Modeling Joint Lives within Families

Olivier Cabrignac, Arthur Charpentier, Ewen Gallic

amse
école d'économie d'aix-marseille
aix-marseille school of economics

Working Papers / Documents de travail

Modeling Joint Lives within Families

Olivier Cabrignac
Arthur Charpentier
Ewen Gallic

WP 2020 - Nr 21

Modeling Joint Lives within Families

by

Olivier Cabrignac

SCOR
5 Avenue Kléber, 75795 Paris, France
ocabrignac@scor.com

Arthur Charpentier

Université du Québec à Montréal (UQAM)
201, avenue du Président-Kennedy,
Montréal (Québec), H2X 3Y7, Canada
charpentier.arthur@uqam.ca

Ewen Gallic

Aix-Marseille Univ., CNRS, EHESS, Centrale Marseille, AMSE
5-7 boulevard Maurice Bourdet CS 50498
13205 Marseille Cedex 01, France
ewen.gallic@univ-amu.fr

June 2020

A. Charpentier acknowledges the support of the Natural Sciences and Engineering Research Council of Canada Grant NSERC-2019-07077. E. Gallic acknowledges the support of the French National Research Agency Grant ANR-17-EURE-0020. Initial work on the dataset was funded by the ACTINFO chair, of the Institut Louis Bachelier. O. Cabrignac points out that this article is not meant to represent the position or opinion of SCOR.

Abstract

Family history is usually seen as a significant factor insurance companies look at when applying for a life insurance policy. Where it is used, family history of cardiovascular diseases, death by cancer, or family history of high blood pressure and diabetes could result in higher premiums or no coverage at all. In this article, we use massive (historical) data to study dependencies between life length within families. If joint life contracts (between a husband and a wife) have been long studied in actuarial literature, little is known about child and parents dependencies. We illustrate those dependencies using 19th century family trees in France, and quantify implications in annuities computations. For parents and children, we observe a modest but significant positive association between life lengths. It yields different estimates for remaining life expectancy, present values of annuities, or whole life insurance guarantee, given information about the parents (such as the number of parents alive). A similar but weaker pattern is observed when using information on grandparents.

JEL: C13; C18; C46; C55; J11; J12; G22; G32

Keywords: annuities; collaborative data; dependence; family history; genealogy; grandparents-grandchildren; information; joint life insurance; parents-children; whole life insurance

The authors thank the participants of the Online International Conference in Actuarial Science, Data Science and Finance (OICA) for stimulating questions. More particularly thanks to Tim J. Boonen and Montserrat Guillen for feedbacks. The authors also wish to thank Jerôme Galichon and Geneanet for kindly providing the dataset.

Contents

1 Introduction

Family history is usually seen as a significant factor insurance companies look at when applying for a life insurance policy.[1] As shown in Figure 1, family history of cardiovascular diseases, cancers (ovarian, breast, colon, lung, etc) or family history of high blood pressure and diabetes are usually asked in medical forms. Family can include parents (father and mother), siblings, spouse, children, and if deceased, the age at death can be asked. As discussed in Cutler and Zeckhauser (1998) or Pardo and Schott (2013), the information provided can yield higher insurance premiums, or no coverage at all.

Assuming that family history should have an impact on premiums means that the risk of the policyholder is correlated with information related to siblings, parents, etc. Frees et al. (1996) considered the case of insurance products with dependent mortality, on spouses, showing that life lenghts with a married couple are (positively) correlated, and that this correlation should reflect in standard joint life annuities. The literature has not limited itself to looking at life course relationships within couples. The idea of the existence of a "*longevity inheritance*", as Pearl (1931) named it, has given rise to a considerable body of work to study longevity ties within families. However, most studies have relied on modest sample sizes.

In this article, we use a massive dataset of family trees to study correlations between life lengths between relatives. This dataset concerns individuals born in metropolitan France at the beginning of the 19th century and their descendants. Due to the historical nature of the data, and in contrast with the individuals on which Frees et al. (1996) focused, the grandchildren of individuals born at the beginning of the 19th century have all died by now. Their age at the time of death is *de facto* observable. We therefore avoid dealing with incomplete and partial data when studying mortality.[2] In a first step, we use this rich data to explore joint mortality within couples. The analysis is based on 135,128 pairs of individuals born in the 19th century. It focuses solely on demographic characteristics. In a second and third steps, knowledge of the descendants makes it possible not only to investigate the links between children and parents, but also between children and their grandparents. To this end, we use 174,318 observations linking individuals to their parents where information about both parents is available, and 59,463 individuals for whom the birth and death dates of the 4 grandparents are known. In each of the three cases, we begin by studying the correlation of life spans. Then we

[1] Where it is allowed, *e.g.*, in North America or in the United Kingdom, for instance. Some data privacy regulation may, however, restrict access to those data. It should also be noted that such practice is not allowed in most European countries, as discussed in Schmitz and Wiesing (2006).

[2] For example, since we face complete data regarding age at death, a convenient way to approximate life expectancy is to simply use the average age at death, no modeling assumption is required here.

move from demographics to life insurance actuarial present values.

In line with what is reported in the literature, a positive association in mortality within couples is observed with our data. This dependence has a significant impact on joint life insurance product, consistent with results already obtained in previous literature (but usually with smaller datasets). Similarly, we also find a small but significant relationship between children and parents life lengths. With both parents still alive when twenty years old, life expectancy is relatively higher than when one or neither parent is still alive. Similar results are observed at older ages (30 and 40 years). Simulations based on historical data show that knowing whether a person's parents are still alive at any age has significant effects from an insurer's perspective. For example, for a young male individual (about 30 years old), knowing that both parents are still alive translates into an increase of about 2 years in his life expectancy relative to a person whose both parents are deceased. And at the same age, annuities should be 4% higher than the average population when both parents are alive, and 4% lower when both parents deceased. For a whole life insurance, the order of magnitude is almost the same (but opposite as insurance premiums are decreasing with life expectancy). Finally, the links are much less important between a person's mortality and that of their grandparents. As a result, the implications in terms of insurance are weaker and more questionable, with $\pm 1\%$, at best.

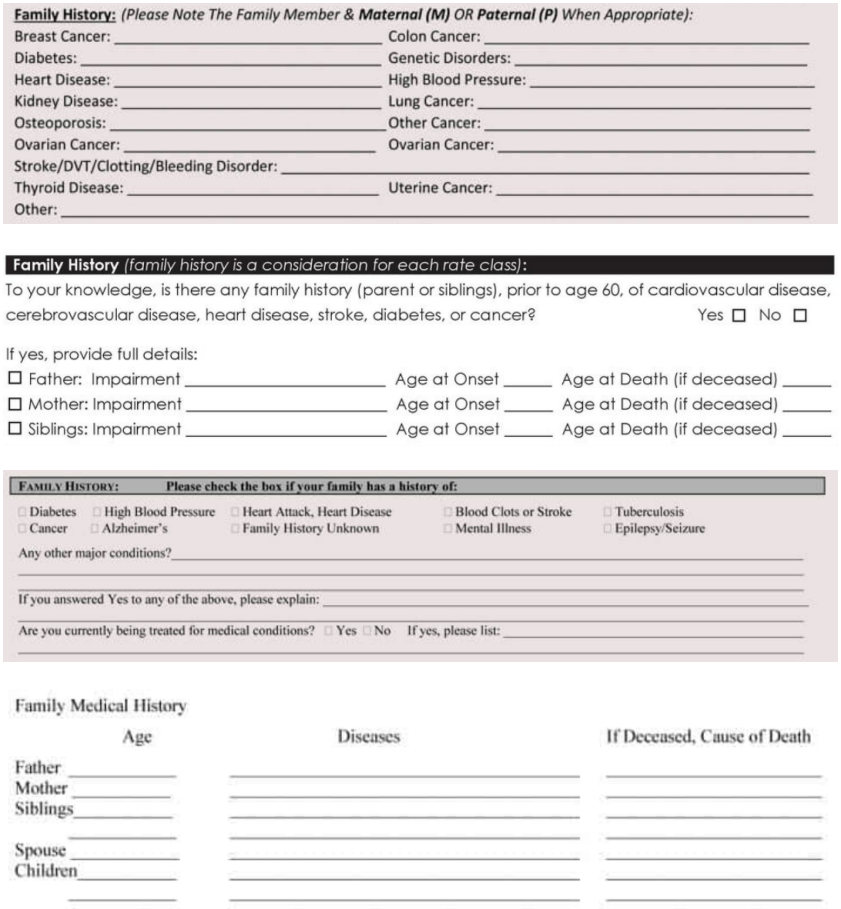
The remainder of the article is structured as follows. Section 2 concisely describes the datasets. Section 3 first recalls classical notations and concepts used when modeling (univariate and then joint) mortality and then presents various insurance products. In Section 4, based on 19th century data, joint life dependencies between husbands and wives is investigated. Section 5 discusses inter-generational dependencies for life lengths. More specifically, the links between an individual's remaining lifetime and information about his or her parents are first studied (these include, for example, determining the remaining life span of a 40 years old person, given that both his or her parents are either still alive, or dead). Then, a similar analysis is conducted to study the relationship between an individual's remaining lifetime and the characteristics of his or her grandparents. Section 6 concludes.

2 Description of the Data

The data used in this paper come from a genealogy website, Geneanet.[3] On this website, each user is invited to build their own family tree and can decide to share it with the rest of the community. We rely on these shared trees. Each user's tree contains, with varying degrees of completeness, information about the family members who make up the tree. This information concerns the events

[3] https://www.geneanet.org/.

Figure 1. Excerpts of medical history classical forms, with the *family medical history* section (source: https://printabletemplates.com/)



Family History: *(Please Note The Family Member &* ***Maternal (M)*** *OR* ***Paternal (P)*** *When Appropriate):*

Breast Cancer: _____ Colon Cancer: _____
Diabetes: _____ Genetic Disorders: _____
Heart Disease: _____ High Blood Pressure: _____
Kidney Disease: _____ Lung Cancer: _____
Osteoporosis: _____ Other Cancer: _____
Ovarian Cancer: _____ Ovarian Cancer: _____
Stroke/DVT/Clotting/Bleeding Disorder: _____
Thyroid Disease: _____ Uterine Cancer: _____
Other: _____

Family History *(family history is a consideration for each rate class)*:

To your knowledge, is there any family history (parent or siblings), prior to age 60, of cardiovascular disease, cerebrovascular disease, heart disease, stroke, diabetes, or cancer? Yes ☐ No ☐

If yes, provide full details:

☐ Father: Impairment _____ Age at Onset _____ Age at Death (if deceased) _____
☐ Mother: Impairment _____ Age at Onset _____ Age at Death (if deceased) _____
☐ Siblings: Impairment _____ Age at Onset _____ Age at Death (if deceased) _____

FAMILY HISTORY: **Please check the box if your family has a history of:**

☐ Diabetes ☐ High Blood Pressure ☐ Heart Attack, Heart Disease ☐ Blood Clots or Stroke ☐ Tuberculosis
☐ Cancer ☐ Alzheimer's ☐ Family History Unknown ☐ Mental Illness ☐ Epilepsy/Seizure

Any other major conditions? _____

If you answered Yes to any of the above, please explain: _____

Are you currently being treated for medical conditions? ☐ Yes ☐ No If yes, please list: _____

Family Medical History

	Age	Diseases	If Deceased, Cause of Death
Father			
Mother			
Siblings			
Spouse			
Children			

that can be found on civil and religious registers, *i.e.*, birth, marriage, if any, and death. More specifically, we are interested in the dates of birth and death of individuals, as well as their family relationships. The sample we have contains individuals born in metropolitan France between 1800 and 1804, their descendants (and their parents) – keeping only those born before 1900.[4] It is well established that information regarding parents is quite well represented, as long as the information exists, but information related to other first-degree relatives (siblings, children) – not to say second-degree relatives (cousins, grand-children) is more sparse. Thus, the extraction we have can have possible bias, as discussed in Charpentier and

[4]The scope of Geneanet's data mainly concerns European countries, and more particularly France.

Gallic (2020a).

As the data focus on individuals born before 1900, all of them are dead by now. We do not face any censored-observation problem. However, some challenges arise with these data. Since each user creates their own family tree, an ancestor common to the trees of multiple users may be present several times in the raw data. To avoid redundancy, we refer to Charpentier and Gallic (2020a,b) for more explanations about the methodology. In a nutshell, an important challenge was to deal with natural overlap of various family trees, and to identity an ancestor appearing in several trees as a unique individual.[5] In Charpentier and Gallic (2020a), we proved that while there were bias in collaborative genealogical data (compared with official demographic data), more specifically on infant and young age mortality, these data provided very accurate mortality information (on force of mortality or remaining life expectancy, for ages higher than 20). In Charpentier and Gallic (2020b), the goal was to study family migration, from a starting ancestor, and the challenge was to get a forward genealogical analysis (from ancestor to descendants) from backward genealogical data (from descendant to ancestors).

Once the trees have been matched and the data cleaned, a reference dataset is obtained containing, for each individual in the trees, information about: their birth and death dates, and a link to their parents' identifier. Two databases can then be created to study: (*i*) the dependencies within couples (in Section 4) , and (*ii*) the inter-generational dependencies for life lengths (in Section 5). The remainder of this section focuses on providing more detail on these two datasets.

2.1 Husband and Wife Dependencies

There is an intensive literature about the positive association of life spans between couples, and more specifically on bereavement. Riley (1983) suggests several sociological explanation of the positive association among life lengths, reinforced by mortality excess following the death of the first one. Some articles also focus on medical (mainly psychological) aspects. For example, 4,486 widowers of 55 years old (and older) have been followed up for nine years since the death of their wives in 1957 in Parkes et al. (1969). In the first six months about 25% of the deaths were from the same diagnostic group as the wife's death, and a higher mortality was observed. The authors concluded that there is no evidence suggesting that the proportion is any different among widows and widowers who have been bereaved for more than six months. Parkes et al. (1969) coined *broken heart syndrom* to describe that short term (positive) correlation. This was confirmed in Jagger and Sutton (1991), but on a smaller dataset (with 344 elderly persons who were living

[5]That process of merging trees had to take into account typos in names, and (partially) missing information about dates.

with a spouse and who were part of a survey of a population of people aged 75 years and over). Nevertheless, Kastenbaum and Costa (1977) claims that a large part of the excess of mortality following the death of the first one can be related to common health problems, rather than psychological trauma following the loss (see also more recently Espinosa and Evans (2008) that reached a similar conclusion).

We create a first set of data from family trees to see if the same type of associations that have been described in the literature between the lifespan of the members of a couple can be observed with our data. It is important to note that the term *couple* used in this study refers to a very specific definition. We do not rely on the definition of *married couple* (from census types of datasets), as in Glick (1955) for example. Like studies based on genealogical data, such as Beeton and Pearson (1901), we study here mothers and fathers. This comes from the structure of the data. Couples are observed as *parents*, and it is rather uncommon with genealogy data to have couples without children. We thus make two assumptions: (i) two people are defined here as a couple when they had a child together, and (ii), we further assume that parents were living together. Consequently, two persons who lived in a union but did not have a child (or without information on possible children available in the data) are not present in the dataset. This could lead to potential biases, but the main purpose of the section devoted to the study of joint mortality within couples is to see if the dataset in our hands provides similar results than previous studies.

Before going any further, it seems important to provide more details on the structure of the data. For individuals observed as *parents*, a table in which each row corresponds to a couple can be created. Each row contains the dates of birth and death of a father and a mother, as shown in the example provided in Table 1 (with only 6 random observations). It should be noted that information for all four dates may not be fully available. Two situations arise. First, the date of birth or death of one or both members of the couple may simply be missing. When such a case occurs, we remove the couple from the data. Secondly, some dates may be incomplete: the month and day may not be stated. In such a case, the date is converted into July 1st of that year, which corresponds to the average date assuming uniform birth over the calendar year[6]. Finally, we end up with a dataset of $n = 135,128$ couples, where husbands and wives were born between 1800 and 1870.

When this dataset is created, it is straightforward to calculate the ages at death of each member of the couple, that are denoted t_{f} and t_{m} for the father and the

[6]In that dataset, 14.3% of dates for men and 16.1% of dates for women were incomplete, usually more for death than birth (6.6% and 7.5% for birth dates for men and women respectively, and 10.2% and 11.3% of death dates). We assume here that the impact on various quantities (including correlation) would be rather small

Table 1. Dataset for the joint life model, father/husband and mother/wife.

	Father/Husband			Mother/Wife		
	birth	death	age	birth	death	age
1	1804-03-18	1880-09-21	76.5	1804-02-29	1854-08-26	50.5
2	1836-12-27	1902-04-06	65.3	1832-02-06	1901-09-11	69.6
3	1804-03-30	1870-01-17	65.8	1800-07-10	1868-07-12	68.0
4	1804-02-22	1876-01-06	71.9	1803-05-04	1881-05-16	78.0
5	1800-09-11	1837-08-15	36.9	1800-00-00	1836-08-11	36.1
6	1800-00-00	1843-03-08	42.7	1800-00-00	1865-03-15	64.7

Note: Random extraction of 6 rows of our entire dataset. In the case where only the year is mentioned, the date is eventually converted into July 1st of that year (mid-year, average date assuming uniform birth over the calendar year).

mother, respectively. As shown in Table 2, husbands and wives have a very similar life expectancy. (about 62 years old). As confirmed in Beltrán-Sánchez et al. (2015) the differences between men and women's life expectancies (admitted as a fact, nowadays) began to emerge only in the late 1800s, in most modern countries. It may be noted that the average ages at death we observe are relatively high for the period; it should be remembered that the individuals studied do not concern the entire population, but only those who have had children. Hence, the average is not driven down by the high infant mortality observed in 19th century France. The age difference in the couples studied shows that men are, on average, slightly older than their wives by two years, albeit with a significant standard deviation. In fact, in a majority of couples (58%), the man is older than his wife.

Table 2. Descriptive statistics for husbands and wives.

Variable	Mean	SD	Min	Max	Q_1	Q_2	Q_3
Husband age at death t_f	63.2	15.5	15	105	52.5	65.4	75.0
Wife age at death t_m	62.4	17.3	15	105	50.1	65.2	75.9
Age difference	-2.2	12.1	-80	80	-5.0	-1.0	1.0

Note: Husbands and wives were born between 1800 and 1870 (n = 135, 128). SD stands for standard deviation, Q_1, Q_2, and Q_3 stand for the first, second, and third empirical quartiles. The age difference indicates the number of years separating the man and the woman. Negative values indicate that the man is older than the woman.

2.2 Ancestors and Children Dependencies

We explained in the introduction that family history can be important in insurance. Almost a century ago, Pearl (1931) mentioned that "*even a business so precise in some particulars as life insurance, which has, from its beginnings, acted on the assumption that the duration of life of an individual's near kinsfolk is of importance in estimating the nature of its risk accepted in insuring him, has made singularly little effort to determine exactly the weight of this factor.*"

We explore family history in two parts, by analysing the relationship between an individual's lifespan and that of: (*i*) his or her parents, and (*ii*) his or her grandparents.

2.2.1 Parents and Children Dependencies

The historical reference on dependencies between life lengths of parents and children is probably Beeton and Pearson (1901) (with also a lot of concern about brothers and sisters - we refer to Bandeen Roche and Liang (1996) for a modern perspective on that issue). In that study, they also use 19th century genealogical data (which would give us an order of comparison, even if most computations were performed on about 1,000 pairs).

The idea of "*inheritance of longevity*" (as Pearl, 1931 named it) received a lot of attention, and has been intensively discussed in the literature over the years, with various types of data. For example, Gudmundsson et al. (2000) based on Icelandic population, concluded that longevity was inherited within families, in their view probably because of shared genes. Hjelmborg et al. (2006) looking at twin data, concluded that genetic influences on the lifespan were minimal before age 60 and only increase after that age. Kowald and Kirkwood (2016), on the other hand, rejected any idea that mortality in old age is genetically programmed. Consistent with that view, a Swedish study of men born in 1913, found that a number of social and behavioural factors measured at age 50 were better predictors of longevity than their parents' survivorship, predicted longevity Wilhelmsen et al. (2011). Caution must be exercised when advancing the idea of genetic transmission, especially when relying on genealogical data that provide no evidence of the veracity of the relationship between individuals. Indeed, as pointed out in Bellis et al. (2005), the median rates of paternal divergence as measured in studies conducted mainly during the second half of the 20th century amount to 3.7%, with some studies even advancing values as high as 30% (considering tests performed on selected populations for reasons other than disputed paternity).

Mayer (1991) studied up to six family pedigrees, from 1650 to 1874. An important issue in the literature is to understand if that correlation (or transmission)

is due to genetics, or associated life styles, or social class. For example, Ruby et al. (2018) explains that "*the majority of that correlation was also captured by correlations among non genetic (in-law) relatives, suggestive of highly assortative mating around life span-influencing factors (genetic and/or environmental)*". Philippe and Opitz (1978) mentions that spouse life spans "*correlate as much or more than those of genetic relatives*". See also Garibotti et al. (2006), Piraino et al. (2014) or Temby and Smith (2014), who try to distinguish genetic effects to socioeconomic status in family history. Philippe (1980) suggests that there could be an overestimation of the possible positive correlation, presents a feature of some *elite sub-class*, but not of the general population of a community. Almost all studies are based on rather small sample. For example, Abbott et al. (1978) studied 7,103 progeny, sons and daughters of 1,766 men or women, who were alive in 1922–1930. Matroos et al. (1984) had 2,370 middle-aged children, while Vaillant (1991) used a cohort of (only) 184 men.

Nevertheless, most studies confirm a significant but weak association within families. Bocquet-Appel and Jakobi (1990) analyzed familial correlations of longevity at Arthez d'Asson, for individuals born between 1686 and 1899. At birth, the correlation is rather small (0.103), but it increases with the age of the son. For example, at 20 years old, the correlation is larger (0.167). As Vaupel (1988) wrote it "*the life spans of parents and children appear only weakly related, even though parents affect their children's longevity through both genetic and environmental influences*" (see also Vaupel et al. (1979) or Vaupel et al. (1998), with similar conclusions).

To address the issues developed in this literature, a second set of data was created from the genealogy data. By using the parents' identifiers, and then those of the parents' parents, it is possible to create a table in which each observation provides the dates of birth and death of individuals, their parents, and their grandparents. As in the case of couples, dates may be missing or incomplete. If the treatment is the exact same for the case of incomplete dates, the treatment of missing dates differs depending on whether the focus is on parents or grandparents. For parents, as for couples, if one of the dates of birth or death is not available, then the individual is removed from the observations. For grandparents, if both dates of birth and death of all grandparents are missing, then the individual is removed from the observations. It should therefore be noted that if the information needed to calculate the age at death of at least one of the grandparents is available, the individual is kept.

The sample thus allows us to focus on the relationship between the age at death of an individual, t_c, and that of his or her parents, t_f and t_m, for the father and the mother, respectively. As reported in Table 3 the sample is composed of $174,318$ observations for which full information on ages at death is available. It consists of two parts of $90,828$ men and $83,490$ women. The average age at death of

individuals is 44.5 years (43.4 for men and 45.7 for women). We wish to compare this measure with that of the parents. This can be done by looking separately at the age at death of each parent, *i.e.*, t_{f} and t_{m}. It is also possible to construct indicators to represent the age at death of parents in the household. Three are under consideration. First, average age at death of the parents, mean$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$, which is equal to 64.5. It can be noted that it is relatively higher than that of the children. This can be explained by the fact that individuals who died at a young age are not taken into account: if parents are present in the data, they have necessarily had children (survivor bias). The sample of parents and children is constructed starting from children born between 1800 and 1900. Second, we look at the age at death of the first to die, min$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$, which quite a bit lower, 56.4 years on average. And third, we consider the age at death of the last survivor, max$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$, which is equal to 72.2 on average.

Table 3. Age at death of the individuals and that of their parents, according to the gender of the children.

	Mean	SD	Min	Max	Q_1	Q_3
	Men ($n = 90,828$)					
Individual (t_{c})	43.4	30.1	0	104.4	10.2	70.0
Father (t_{f})	64.7	14.6	15	104.0	54.8	75.7
Mother (t_{m})	64.2	16.4	15	104.8	53.1	76.8
First to die (min$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	56.6	14.7	15	100.0	45.6	68.2
Last survivor (max$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	72.3	12.0	15	104.8	65.7	80.8
Average parents (mean$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	64.4	11.8	15	100.9	56.5	73.3
	Women ($n = 83,490$)					
Individual (t_{c})	45.7	32.0	0	104.7	9.9	74.7
Father (t_{f})	64.3	14.8	15	104.0	54.2	75.6
Mother (t_{m})	64.0	16.6	15	104.4	52.5	76.7
First to die (min$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	56.2	14.8	15	100.8	45.1	67.9
Last survivor (max$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	72.1	12.2	15	104.4	65.3	80.8
Average parents (mean$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	64.2	11.9	15	101.4	56.1	73.0
	Men & Women ($n = 174,318$)					
Individual (t_{c})	44.5	31.1	0	104.7	10.0	72.2
Father (t_{f})	64.5	14.7	15	104.0	54.5	75.6
Mother (t_{m})	64.1	16.5	15	104.8	52.8	76.7
First to die (min$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	56.4	14.7	15	100.8	45.3	68.0
Last survivor (max$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	72.2	12.1	15	104.8	65.5	80.8
Average parents (mean$\{t_{\mathrm{f}}, t_{\mathrm{m}}\}$)	64.3	11.8	15	101.4	56.3	73.2

Note: this table provides descriptive statistics of the ages contained in the dataset of children and parents. n stands for the number of observations, SD is the standard deviation, Q_1 and Q_3 are the first and third empirical quartiles, respectively.

Table 4 shows the distribution of the number of individuals by decade. Because of the way in which the sample was constructed, starting with individuals born in France between 1800 and 1804, a relatively high proportion of individuals in the raw data belong to the cohorts (1790 – 1800] and (1800 – 1810].[7] However, incomplete information on the dates of birth and death of both parents for these people is only rarely available. As a result, the data used in the analysis contain few individuals from these cohorts. Also due to the constitution of the sample, the number of individuals in the (1810, 1820) cohort is very small, since we must wait until individuals born between 1800 and 1804 have had children before they can take on the role of parents in the database. For subsequent cohorts the observed average values of the ages at death of the children or their parents remain within the same orders of magnitude as for the rest of the sample.

Table 4. Information about individuals per cohort.

Cohort	n	Age individual t_c	Age father t_f	Age mother t_m	Prop. Women (%)
(1790,1800]	476	42.5	62.0	59.4	47.7
(1800,1810]	3, 332	39.9	62.8	59.9	47.8
(1810,1820]	250	47.4	58.7	58.4	42.8
(1820,1830]	30, 445	42.3	63.4	62.3	46.5
(1830,1840]	41, 345	38.8	65.2	64.4	46.9
(1840,1850]	12, 492	36.9	66.4	66.1	47.4
(1850,1860]	11, 409	40.6	62.7	61.4	48.7
(1860,1870]	19, 311	42.8	63.1	62.1	49.2
(1870,1880]	14, 813	48.8	64.2	63.2	50.0
(1880,1890]	15, 862	53.2	64.8	65.4	48.8
(1890,1900]	24, 583	56.2	66.0	68.3	48.4

Note: This table reports some key descriptive statistics for the individuals contained in the dataset used to study the dependence between children and parents. n stands for the number of observations, Prop. Women gives the proportion of women among the individuals, for each cohort.

2.2.2 Grandparents and Children Dependencies

As mentioned recently in Choi (2020), "*little is known about whether and how intergenerational relationships influence older adult mortality*", especially between children and their grandparents. Our sample makes it possible to investigate this question. We adopt the following notations: t_{gff} and t_{gmf} for the grandfather and grandmother on the father's side, and t_{gfm} and t_{gmm} for the grandfather and grandmother on the mother's side.

As with the relationship between parents and children, Table 5 reports some descriptive statistics on the age at death of the grand parents, keeping only the

[7] The 1790-1800 cohort is composed of individuals born in 1800 only.

individuals for whom birth and death dates are known for at least one grandparent (which does not necessarily imply that the birth and death dates of both *parents* are complete).[8] As can be seen, the distribution of males is higher than that of females (831,479 versus 740,461). The average age at death of grandchildren is 43.7 (41.9 for men and 45.8 for women). As is the case for what is observed with parents, the age at death of grandparents is much higher in our data, ranging from 62.8 to 64.6 years on average. This difference is mainly explained by some survivor bias: grandparents have been parents, so at least, they reach the 20's, while their grandchildren include a lot of individuals who died very early. For instance, 25% of the grand children did not live beyond 6 years old. The three indicators that make it possible to subsequently synthesize the relationship between an individual's lifespan and that of his or her grandparents (*i.e.*, last survivor, first to die and average age at death) provide the same orders of magnitude as those for parents.

[8]It should be noted that the number of individuals for whom information on all four grandparents is available is small relative to the sample size: 31,096 men and 28,367 women. See Table C3 in the Appendix. On average, the raw data provide information only on 1.6 grandparents (1st quartile is 1 and 3rd quartile is 2).

Table 5. Age at death of the individuals and age at death of their grandparents, according to the gender of the grandchildren.

	Mean	SD	Min	Max	Q_1	Q_2	Q_3	No. Missing
	Men (n = 831, 479)							
Individual (t_c)	41.9	31.0	0	104.7	5.4	45.3	70.4	0
Maternal grandfather	64.1	14.6	15	104.7	54.1	66.1	75.2	526,191
Maternal grandmother	62.8	16.0	15	104.4	51.8	65.3	75.1	511,685
Paternal grandfather	64.6	14.3	15	104.7	55.1	66.7	75.4	492,622
Paternal grandmother	63.3	15.7	15	104.4	52.8	65.8	75.2	489,805
Last survivor	67.6	14.4	15	104.7	59.3	70.3	78.1	0
First to die	59.6	15.5	15	104.7	48.2	61.0	71.4	0
Average grandparents	63.6	13.6	15	104.7	55.1	65.1	73.5	0
	Women (n = 740, 461)							
Individual (t_c)	45.8	33.1	0	104.9	7.0	52.5	76.4	0
Maternal grandfather	64.1	14.6	15	104.7	54.2	66.1	75.1	461,575
Maternal grandmother	62.8	16.0	15	104.8	51.9	65.3	75.1	446,549
Paternal grandfather	64.6	14.4	15	104.7	55.0	66.6	75.5	444,459
Paternal grandmother	63.3	15.7	15	104.9	52.9	65.8	75.2	439,981
Last survivor	67.7	14.4	15	104.9	59.4	70.3	78.2	0
First to die	59.5	15.5	15	104.9	48.1	60.9	71.3	0
Average grandparents	63.6	13.6	15	104.9	55.1	65.1	73.5	0
	Men & Women (n = 1, 571, 940)							
Individual (t_c)	43.7	32.0	0	104.9	6.1	48.4	73.3	0
Maternal grandfather	64.1	14.6	15	104.7	54.2	66.1	75.2	987,766
Maternal grandmother	62.8	16.0	15	104.8	51.8	65.3	75.1	958,234
Paternal grandfather	64.6	14.4	15	104.7	55.0	66.7	75.4	937,081
Paternal grandmother	63.3	15.7	15	104.9	52.8	65.8	75.2	929,786
Last survivor	67.6	14.4	15	104.9	59.4	70.3	78.2	0
First to die	59.5	15.5	15	104.9	48.1	61.0	71.4	0
Average grandparents	63.6	13.6	15	104.9	55.1	65.1	73.5	0

Note: this table provides descriptive statistics of the ages contained in the dataset of grandchildren and their grandparents. n stands for the number of observations, SD is the standard deviation, Q_1, Q_2 and Q_3 are the first, second, and third empirical quartiles, respectively, and No. Missing refers to the number of missing values.

Table 6 gives a better idea of the distribution of the data according to the birth cohort of individuals for whom relationships with their grandparents are studied. The table gives the number of observations for each cohort as well as the average age at death of grandchildren (regardless of gender) and each grandparent. The upper part gives, the information for grandchildren for whom the birth and death dates of at least one grandparent are known. The lower part is restricted to grandchildren whose birth and death dates of all four grandparents are known. Again, due to the initial sample design, very few people are present for the first three cohorts (however, there are still a substantial number of individuals in the sample, with

$n = 59,463$ individuals for whom full detailed information about the four grandparents is available).

Table 6. Information about individuals per cohort for the dependencies between individuals and their grandparents when partial information is known on grandparents (top) and when all the grand parents are known (bottom).

Cohort	n	t_c	t_{mgf}	t_{mgm}	t_{pgf}	t_{pgm}	Prop. Women (%)
	All individuals with information on at least one grandparent						
(1790,1800]	1,277	45.7	66.7	65.6	66.8	65.0	46.9
(1800,1810]	6,375	42.8	66.9	66.3	67.4	66.5	48.2
(1810,1820]	5,611	42.8	61.6	60.3	62.9	61.4	47.7
(1820,1830]	178,975	40.5	63.8	62.2	63.8	62.3	46.8
(1830,1840]	252,818	37.1	64.6	62.9	64.8	63.0	46.6
(1840,1850]	119,729	35.5	64.1	62.3	65.2	63.2	46.9
(1850,1860]	145,099	36.9	63.3	62.5	63.8	62.4	46.8
(1860,1870]	216,136	40.3	64.8	63.9	64.8	63.8	46.8
(1870,1880]	201,357	47.6	65.1	64.0	65.6	64.7	47.6
(1880,1890]	202,352	51.5	63.4	62.4	64.5	63.6	47.8
(1890,1900]	242,211	54.7	63.3	62.6	63.8	63.4	47.4
	All four grandparents known						
(1790,1800]	1	52.4	59.2	72.5	56.8	46.6	0.0
(1800,1810]	9	47.4	69.2	56.8	66.9	64.6	44.4
(1810,1820]	128	45.8	61.4	60.0	60.6	61.6	38.3
(1820,1830]	19,468	40.9	64.1	63.0	63.8	62.7	47.1
(1830,1840]	26,651	37.7	64.8	63.4	64.8	63.1	47.2
(1840,1850]	7,093	35.2	65.1	63.4	65.1	63.3	48.1
(1850,1860]	1,229	36.7	65.7	63.8	63.5	62.6	51.3
(1860,1870]	2,090	38.5	65.1	64.3	63.8	63.1	51.0
(1870,1880]	1,294	46.4	65.3	64.1	64.5	63.0	52.9
(1880,1890]	610	49.9	63.5	62.1	63.6	61.6	51.6
(1890,1900]	890	54.5	63.1	62.1	62.5	61.8	50.6

Note: This table reports some key descriptive statistics for the individuals contained in the dataset used to study the dependence between children and grandparents. n stands for the number of observations, t_c, t_{gfm}, t_{gmm}, t_{gff}, and t_{gmf} stand for the age at death of individuals, maternal grandfather, maternal grandmother, paternal grandfather, paternal grandmother, respectively. Prop. Women gives the proportion of women among the individuals, for each cohort.

3 From Mortality Models to Insurance Premiums

In this section, we introduce general notation that we will use when discussing the impact of joint life dependencies, withing families. We start with general notations to model univariate mortality, and then introduce copulas which describe the joint distribution. Inference is also discussed in that first part. Then, we present various insurance products, that we will price in various contexts, such as a life insurance and pension annuities (as well as joint life related guarantees). We might stress here that we focus on contemporary insurance guarantees: the goal is not to discuss historical insurance prices in the early 19th century, but to understand the impact of possible dependencies between family relatives on insurance prices based on the massive (historical) data we have.[9]

3.1 Univariate Mortality

The lifetime of a newborn is modelled by a positive variable T, with cumulative distribution function F (with $F(t) = \mathbb{P}[T \leq t]$ for any positive t) and survival function S (with $S(t) = \mathbb{P}[T > t]$). Let T_x denote the remaining lifetime of the person at age x, in the sense that $T_x = (T - x)|T > x$. Conditional cumulative distribution function is F_x (with $F_x(t) = \mathbb{P}[T_x \leq t] = \mathbb{P}[T - x \leq t|T > x]$) which is also denoted ${}_tq_x$ in actuarial literature, while the survival distribution function is S_x (with $S_x(t) = \mathbb{P}[T_x > t] = \mathbb{P}[T - x > t|T > x]$) which is also denoted ${}_tp_x$. One can write, for $t > 0$

$$ {}_tp_x = \mathbb{P}[T - x > t|T > x] = \frac{\mathbb{P}[T > t + x]}{\mathbb{P}[T > x]} = \frac{S(x + t)}{S(x)},$$

with the convention that ${}_0p_x = 1$. Assuming that T is an absolutely continuous random variable allows us to consider the density of $T(x)$, denoted f_x which satisfies

$$f_x(t) = {}_tp_x \cdot \mu(x + t),$$

where μ is the hazard rate, also called force of mortality in demographic and actuarial applications. In the life table terminology, L_x denotes the size of a cohort-type group, at age x, with $L_0 = 100,000$. Note that $L_x = L_0 \cdot {}_xp_0$.

And finally, curtate life expectancy for individual (x) is defined as

$$e_x = \mathbb{E}\big(\lfloor T_x \rfloor\big) = \mathbb{E}\big(\lfloor T - x \rfloor | T > x\big) = \sum_{t=0}^{\infty} {}_tp_x \cdot q_{x+t} = \sum_{t=1}^{\infty} {}_tp_x,$$

[9] Due to recent regulation on personal data in most countries, running such a study on contemporary data would be much more difficult. If mortality *per se* changed a lot, we study the impact of dependencies on insurance prices based on order of magnitudes of correlations obtained on historical data.

for some integer $x \in \mathbb{N}$, using notations from Bowers et al. (1997), where we count here the expected number of future years completed by (x) prior to death (to contrast, the complete expectation of life, $\mathbb{E}(T_x)$, is denoted $\mathring{e}_x$).

In the context of small amount of data, it is natural to use parametric models to compute complex quantities. This will be the case with joint life models, to derive more robust estimates (but probably more model sensitive). Classical parametric models for mortality are Gompertz distribution, with $\mu(x) = Ae^{Bx}$ (from Gompertz, 1825) or Beard with $\mu(x) = Ae^{B^x}/(1+KAe^{B^x})$ (from Beard, 1971). Carriere (1992) suggested to use mixtures of distribution, with $S(x) = \psi_1 S_1(x) + \psi_2 S_2(x) + \psi_3 S_3(x)$ from standard survival families (that will be the Carriere model). Finally, the Heligman-Pollard with $q(x)/p(x) = A^{(x+B)^C} + De^{-E(\log x - \log F)^2} + GH^x$ (from Heligman and Pollard, 1980). The later was initially fitted on Australian mortality and performed fairly well at all age. That is also what we observe on our data. The first part $A^{(x+B)^C}$ is a rapidly declining exponential, that reflects the fall in mortality during the early childhood years (this component of mortality has three parameters, A, which is nearly equal to $1 - q_0$, B which is a location factor, and C which measures the rate of mortality decline in childhood). The third term GH^x is the Gompertz model, that reflects the near geometric rise in mortality at the adult ages. And finally, the second term $De^{-E(\log x - \log F)^2}$ reminds of a lognormal model, that reflects accident mortality (as discussed in Heligman and Pollard, 1980, F indicates location, E the spread, and D the severity). The adjustment of these four distributions on male mortality based on our data can be visualized graphically. In Figure 2, we can compare the parametric models to raw values (force of mortality $\mu(x)$ and survival probabilities ${}_xp_0$ or L_x/L_0). The mixed distribution suggested in Carriere (1992) provides a very good fit (and not the other standard models). Computations were performed using the `MortalityLaws` R package (see Pascariu, 2020)

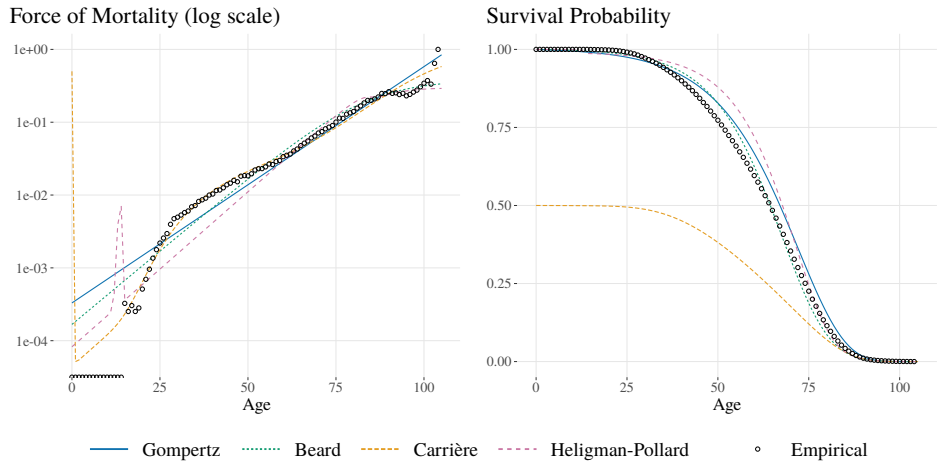
3.2 Modeling the Dependence Structure

A classical model to describe the joint distribution of a pair of lifetimes (T_1, T_2), – for remaining lifetimes of two individuals, which could be either (T_f, T_m) for the father and the mother, or (T_m, T_c) for the mother and a child of hers – is to use a copula function, so that the joint survival function

$$S(t_1, t_2) = \mathbb{P}[T_1 > t_1, T_2 > t_2] = C\big(S_1(t_1), S_2(t_2)\big),$$

where $C : [0,1]^2 \to [0,1]$ is a copula function (see Joe, 1997 or Nelsen, 1999, but in the context of survival lifetimes, using the survival copula makes more

Figure 2. Marginal mortality, with force of mortality for males on the left, μ_x, and the survival function on the right.



sense.).[10] Some parametric family can be considered for C. The independent and the comonotonic copulas are defined as $C_\perp(u,v) = uv$ and $C_+(u,v) = \min\{u,v\}$. In Appendix A, popular parametric copulas are mentioned. In a nutshell, for inference, we use here pseudo observations based on ranks, $(\widehat{u}_i, \widehat{v}_i)$ where

$$\widehat{u}_i = \widehat{S}_1(t_{1,i}) \text{ where } \widehat{S}_1(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}(t_{1,j} > t),$$

for the first type of individuals (say *father* for joint life in a couple), and a similar expression for the second type[11].

We define the empirical copula $\widehat{C}_n$ as the cumulative distribution function of $(\widehat{u}_i, \widehat{v}_i)$'s

$$\widehat{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\big(\widehat{u}_i \le u, \widehat{v}_i \le v\big),$$

or some smooth version $\widetilde{C}_n(u,v)$ using some probit transformation, as in Geenens et al. (2017). In the case where copulas are non-symmetric, remember that since we look at the survival copula, the lower corner $(0,0)$ corresponds to small probabilities (and therefore large ages).

[10] Note that C is usually called the survival copula of the pair (T_1, T_2).

[11] Here we assume that data consist in pairs $(t_{1,i}, t_{2,i})$ of ages at dead, but actually, we have the *dates* of deaths, so a dynamic model could be considered, this point will be addressed in the next section.

3.3 Impact on Annuities and Life Insurance Premiums

In the section, we present various insurance products and their notations. These products will allow us to compute various quantities at time t, when the insurance contract is signed. In the case of a single life guarantee, we consider an individual age x (at time t), while for multiple lives (here father and mother), both have age x_f and x_m respectively. As discussed in the next section, most quantities are rather stable over time, so notation t will not be used here. Furthermore, as mentioned at the beginning of that section, we consider here 'contemporary' insurance guarantees, even if we use historical data.

3.3.1 Notations for Single Life Guarantees

For financial application, let v denote the discount factor associated with the (constant) annual rate i, in the sense that $v = (1+i)^{-1}$. Formally, we compute the expected present value, of future cash flow, for a sequence $c_1, \ldots, c_n$, at time $t_1, \ldots, t_n$, where the expected value is computed related to probabilities of paying, denoted generally $p_1, \ldots, p_n$,

$$\sum_{k=1}^{n} v^{t_i} c_i p_i.$$

For example, whole life insurance are contractual guarantees that promise a fixed amount (a normalized cash flow of \$1) *at the time of death*. The expected present value of such a contract is

$$\overline{A}_x = \sum_{k=0}^{\infty} v^k \, {}_kp_x q_{x+k},$$

or, when payment is made at the end of the year of the death

$$A_x = \sum_{k=0}^{\infty} v^{k+1} \, {}_kp_x q_{x+k}.$$

The n-year term insurance provides a fixed amount only if the death occurs in the next n years, and the expected present value

$$A^1_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^{k+1} \, {}_kp_x q_{x+k},$$

again, with a payment at the time of death. A n-year endowment guarantees a payment at the end of the n years if the person survives. The expected present value is here

$${}_nE_x = A_{x:\overline{n}|}^{\ 1} = v^n \, {}_np_x.$$

For example, annuities are contractual guarantees that promise a *periodic income* (usually annually, with a normalized cash flow of \$1) over the lifetimes of individuals. The actuarial present value of the annuity of an individual age (x) is

$$a_x = \sum_{k=0}^{\infty} v^{k+1}{}_k p_x,$$

where payment is made at the end of the year if individual is still alive (as in Bowers et al. (1997), $\ddot{a}_x$ is used for payments done at the beginning of the year – also called *annuity-due*). Such a contract pays \$1 at the end of the years 1, 2, 3, ... as long as the individuals is alive. As previously, an n-year temporary life annuity guarantees yearly unit cash flows, until year n, over the lifetimes of individual (x)

$$a_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^{k+1}{}_k p_x.$$

We refer to Bowers et al. (1997) for technical distinctions among the various subtleties. Here, we simply compare life insurance expected present values, and annuities, in various scenarios. To that end, use numerical tools developed for the `lifecontingencies` R package, presented in Spedicato (2013).

3.3.2 Notations for Conditional Single Life Guarantees

In Section 5, we will consider single life guarantees, using family history as conditional information, at the time of signature. For instance, we can consider the actuarial present value of the annuity of an individual age (x), given some information $\star$ about parents or grand-parents, is

$$a_x^{\star} = \sum_{k=0}^{\infty} v^{k+1}{}_k p_x^{\star},$$

where $\star$ could mean that both parents are still alive when the child has age x, for example, or that only one of them is alive.

3.3.3 Notations for Multiple Life Guarantees

In section 4, we will also consider joint and survivor annuities options on two live contracts. More precisely, as discussed earlier, we consider the case of married couples, where the two lives are the one of the father (denoted f) and the one of the mother (denoted m). Let $T(x_f, x_m)$ denote the joint life status and $T(\overline{x}_f, \overline{x}_m)$ the last survivor, defined as

$$T(x_f, x_m) = \min\{T_m(x_f), T_f(x_m)\} \text{ and } T(\overline{x}_f, \overline{x}_m) = \max\{T_f(x_f), T_m(x_m)\}.$$

Survival probabilities are

$$ {}_tp_{x_f,x_m} = \mathbb{P}[T(x_f,x_m) > t] \text{ and } {}_tp_{\overline{x}_f,\overline{x}_m} = \mathbb{P}[T(\overline{x}_f,\overline{x}_m) > t], $$

while curtate life expectancies are

$$ e_{x_f,x_m} = \mathbb{E}\big(\lfloor T(x_f,x_m)\rfloor\big) = \sum_{t=1}^{\infty} {}_tp_{x_f,x_m} \text{ and } e_{\overline{x}_f,\overline{x}_m} = \mathbb{E}[T(\overline{x}_f,\overline{x}_m)] = \sum_{t=1}^{\infty} {}_tp_{\overline{x}_f,\overline{x}_m}. $$

A *joint-life annuity* pays benefits until the death of the first of the two annuitants, $T(x_f,x_m)$, for a husband/father (x_f) and a wife/mother (x_m). The standard joint-life annuity pays \$1 at the end of the years 1, 2, 3, ... as long as both spouses survive. Its actuarial present value is

$$ a_{x_f,x_m} = \sum_{k=1}^{\infty} v^k {}_kp_{x_f,x_m}. $$

A *last-survivor annuity* pays a certain amount until the second (and last) death, $T(\overline{x}_f,\overline{x}_m)$. The standard joint-life annuity pays \$1 at the end of the years 1, 2, 3, ... as long as long as either spouses survives. Its actuarial present value is

$$ a_{\overline{x}_f,\overline{x}_m} = \sum_{k=1}^{\infty} v^k {}_kp_{\overline{x}_f,\overline{x}_m}. $$

A *reversion annuity* starts after the first death $T(x_m,x_f)$ until the last one $T(\overline{x}_f,\overline{x}_m)$. Some one-way reversion can be considered, for instance from the husband to the wife, also called *widow's pension*: payments start at the death of (x_f) until the death of (x_m), and no payment is made if (x_m) dies before (x_f). The actuarial present value of the *reversion annuity* is $a_{x_f,x_m} - a_{\overline{x}_f,\overline{x}_m}$, while the actuarial present value of the *widow's pension* is $a_{x_m|x_f} = a_{x_m} - a_{x_f,x_m}$.

4 Husband and Wife Dependencies

This section highlights, as observed in the literature, the positive links between the life expectancy of the members of a couple, and then examines the impacts on annuities and life insurance premiums.

4.1 Empirical Relationship Between Lifespans Within Couples

The nonparametric[12] estimation of the (survival) copula of remaining lifetimes (T_{x_f},T_{x_m}), can be seen in Figure 3. The nonparametric estimator suggest very

[12]Estimated parametric densities – Gaussian, Clayton and Gumbel – are mentioned in the Appendices.

similar behavior in the lower and in the upper tails, which would disqualify Gumbel and Clayton copula. Based on almost 15,000 individuals (but unfortunately most were censored data since a lot of people were still alive), Frees et al. (1996) suggested to use Frank copula (exhibiting symmetric dependence between the lower and the upper tail), with $\widehat{\theta} = 3.367$, corresponding to a 0.5 Spearman correlation. Denuit et al. (2001) selected at random two cemeteries in Brussels (Koekelberg and Ixelles / Elsene) and they collected the ages at death of 533 couples buried there. Those data are very close to the ones we have, and they observed a 0.139 Spearman correlation, and they used Gumbel copula, with parameter $\widehat{\theta} = 0.104$. In our data – see Figure 4 – Spearman correlation was 0.156, with a 95% confidence interval $(0.151; 0, 161)$.

The link between the age at death of a person and that of their spouse can be visualized in a simple way through several methods. Figure 5 shows a woman's age at death as a function of her husband's age at death as well as the relationship in the other direction. The positive correlation between these two quantities can thus be seen graphically, as in Beeton and Pearson (1901).

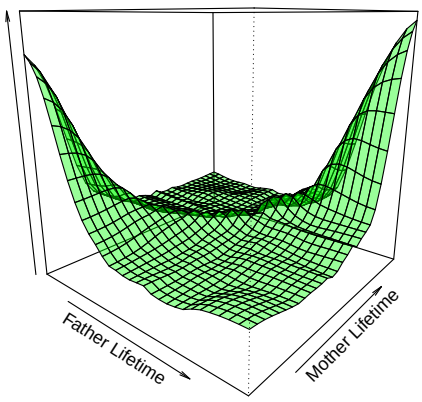
4.2 Annuities and Life Insurance Premiums Within Couples

As mentioned previously, the positive relationship exhibited between the lifetime of the members of a couple can be used to derive bounds for most actuarial quantities. To that end, we consider an annuity and a life insurance signed by a man. We compare their present value according to the age of the annuitant and distinguishing between cases where the policyholder's wife is alive or deceased at the time of signature. The results are shown in Figure 6, for an annuity (on the left) and for a life insurance (on the right). The values are expressed relative to the case where information on the wife's status is not taken into account, represented by the blue dashed line. This reference situation is compared to those where the current values of the two types of contracts are calculated by separating the men whose wives are still alive when the contract is signed (black solid line) from those whose wives are deceased (pink dashed line). In Figure 6 (and all figures where relative differences are computed), the baseline is the entire population. For a male, age 60, the expected value of the pension should be 2% larger if his wife is still alive, for example, but the premium of a life insurance is only 5‰ lower.

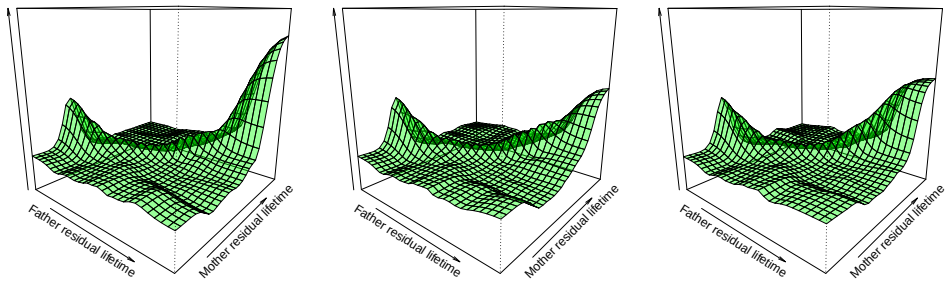
Figure 7 compares the cost of a standard widow's annuity $a_{\mathrm{m|f}}$ relative to the cost of the annuity under the assumption of (statistically) independence. As explained in the appendix, since we have a positive relationship between joint lives, widow's pension should be lower than under the assumption that joint lives are independent. For a mother in her 30's, the value of the widow's pension should be about 10% lower (than the independent case), while it should be 7% lower if she is in her 60's.

Figure 3. Nonparametric estimate of the copula density for (T_f, T_m), given some restrictions on x.

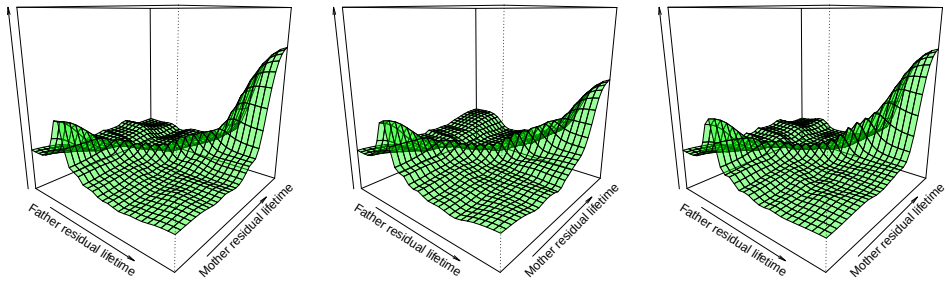
(a) (T_f, T_m) without restrictions on x



(b) $T_f \geq 25$ (c) $T_f \geq 35$ (d) $T_f \geq 45$

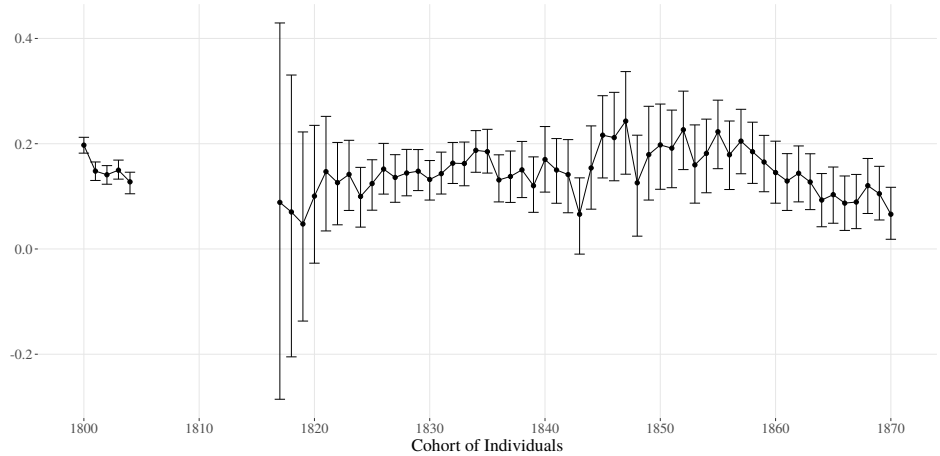


(e) $T_m \geq 25$ (f) $T_m \geq 35$ (g) $T_m \geq 45$



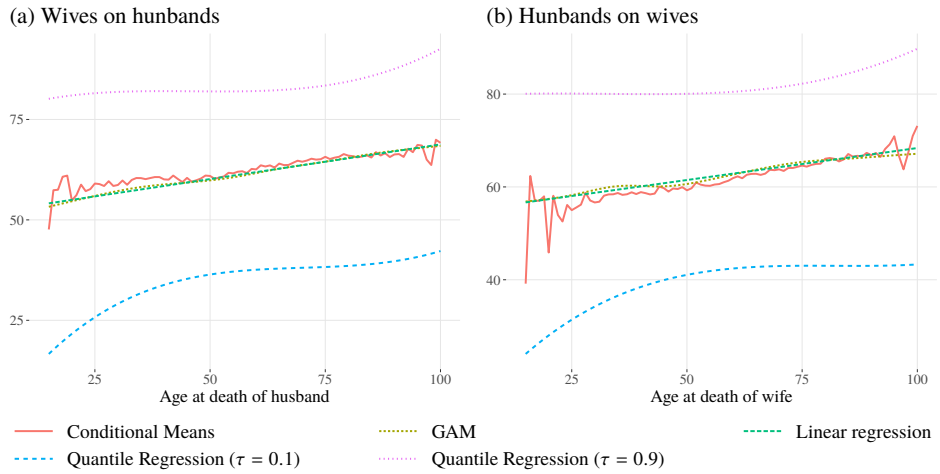
Note: On top, nonparametric estimate of the (survical) copula density for (T_f, T_m). Below, on top, copulas of (T_f, T_m) given $T_f \geq x$ (and that both are still alive), for $x = 25, 35, 45$. At the bottom, copulas of (T_f, T_m) given $T_m \geq x$ (and that both are still alive), for $x = 25, 35, 45$. The z-axis for copula density graphs is always $[0, 5]$, which allows us to compare all distributions.

Figure 4. Spearman correlation between age at death of the spouses, by year of birth of the husband.



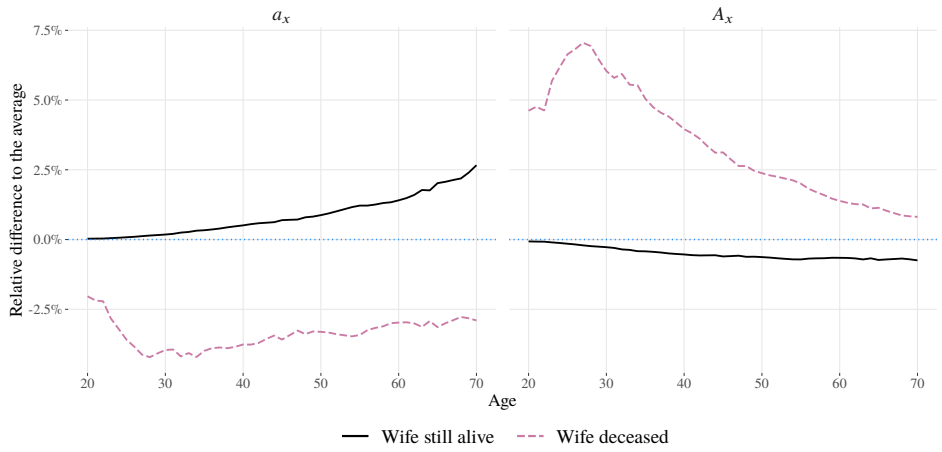
Note: The dots represent the estimated Spearman correlation between the age at death of spouses for each cohort, using the birth year of the father to define the cohorts. The bars correspond to 95% bootstrap confidence interval (based on 1, 000 resamples).

Figure 5. Age at death of women as a function of age at death of their husband (left) and age at death of men as a function of the age at death of their wife (right).



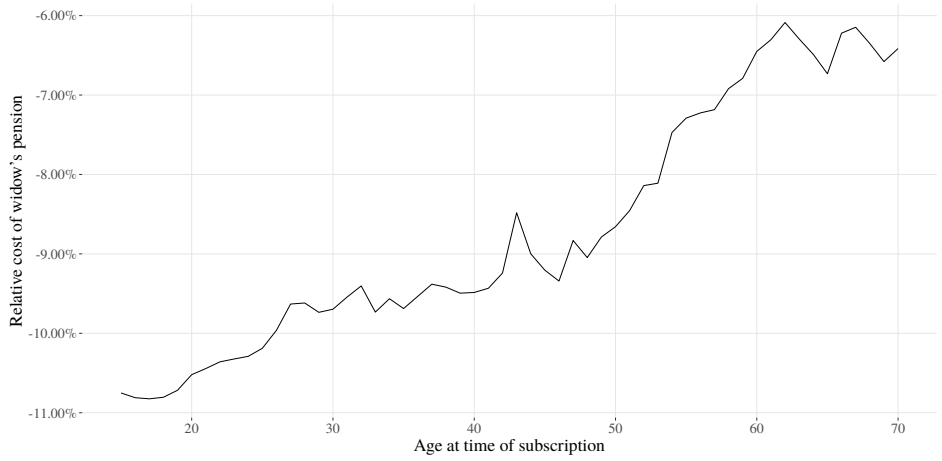
Note: The blue and purple dashed lines correspond to the first and ninth deciles of quantile regression of age at death of an individual as a function of the age at death of their spouse.

Figure 6. Relative difference to the average (in %) of present value of an annuity (left) and expected present value for a life insurance (right) depending on the age of the annuitant and on the death status of the wife at the time of the contract.



Note: the horizontal blue dashed line corresponds to the average value calculated on all individuals, regardless of the death status of the spouse. The annuities are calculated for 100 terms and the expected present value for life insurance are calculated for 100-year coverage. The interest rate is assumed to be 3%.

Figure 7. Widow's pension, $a_{\mathrm{m|f}}$ (relative to independent case $a^{\perp}_{\mathrm{m|f}}$), as a function of x_{m}.



5 Parents and Children Dependencies

This section first presents the results of the analysis of the links between the lifespan of an individual and that of his or her parents, and then considers the links with the

lifespan of grandparents.

5.1 Children Conditional on Parents

As noted in Section 2, several parametric models were fitted to estimate individual mortality. The adjustments of the Gompertz, Beard, Carrière and Hellingman-Pollard distributions on the force of mortality and survival probabilty, respectively, for men, women, fathers and mothers are displayed in the Appendix, in Figures B1 and B2. The graphs also report the observed values. The Carrière model is the one that best fits the data.

5.1.1 Empirical Evidence of the Relationship Between an Individual's Lifespan and that of his or her Parents

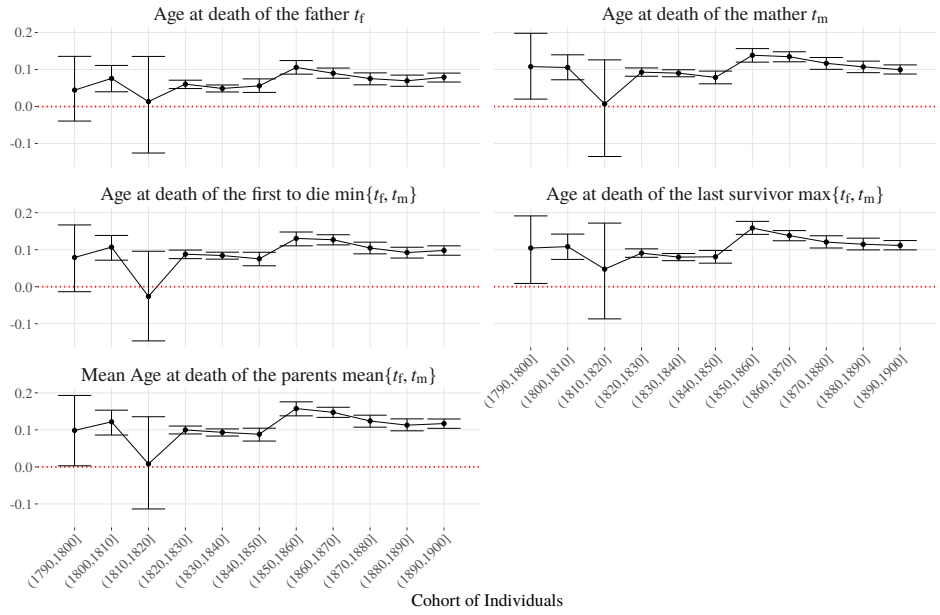
As was done in the case of couples, a first way to visualize the relationship between the age at death of children and their parent's age at death is to look at the correlation between t_c and various variables: the age at death of the father t_f, the age at death of the mother t_m, the age at death of the first to die $\min\{t_f, t_m\}$, the age at death of the last survivor $\max\{t_f, t_m\}$, and the average age at death of the parents $\text{mean}\{t_f, t_m\}$. As can be seen in Figure 8, this correlation is positive, albeit relatively weak and appears to be constant over cohorts. Overall, regardless of the cohorts, the Spearman correlation between t_c and $\text{mean}\{t_f, t_m\}$ is 0.125, with a 95% bootstrap confidence interval of $[0.121; 0.130]$.

These small links between parents and children are also seen in Figure 9, which shows the relationships between t_c and the same age at death variables as previously, this time using the copula estimation results. Regardless of the variable selected for the parents, the surfaces look the same.

The evolution of the residual life expectancy e_x of men given some information about their parents can be visualized in Figure 10. Various situations are compared concerning the death's status of the parents at a given age of their son (at 20 years of age for the graphs on the left, 30 years of age for those in the middle and 40 years of age for those on the right): information not accounted for (baseline situation), both parents still alive, the father deceased and the mother still alive, the mother deceased and the father still alive, one of the two parents still alive regardless of gender, and both parents deceased. The upper graphs show the residual life expectancy expressed in years, while the lower graphs allow an easier comparison with the reference situation by showing the difference in years on the residual life expectancy of sons compared to the reference situation in which the information on the death of the parents is not taken into account.

Irrespective of the age at which the information regarding parent's death is

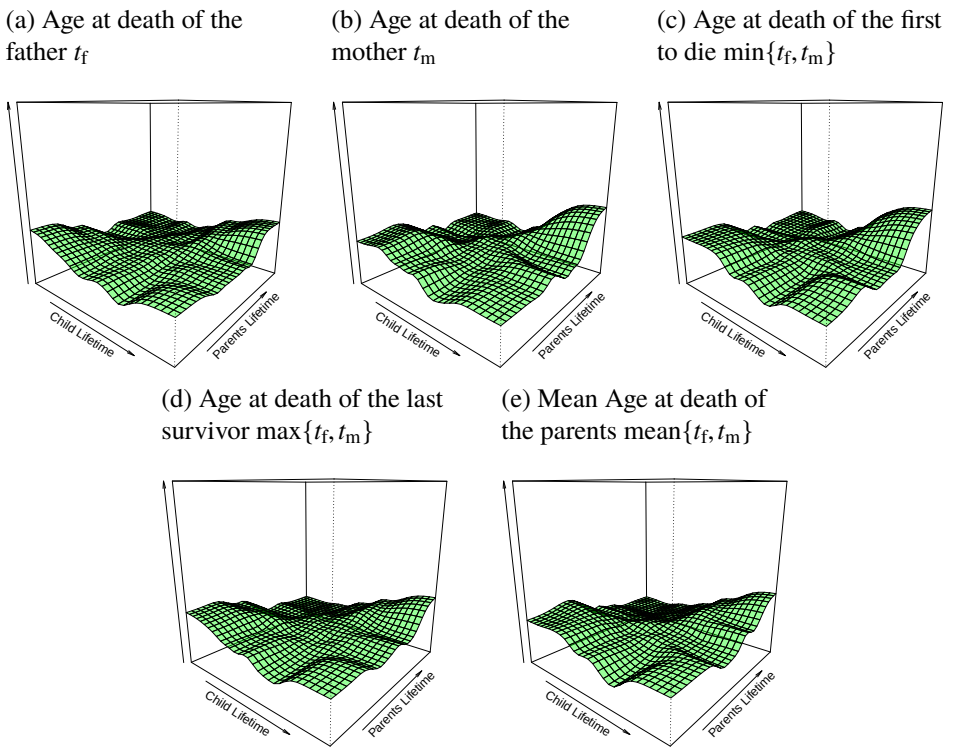
Figure 8. Spearman correlation between age at death of individuals t_c and age at death of their parents.



Note: The dots represent the estimated Spearman correlation between the age at death of a child t_c and that of their parents. The bars correspond to 95% bootstrap confidence interval. The red horizontal line corresponds to a value of 0. The correlations for the (1790, 1800] and (1810, 1820] cohorts are calculated on 476 and 250 couples of parents only. For the other cohorts, the correlations are calculated from a much larger number of observations, ranging from 3, 332 (for the (1800, 1810] cohort) to 41, 345 (for the (1830, 1840] cohort).

looked upon, *i.e.*, 20, 30 or 40 years old, it can be noted that the residual life expectancy curve for males whose parents are both still alive is systematically above the other curves, while the curve for children with both parents dead is systematically below the other curves. The difference in life expectancy is, however, relatively small and lessens over the years. As reported in Table B2, a male child whose both parents were still alive when he was 20 years old had a life expectancy of 39.7 years, compared to only 37.0 if both parents were deceased. In comparison, the residual life expectancy of a 20 years old man, without taking into account information about his parents, is 39.1 years. At age 30, the difference was much smaller: compared to the baseline value of 33.3, a male individual whose parents were both alive at that time was expected to live an additional 0.8 year while a man whose parents were both deceased was expected to live 1.6 year less. Lastly, at age 40, an man was expected to live another 26.7 years, a bit more (1.3 year) if his parents were still both alive at that age and slightly less if both parents were deceased (about one

Figure 9. Nonparametric estimation of the copula densities.



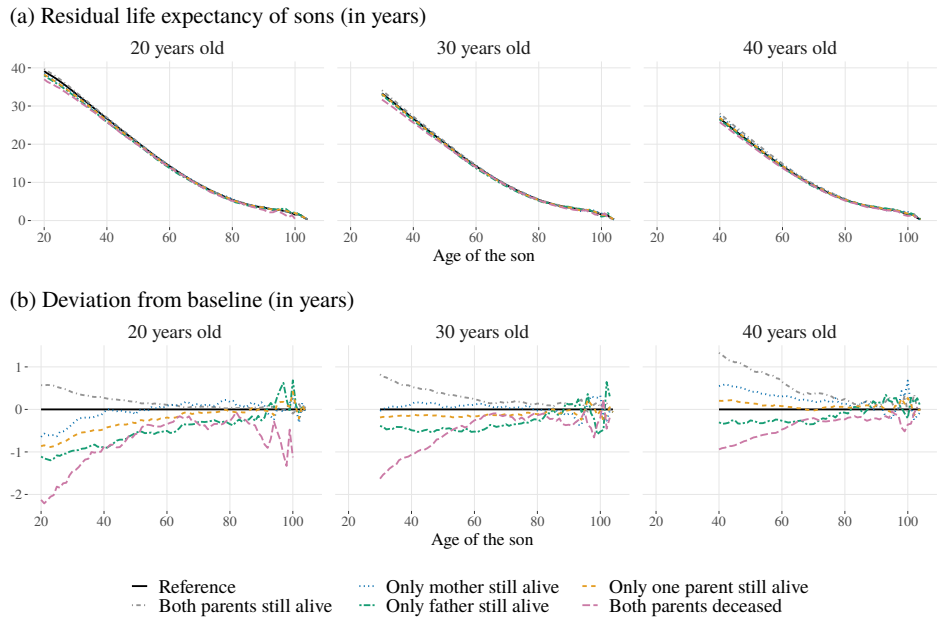
Note: these graphs show the dependence between the age at death of a children (t_c) and the age at death of a parent, considering different possible measures for the parent: the age at death of the father t_f, the age at death of the mother t_m, the age at death of the first to die $\min\{t_f, t_m\}$ and the age at death of the last survivor $\max\{t_f, t_m\}$.

year). Similar patterns are observed for female children.

5.1.2 Annuities and Life Insurance Premiums Accounting for the Status of Parents

In Figure 11, we observe the evolution of the present value of an annuity and the life insurance, as a function of the age of the insured, x, given information about his or her parents when buying the insurance contract. It can be noted that (empirical) monotonicity is consistent with theoretical results (decreasing with x for the pension and increasing for the life insurance), and the ordering of the three cases (both parents deceased, one parent still alive and both parents still alive) is consistent with the positive dependence between all lifespans. Figure 12 displays the relative difference (to the average baseline, as previously). It shows that present

Figure 10. Residuals life expectancy depending on the death status of the parents, at different times in the lives of men.



Note: each panel reports the residual life expectancy of individuals (only men) according to their parents' death status (information not accounted for – reference –, both still alive, only the father still alive, only the mother still alive, both deceased), at different times in the life of the individuals (at 20, 30, and 40 years old). Left panels thus indicates the residual life expectancy of men when both parents are still alive when they are 20 years old, when only one of them is alive, and so on. Top panels show the residuals life expectancy expressed in years. Bottom panels show the relative difference to the reference, expressed in years.

value of the annuity is consistently 3% lower when both parents are deceased, whatever the age. If both parents are alive, the difference of the present value of that annuity is increasing with x, and is 3.7% higher for an insured age 40.

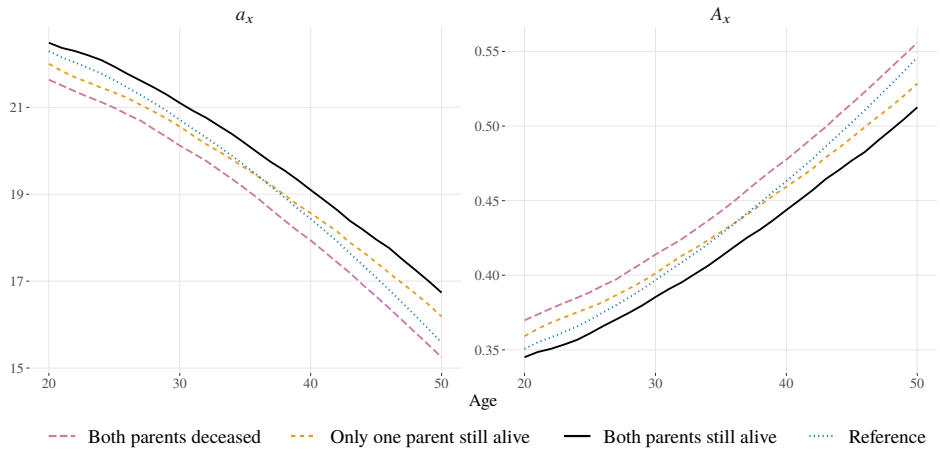
5.2 Children Conditional on Grand-Parents

We now extend our previous work to grandparents, where insured provide information about his or her four grandparents.

5.2.1 Empirical Evidence of the Relationship Between an Individual's Lifespan and that of his or her Grandparents

Spearman's correlation between the age at death of individuals t_c and the average age at death of their grandparents mean$\{t_{\text{gfm}}, t_{\text{gmm}}, t_{\text{gff}}, t_{\text{gmf}}\}$ is even weaker than with

Figure 11. Present value of an annuity (left) and expected present value for a life insurance (right) depending on the age of the annuitant and on how many parents are still alive at the time of the contract. The interest rate is assumed to be 3%.



Note: the annuities are calculated for 100 terms, the expected present value for life insurance are calculated for 100-year coverage.

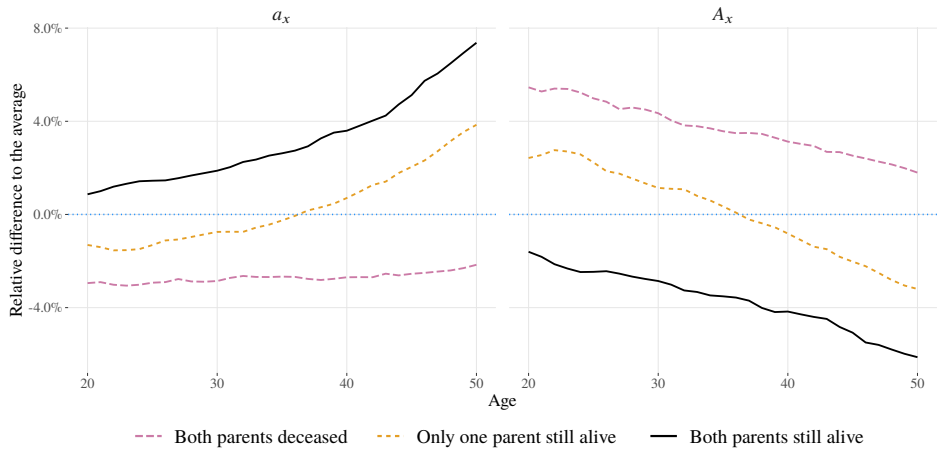
the average age at death of their parents mean$\{t_f, t_m\}$: 0.0251, with a 95% bootstrap confidence interval equal to $[0.0235, 0.0266]$. However, this confidence interval stresses that this correlation, although tenuous, is sinigicatively different from zero. In addition, as shown in Figure 13, this correlation appears to be relatively stable over time. The same Figure also shows that the correlation between the age at death of an individual and that of grandparents is stable over time when the latter is measured by using only the age at death of the first or last to die.

In a similar way to what was presented for the parents, the relationship between t_c and mean$\{t_{gfm}, t_{gmm}, t_{gff}, t_{gmf}\}$ can be studied using the results of copula estimates. These are graphed in Figure 14.

A comparison of the remaining life expectancy of an individual (male) according to the number of his grandparents still alive when he is 10, 15 or 20 years old is shown in Figure 15, taking as a reference the situation in which knowledge of this information is not taken into account.[13] When all four grandparents of a person (male) are still alive when he is 10 years old, the figure shows that his remaining life expectancy (grey dot-dashed line) is relatively higher than average (black solid

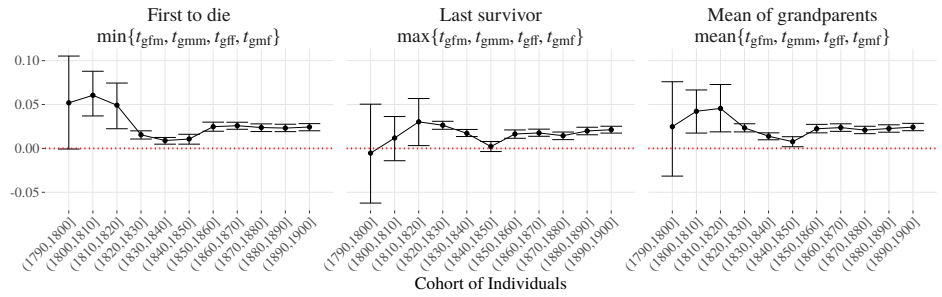
[13]Because of the relatively low life expectancy in the 19th century, it is unfortunately not possible to form groups of individuals with 3 or 4 grandparents when the first ones are 30 or 40 years old. This is why the ages of interest here (10, 15 and 20 years) are lower than in the previous comparison with information on parents (20, 30 and 40 years).

Figure 12. Relative difference to the average (in %) of present value of an annuity (left) and expected present value for a life insurance (right) depending on the age of the annuitant and on how many parents are still alive at the time of the contract.



Note: the horizontal blue dashed line corresponds to the average value calculated on all individuals, regardless of the death status of the parents. The annuities are calculated for 100 terms and the expected present value for life insurance are calculated for 100-year coverage. The interest rate is assumed to be 3%.

Figure 13. Spearman correlation between age at death of individuals t_c and age at death of their grandparents.

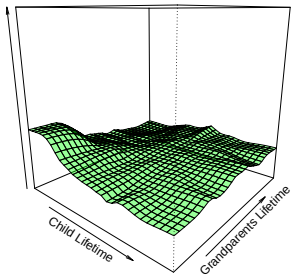


Note: The dots represent the estimated Spearman correlation between the age at death of a child t_c and that of their grandparents. The bars correspond to 95% bootstrap confidence interval.

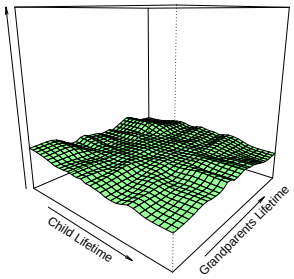
line). On the contrary, when all four grandparents are dead, then the remaining life of the grandson (pink dashed line) is lower than the average, although the absolute difference is not as large. For both cases, the absolute deviation from the average decreases over the years. This difference between the two groups is much less marked if the number of grandparents still alive when the children are 15 years old

Figure 14. Nonparametric estimation of the copula densities for the grandparents.

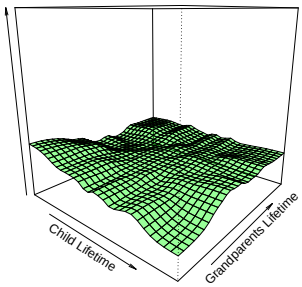
(a) Age at death of the first to die $\min\{t_{\text{gfm}}, t_{\text{gmm}}, t_{\text{gff}}, t_{\text{gmf}}\}$



(b) Age at death of the last survivor $\max\{t_{\text{gfm}}, t_{\text{gmm}}, t_{\text{gff}}, t_{\text{gmf}}\}$



(c) Mean Age at death of the grandparents $\text{mean}\{t_{\text{gfm}}, t_{\text{gmm}}, t_{\text{gff}}, t_{\text{gmf}}\}$



Note: these graphs show the dependence between the age at death of a children (t_c) and the age at death of their grandparents, considering different possible measures for the grandparents: the age at death of the first to die $\min\{t_{\text{gfm}}, t_{\text{gmm}}, t_{\text{gff}}, t_{\text{gmf}}\}$, the age at death of the last survivor $\max\{t_{\text{gfm}}, t_{\text{gmm}}, t_{\text{gff}}, t_{\text{gmf}}\}$ or the average age at death of the grandparents $\text{mean}\{t_{\text{gfm}}, t_{\text{gmm}}, t_{\text{gff}}, t_{\text{gmf}}\}$.

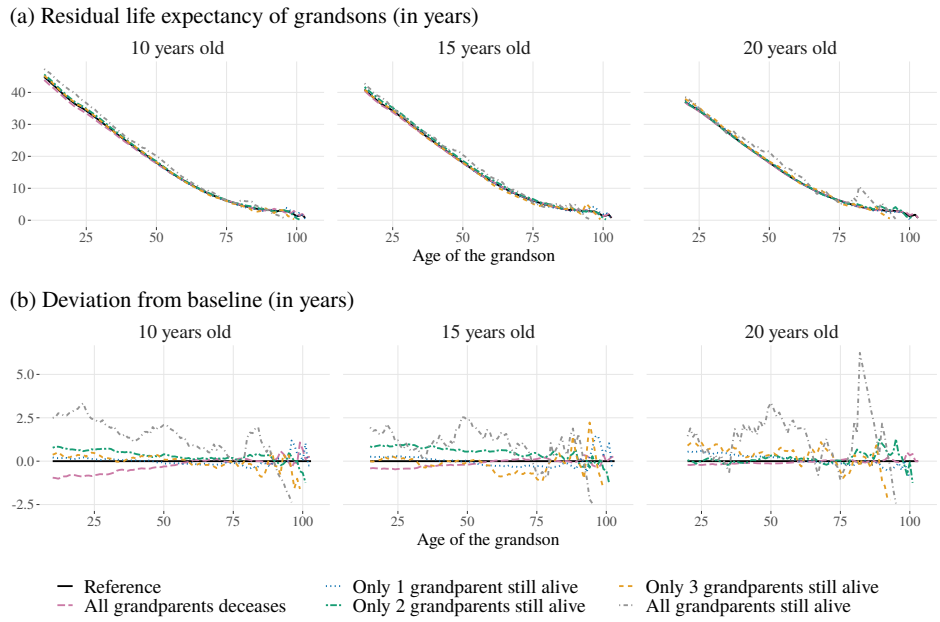
is considered, but it should be noted that those who still have their four grandparents at that age still have a relatively higher remaining life expectancy than the average.

5.2.2 Annuities and Life Insurance Premiums Accounting for the Status of Grandparents

In this section, computations are based on a (much) smaller dataset, where we kept individuals for whom information about the four grand-parents was available.

Figure 16 is the analogous of Figure 11, where the evolution of the present value of an annuity and the life insurance is represented as a function of the age of the insured, x, given information about his or her grandparents when buying the insurance contract. Again, the (empirical) monotonicity is consistent with theoretical results (decreasing with x for the pension and increasing for the life insurance), and the ordering of the three cases (all grandparents deceased, one or two still alive and three or four still alive) is consistent with the overall positive dependence between all lifespans. But here, as seen in Figure 17, the relative difference is much smaller, (at most) 2%, except perhaps when three or four grands parents are still alive and when x is 'large' (but in that case, the number of observations is much smaller, and the difference probably not significant).

Figure 15. Residual life expectancy depending on the death status of the grandparents, at different times in the lives of men.

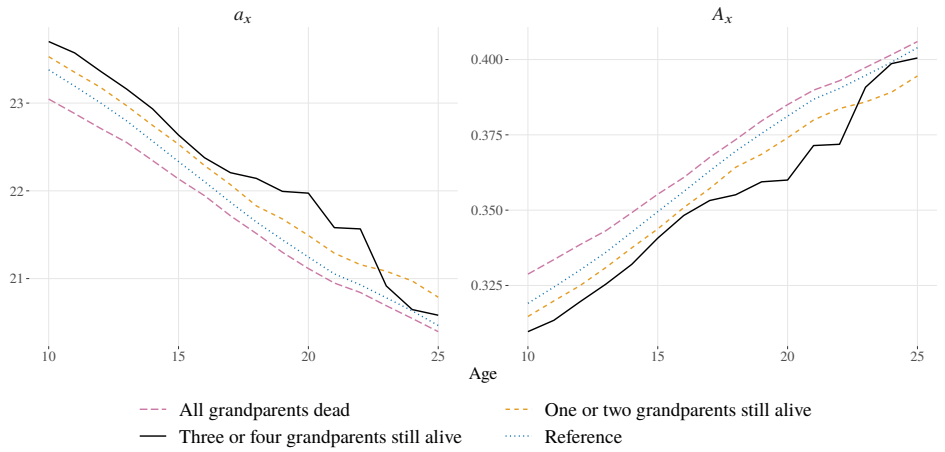


Note: each panel reports the residual life expectancy of individuals (only men) according to their grandparents' death status (information not accounted for – reference –, all four still alive, only one still alive, two still alive, three still alive, all four deceased), at different times in the life of the grandsons (at 10, 15, and 20 years old). Left panels thus indicates the residual life expectancy of grandsons when all four grandparents are still alive when they are 10 years old, when only one of the grandparents is still alive, and so on. Top panels show the residuals life expectancy expressed in years. Bottom panels show the relative difference to the reference, expressed in years.

6 Conclusion

In this article, we use collaborative genealogy data to study joint mortality within families. Using 135,128 observations of couples from people born during the 19th Century in France, we observe well-known results from the literature on dependency in mortality. Then we look at the potential links between individuals and their parents regarding mortality. To do this, contrary to what is usually observed in the literature, we rely on a substantial volume of data: 174,318 observations. We find results similar to those obtained in studies with smaller samples, *i.e.*, a very weak but significant association between an individual's lifespan and that of his or her parents. Our data also allow us to take the study a step further by looking at the relationship between an individual's longevity and that of his or her grandparents. As with the parents, a very weak but significant positive association is observed.

Figure 16. Present value of an annuity (left), expected present value for a life insurance (middle) and for an endowment (right) depending on the age of the annuitant and on how many grandparents are still alive at the time of the contract.



Note: the annuities are calculated for 100 terms, the expected present value for life insurance and for endowment are calculated for 100-year coverage. The interest rate is assumed to be 3%.

We then look at the potential implications for annuities and life insurance.

Here, we consider family history only through the prism of the age at death of ancestors. Furthermore, we do not try to explain the nature of the correlation, and it is rather difficult to understand what is due to genetics and heredity on the one side, and environmental factors on the other. This is all the more difficult since the reason for death is not stated in such a dataset. We can still wonder if family history (about the age at death of ancestors) is an important information: since it has a rather small impact (as we proved in this article), if the cause is mainly environmental, the correlation can be substantially captured by other common variables (such as the wealth).

Figure 17. Relative difference to the average (in %) of present value of an annuity (left), expected present value for a life insurance (middle) and for an endowment (right) depending on the age of the annuitant and on how many grandparents are still alive at the time of the contract.

Note: the horizontal blue dashed line corresponds to the average value calculated on all individuals, regardless of the death status of the grandparents. The annuities are calculated for 100 terms, the expected present value for life insurance and for endowment are calculated for 100-year coverage. The interest rate is assumed to be 3%.

A Models for Joint Life Dependencies

A.1 Parametric Copulas

In order to provide a more accurate comparison with related literature, we consider some popular parametric copulas, in this section: Clayton copula is defined as

$$C_{\mathrm{c},\theta}(u,v) = \left[u^{-\theta} + v^{-\theta} - 1\right]^{-1/\theta},$$

(with $\theta \geq 0$ from Clayton, 1978), Gumbel copula

$$C_{\mathrm{g},\theta}(u,v) = \exp\left[-\left((-\log(u))^{\theta} + (-\log(v))^{\theta}\right)^{1/\theta}\right],$$

(with $\theta \geq 1$, from Clayton, 1978) the normal (or Gaussian) copula

$$C_{\mathrm{n},\theta}(u,v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) \mathrm{d}y\mathrm{d}x,$$

(with $\rho \in (-1,1)$), and finally Frank copula (from Frank (1979))

$$C_{\mathrm{c},\theta}(u,v) = -\frac{1}{\theta}\log\left[1 + \frac{(\exp(-\theta u) - 1)(\exp(-\theta v) - 1)}{\exp(-\theta) - 1}\right],$$

(with $\theta \in \mathbb{R}$, with the independent copula with θ equals zero).

To estimate the parameters of the copula, instead of using the IFM method of Joe and Xu (1996), we prefer the omnibus semiparametric procedure described in Oakes (1989), where the copula is fitted on non-parametric pseudo observations $(\widehat{u}_{f,i}, \widehat{u}_{m,i})$ where

$$\widehat{u}_{f,i} = \widehat{S}_f(x_{f,i}) \text{ where } \widehat{S}_f(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}(x_{f,j} > x),$$

for fathers, and a similar expression for mothers.

We define the empirical copula $\widehat{C}_n$ as the cumulative distribution function of $(\widehat{u}_{f,i}, \widehat{u}_{m,i})$'s

$$\widehat{C}_n(u,v) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left(\widehat{u}_{f,i} \leq u, \widehat{u}_{m,i} \leq v\right),$$

or some smooth version $\widetilde{C}_n(u,v)$ using some probit transformation, as in Geenens et al. (2017).

If a positive dependence is observed, in the sense defined by Lehmann (1966), it can be used to derive bounds for most actuarial quantities.

A.1.1 Positive Association between Lifetimes

As in Lehmann (1966) – see also Scarsini and Shaked (1996) for an exhaustive survey – X and Y are said to be positively quadrant dependent (PQD) if and only if

$$F_X(x) \cdot F_y(y) \leq \mathbb{P}[X \leq x, Y \leq y] \text{ for all } x, y \in \mathbb{R}_+,$$

or equivalently

$$S_X(x) \cdot S_y(y) \leq \mathbb{P}[X > x, Y > y] \text{ for all } x, y \in \mathbb{R}_+.$$

The later can be written equivalently

$$C_\perp(u, v) \leq C(u, v) \text{ for all } u, v \in [0, 1].$$

An interesting interpretation of PQD association of lifetimes is given in Denuit and Scaillet (2004): if lifetimes of a husband age (x_m) and spouse age (x_f) are positively quadrant dependent, then for all $t \in \mathbb{R}_+$,

$$\mathbb{E}[T_{x_f} | T_{x_m} > t] \geq \mathbb{E}[T_{x_f}] \text{ and } \mathbb{E}[T_{x_m} | T_{x_f} > t] \geq \mathbb{E}[T_{x_m}].$$

The interpretation of those inequality is that knowing that one of the two spouses is still alive, at some time, increases the remaining lifetime of the other one.

Gijbels et al. (2010) compares several test for positive quadrant dependence, and in this section, we use a Kolmogorov-Smirnov test, as in Scaillet (2005). More specifically, we want to test $H_0 : C \geq C_\perp$, and we use

$$S_n = \sqrt{n} \sup_{(u,v)} \{C_\perp(u, v) - \widehat{C}_n(u, v)\},$$

where $\widehat{C}_n$ is the empirical copula.[14] Using 5,000 bootstrap samples, we obtain a p-value lower than 1‰, with 90% chance. So we can claim, with strong confidence, that in our data, joint lifes are PQD.

[14] and the p-value is approximated using standard bootstrap techniques, as described in section 3.2. of Scaillet (2005). More specifically, if $\widehat{C}_n^\star$ is the empirical copula built from a bootstrap sample, define

$$S_n^\star = \sqrt{n} \sup_{(u,v)} \{\widehat{C}_n^\star(u, v) - \widehat{C}_n(u, v)\}$$

and then $p_n^\star = \mathbb{P}[S_n^\star > S_n]$. In the application, a 500×500 uniform grid is used to approximate the supremium on the unit-square.

A.1.2 From Joint Distributions to Insurance Premiums

Consider two positive random variables X and Y, with marginal cumulative distributions F_X and F_Y respectively, and survival functions S_X and S_Y. From Fréchet (1960), without any further assumption,

$$\max\{0, F_X(x) + F_y(y) - 1\} \leq \mathbb{P}[X \leq x, Y \leq y] \leq \min\{F_X(x), F_y(y)\},$$

or equivalently

$$\max\{0, S_X(x) + S_y(y) - 1\} \leq \mathbb{P}[X > x, Y > y] \leq \min\{S_X(x), S_y(y)\},$$

for all $x, y \in \mathbb{R}_+$. In the context of joint lifes, the later can be written

$$\max\{0, {}_tp_{x_m} + {}_tp_{x_f} - 1\} \leq {}_tp_{x_m,x_f} \leq \min\{{}_tp_{x_m}, {}_tp_{x_f}\}, \tag{1}$$

for all time $t \in \mathbb{R}_+$. The upper bound is obtained when the associated copula is C^+.

Sharper bounds can be derived, at least for the lower bound, assuming some positive association between variables, namely the PQD property. In that case, the lower bounds corresponds to the independent case, and the associated copula is $C^\perp$. Thus, as a consequence, if lifetimes of a husband age (x_m) and spouse age (x_f) are positively quadrant dependent, then Equation (1) becomes

$${}_tp_{x_m} \cdot {}_tp_{x_f} \leq {}_tp_{x_m,x_f} \leq \min\{{}_tp_{x_m}, {}_tp_{x_f}\}, \tag{2}$$

Consider some quantity of interest $\mathcal{I}$ (that could be the life expectancy, or some annuity), so that $\mathcal{I}(X,Y)$ can be written $\mathbb{E}[\varphi(X,Y)]$, where φ is a supermodular function, in the sense that

$$\varphi(x_1, y_1) + \varphi(x_2, y_2) \geq \varphi(x_1, y_2) + \varphi(x_2, y_1) \text{ for all } x_2 \geq x_1, y_2 \geq y_1. \tag{3}$$

As proved in Lorentz (1953) and Cambanis et al. (1976), comonotonic vectors maximize $\mathcal{I}(X,Y)$, in the sense that

$$\mathcal{I}(F_X^{-1}(U), F_Y^{-1}(1-U)) \leq \mathcal{I}(X,Y) \leq \mathcal{I}(F_X^{-1}(U), F_Y^{-1}(U)),$$

where U is uniformly distributed. And in the case where (X,Y) are positively quandrant dependent,

$$\mathcal{I}(X^\perp, Y^\perp) \leq \mathcal{I}(X,Y) \leq \mathcal{I}(X^+, Y^+),$$

where $(X^\perp, Y^\perp)$ denotes an independent version of vector (X,Y), in the sense that $X^\perp$ has the same distribution as X, $Y^\perp$ has the same distribution as Y, and the copula of $(X^\perp, Y^\perp)$ is $C^\perp$ – and similarly for some (X^+, Y^+). In the case where φ is a supermodular function (with a *less or equal* instead of a *greater or equal* in Equation 3), bounds are inverted. As mentioned in Carriere and Chan (1986) and Denuit and Scaillet (2004), those bounds appear when calculating various annuities.

A.2 Bounds for Insurance Premiums

As discussed previously assuming positive association between life times $T(x_f)$ and $T(x_m)$, the independence case and the comonotonic cases will provide lower and upper bounds for various quantities. Hence, in the independent case,

$$a^{\perp}_{\overline{x}_f,\overline{x}_m} = \sum_{k=1}^{\infty} v^k \left({}_kp_{x_f} + {}_kp_{x_m} - {}_kp_{x_f} \cdot {}_kp_{x_m} \right),$$

$$a^{\perp}_{x_f,x_m} = \sum_{k=1}^{\infty} v^k {}_kp_{x_f} \cdot {}_kp_{x_m},$$

$$a^{\perp}_{x_f|x_f} = \sum_{k=1}^{\infty} v^k {}_kp_{x_m} - \sum_{k=1}^{\infty} v^k {}_kp_{x_f} \cdot {}_kp_{x_m},$$

while in the perfectly correlated case,

$$a^{+}_{\overline{x}_f,\overline{x}_m} = \sum_{k=1}^{\infty} v^k \left(1 - \min\{ {}_kp_{x_f}, {}_kp_{x_m} \} \right),$$

$$a^{+}_{x_f,x_m} = \sum_{k=1}^{\infty} v^k \min\{ {}_kp_{x_f}, {}_kp_{x_m} \},$$

$$a^{+}_{x_m|x_f} = \sum_{k=1}^{\infty} v^k {}_kp_{x_m} - \sum_{k=1}^{\infty} v^k \min\{ {}_kp_{x_f}, {}_kp_{x_m} \}.$$

And as shown in Denuit and Scaillet (2004), if lifetimes are positively quadrant dependent,

$$a^{\perp}_{x_f,x_m} \leq a_{x_f,x_m} \leq a^{+}_{x_f,x_m},$$

while

$$a^{+}_{\overline{x}_f,\overline{x}_m} \leq a_{\overline{x}_f,\overline{x}_m} \leq a^{\perp}_{\overline{x}_f,\overline{x}_m},$$

and

$$a^{+}_{x_m|x_f} \leq a_{x_m|x_f} \leq a^{\perp}_{x_m|x_f}.$$

Thus, for the last-survivor and the widow's pension, the independence assumption is conservative (as soon as lifetimes are positively associated). In that case, using the independence assumption for pricing those annuities will incorporate some safety loading (as in Figure 7 where we plot the present value of a widow's pension, $a_{\mathrm{m|f}}$ (relative to independent case $a^{\perp}_{\mathrm{m|f}}$), as a function of x_{m}.).

B Parents and Children Dependencies

Figure B1. Force of mortality (log scale) μ_x for individuals depending on their gender (top), and for their parents (bottom).

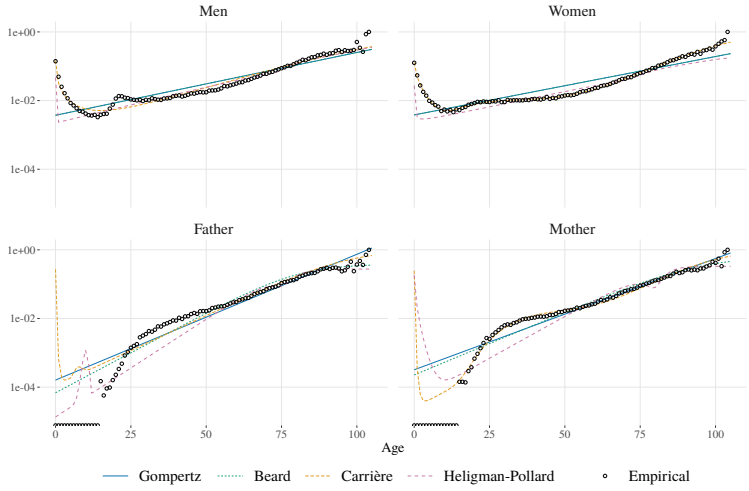


Figure B2. Survival function for individuals depending on their gender (top), and for their parents (bottom).

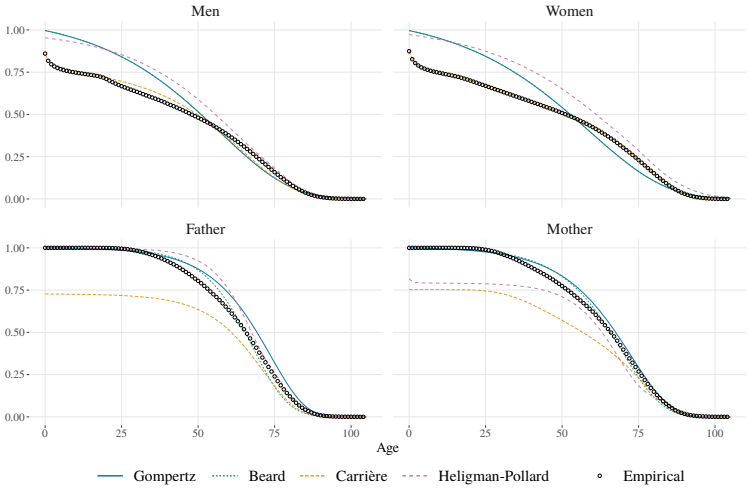
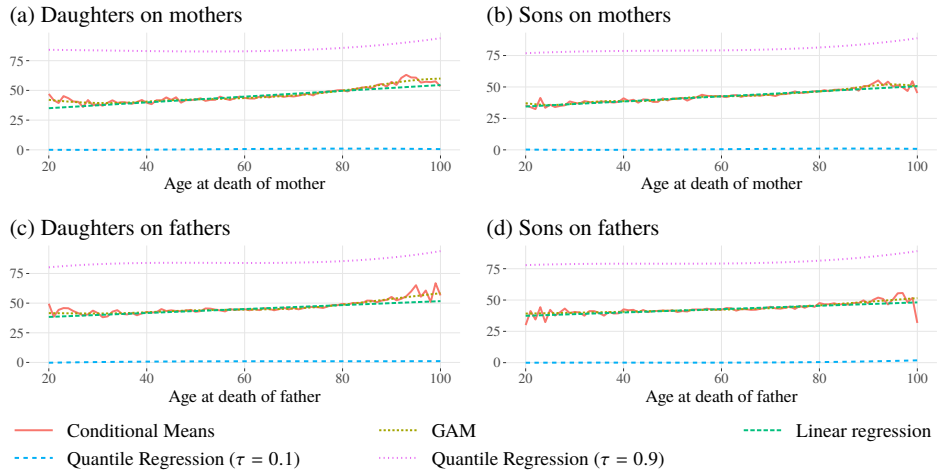


Figure B3. Age of individuals (daughters and sons) as a function of age of parents.



Note: The blue and purple dashed lines correspond to the first and ninth deciles of quantile regression of age at death of an individual as a function of the age at death of their spouse.

Table B1. Linear regression coefficients of children's age at death as a function of parent's age at death.

	Variables		
Relationship	Intercept	Age of the parents	R^2
Daughter vs. Mother	30.17 [29.31,31.03]	0.243 [0.230,0.256]	0.0158
Son vs. Mother	30.39 [29.60,31.18]	0.202 [0.190,0.214]	0.0120
Daughter vs. Father	35.03 [34.06,36.00]	0.167 [0.152,0.181]	0.0059
Son vs. Father	34.71 [33.82,35.59]	0.134 [0.120,0.147]	0.0042

Note: The 95% confidence intervals for each coefficient are provided between brackets next to the estimates.

Table B2. Residual life expectancy of children depending on information on parents.

		Men		Women	
x	Information used	$e_{x,\text{male}}$	Dev. from Ref.	$e_{x,\text{female}}$	Dev. from Ref.
	Information on parents when the child is 20 years old				
20	Reference	39.13	0.00	43.17	0.00
20	Both parents still alive	39.70	0.57	43.69	0.51
20	Only mother still alive	38.48	-0.65	43.12	-0.05
20	Only father still alive	38.02	-1.11	41.37	-1.81
20	Only one parent still alive	38.27	-0.86	42.30	-0.87
20	Both parents deceased	37.00	-2.13	42.03	-1.15
30	Reference	33.31	0.00	36.87	0.00
30	Both parents still alive	33.71	0.40	37.27	0.40
30	Only mother still alive	32.97	-0.34	36.73	-0.14
30	Only father still alive	32.34	-0.97	35.20	-1.66
30	Only one parent still alive	32.67	-0.64	36.02	-0.84
30	Both parents deceased	31.85	-1.45	36.71	-0.16
40	Reference	26.71	0.00	30.32	0.00
40	Both parents still alive	26.98	0.27	30.60	0.28
40	Only mother still alive	26.64	-0.07	30.21	-0.11
40	Only father still alive	25.81	-0.90	29.13	-1.19
40	Only one parent still alive	26.25	-0.46	29.71	-0.60
40	Both parents deceased	25.84	-0.87	30.22	-0.10
	Information on parents when the child is 30 years old				
30	Reference	33.31	0.00	36.87	0.00
30	Both parents still alive	34.13	0.82	37.99	1.12
30	Only mother still alive	33.27	-0.03	37.07	0.20
30	Only father still alive	32.92	-0.39	35.01	-1.86
30	Only one parent still alive	33.12	-0.19	36.18	-0.69
30	Both parents deceased	31.68	-1.63	35.78	-1.09
40	Reference	26.71	0.00	30.32	0.00
40	Both parents still alive	27.25	0.54	31.12	0.80
40	Only mother still alive	26.84	0.13	30.53	0.21
40	Only father still alive	26.22	-0.49	28.70	-1.62
40	Only one parent still alive	26.57	-0.14	29.74	-0.58
40	Both parents deceased	25.65	-1.06	29.68	-0.64
	Information on parents when the child is 40 years old				
40	Reference	26.71	0.00	30.32	0.00
40	Both parents still alive	28.04	1.33	31.90	1.58
40	Only mother still alive	27.26	0.55	31.54	1.22
40	Only father still alive	26.41	-0.30	29.30	-1.02
40	Only one parent still alive	26.91	0.20	30.66	0.34
40	Both parents deceased	25.77	-0.94	29.10	-1.22

Note: The reference situation is one in which no information regarding the death's status of the parents is accounted for. x is the age of children and e_x the corresponding residual life expectancy expressed in years. The deviation from the reference is the difference between life expectancy when some information on parents death is used and life expectancy when such information is not accounted for.

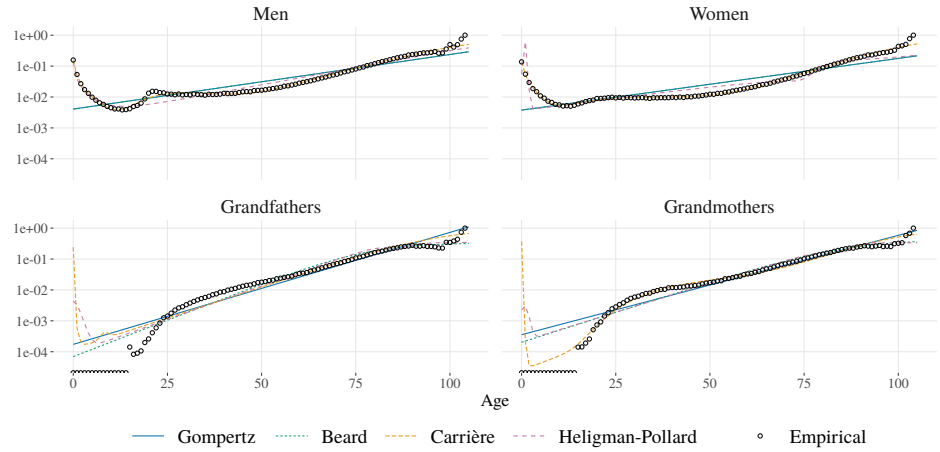
C Grandparents and Children Dependencies

Table C3. Age at death of the individuals and age of their grandparents, according to the gender of the grandchildren, keeping only grandchildren whose four grandparents are known.

	Mean	SD	Min	Max	Q_1	Q_2	Q_3	No Missing
	Men ($n = 31,096$)							
Individual (t_c)	39.1	29.5	0.0	104.0	3.8	43.3	65.7	0
Maternal grandfather	64.6	14.3	15.0	104.1	55.2	66.7	75.4	0
Maternal grandmother	63.3	15.3	15.0	104.0	53.0	65.8	75.0	0
Paternal grandfather	64.4	14.2	15.0	104.1	55.1	66.5	75.2	0
Paternal grandmother	62.9	15.2	15.0	102.5	53.0	65.1	74.5	0
Last survivor	77.5	8.2	15.0	104.1	72.7	78.1	82.9	0
First to die	48.4	12.2	15.0	100.0	39.2	48.1	57.6	0
Average grandparents	63.8	8.0	15.0	100.0	58.6	64.2	69.5	0
	Women ($n = 28,367$)							
Individual (t_c)	38.9	30.1	0.0	104.4	4.2	40.4	66.6	0
Maternal grandfather	64.7	14.3	15.7	104.2	55.3	66.9	75.5	0
Maternal grandmother	63.2	15.3	15.2	104.0	53.0	65.7	74.9	0
Paternal grandfather	64.4	14.3	15.0	104.1	54.9	66.3	75.3	0
Paternal grandmother	62.9	15.2	16.0	102.6	52.7	65.2	74.6	0
Last survivor	77.5	8.2	25.0	104.2	72.6	78.1	83.0	0
First to die	48.4	12.2	15.0	102.0	39.2	48.1	57.6	0
Average grandparents	63.8	8.1	25.0	102.5	58.5	64.1	69.6	0
	Men & Women ($n = 59,463$)							
Individual (t_c)	39.0	29.8	0.0	104.4	4.0	42.0	66.1	0
Maternal grandfather	64.6	14.3	15.0	104.2	55.2	66.7	75.4	0
Maternal grandmother	63.3	15.3	15.0	104.0	53.0	65.7	74.9	0
Paternal grandfather	64.4	14.3	15.0	104.1	55.0	66.4	75.3	0
Paternal grandmother	62.9	15.2	15.0	102.6	52.9	65.1	74.6	0
Last survivor	77.5	8.2	15.0	104.2	72.7	78.1	83.0	0
First to die	48.4	12.2	15.0	102.0	39.2	48.1	57.6	0
Average grandparents	63.8	8.0	15.0	102.5	58.6	64.1	69.5	0

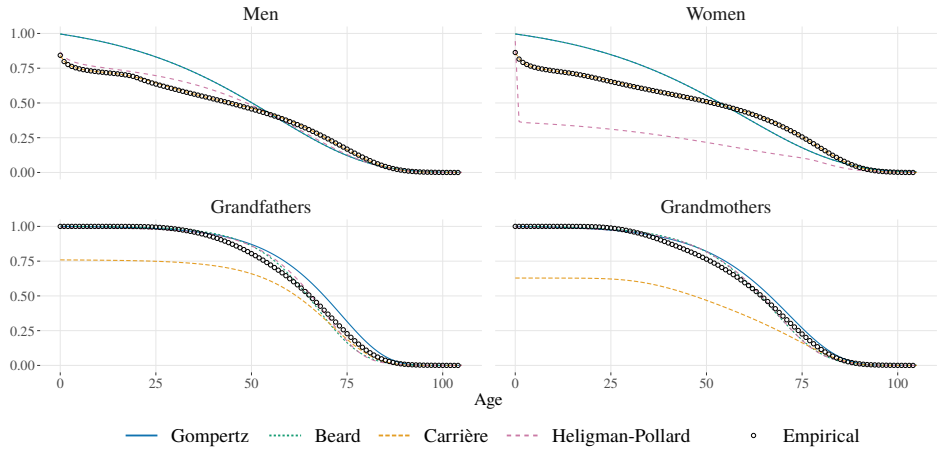
Note: this table provides descriptive statistics of the ages contained in the dataset of grandchildren and their grandparents. n stands for the number of observations, SD is the standard deviation, Q_1, Q_2, and Q_3 are the first, second, and third empirical quartiles, respectively, and No. Missing refers to the number of missing values.

Figure C4. Force of mortality (log scale) μ_x for individuals depending on their gender (top), and for their grandparents (bottom).



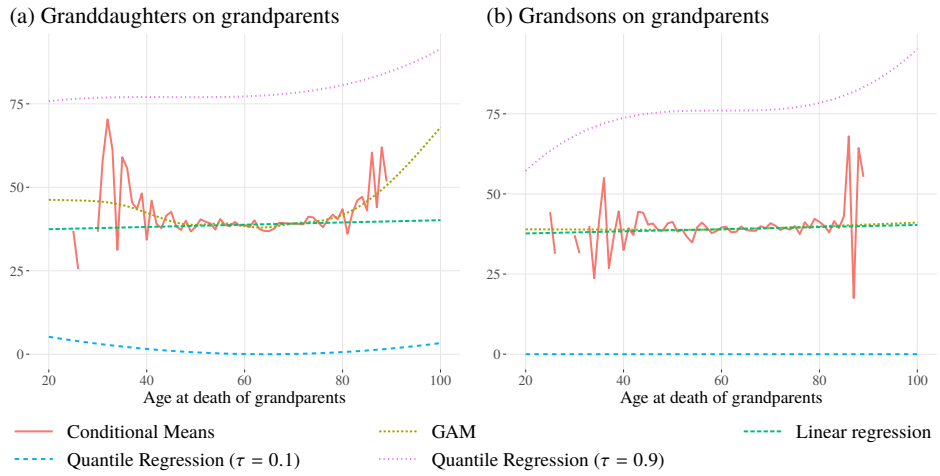
Note:

Figure C5. Survival function for individuals depending on their gender (top), and for their grandparents (bottom).



Note:

Figure C6. Age of individuals (daughters and sons) as a function of age of parents.



Note:

Table C4. Linear regression coefficients of grandchildren's age at death as a function of grandparent's age at death.

	Variables		
Grandchildren	Intercept	Age of the parents	R^2
Granddaughter	37.113 [34.32,39.91]	0.0280 [-0.0154,0.0715]	0.000056
Grandson	37.056 [34.43,39.69]	0.0325 [-0.0084,0.0734]	0.000078

Note: The 95% confidence intervals for each coefficient are provided between brackets next to the estimates.

Table C5. Residual life expectancy of grandchildren depending on information on grandparents.

		Men		Women	
x	Information used	$e_{x,\text{male}}$	Dev. from Ref.	$e_{x,\text{female}}$	Dev. from Ref.
	Information on grandparents when the grandson is 10 years old				
10	Reference	44.84	0.00	44.60	0.00
10	All grandparents deceases	43.92	-0.92	43.29	-1.30
10	Only 1 grandparent still alive	45.06	0.22	45.04	0.44
10	Only 2 grandparents still alive	45.55	0.71	45.28	0.68
10	Only 3 grandparents still alive	45.25	0.41	45.82	1.23
10	All grandparents still alive	47.26	2.42	47.38	2.78
15	Reference	40.88	0.00	41.06	0.00
15	All grandparents deceases	40.03	-0.85	39.73	-1.33
15	Only 1 grandparent still alive	41.07	0.19	41.58	0.52
15	Only 2 grandparents still alive	41.54	0.66	41.66	0.60
15	Only 3 grandparents still alive	41.18	0.30	42.46	1.40
15	All grandparents still alive	43.44	2.56	43.20	2.13
20	Reference	37.08	0.00	37.71	0.00
20	All grandparents deceases	36.23	-0.86	36.35	-1.35
20	Only 1 grandparent still alive	37.26	0.18	38.27	0.57
20	Only 2 grandparents still alive	37.67	0.59	38.24	0.53
20	Only 3 grandparents still alive	37.46	0.38	39.14	1.44
20	All grandparents still alive	40.24	3.16	40.03	2.32
	Information on grandparents when the grandson is 15 years old				
15	Reference	40.88	0.00	41.06	0.00
15	All grandparents deceases	40.48	-0.40	40.20	-0.86
15	Only 1 grandparent still alive	41.12	0.24	41.63	0.57
15	Only 2 grandparents still alive	41.70	0.82	42.34	1.28
15	Only 3 grandparents still alive	40.94	0.06	42.67	1.61
15	All grandparents still alive	42.69	1.82	43.33	2.27
20	Reference	37.08	0.00	37.71	0.00
20	All grandparents deceases	36.68	-0.41	36.85	-0.86
20	Only 1 grandparent still alive	37.28	0.19	38.21	0.51
20	Only 2 grandparents still alive	37.94	0.85	38.99	1.29
20	Only 3 grandparents still alive	37.42	0.33	39.44	1.74
20	All grandparents still alive	38.67	1.59	40.92	3.21
	Information on grandparents when the grandson is 20 years old				
20	Reference	37.08	0.00	37.71	0.00
20	All grandparents deceases	36.88	-0.20	37.17	-0.54
20	Only 1 grandparent still alive	37.59	0.51	38.49	0.78
20	Only 2 grandparents still alive	37.14	0.06	39.26	1.55
20	Only 3 grandparents still alive	37.98	0.90	41.02	3.32
20	All grandparents still alive	38.51	1.43	42.10	4.40

Note: The reference situation is one in which no information regarding the death's status of the grandparents is accounted for. x is the age of grandchildren and e_x the corresponding residual life expectancy expressed in years. The deviation from the reference is the difference between life expectancy when some information on grandparents death is used and life expectancy when such information is not accounted for.

References

Abbott, M. H., Abbey, H., Bolling, D. R., Murphy, E. A., and Opitz, J. M. (1978). The familial component in longevity—a study of offspring of nonagenarians: Iii. intrafamilial studies. *American Journal of Medical Genetics*, 2(2):105–120.

Bandeen Roche, K. and Liang, K. (1996). Modelling failure-time associations in data with multiple levels of clustering. *Biometrika*, 83(1):29–39.

Beard, R. E. (1971). Some aspects of theories of mortality, cause of death analysis, forecasting and stochastic processes. *Biological aspects of demography*, 999:57–68.

Beeton, M. and Pearson, K. (1901). On the inheritance of the duration of life, and on the intensity of natural selection in man. *Biometrika*, 1(1):50–89.

Bellis, M. A., Hughes, K., Hughes, S., and Ashton, J. R. (2005). Measuring paternal discrepancy and its public health consequences. *Journal of Epidemiology & Community Health*, 59(9):749–754.

Beltrán-Sánchez, H., Finch, C. E., and Crimmins, E. M. (2015). Twentieth century surge of excess adult male mortality. *Proceedings of the National Academy of Sciences*, 112(29):8993–8998.

Bocquet-Appel, J.-P. and Jakobi, L. (1990). Familial transmission of longevity. *Annals of Human Biology*, 17(2):81–95.

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., and Nesbitt, C. J. (1997). *Actuarial Mathematics*. Society of Actuaries, Shaumburg, IL, second edition.

Cambanis, S., Simons, G., and Stout, W. (1976). Inequalities for $\mathbb{E}[k(x, y)]$ when the marginals are fixed. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 36(1):285–294.

Carriere, J. F. (1992). Parametric models for life tables. *Transactions of the Society of Actuaries*, 44:77–99.

Carriere, J. F. and Chan, L. (1986). The bounds of bivariate distributions that limit the value of last-survivor annuities. *Transactions of the Society of Actuaries*, 38:51–74.

Charpentier, A. and Gallic, E. (2020a). La démographie historique peut-elle tirer profit des données collaboratives des sites de généalogie ? *Population*, to appear.

Charpentier, A. and Gallic, E. (2020b). Using collaborative genealogy data to study migration: a research note. *The History of the Family*, 25(1):1–21.

Choi, S.-W. E. (2020). Grandparenting and mortality: How does race-ethnicity matter? *Journal of Health and Social Behavior*.

Clayton, D. G. (1978). A model for association in bivariate life tables and its applications in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika*, 65(1):141–151.

Cutler, D. M. and Zeckhauser, R. J. (1998). Adverse selection in health insurance. *Forum for Health Economics & Policy*, 1(1).

Denuit, M., Dhaene, J., Le Bailly De Tilleghem, C., and Teghem, S. (2001). Measuring the impact of a dependence among insured life lengths. *Belgian Actuarial Bulletin*, 1(1):18–39.

Denuit, M. and Scaillet, O. (2004). Nonparametric Tests for Positive Quadrant Dependence. *Journal of Financial Econometrics*, 2(3):422–450.

Espinosa, J. and Evans, W. N. (2008). Heightened mortality after the death of a spouse: Marriage protection or marriage selection? *Journal of Health Economics*, 27(5):1326 – 1342.

Frank, M. (1979). On the simultaneous associativity of $f(x, y)$ and $x + y - f(x, y)$. *Aequationes Mathematicae*, 19:194–226.

Frees, E. W., Carriere, J., and Valdez, E. (1996). Annuity valuation with dependent mortality. *The Journal of Risk and Insurance*, 63(2):229–261.

Fréchet, M. (1960). Sur les tableaux dont les marges et des bornes sont données. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 28(1/2):10–32.

Garibotti, G., Smith, K. R., Kerber, R. A., and Boucher, K. M. (2006). Longevity and Correlated Frailty in Multigenerational Families. *The Journals of Gerontology: Series A*, 61(12):1253–1261.

Geenens, G., Charpentier, A., and Paindaveine, D. (2017). Probit transformation for nonparametric kernel estimation of the copula density. *Bernoulli*, 23(3):1848–1873.

Gijbels, I., Omelka, M., and Sznajder, D. (2010). Positive quadrant dependence tests for copulas. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 38(4):555–581.

Glick, P. C. (1955). The life cycle of the family. *Marriage and Family Living*, 17(1):3–9.

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.

Gudmundsson, H., Gudbjartsson, D. F., Kong, A., Gudbjartsson, H., Frigge, M., Gulcher, J. R., and Stefánsson, K. (2000). Inheritance of human longevity in iceland. *European Journal of Human Genetics*, 8:743–749.

Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01):49–80.

Hjelmborg, J., Iachine, I., Skytthe, A., Vaupel, J. W., McGue, M., Koskenvuo, M., Kaprio, J., Pedersen, N. L., and Christensen, K. (2006). Genetic influence on human lifespan and longevity. *Human Genetics*, 119:312.

Jagger, C. and Sutton, C. (1991). Death after marital bereavement–is the risk increased? *Statistics in Medecine*, 10(3):395–404.

Joe, H. (1997). *Multivariate Models and Multivariate Dependence Concepts*, volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models.

Kastenbaum, R. and Costa, P. T. (1977). Psychological perspectives on death. *Annual Review of Psychology*, 28(1):225–249. PMID: 324377.

Kowald, A. and Kirkwood, T. B. L. (2016). Can aging be programmed? a critical literature review. *Aging Cell*, 15(6):986–998.

Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.*, 37:1137–1153.

Lorentz, G. G. (1953). An inequality for rearrangements. *The American Mathematical Monthly*, 60(3):176–179.

Matroos, A. W., van der Heide-Wessel, C., and van der Heide, R. M. (1984). Parental survival, an independent predictor of longevity in middle-aged persons. *American Journal of Epidemiology*, 119(5):742–750.

Mayer, P. J. (1991). Inheritance of longevity evinces no secular trend among members of six new england families born 1650–1874. *American Journal of Human Biology*, 3(1):49–58.

Nelsen, R. B. (1999). *An Introduction to Copulas*. Springer Verlag, New York.

Oakes, D. (1989). Bivariate survival models induced by frailties. *Journal of the American Statistical Association*, 84(406):487–493.

Pardo, C. and Schott, W. (2013). Health insurance selection in Chile: a cross-sectional and panel analysis. *Health Policy and Planning*, 29(3):302–312.

Parkes, C., Benjamin, B., and Fitzgerald, R. G. (1969). Broken heart: A statistical study of increased mortality among widowers. *The British Medical Journal*, 1(1):740–743.

Pascariu, M. D. (2020). *MortalityLaws: Parametric Mortality Models, Life Tables and HMD*. R package version 1.8.4.

Pearl, R. (1931). Studies on human longevity. iv. the inheritance of longevity. preliminary report. *Human Biology*, 3(2):245. Dernière mise à jour - 2013-02-24.

Philippe, P. (1980). Longevity: Some familial correlates. *Social Biology*, 27(3):211–219.

Philippe, P. and Opitz, J. M. (1978). Familial correlations of longevity: An isolate-based study. *American Journal of Medical Genetics*, 2(2):121–129.

Piraino, P., Muller, S., Cilliers, J., and Fourie, J. (2014). The transmission of longevity across generations: The case of the settler cape colony. *Research in Social Stratification and Mobility*, 35:105 – 119. Inequality Across Multiple Generations.

Riley, J. W. (1983). Dying and the meanings of death: Sociological inquiries. *Annual Review of Sociology*, 9(1):191–216.

Ruby, J. G., Wright, K. M., Rand, K. A., Kermany, A., Noto, K., Curtis, D., Varner, N., Garrigan, D., Slinkov, D., Dorfman, I., Granka, J. M., Byrnes, J., Myres, N., and Ball, C. (2018). Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics*, 210(3):1109–1124.

Scaillet, O. (2005). A kolmogorov-smirnov type test for positive quadrant dependence. *Canadian Journal of Statistics*, 33(3):415–427.

Scarsini, M. and Shaked, M. (1996). Positive dependence orders: A survey. In Heyde, C. C., Prohorov, Y. V., Pyke, R., and Rachev, S. T., editors, *Athens Conference on Applied Probability and Time Series Analysis*, pages 70–91, New York, NY. Springer New York.

Schmitz, D. and Wiesing, U. (2006). Just a family medical history? *Britsh Medical Journal*, 332(7536):297–299.

Spedicato, G. (2013). The lifecontingencies package: Performing financial and actuarial mathematics calculations in r. *Journal of Statistical Software, Articles*, 55(10):1–36.

Temby, O. F. and Smith, K. R. (2014). The association between adult mortality risk and family history of longevity: the moderating effects of socioeconomic status. *Journal of Biosocial Science*, 46(6):703–716.

Vaillant, G. E. (1991). The Association of Ancestral Longevity With Successful Aging. *Journal of Gerontology*, 46(6):P292–P298.

Vaupel, J. W. (1988). Inherited frailty and longevity. *Demography*, 25(2):277–287.

Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., Iachine, I. A., Kannisto, V., Khazaeli, A. A., Liedo, P., Longo, V. D., Zeng, Y., Manton, K. G., and Curtsinger, J. W. (1998). Biodemographic trajectories of longevity. *Science*, 280(5365):855–860.

Vaupel, J. W., Manton, K. G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*, 16(3):439–454.

Wilhelmsen, L., Svärdsudd, K., Eriksson, H., Rosengren, A., Hansson, P.-O., Welin, C., Odén, A., and Welin, L. (2011). Factors associated with reaching 90 years of age: a study of men born in 1913 in gothenburg, sweden. *Journal of Internal Medicine*, 269(4):441–451.