individual life course (Elo (2009)). On the other hand, occupation, income or wealth allow to take into account latter parts of the life course and might allow to capture impacts of public policies better than educational attainment measurements (National Research Council and Committee on Population and others (2011)). However, the variability of the occupational status through the life course and the difficulty of assigning an occupational group to individuals is important when studying socioeconomic gradients by occupational rankings. For instance, the issue of assigning an occupational group to individuals who are not in the labor force or retirees is classical. Interpretations can also differ significantly depending on the period in the life course at which occupational status is measured.

Additional complexity is also generated by the potential bidirectionality of causation, especially concerning economic measures of SES such as income or wealth, for which causal pathways are debated. Evidence from the economic literature has shown that ill health can also lead to a decrease in income or wealth. This is particularly true in countries like the United States with poorer national health care coverage than most Western Europe countries, and where poor health is a significant contributor to bankruptcy (Himmelstein et al. (2009)), retirement or unemployment (Smith (2007), cited in National Research Council and Committee on Population and others (2011), Case and Deaton (2005), cited in Cutler et al. (2006)).

Explaining the socioeconomic gradient in health and mortality

The difficulty in interpreting results of empirical measurements of the SES gradient in mortality reflects our little understanding of the risk factors that underlie the repercussions of socioeconomic inequalities on health and mortality. Theories explaining the SES gradient are still being debated, and their testing is often not straightforward and not unbiased as far as the measurements are concerned (see below the discussion on absolute versus relative measures). Furthermore, the impact of inequalities on aggregated variables or on the interpretation in terms of public policy can differ substantially according to different theories.

The mechanisms through which SES is assumed to generate inequalities in health and longevity are usually grouped in three broad categories: material, behavioral and psychosocial (Cutler et al. (2006)).

Material risk factors Maybe one of the most natural explanation of socioeconomic differences in health is that wealthier individuals have better access to health care, even

in countries with national health care coverage where potential two-tiered systems can also create inequalities. Individuals with a higher income can also maintain a healthier lifestyle, being able to buy expensive organic food or pay for gym memberships. However, access to health care or material resources does not appear to be the primary factor explaining the SES gradient (National Research Council and Committee on Population and others (2011),Cutler et al. (2006)). For instance, the education gradient in the U.S steepened between the sixties and the eighties, even though the Medicare program was enacted in 1965 (Pappas et al. (1993), cited in Elo (2009)).

Behavioral risk factors The second explanation is that individuals with higher educational attainment are more likely to adopt healthier behaviors and to avoid risks. By accumulating knowledge, skills and ressources, individuals who are higher on the SES ladder should be able to take better advantage of new health knowledge and technological innovations, as well as to turn more rapidly toward healthier behaviors. This behavioral explanation of socioeconomic inequalities is linked to the theory of Link and Phelan of fundamental causes (Link and Phelan (1995), Phelan et al. (2010)). The aim of Link and Phelan's theory is to explain the persistence of pervasive effects of social class inequalities through time, despite dramatic changes in diseases and risk factors. According to the theory, the accumulation among other resources of so-called human capital allows more educated individuals to use resources and develop better protective strategies, whenever they can and no matter what the risks are. Let us take the example of smoking, described in (Link (2008)). When first evidence linking smoking to lung cancer emerged in the fifties, smoking was not correlated to SES. But as the knowledge of the harm caused by smoking spread, strong inequalities in smoking behavior appeared, reflected in the fact that more educated individuals quit smoking earlier. However, a number of studies (see National Research Council and Committee on Population and others (2011) for examples) have shown that if behavioral differences play a significant role in explaining the SES gradient in mortality, it does not explain everything, and may not even account for the major part of the differentials. For instance, the famous study of Whitehall civil servants (Marmot (1994)) showed that health differentials subsisted even when factors such as smoking or drinking were controlled.

Psychosocial Factors Another prominent and rather recent theory explaining socioeconomic differentials in mortality is that health is impacted by the SES through pyschosocial factors (Cutler et al. (2006); Wilkinson and Pickett (2009)). Among pyschosocial factors are stress, anxiety, depression or anger. Accumulated exposure to stress has received particular attention in literature, due to its pervasive effects on health. Indeed, prolonged exposure to chronic stress affects multiple physiological systems by shifting priorities

from systems such as the immune, digestive or cardiovascular systems in favor of systems responding to threat or danger (Wilkinson and Pickett (2009)), and by leading to a state of so-called "allostatic load". The link between low social status and stress has been supported by a number of studies on primates. For instance, Sapolsky (2005, 2004) showed that among wild baboons, subordinate animals presented higher level of glucocorticoids, a hormone with a central role in stress response.

The impact on aggregated variables

On an aggregated level, socioeconomic inequalities impact national mortality not only through the importance of the SES gradient, but also through the composition of the population and its heterogeneity. For instance, in a study based on the comparison of the United States with 14 European countries, Avendano et al. (2010) observed that the unusually high educational gradient in mortality in the United States seems to be counterbalanced by an attractive educational distribution. As a consequence, they found out that the Relative Index of Inequality (RII)[15] of the United States was not especially high in comparison with other countries with an educational gradient of lower magnitude. The age structure of the population also plays a determining role, and the socioeconomic composition of different age classes can vary a lot (see Chapter 4 for a more detailed discussion on this subject).

From a material point of view, the relationship between income and health or mortality was initially thought of as a curvi-linear relation (Preston (1975); Rodgers (1979)). According to this analysis, a reditribution of income from the wealthiest groups to the poorest would result in improving the health of the poor rather than endanger the health of the wealthy. This non-linear relationship shows that the impact of inequality at the aggregated level of a country is not trivial. For instance, if a country experiences a high level of income inequalities, the overall mortality in the country can be higher than in a country with the same average level of income but with a lower level of inequality.

But recent studies, based on the pyschosocial explanation of the SES gradient, seem to indicate that the relation between aggregated mortality and inequality is even more complicated. They argue that the presence of inequality itself impacts the health and mortality of individuals. For instance, Wilkinson et al. (2009) studied the association between life expectancy and income inequality[16] among the 50 richest countries of more

[15]The RII is an index of inequality which takes into account differences in mortality as well the populations composition, see Regidor (2004) for details on the computation of the index.

[16]Income inequality was measured in each country as the ratio of income of the poorest 20% to the richest 20%.

than 3 million inhabitants, and found out a correlation of 0.44 between life expectancy and the level of inequalities, while no significant association was found between life expectancy and the average income. These results suggest that health and mortality are impacted by the *relative* social position of individuals, rather than their absolute material living standards (Pickett and Wilkinson (2015); Wilkinson and Pickett (2009)). This is closely linked to the theory of psychosocial factors, which assumes that it is the relative social ranking which determines the level of exposure to psychosocial problems and the ability to cope with them. Wilkinson and Pickett (2009) go even further and argue that inequalities affect not only individuals at the bottom of the socioeconomic ladder, but the vast majority of the population. For instance, Wilkinson and Pickett (2008) compared standardized mortality rates in counties of the 25 more equal states in the US and in counties of the 25 less equal states[17] (see Figure 5.3). They found out that for counties with the same median income, mortality rates were higher in counties in the more equal states than in counties in less equal states. The relation held at all levels of median income, with more important differences for counties with lower median income. When measuring the impact of inequality on health, the size of the area appears to be an important variable to take into account. On the one hand, the relationship between health and inequalities, when measured at the level of large areas such as states of big regions, seems to be fairly strong. On the other hand, Wilkinson and Pickett (2009) note that at the level of smaller areas such as neighborhoods, the average level of income seems to matter more than one's relative social position in the neighborhood. This "neighborhood effect" has been studied by many authors and constitutes a field of research in itself (se e.g Diez Roux (2007); Diez Roux and Mair (2010); Kawachi and Berkman (2003); Nandi and Kawachi (2011)). Societal inequalities, neighborhood environment and individual socioeconomic characteristics thus impact health and mortality at multiple scales, making the analysis of factors responsible for poorer health highly difficult. It is even more difficult to understand what happens at the aggregated level.

[17] The measure of inequality was based on the Gini coefficient of household income.

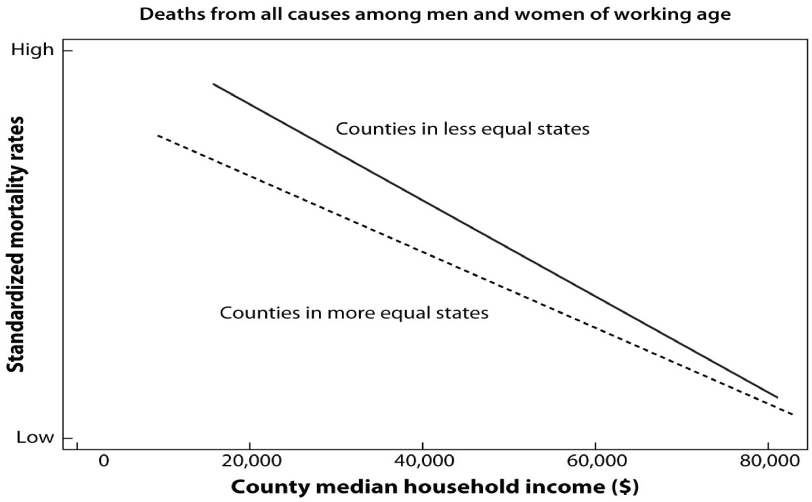


Fig. 5.3 Relationship between median county income and standardized mortality rates among working-age individuals, reproduced from Wilkinson and Pickett (2009) (Figure 11)

Discussion

The problems surrounding the measure of the SES is revealing of the issues at hand. The interpretation of data across time and places is a delicate matter. As illustrated by the major changes in the composition of women occupational subgroups, the effects of composition changes have to be carefully addressed to. Besides, it is rather unlikely that a single measure of SES, at only one point in the life course of individuals, could capture accurately the many pathways by which social status can affect health and mortality (Elo (2009)). However, there are many limitations in the ability to obtain reliable data from multiple measures of SES.

There are often limited opportunities for the empirical testing of complex theories such as the fundamental cause theory or the theory of psychosocial factors. The design of empirical test is not straightforward, to say the least. Natural experiments, such as the evolution of smoking behavior, or experiments on non-human populations, such as Saplosky's study of baboons, can give valuable insights on theories. However, as stated in the conclusion of the report of National Research Council and Committee on Population and others (2011), " it is sometimes difficult, expensive, and ethically challenging to alter individual behavior".

Pathways involved in translating SES into mortality outcomes can differ substantially according to the theory taken into account. Moreover, the impact of these underlying mechanisms on aggregated variables can also differ a lot, ranging from composition effects due to the curvilinear relation between material resources and longevity, to the global

(non-linear) effects of the social stratification on individuals. From the fundamental causes point of view, new advances in knowledge and technology related to health will probably increase the SES gradient in health and mortality (Cutler et al. (2006); Phelan et al. (2010)).

This illustrates how different underlying factors explaining the SES gradient can influence our views on the impact of socioeconomic inequalities on aggregated quantities, and in turn influence choices of public policies. Different types of policy recommendation can be made, according to the underlying factors or measures of SES which are considered to be most prominent. For instance, Phelan et al. (2010) recommend two types of public interventions. The first type focuses on reducing socioeconomic inequality itself in order to redistribute resources and knowledge. This would be consistent with the views of Wilkinson and Pickett (2009) on the general impact of inequality in a country. The second type of recommendations falls into the domain of public health. Governments should be careful and design interventions which do not increase inequalities, by favouring for instance health interventions which would benefit everyone automatically.

We believe that the dynamic modeling of the evolution of the population may help to address these issues. A fine-grained modeling of the population dynamic could help to evaluate the impact of changes in the composition of socioeconomic subgroups. In addition, modeling the population dynamics can serve as a simulation tool in order to take into account various measures of SES, when empirical data are limited. It can also be used to test hypotheses regarding which aspects of SES are the most important for reducing socioeconomic inequalities in health and mortality. By using population simulation as an experimenting tool when real life experiences are not possible, theories can be tested by comparing the aggregated outcomes produced by the model to what is observed in reality.

However, the above paragraph shows us the complexity of the phenomena involved. Socioeconomic inequalities impact health and mortality through complicated pathways. Phenomena are often non-reproducible - risk factors, as well as the economic, social or demographic environment have changed dramatically over the recent years - with effects which are often delayed. Furthermore, findings suggest that the impact of socioeconomic inequalities is highly non-linear. Individual characteristics do not fully explain the longevity of individuals. Mechanisms acting at different scales appear to be equally important. For instance, the neighborhood effect, the relative social position of individuals or the global level of inequality in society are also important factors to take into account. From these examples, it is quite easy to see the modeling challenges brought about by the new paradigm of the second demographic transition. At yet, there is also an urgent

need for complex population models, for a better understanding of the observed data, as well as to serve as an alternative when empirical testing is not possible.

5.4 Modeling complex population evolutions

Before the 1980s, demographic models were principally focused on the macro-level, and used aggregate data to produce average indicators. In view of the previous considerations, producing a pertinent modeling directly at the macro-level appears to be a more and more complicated-if not impossible-task. Hence, demographic models have increasingly shifted towards a finer-grained modeling of the population in the last decades.
There is thus an intrinsic interest in describing the variability and heterogeneity of the population on a more detailed level, in order to obtain *macro*-outcomes by aggregation, to be used forecasting/projections and/or policy recommendations, or in a broader sense for the analysis of social economic policies.
Over the last two decades, the increase of computing power and improvements in numerical methods have made it possible to study rich heterogeneous individual models. Indeed, a wide variety of models simulating individual behavior have been developed for different purposes and used in different domains. In this section, we give an overview of two types of models widely used in demography: Standard microsimulation models (MSMs) and Agent based models (ABMs) which are derived from the idea of Orcutt (1957) (see Morand et al. (2010)).

5.4.1 Dynamic microsimulation

Microsimulation models

Microsimulation issues A dynamic microsimulation model provides a simulation tool of individual trajectories in order to obtain macro outcomes by aggregation. It provides a way of combining different processes (biological, cognitive, social) describing the lives of people who evolve over time. One main feature of this class of model is its capacity to interpret macro level changes, represented by macroeconomic complex quantities (or indicators) (e.g life expectancy, mortality rate,...), resulting from the simulation of the dynamic life courses of individuals, also called micro units. A dynamic microsimulation model, usually relying on an important amount of empirical data, is parametrized with micro-econometrics and statistical methods (Spielauer (2011)).

Examples of microsimulation models The history of microsimulation in social sciences goes back to the work of Orcutt (1957), who developed so-called data-driven dynamic microsimulation models. Following the original model of Orcutt, the first large-scale dynamic microsimulation model called DYNASIM[18] was developed for the forecasting of the US population up to 2030. This model considered different demographic and economic scenarios, meant to analyse the socioeconomic status and behavior of individuals and families in the US (cost of teenage childbearing for the public sector, unemployment compensations and welfare programs...). Since then, most statistical or demographical government bodies in developed countries have used their own microsimulation models, developed for different purposes. A comprehensive description of various microsimulation models can be found in the surveys of Li and O'Donoghue (2013); Morand et al. (2010); Zaidi and Rake (2001). For instance DYNACAN in Canada was designed to model the Canada Pension Plan (CPP) and analyze its contributions and benefits at individual and family level. In Australia, DYNAMOD was developed to carry out a projection of the outlook of Australian population until the year 2050. In Europe, the MICMAC project[19] was implemented by a consortium of research centers whose objective is to provide demographic projections concerning detailed population categories, that are required for the design of sustainable (elderly) health care and pension systems in the European Union. The specificity of the micmac consists in providing a micro-macro modeling of the population, with micro level projections that are consistent with the projections made by the macro model.
In France, the INSEE developed different versions of a microsimulation model, the current version being DESTINIE 2 (Blanchet et al. (2009)). This model is used for instance to measure the efficiency of reforms on state pension systems, and is based on a representative sample of the national population.

A dynamic microsimulation exercise

A demographic micro-model can be viewed as a population database, which stores dynamically information (characteristics) on all members (individuals) of the heterogeneous population (Willekens (2005)). Zinn (2011) gives the main steps of a microsimulation exercise which consist of:
(i) *State space and state variables:* The state space is composed of all the combinations of the values (attributes) of individuals' characteristics, called state variables. Age, sex,

[18]Since then, updated versions were developed, with for example DYNASIM3 in 2004 (Li and O'Donoghue (2013)).
[19]The MICMAC project is documented in Willekens (2005).

marital status, fertility and mortality status, education or emigration are examples of state variables. An example of a state is given by the possible values of state variables: (Female, Married, 1 Child, Alive, Not emigrated, Lower secondary school)[20].

(ii) *Transition rates:* Events occurring during the life course of individuals are characterized by individual hazard functions, or *individual transition rates / probabilities*. Each of the transition probabilities is related to an event, i.e. a change in one of the state variables of the individual. These probabilities are estimated conditionally on demographic covariates (i.e explanatory variables such as gender, age, educational attainment, children born, ethnicity), and other risk factors that affect the rate of occurrence of some events (environmental covariates that provide external information on the common (random) environment) (Spielauer (2011)). In microsimulation models, the covariates are often estimated by using logit models (see Zinn (2011)).

(iii) *Dynamic simulation:* Dynamic simulation aims at *predicting the future state* of the population, by making the distinction between events influencing the population itself and those affected by it (population ageing, concentration of wealth, sustainability of social policies...).

(iv) *Internal consistency:* Microsimulation models can handle links between individuals, which can be qualified as "internal consistency" (Van Imhoff and Post (1998)). Individuals can be grouped together in the database into "families" (for instance, if they are married or related). When a state variable of an individual in the group changes, the state variables of the other members are updated if needed. For example, this can be the case when such events as marriage, divorce or a child leaving the parental home take place.

(v) *Output of microsimulation exercise and representation:* The output of a dynamic microsimulation model is a simulated database with longitudinal information, e.g. in the form of individual virtual biographies, viewed as a sequence of state variables. The effects of different factors can be revealed more clearly when grouping individuals with life courses embedded in similar historical context. Usually, individuals are grouped in cohorts (individuals with the same age) or in generations. The aggregation of individual biographies of the same cohort yields a bottom-up estimate of the so-called cohort biography. Nevertheless, in the presence of interactions, all the biographies have to be simulated simultaneously, which is challenging for large populations.

[20]The state variables Dead and Emigrated are considered as absorbing states, i.e. once they have been entered, they will never be left again.

Sources of randomness

Microsimulation models are subject to several sources of uncertainty and randomness which have been discussed in detail in the work of Van Imhoff and Post (1998). The so-called "inherent randomness" is due to the nature of Monte Carlo random experiments (different simulations produce variable sets of outcomes). This type of randomness can be diminished when simulating large populations (increasing the number of individuals in the database) or repeating random experiments many times to average the results, which implies important computational cost[21]. In the presence of interactions, one should be careful since the two techniques are not equivalent. Such is the case of agent based models, which are discussed in the next section.

The starting population, which is the initial database for the microsimulation model, can be either a sample of the population based on survey data, or a synthetic population created by gathering data from different information sources. This initial population is subject to random variations and sampling errors. Moreover, at the individual level, the state variables and the covariates must be known before starting the simulation, and their joint distribution within the initial database is random. Van Imhoff and Post (1998) note that any deviation of the sample distribution may impact future projections.

These previous sources of randomness can be mitigated by increasing the size of the database and are probably less important in comparison with the so-called specification randomness. The outputs of a microsimulation model can be subject to a high degree of randomness when an important number of covariates are included. Indeed, there are calibrating errors resulting from the estimation of empirical data each relationship between probabilities of the model and covariates. Moreover, each additional covariate requires an extra set of Monte Carlo experiments, with a corresponding increase in Monte Carlo randomness.

The *specification randomness* can be reduced by using sorting or alignment methods, a calibration technique that consists in selecting the simulated life course in such a way that the micro model respects some macro properties, including the property of producing the expected values. In DESTINIE 2, this alignment is ensured by adjusting the individual transition rates to obtain the annual number of births, deaths and migration consistent with some macro projections (Blanchet et al. (2009)).

[21] Various techniques to accelerate Monte Carlo simulation coupled with variance reduction has been developed in many areas.

Discussion

In many cases, behaviors are more stable or better understood on the micro level than on aggregated levels that are affected by structural changes when the number or size of the micro-units in the population changes. Thus, microsimulation models are well suited to explain processes resulting from the actions and interactions of a large number of micro-units. For instance, according to Spielauer (2011), an increase of graduation rates[22] at the macro level can lie entirely in the changing composition of the parents' generations, and not necessarily in a change of individuals' behaviors.
In order to produce more micro-level explanations for population change, microsimulation models require an increasing amount of high quality data to be collected. Silverman et al. (2011) point out the "Over-dependence on potentially immense sets of data" of microsimulation models and the expensive data collection required to provide inputs for those models. Most of the time, only large entities such as national or international institutions are able to complete this demanding task. The size of the samples also has an impact on the run time of the model; the larger the sample's size is, the longer the run speed will be, which will result in a trade-off.
Silverman et al. (2011) argue in favour of the use of more abstract computational models rather than on highly data-driven research. More recently, Agent Based Models (ABM), which also derived from individual-based models, have been increasingly applied in various areas to analyze macro level phenomena gathered from micro units. These models emphasize interactions between individuals through behavioral rules and individual strategies. In this context, Zinn (2017) stressed the importance of incorporating behavioral rules through ABM models (e.g kinship, mate matching models..) since demographic microsimulation is well suited for population projection, if only the model considers independent entities.

5.4.2 Agent Based Models (ABM)

What is ABM?

The main purpose of the Agent Based Models (ABM) is to explain macroscopic regularities by replicating the behavior of complex, real-world systems with dynamical systems of interacting agents based on the so-called bottom-up approach (Billari (2006); Tesfatsion (2002)). ABM consists basically of the *simulation* of interactions of *autonomous agents* i.e independent individuals (which can be households, organizations, companies, or

[22]Graduation rate represents the estimated percentage of people who will graduate from a specific level of education over their lifetime.

nations...depending on the application). As in microsimulation, agents are defined by their attributes. Each single agent is also defined by *behavioral rules*, which can be simple or complex (e.g utility optimization, complex social patterns...), deterministic or stochastic, on whose basis she/he interacts with other agents and with the simulated environment (Billari (2006); Morand et al. (2010)).

Applications The applications of ABM range from social, economic or political sciences to demography (Billari (2006)). For instance, Tesfatsion (2002) used Agent-based Computational economies (ACE) in order to model decentralized economic markets through the interaction of autonomous agents. In demography, ABM are used in Diaz et al. (2011) to explain trends in fertility by simple local interactions, in order to solve the difficult problem of age-specific projection of fertility rates. Billari et al. (2007) developed an ABM based on the interaction between heterogeneous potential partners, which typically takes place in the marriage market (partnership formation) and which is called "The Wedding Ring model". The purpose of this model is to study the age pattern of marriage using a bottom-up approach. This model was implemented using the software package NetLogo (Wilensky (1999)) which is designed for constructing and exploring multilevel systems[23]. Burke and Heiland (2006) suggested the use of an agent based model to explain the differences in obesity rates between women with different educational attainment in the United States. The model integrates biological complex agents (variation of women's metabolism) interacting within a social group, and is able to reproduce the fact that better educated women experience on average lower weights and smaller dispersion of weights. For more examples of agent based models applications, we refer to the work of Morand et al. (2010) that details different examples of ABM in spatial demography, family demography and historical demography. The book of Billari (2006) also presents various applications of agent-based computational modeling, in particular in demography.

A dynamic exercise of an ABM

The key defining feature of an ABM model is the interactions between heterogeneous individuals. Moreover, an agent based model is grounded on a dynamic simulation, which means that agents adapt dynamically to changes in the simulated environment. They act and react with other agents in this environment at different spatial and temporal scales (Billari (2006)). This contrasts with Microsimulation models (MSM) which rely

[23]Multi-level agent based models integrate different levels (complementary points of view) of representation of agents with respect to time, space and behavior.

on transition rates that are determined a priori (and once).
Agent based models are based on some rules, or heuristics, which can be either deterministic or stochastic, and which determine the decision-making process. For example, in an agent based marriage market model, the appropriate partner can be chosen as the one who has the most similar education level to the considered agent, or an ideal age difference (Billari et al. (2007)).
Besides, in comparison with Microsimulation models, which operate on a realistic scale (real data), but use very simple matching algorithms (often a Monte Carlo "roll the dice" styled decision rule), agent based models use small and artificial data sets, but show more complexity in modeling how the agents viewed and chose partners.

Limitations

The design of agent based model needs a certain level of expertise in the determining of behavioral rules. Furthermore, when modeling large systems (large number of agents), computational time rises considerably. Indeed, ABM models are not designed for extensive simulations.
The parameters of an agent based model can be either calibrated using accurate data, or consider sensitivity analysis incorporating some level of comparison with actual data. For instance, Hills and Todd (2008) compare the results of their Agent-Based Marriage and Divorce Model (MADAM) to real age-at-marriage distributions. But the outputs of an agent based model also depend on the "internal" structure of the model, determined by the behavioral rules (Gianluca (2014)). Consequently, the strategies to calibrate parameters, and to overcome the problem of dependency on the model's structure, rely on available empirical information. It is important to note that ABMs are designed to focus on process related factors or on the demonstration of emergent properties, rather than to make projections.

5.4.3 Conclusion

The interest of dynamic microsimulation is to constitute both a modeling exercise, and an exercise to run the model and experiment with it (Spielauer (2011)). In addition to helping to test theory or to picture the future, the exercise may be used as a simulator by policy makers (or citizens) or for a better assessment of the impact of public policies. The results/outputs of microsimulation models are population projections rather than forecasts, which is what would happen if the assumptions and scenarios chosen were to prove correct on what the future will probably be.

The discussion on demographic modeling demonstrated that Microsimulation models (MSM) strongly depend on data Silverman et al. (2011). Then, it faces pragmatic challenges in collecting and cleaning data, in addition to the different sources of randomness discussed above. In parallel with the spread of microsimulation models, there is a growing interest in Agent Based Models, which are suited to model complex systems that take full account of interactions between heterogeneous agents. The major difficulty in using of Agent based model is the absence of theoretical model. Indeed, there is no codified set of recommendations or practices on how to use these models within a program of empirical research. It is essentially based on the cognition and expertise of the developer. In this context, new hybrid applications (combining MSM and ABM models) have been recently proposed in literature. For instance, Grow and Van Bavel (2016) present many examples that combine MSM and ABM in demographic models. These new models aim at describing the heterogeneous movements, interactions and behaviors of a large number of individuals within a complex social system at a fine spatial scale. For instance, Zinn (2017) uses a combination of MSM and ABM for modeling individuals and couples life courses by integrating social relations and interactions. The efficiency of these combined and "sophisticated" models to overcome the loopholes of the simple models is an open issue.

5.5 Conclusion and perspectives

Facing all these modeling challenges, we advocate the development of a new mathematical theoretical framework for the modeling of complex population dynamics in demography. As we have seen in Section 5.3, a number of questions cannot be answered by the sole study of data, and models allow us to generate and experiment with varyous scenarios, so as to test theories or causal links for instance. Theoretical models can help us "to escape from the tyranny of data", as claimed by Silverman et al. (2011).
On the other hand, empirical evidence point out a number of key issues which cannot be overlooked, and which demonstrate the "inextricable complexity" of dynamic modeling of realistic human populations. Variables such as mortality or fertility rates are by no means stationary; populations are more and more heterogeneous, with socioeconomic inequality playing an important role at several levels (individual, neighborhood and societal); interactions between individuals and their environment are bidirectional. These are just a few examples illustrating the complexity of modeling. An adapted mathematical framework could contribute significantly to better understand aggregation issues and find out adequate policy recommendations, in concordance with this new paradigm of

heterogeneity and non-linearity. More specifically, theoretical models often allow us to reduce complexity by deriving and/or justifying approximations in population dynamics. By changing point of view, data can also be represented differently,and thus permit to go beyond what is usually done.

The historical analysis of these two centuries of demographic transitions show that populations have experienced dramatic changes and upheavals. But we can also see, a number of phenomena and timescales present remarkable regularities. These profound regularities, or "fundamental causes", have been noted by several authors, in very different contexts. In our opinion, the identification and understanding of these regularities or cycles is fundamental.

Age is also a critical dimension when studying human population dynamics. The age structure of a population generates a lot of complexity in the representation and statistical analysis of data. This so-called Age Period Cohort (APC) problem has been well documented in statistical literature, and should be a main focus in the dynamical modeling of populations. Furthermore, the human life cycle is composed of very different periods, with transition rates of a different order and phenomena of a different nature at each stage. Understanding how to take into account this heterogeneity in age is a critical point. The notion of age itself changes over time. Individuals seem to have rejuvenated, in the sense that today's 65-year-olds are "much younger" than individuals of the same age thirty years ago.

Thus, the shift in paradigm observed in recent demographic trends has highlighted a number of new issues which force us to reconsider many aspects of the traditional modeling of human populations. Multiple questions are still open, with difficult challenges ahead, but also exciting perspectives for the future.

References for Chapter 1

Alai, D. H., Arnold, S., Bajekal, M., and Villegas, A. M. (2017). Causal Mortality by Socioeconomic Circumstances: A Model to Assess the Impact of Policy Options on Inequalities in Life Expectancy. In *Conference proceedings of the Society of Actuaries 2017 Living to 100 Symposium*.

Anderson, D. F. and Kurtz, T. G. (2015). *Stochastic Analysis of Biochemical Systems*. Springer.

Auger, P., Charles, S., Viala, M., and Poggiale, J.-C. (2000). Aggregation and emergence in ecological modelling: integration of ecological levels. *Ecological Modelling*, 127(1):11 – 20.

Auger, P., Poggiale, J., and Sánchez, E. (2012). A review on spatial aggregation methods involving several time scales. *Ecological Complexity*, 10:12–25.

Barrieu, P., Bensusan, H., El Karoui, N., Hillairet, C., Loisel, S., Ravanelli, C., and Salhi, Y. (2012). Understanding, modelling and managing longevity risk: key issues and main challenges. *Scandinavian actuarial journal*, 2012(3):203–231.

Bensusan, H. (2010). *Risques de taux et de longévité: Modélisation dynamique et applications aux produits dérivés et à l'assurance vie*. PhD thesis, École Polytechnique.

Bhaskaran, B. (1986). Almost sure comparison of birth and death processes with application to m/m/s queueing systems. *Queueing Systems*, 1(1):103–127.

Bloom, D. E. and Luca, D. L. (2016). The global demography of aging: Facts, explanations, future. *Handbook of the Economics of Population Aging*, 1:3–56.

Boumezoued, A. (2016). *Approches micro-macro des dynamiques de populations hétérogènes structurées par âge. Application aux processus auto-excitants et à la démographie*. PhD thesis, Université Pierre et Marie Curie.

Cairns, A., Blake, D., and Dowd, K. (2006). A Two-Factor Model for Stochastic Mortality with Parameter Uncertainty: Theory and Calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Cambois, E., Meslé, F., and Pison, G. (2009). L'allongement de la vie et ses conséquences en France . *Regards croisés sur l'économie*, 5.

Canning, D. (2011). The causes and consequences of demographic transition. *Population studies*, 65(3):353–361.

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

Christensen, K., Doblhammer, G., Rau, R., and Vaupel, J. W. (2009). Ageing populations: the challenges ahead. *The lancet*, 374(9696):1196–1208.

Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The determinants of mortality. *The Journal of Economic Perspectives*, 20(3):97–120.

d'Albis, H. and Collard, F. (2013). Age groups and the measure of population aging. *Demographic Research*, 29(23):617–640.

Dowd, J. B. and Hamoudi, A. (2014). Is life expectancy really falling for groups of low socio-economic status? lagged selection bias and artefactual trends in mortality. *International Journal of Epidemiology*, 43(4):983–988.

Edwards, R. D. and Tuljapurkar, S. (2005). Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review*, 31(4):645–674.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology*, 35:553–572.

Eyler, J. M. (1973). William farr on the cholera: the sanitarian's disease theory and the statistician's method. *Journal of the history of medicine and allied sciences*, 28(2):79.

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

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–583.

Haberman, S., Kaishev, V., Millossovich, P., and Villegas, A. (2014). Longevity Basis Risk A methodology for assessing basis risk. Technical report, Institute and Faculty of Actuaries (IFoA), Life and Longevity Markets Association (LLMA).

Huber, M. (2012). Spatial birth–death swap chains. *Bernoulli*, 18(3):1031–1041.

Jarner, S. F. and Kryger, E. M. (2011). Modelling adult mortality in small populations: The saint model. *Astin Bulletin*, 41(02):377–418.

Kallenberg, O. (2017). *Random Measures, Theory and Applications*, volume 77. Springer.

Kirkwood, T. B. (2015). Deciphering death: a commentary on gompertz (1825) 'on the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies'. *Phil. Trans. R. Soc. B*, 370(1666):20140379.

Kurtz, T. G. (1992). Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis*, pages 186–209. Springer.

Lee, R. and Carter, L. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659–671.

Lee, R. D. and Reher, D. S. (2011). *Demographic transition and its consequences*. JSTOR.

Lesthaeghe, R. (2014). The second demographic transition: A concise overview of its development. *Proceedings of the National Academy of Sciences*, 111(51):18112–18115.

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

Ludkovski, M., Risk, J., and Zail, H. (2016). Gaussian process models for mortality rates and improvement factors. *arXiv:1608.08291*.

Marvá, M., Moussaouí, A., de la Parra, R. B., and Auger, P. (2013). A density-dependent model describing age-structured population dynamics using hawk–dove tactics. *Journal of Difference Equations and Applications*, 19(6):1022–1034.

Massoulié, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic processes and their applications*, 75(1):1–30.

Méléard, S. and Tran, V. (2009). Trait substitution sequence process and canonical equation for age-structured populations. *Journal of mathematical biology*, 58(6):881–921.

Mireaux, E. (1962). Un chirurgien sociologue: Louis-René Villermé. *Revue des deux mondes*.

National Research Council and Committee on Population (2011). *Explaining divergent levels of longevity in high-income countries*. National Academies Press.

Oeppen, J. and Vaupel, J. W. (2002). Broken limits to life expectancy. *Science*, 296(5570):1029–1031.

Olshansky, S. J., Antonucci, T., Berkman, L., Binstock, R. H., Boersch-Supan, A., Cacioppo, J. T., Carnes, B. A., Carstensen, L. L., Fried, L. P., Goldman, D. P., Jackson, J., Kohli, M., Rother, J., Zheng, Y., and Rowe, J. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. *Health Affairs*, 31(8):1803–1813.

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population studies*, 29(2):231–248.

Reher, D. S. (2011). Economic and social implications of the demographic transition. *Population and Development Review*, 37:11–33.

Renshaw, A. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. 38(3):556–570.

Rolski, T. and Szekli, R. (1991). Stochastic ordering and thinning of point processes. *Stochastic processes and their applications*, 37(2):299–312.

Shang, H. L. and Haberman, S. (2017). Grouped multivariate and functional time series forecasting: An application to annuity pricing. *Insurance: Mathematics and Economics*.

Shang, H. L. and Hyndman, R. J. (2017). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, 26(2):330–343.

Silverman, E., Bijak, J., and Noble, J. (2011). Feeding the beast: can computational demographic models free us from the tyranny of data? In Lenaerts, T., Giacobini, M., Bersini, H., Bourgine, P. , Dorigo, M. and Doursat, R., editor, *Advances in Artificial Life, ECAL 2011*, pages 747–754. MIT Press.

Vallin, J. and Meslé, F. (2001). *Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle*. Éditions de l'Institut national d'études démographiques.

Vaupel, J. W. and Loichinger, E. (2006). Redistributing work in aging europe. *Science*, 312(5782):1911–1913.

Villegas, A. M. and Haberman, S. (2014). On the Modeling and Forecasting of Socioeconomic Mortality Differentials: An Application to Deprivation and Mortality in England. *North American Actuarial Journal*, 18(1):168–193.

Villermé, L. (1830). De la mortalité dans les divers quartiers de la ville de Paris, et des causes qui la rendent très différentes dans plusieurs d'entre eux, ainsi que dans les divers quartiers de beaucoup de grandes villes. *Annales d'hygiène publique et de médecine légale*, 3:294–321.

World Health Organization (2015). World report on ageing and health. Technical report, World Health Organization.

Yin, G. and Zhang, Q. (2012). *Continuous-time Markov chains and applications: a two-time-scale approach*, volume 37. Springer Science & Business Media.

Zinn, S. (2017). Simulating synthetic life courses of individuals and couples, and mate matching. In *Agent-Based Modelling in Population Studies*, pages 113–157. Springer.

References for Chapter 2

Anderson, D. and Kurtz, T. (2015). *Stochastic Analysis of Biochemical Systems*. Springer.

Auger, P., Charles, S., Viala, M., and Poggiale, J.-C. (2000). Aggregation and emergence in ecological modelling: integration of ecological levels. *Ecological Modelling*, 127(1):11 – 20.

Auger, P. and Pontier, D. (1998). Fast game theory coupled to slow population dynamics: the case of domestic cat populations. *Mathematical Biosciences*, 148(1):65 – 82.

Bansaye, V. and Méléard, S. (2015). Stochastic models for structured populations. *Mathematical Biosciences Institute Lecture Series. Springer edition.*

Bezborodov, V. (2015). Markov birth-and-death dynamics of populations. *arXiv preprint arXiv:1502.06783*.

Bhaskaran, B. (1986). Almost sure comparison of birth and death processes with application to m/m/s queueing systems. *Queueing Systems*, 1(1):103–127.

Brémaud, P. and Massoulié, L. (1996). Stability of nonlinear hawkes processes. *The Annals of Probability*, pages 1563–1588.

Cox, D. R. and Miller, H. D. (1977). *The theory of stochastic processes*, volume 134. CRC Press.

Feller, W. (1968). *An introduction to probability theory and its applications: volume I*, volume 3. John Wiley & Sons New York.

Florens, J. and Fougere, D. (1996). Noncausality in continuous time. *Econometrica: Journal of the Econometric Society*, pages 1195–1212.

Fournier, N. and Méléard, S. (2004). A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *Ann. Appl. Probab.*, 14(4):1880–1919.

Garcia, N. and Kurtz, T. (2008). Spatial point processes and the projection method. In *In and Out of Equilibrium 2*, pages 271–298. Springer.

Garcia, N. L. and Kurtz, T. G. (2006). Spatial birth and death processes as solutions of stochastic equations. *ALEA Lat. Am. J. Probab. Math. Stat*, page 2249658.

Gjessing, H., Røysland, K.and Pena, E., and Aalen, O. (2010). Recurrent events and the exploding cox model. *Lifetime data analysis*, 16(4):525–546.

Huber, M. (2012). Spatial birth–death swap chains. *Bernoulli*, 18(3):1031–1041.

Jacobsen, M. (1982). *Statistical analysis of counting processes*, volume 12. Springer, lectures note in statistics edition.

Jacod, J. and Shiryaev, A. (1987). *Limit theorems for stochastic processes*. Springer.

Lewis, P. and Shedler, G. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly*, 26(3):403–413.

Marvá, M., Moussaouì, A., de la Parra, R. B., and Auger, P. (2013). A density-dependent model describing age-structured population dynamics using hawk–dove tactics. *Journal of Difference Equations and Applications*, 19(6):1022–1034.

Massoulié, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic processes and their applications*, 75(1):1–30.

Nassar, E. and Pardoux, E. (2017). On the large-time behaviour of the solution of a stochastic differential equation driven by a poisson point process. *Advances in Applied Probability*, 49(2):344–367.

Preston, C. (1975). Spatial birth and death processes. *Advances in applied probability*, 7(3):465–466.

Rolski, T. and Szekli, R. (1991). Stochastic ordering and thinning of point processes. *Stochastic processes and their applications*, 37(2):299–312.

References for Chapter 3

Aldous, D. J., Eagleson, G., et al. (1978). On mixing and stability of limit theorems. *The Annals of Probability*, 6(2):325–331.

Auger, P., Charles, S., Viala, M., and Poggiale, J.-C. (2000). Aggregation and emergence in ecological modelling: integration of ecological levels. *Ecological Modelling*, 127(1):11 – 20.

Auger, P., Poggiale, J., and Sánchez, E. (2012). A review on spatial aggregation methods involving several time scales. *Ecological Complexity*, 10:12–25.

Häusler, E. and Luschgy, H. (2015). *Stable Convergence and Stable Limit Theorems*, volume 74. Springer.

Jacod, J. (1987). Sur la convergence des processus ponctuels. *Probability theory and related fields*, 76(4):573–586.

Jacod, J. and Mémin, J. (1981). Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité. *Séminaire de Probabilités XV 1979/80*, pages 529–546.

Jacod, J. and Shiryaev, A. (1987). *Limit theorems for stochastic processes*. Springer.

Jajte, R. and Paszkiewicz, A. (1999). Conditioning and weak convergence. *Probability and mathematical statistics-Worclaw University*, 19(2):453–461.

Kallenberg, O. (1975). *Random Measures*. Schriftenreihe. Akad.-Verlag.

Kallenberg, O. (2006). *Foundations of modern probability*. Springer Science & Business Media.

Kallenberg, O. (2017). *Random Measures, Theory and Applications*, volume 77. Springer.

Kurtz, T. G. (1992). Averaging for martingale problems and stochastic approximation. In *Applied Stochastic Analysis*, pages 186–209. Springer.

Marvá, M., Moussaouí, A., de la Parra, R. B., and Auger, P. (2013). A density-dependent model describing age-structured population dynamics using hawk–dove tactics. *Journal of Difference Equations and Applications*, 19(6):1022–1034.

Sánchez, E., Arino, O., Auger, P., and de la Parra, R. (2000). A singular perturbation in an age-structured population model. *SIAM Journal on Applied Mathematics*, 60(2):408–436.

Taylor, J. E. and Véber, A. (2009). Coalescent processes in subdivided populations subject to recurrent mass extinctions. *Electron. J. Probab*, 14:242–288.

Yin, G. and Zhang, H. (2004). Two-time-scale markov chains and applications to quasi-birth-death queues. *SIAM journal on applied mathematics*, 65(2):567–586.

Yin, G. and Zhang, Q. (2012). *Continuous-time Markov chains and applications: a two-time-scale approach*, volume 37. Springer Science & Business Media.

References for Chapter 4

Alai, D. H., Arnold, S., Bajekal, M., and Villegas, A. M. (2017). Causal Mortality by Socioeconomic Circumstances: A Model to Assess the Impact of Policy Options on Inequalities in Life Expectancy. In *Conference proceedings of the Society of Actuaries 2017 Living to 100 Symposium*.

Arnold, S., Boumezoued, A., Labit Hardy, H., and El Karoui, N. (2016). Cause-of-death mortality: What can be learned from population dynamics? *Insurance: Mathematics and Economics, in press*.

Bajekal, M. (2005). Healthy life expectancy by area deprivation: magnitude and trends in England, 1994-1999. *Health Statistics Quarterly*, 25:18.

Bajekal, M., Scholes, S., O'Flaherty, M., Raine, R., Norman, P., and Capewell, S. (2013). Implications of using a fixed imd quintile allocation for small areas in england from 1981 to 2007. *PLoS One*, 8(3).

Bensusan, H. (2010). *Risques de taux et de longévité: Modélisation dynamique et applications aux produits dérivés et à l'assurance vie*. PhD thesis, École Polytechnique.

Boumezoued, A. (2016). *Approches micro-macro des dynamiques de populations hétérogènes structurées par âge. Application aux processus auto-excitants et à la démographie.* PhD thesis, Université Pierre et Marie Curie.

Chiang, C. (1968). Introduction to stochastic processes in biostatistics. *John Wiley and Sons, New York*.

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

Department for Communities and Local Government (2015). The English Indices of Deprivation 2015 - Statistical release.

Diez Roux, A. V. and Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186(1):125–145.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology*, 35:553–572.

Ferriere, R. and Tran, V. (2009). Stochastic and deterministic models for age-structured populations with genetically variable traits. In *ESAIM: Proceedings*, volume 27, pages 289–310.

Guildea, Z., Fone, D. L., Dunstan, F., Sibert, J., and Cartlidge, P. (2001). Social deprivation and the causes of stillbirth and infant mortality. *Archives of Disease in Childhood*, 84(4):307–310.

Haberman, S., Kaishev, V., Millossovich, P., and Villegas, A. (2014). Longevity Basis Risk A methodology for assessing basis risk.

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

Iannelli, M., Martcheva, M., and Milner, F. A. (2005). *Gender-structured population modeling: mathematical methods, numerics, and simulations*, volume 31. Siam.

Jarner, S. F. and Kryger, E. M. (2011). Modelling adult mortality in small populations: The saint model. *Astin Bulletin*, 41(02):377–418.

Keyfitz, N. and Caswell, H. (2005). *Applied mathematical demography*, volume 47. Springer.

Labit Hardy, H. (2016). *Impacts of cause-of-death mortality changes: A population dynamic approach*. PhD thesis, Université de Lausanne.

Lu, J., Wong, W., and Bajekal, M. (2014). Mortality improvement by socio-economic circumstances in England (1982 to 2006). *British Actuarial Journal*, 19(01):1–35.

McKendrick, A. (1926). Application of mathematics to medical problems. *Proc. Edin. Math. Soc.*, 54:98–130.

Meyricke, R. and Sherris, M. (2013). The determinants of mortality heterogeneity and implications for pricing annuities. *Insurance: Mathematics and Economics*, 53(2):379–387.

Nandi, A. and Kawachi, I. (2011). Neighborhood effects on mortality. In *International Handbook of Adult Mortality*, pages 413–439. Springer.

National Research Council and Committee on Population (2011). *Explaining divergent levels of longevity in high-income countries*. National Academies Press.

Noble, M., Mclennan, D., Wilkinson, K., Whitworth, A., Exley, S., Barnes, H., Dibben, C., McLennan, D., et al. (2007). The english indices of deprivation 2007.

Norman, P. and Darlington-Pollock, F. (2016). The changing geography of deprivation in great britain: Exploiting small area census data, 1971 to 2011. *In: Chapter 30 in Stillwell, J. & Duke-Williams, O. Unlocking the UK 2011 Census: Handbook of Census Resources, Methods and Applications*.

Oakley, L., Maconochie, N., Doyle, P., Dattani, N., and Moser, K. (2009). Multivariate analysis of infant death in England and Wales in 2005-06, with focus on socio-economic status and deprivation. *Health Statistics Quarterly*, 42(1):22–39.

Office for National Statistics (2012). *2011 Census: Population and Household Estimates for Small Areas in England and Wales, March 2011*. Office for National Statistics.

Office for National Statistics (2015). Live births, stillbirths, and the intensity of childbearing measured by the total fertility rate. Access date: May 2016.

Pelovska, G. and Iannelli, M. (2006). Numerical methods for the Lotka-Mckendrick's equation. *Journal of computational and applied mathematics*, 197(2):534–557.

Shang, H. L. and Haberman, S. (2017). Grouped multivariate and functional time series forecasting: An application to annuity pricing. *Insurance: Mathematics and Economics*.

Shang, H. L. and Hyndman, R. J. (2017). Grouped functional time series forecasting: An application to age-specific mortality rates. *Journal of Computational and Graphical Statistics*, 26(2):330–343.

The Human Mortality Database (2016). University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). English data, February 2016.

Townsend, P. (1979). *Poverty in the United Kingdom*. Allen Lane and Penguin Books.

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.

Villegas, A. M. (2015). *Mortality: Modelling, Socio-Economic Differences and Basis Risk*. PhD thesis, Cass Business School.

Villegas, A. M. and Haberman, S. (2014). On the modeling and forecasting of socioeconomic mortality differentials: An application to deprivation and mortality in England. *North American Actuarial Journal*, 18(1):168–193.

Von Foerster, H. (1959). *The Kinetics of Cellular Proliferation*. Grune & Stratton.

Webb, G. F. (1985). *Theory of nonlinear age-dependent population dynamics*. CRC Press.

Willets, R. (2004). The cohort effect: insights and explanations. *British Actuarial Journal*, 10(04):833–877.

References for Chapter 5

Acemoglu, D. and Johnson, S. (2007). Disease and development: the effect of life expectancy on economic growth. *Journal of political Economy*, 115(6):925–985.

Acemoglu, D. and Johnson, S. (2014). Disease and development: a reply to Bloom, Canning, and Fink. *Journal of Political Economy*, 122(6):1367–1375.

Ariès, P. (1980). Two successive motivations for the declining birth rate in the west. *Population and Development Review*, pages 645–650.

Avendano, M., Kok, R., Glymour, M., Berkman, L., Kawachi, I., Kunst, A., and Mackenbach, J. (2010). Do Americans Have Higher Mortality Than Europeans at All Levels of the Education Distribution?: A Comparison of the United States and 14 European Countries. In Council, N. R., editor, *International Differences in Mortality at Older Ages: Dimensions and Sources*. Washington, DC: The National Academies Press.

Billari, F. (2006). *Agent-Based Computational Modelling: Applications in Demography, Social, Economic and Environmental Sciences*. Contributions to Economics. Physica-Verlag HD.

Billari, F., Aparicio Diaz, B., Fent, T., and Prskawetz, A. (2007). The "Wedding-Ring": An agent-based marriage model based on social interaction. *Demographic Research*, 17(3):59–82.

Blanchet, D., Buffeteau, S., Crenner, E., and Le Minez, S. (2009). The destinie 2 microsimulation model: overview and illustrative results. In *2 nd IMA conference, Ottawa*.

Blanpain, N. (2011). L'espérance de vie s'accroît, les inégalités sociales face à la mort demeurent. *Insee première*, 1372(4).

Bloom, D. E. and Canning, D. (2007). Commentary: The preston curve 30 years on: still sparking fires. *International Journal of Epidemiology*, 36(3):498–499.

Bloom, D. E., Canning, D., and Fink, G. (2014). Disease and development revisited. *Journal of Political Economy*, 122(6):1355–1366.

Bloom, D. E. and Luca, D. L. (2016). The global demography of aging: Facts, explanations, future. *Handbook of the Economics of Population Aging*, 1:3–56.

Bongaarts, J. (2009). Human Population Growth and the Demographic Transition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1532):2985–2990.

Bongaarts, J. (2014). Trends in causes of death in low-mortality countries: Implications for mortality projections. *Population and Development Review*, 40(2):189–212.

Brody, H., Rip, M. R., Vinten-Johansen, P., Paneth, N., and Rachman, S. (2000). Map-making and myth-making in broad street: the london cholera epidemic, 1854. *The Lancet*, 356(9223):64–68.

Burke, M. A. and Heiland, F. (2006). *The Strength of Social Interactions and Obesity among Women*. Physica-Verlag HD.

Cambois, E., Meslé, F., and Pison, G. (2009). L'allongement de la vie et ses conséquences en France . *Regards croisés sur l'économie*, 5.

Canning, D. (2011). The causes and consequences of demographic transition. *Population studies*, 65(3):353–361.

Case, A. and Deaton, A. S. (2005). Broken down by work and sex: How our health declines. In *Analyses in the Economics of Aging*, pages 185–212. University of Chicago Press.

Chalklin, C. (2001). *The rise of the English town, 1650-1850*, volume 43. Cambridge University Press.

Cicolella, A. (2010). Santé et environnement : la 2e révolution de santé publique. *Santé Publique*, 22(3):343–351.

Cutler, D., Deaton, A., and Lleras-Muney, A. (2006). The determinants of mortality. *The Journal of Economic Perspectives*, 20(3):97–120.

Cutler, D. and Miller, G. (2005). The role of public health improvements in health advances: the twentieth-century united states. *Demography*, 42(1):1–22.

Deaton, A. (2003). Health, Inequality, and Economic Development . *Journal of Economic Literature*, 41(1):113–158.

Diaz, B., Fent, T., Fürnkranz-Prskawetz, A., and Bernardi, L. (2011). Transition to parenthood: The role of social interaction and endogenous networks. *Demography*, 48(2):559–579.

Diez Roux, A.-V. (2007). Neighborhoods and health: where are we and were do we go from here? *Revue d'epidemiologie et de sante publique*, 55(1):13–21.

Diez Roux, A. V. and Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186(1):125–145.

Dupaquier, M. and Lewes, F. (1989). Cholera in england during the nineteenth century: medicine as a test of the validity of statistics. *Annales de démographie historique*, pages 215–21.

Easterlin, R. (1980). *Birth and fortune: the impact of numbers on personal welfare*. Basic Books.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology*, 35:553–572.

Eyler, J. M. (1973). William farr on the cholera: the sanitarian's disease theory and the statistician's method. *Journal of the history of medicine and allied sciences*, 28(2):79.

Fabre, G. (1993). Conflits d'imaginaires en temps d'épidémie. *Communications*, 57(1):43–69.

Farr, W. (1852). Report on the mortality of cholera in england, 1848-49. Technical report, General Register Office, Royal College of Physicians of London.

Fidler, D. (2001). The globalization of public health: the first 100 years of international health diplomacy. *Bulletin of the World Health Organization*, 79(9):842–849.

Gianluca, M. (2014). Potentialités et limites de la simulation multi-agents : une introduction. *Revue française de sociologie*, 55(4):653–688.

Glei, D. A., Meslé, F., and Vallin, J. (2010). Diverging trends in life expectancy at age 50: A look at causes of death. *International differences in mortality at older ages: Dimensions and sources*, pages 2–1.

Grow, A. and Van Bavel, J. (2016). *Agent-Based Modelling in Population Studies: Concepts, Methods, and Applications*. The Springer Series on Demographic Methods and Population Analysis. Springer International Publishing.

Haberman, S., Kaishev, V., Millossovich, P., and Villegas, A. (2014). Longevity Basis Risk A methodology for assessing basis risk.

Hardy, A. (1984). Water and the search for public health in london in the eighteenth and nineteenth centuries. *Medical History*, 28(3):250?282.

Haupt, H. and Laroche, F. (1993). *Histoire sociale de la France depuis 1789*. Bibliothèque allemande. Éditions de la Maison des sciences de l'homme, Paris.

Hills, T. and Todd, P. (2008). Population heterogeneity and individual differences in an assortative agent-based marriage and divorce model (madam) using search with relaxing expectations. *Journal of Artificial Societies and Social Simulation*, 11(4):5.

Himmelstein, D. U., Thorne, D., and Warren, Eand Woolhandler, S. (2009). Medical bankruptcy in the United States, 2007: results of a national study. *The American journal of medicine*, 122(8):741–746.

Huber, V. (2006). The unification of the globe by disease? the international sanitary conferences on cholera, 1851-1894. *The Historical Journal*, 49(2):453–476.

Jardin, A. and Tudesq, A. (1983). *Restoration and Reaction, 1815-1848*. Cambridge history of modern France. Cambridge University Press.

Kawachi, I. and Berkman, L. F. (2003). *Neighborhoods and health*. Oxford University Press.

Koch, T. (2014). 1831: the map that launched the idea of global health. *International Journal of Epidemiology*, 43(4):1014–1020.

Kudlick, C. (1996). *Cholera in Post-revolutionary Paris: A Cultural History*. Studies on the History of Soci. University of California Press.

Le Mée, R. (1998). Le choléra et la question des logements insalubres à Paris (1832-1849). *Population*, 53(1):379–397.

Lee, R. D. and Reher, D. S. (2011). *Demographic transition and its consequences*. JSTOR.

Lesthaeghe, R. (2010). The unfolding story of the second demographic transition. *Population and development review*, 36(2):211–251.

Lesthaeghe, R. (2014). The second demographic transition: A concise overview of its development. *Proceedings of the National Academy of Sciences*, 111(51):18112–18115.

Lesthaeghe, R. and Van de Kaa, D. J. (1986). Twee demografische transities. *Bevolking: groei en krimp*, pages 9–24.

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

Link, B. G. (2008). Epidemiological sociology and the social shaping of population health. *Journal of Health and Social Behavior*, 49(4):367–384.

Link, B. G. and Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of health and social behavior*, pages 80–94.

Lu, J., Wong, W., and Bajekal, M. (2014). Mortality improvement by socio-economic circumstances in England (1982 to 2006). *British Actuarial Journal*, 19(01):1–35.

Marmot, M. G. (1994). Social differentials in health within and between populations. *Daedalus*, pages 197–216.

Meara, E. R., Richards, S., and Cutler, D. M. (2008). The gap gets bigger: changes in mortality and life expectancy, by education, 1981–2000. *Health Affairs*, 27(2):350–360.

Mireaux, E. (1962). Un chirurgien sociologue: Louis-René Villermé. *Revue des deux mondes*.

Morand, E., Toulemon, L., S., P., R., B., and Billari, F. (2010). Demographic modelling: the state of the art. In *SustainCity Working Paper, 2.1a, Ined, Paris.*

Moreau de Jones, A. (1831). *Rapport au Conseil supérieur de santé sur le Cholera-Morbus Pestilien*. BNF/Gallica.

Nandi, A. and Kawachi, I. (2011). Neighborhood effects on mortality. In *International Handbook of Adult Mortality*, pages 413–439. Springer.

National Research Council and Committee on Population and others (2011). *Explaining divergent levels of longevity in high-income countries*. National Academies Press.

Olshansky, S. J., Passaro, D. J., Hershow, R. C., Layden, J., Carnes, B. A., Brody, J., Hayflick, L., Butler, R. N., Allison, D. B., and Ludwig, D. S. (2005). A potential decline in life expectancy in the United States in the 21st century. *New England Journal of Medicine*, 352(11):1138–1145.

Orcutt, G. H. (1957). A new type of socio-economic system. *The Review of Economics and Statistics*, 39(2):116–123.

Paillar, H. (1832). *Histoire statistique du cholèra-morbus qui a régné en France en 1832 par H. Paillard*. chez l'auter, a l'hotel-dieu, et chez J.-B. Baillière.

Pappas, G., Queen, S., Hadden, W., and Fisher, G. (1993). The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *New England journal of medicine*, 329(2):103–109.

Phelan, J. C., Link, B. G., and Tehranifar, P. (2010). Social conditions as fundamental causes of health inequalities: theory, evidence, and policy implications. *Journal of health and social behavior*, 51(1_suppl):S28–S40.

Pickett, K. E. and Wilkinson, R. G. (2015). Income inequality and health: a causal review. *Social Science & Medicine*, 128:316–326.

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population studies*, 29(2):231–248.

Raux, C. (2014). C'est du propre! la salubrité publique à paris au xixe siécle. *Collections Gallica BNF*.

Regidor, E. (2004). Measures of health inequalities: part 2. *Journal of Epidemiology & Community Health*, 58(11):900–903.

Reher, D. S. (2011). Economic and social implications of the demographic transition. *Population and Development Review*, 37:11–33.

Rodgers, G. B. (1979). Income and inequality as determinants of mortality: an international cross-section analysis. *Population studies*, 33(2):343–351.

Sapolsky, R. (2005). Sick of poverty. *Scientific American*, 293(6):92–99.

Sapolsky, R. M. (2004). Social status and health in humans and other animals. *Annu. Rev. Anthropol.*, 33:393–418.

Silverman, E., Bijak, J., and Noble, J. (2011). Feeding the beast: can computational demographic models free us from the tyranny of data? In Lenaerts, T., Giacobini, M., Bersini, H., Bourgine, P., Dorigo, M., and Doursat, R., editors, *Advances in Artificial Life, ECAL 2011*, pages 747–754. MIT Press.

Smith, J. P. (2007). The impact of socioeconomic status on health over the life-course. *Journal of Human Resources*, 42(4):739–764.

Spielauer, M. (2011). What is social science microsimulation? *Social Science Computer Review*, 29(1):9–20.

Stewart, S. T., Cutler, D. M., and Rosen, A. B. (2009). Forecasting the effects of obesity and smoking on US life expectancy. *New England Journal of Medicine*, 361(23):2252–2260.

Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial life*, 8(1):55–82.

Underwood, E. A. (1948). The History of Cholera in Great Britain. *Proceedings of the Royal Society of Medicine*, 41(3):165–173.

Van de Kaa, D. J. (2010). Demographic transition. *Encyclopedia of Life Support Systems*, 1:65–103.

Van Imhoff, E. and Post, W. (1998). Microsimulation methods for population projection. *Population*, 10(1):97–136.

Villermé, L. (1830). De la mortalité dans les divers quartiers de la ville de Paris, et des causes qui la rendent très différentes dans plusieurs d'entre eux, ainsi que dans les divers quartiers de beaucoup de grandes villes. *Annales d'hygiène publique et de médecine légale*, 3:294–321.

Wilensky, U. (1999). Netlogo (and netlogo user manual). Technical report, Center for Connected Learning and Computer-Based Modeling, Northwestern University.

Wilkinson, R. G., Pickett, K., et al. (2009). *The spirit level: Why more equal societies almost always do better*, volume 6. JSTOR.

Wilkinson, R. G. and Pickett, K. E. (2008). Income inequality and socioeconomic gradients in mortality. *American Journal of Public Health*, 98(4):699–704.

Wilkinson, R. G. and Pickett, K. E. (2009). Income inequality and social dysfunction. *Annual Review of Sociology*, 35:493–511.

Willekens, F. (2005). Biographic forecasting: bridging the micro-macro gap in population forecasting. *New Zealand Population Review*, 31:77–124.

World Health Organization (2015). World report on ageing and health. Technical report, World Health Organization.

Zaidi, A. and Rake, K. (2001). Dynamic microsimulation models: A review and some lessons for SAGE. Technical report, ESRC SAGE Research Group.

Zinn, S. (2011). *A Continuous-Time Microsimulation and First Steps Towards a Multi-Level Approach in Demography*. PhD thesis.

Zinn, S. (2017). Simulating synthetic life courses of individuals and couples, and mate matching. In *Agent-Based Modelling in Population Studies*, pages 113–157. Springer.