Family History and Life Insurance

Arthur Charpentier[1], Ewen Gallic[2] & Olivier Cabrignac[3]

[1] UQAM & JRI AXA Research Fund, [2] AMSE & [3] SCOR

CMStatistics (ERCIM WG on Computational and Methodological Statistics), December 2021

https://arxiv.org/abs/2006.08446

'Family History' & Insurance Forms

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			

Agenda

- Motivations
- Existing Literature
- Longitudinal & Collaborative Data
- Genealogical Data
- ‘Family History’ & Life Insurance
- Husband-Wife
- Children-Parents
- Grand Children-Grandparents

Using genealogical trees to understand dependencies in life spans and quantify the impact on (life related) insurance premiums

Literature on Family and Insurance

- Parkes et al. (1969) 4,486 widowers of 55 yearsold (and older) to confirm the *broken heart syndrom*
- Frees et al. (1996): 14,947 insurance contracts, Canadian insurance company, in force in 1988-1993

  → censoring problem

  used also in Carriere (1997), Youn and Shemyakin (1999), Shemyakin and Youn (2001)

  in Luciano et al. (2008), subset of 11,454 contracts, born before 1920 (male) and 1923 (female)
- Denuit et al. (2001): selected two cemeteries in Brussels (Koekelberg and Ixelles / Elsene) and collected the ages at death of 533 couples buried there

Longitudinal Data

Longitudinal data have been used in many demographic projects

- Matthijs and Moreels (2010) (COR$*$), Antwerp, Belgium, 1846–1920, $\approx 125k$ events, $\approx 57k$ individuals
- Mandemakers (2000), Netherlands, 1812–1922, $\approx 77k$ indivivuals
- Bouchard et al. (1989) (BALSAC), Québec, Canada, since 17^{th} century, $\approx 2M$ events, $\approx 575k$ individuals
- Bean et al. (1978) , mainly Utah, USA, since 18^{th} century, $\approx 1.2M$ individuals

Collaborative Data

as well as collaborative data

- Fire and Elovici (2015) with data from `WikiTree.com` $+1M$ profiles (unknown number of individuals)
- Cummins (2017) with data from `FamilySearch.org`, $+1.3M$ individuals
- Gergaud et al. (2016) with biography from wikipedia, $+1.2M$ individuals
- Kaplanis et al. (2018) with data from `Geni.com`, $13M$ individuals

Genealogical Data

Charpentier and Gallic (2020a) comparing our collaborative based dataset (238,009 users, 1,547,086 individual born in $[1800, 1805)$), with official historical data

	ID_user	ID_np	ID_num	Name tabular	Surname	Sex	Date_b
1	daage	besnard|jean|1	575	BESNARD	Jean	1	18000227
2	denisgallienne	besnard|louis|1	22771	BESNARD	Louis	1	18040603
3	domiassi	besnard|jean|	1748	BESNARD	Jean	1	18000227
4	dutheilfr	besnard|pierre|	729	BESNARD	Pierre	1	18001221
5	dvivier1	besnard|louis|1	65196	BESNARD	Louis	1	18001215

	Date_d	Type	Location	Lat	Long	ID_num_m	ID_num_p
1	16810000	NM	Longué, 0180	47.37806	-0.10806	4457	574
2	18831027	ND	Cunault, 49350	47.30833	-0.15389	994	1620
3	18560000	NM	Longué, 49180	47.37806	-0.10806		
4		N	Gennes, 49350	47.34083	-0.23278	99	59
5	18490717	N	Pommeraye, 49244	47.35528	-0.86028	43116	4063

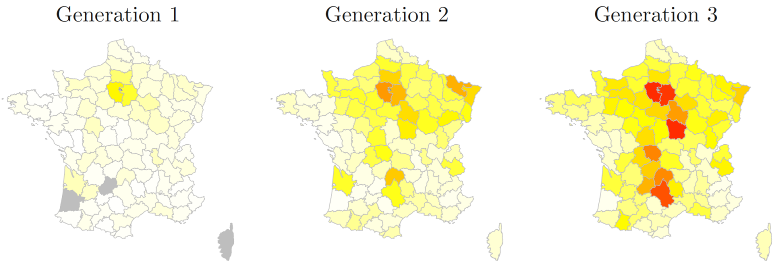
with children, up to 3 generations

- 402 190 children
- 286 071 grand-children
- 222 103 grand-grand-children

Intensive study on exhaustivity & consistency of data

Genealogical Data

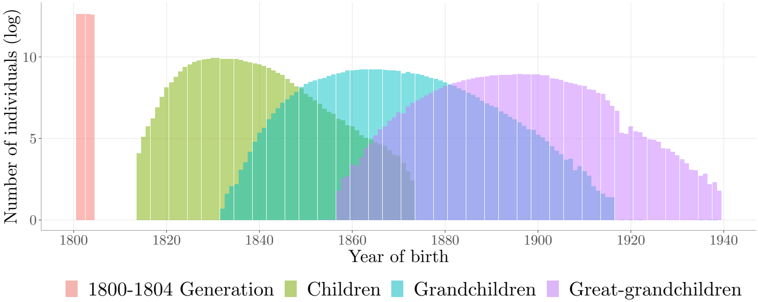
Charpentier and Gallic (2020b) on generational migration



(here Generation 0 was born in Paris)

Genealogical Data & “Generations”

Initial starting generation ■ (born in $[1800, 1805)$),
children ■ (born $\sim [1815, 1870)$),
grand-children ■ (born $\sim [1830, 1915)$),
grand-grand-children ■ (born $\sim [1850, 1940)$)



Demographic & Insurance Notations

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

curtate life expectancy for T_x is defined as

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

actuarial present value of the annuity of an individual age (x) is

$$ a_x = \sum_{k=1}^{\infty} \nu^k \, {}_kp_x \text{ or } a_{x:\overline{n}|} = \sum_{k=1}^{n} \nu^k \, {}_kp_x, $$

and whole life insurance (see Bowers et al. (1997))

$$ A_x = \sum_{k=1}^{\infty} \nu^k \, {}_kp_x \cdot q_{x+k} \text{ or } A^1_{x:\overline{n}|} = \sum_{k=1}^{n} \nu^k \, {}_kp_x \cdot q_{x+k}. $$

Historical Mortality

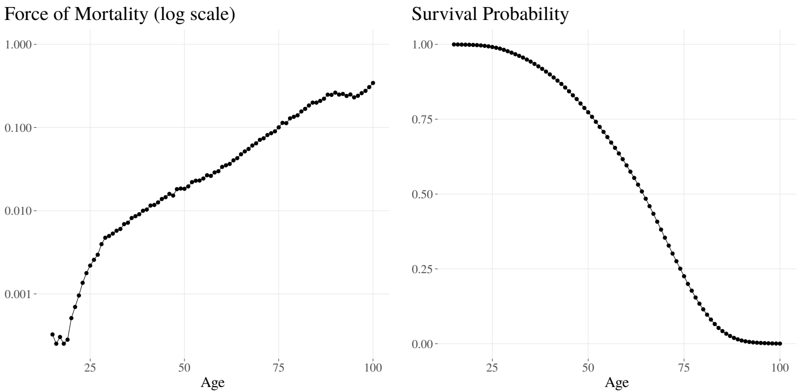
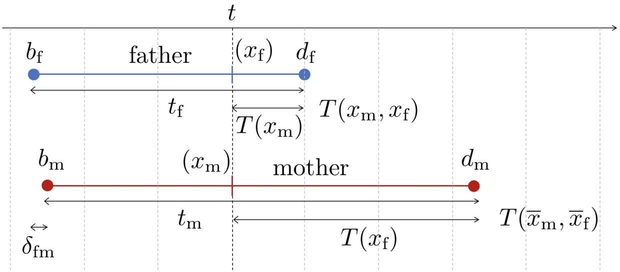


Figure 1: Survival distribution ${}_tp_0 = \mathbb{P}[T > t]$ and force of mortality ${}_1q_x = \mathbb{P}[T \leq x+1 | T > x]$ (log scale), against historical data.

Husband-Wife dependencies

i	birth (b_f) $b_{f,i}$	death (d_f) $d_{f,i}$	age (t_f) $t_{f,i}$	birth (b_m) $b_{m,i}$	death (d_m) $d_{m,i}$	age (t_m) $t_{m,i}$
1	1800-05-04	1835-02-22	34.80356	1762-07-01	1838-01-19	75.55099
2	1778-02-09	1841-02-02	62.97878	1758-07-05	1825-08-03	67.07734
3	1771-01-18	1807-01-17	35.99452	1752-12-28	1815-10-31	62.83641
4	1768-07-01	1814-10-15	46.28611	1768-07-01	1830-12-06	62.42847
5	1766-07-01	1848-01-12	81.53046	1767-02-10	1851-04-22	84.19165
6	1769-06-28	1836-08-28	67.16496	1773-12-17	1825-02-15	51.16222

Table 1: Dataset for the joint life model, father/husband (f) and mother/spouse (m)



Husband-Wife dependencies - Temporal Stability

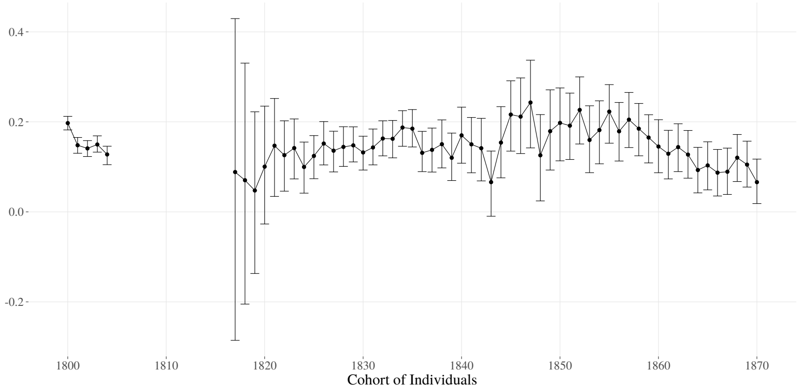


Figure 2: Spearman correlation (T_f, T_m) - per year of birth of the father.

Husband-Wife dependencies

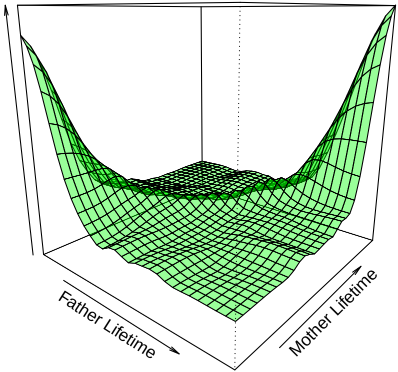


Figure 3: Nonparametric estimation of the copula density, (T_f, T_m).

(using Geenens et al. (2017) estimate)
Here $\widehat{\rho_S} = 0.168$, 95% confidence interval $(0.166; 0, 171)$

Husband-Wife dependencies

Multiple life quantities, e.g. annuities and (whole) life insurance,

$$a_x = \sum_{k=1}^{\infty} \nu^k {}_k p_{x_f} - \sum_{k=1}^{\infty} \nu^k {}_k p_{x_f,x_m}, \quad \text{and} \quad A_x = \sum_{k=1}^{\infty} \nu^k {}_k p_{x_f} - \sum_{k=1}^{\infty} \nu^k {}_k p_{x_f,x_m}$$

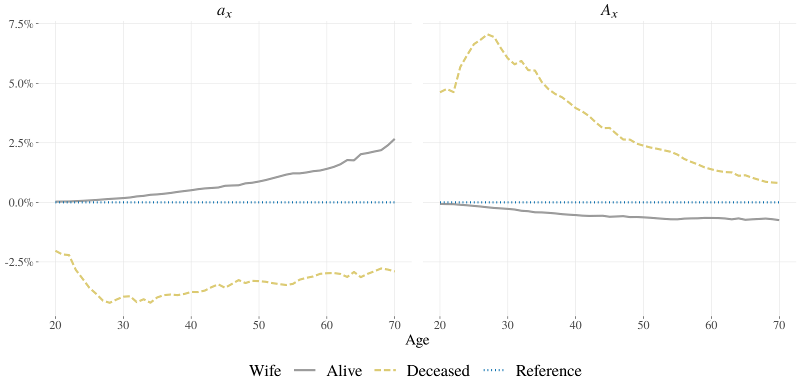


Figure 4: Annuities a_x and (whole) life insurance A_x.

Husband-Wife dependencies

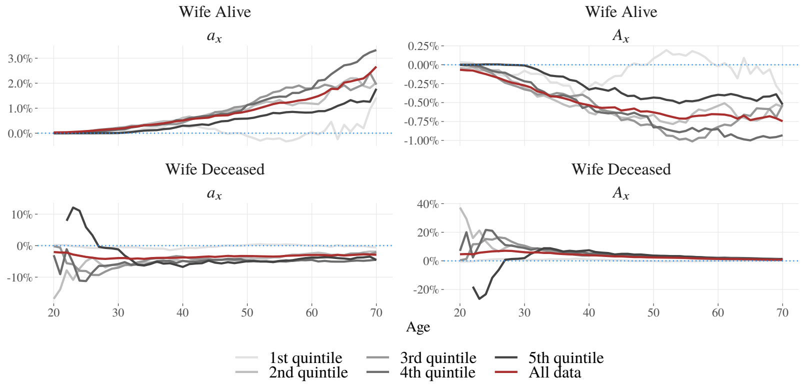


Figure 5: Relative difference to the average (in %) of present value of an annuity (left) and expected present value for a life insurance (right) depending on the age difference between the annuitant and his wife and on the death status (alive on top, deceased at the bottom) of his wife at the time the contract is purchased.

Husband-Wife dependencies

Multiple life quantities, e.g. widow's pension,

$$a_{\mathsf{m}|\mathsf{f}} = \sum_{k=1}^{\infty} \nu^k {}_k p_{x_\mathsf{f}} - \sum_{k=1}^{\infty} \nu^k {}_k p_{x_\mathsf{f},x_\mathsf{m}}, \quad \text{where } {}_t p_{x_\mathsf{f},x_\mathsf{m}} = \mathbb{P}\big[T_{x_\mathsf{f}} > t, T_{x_\mathsf{m}} > t,\big]$$

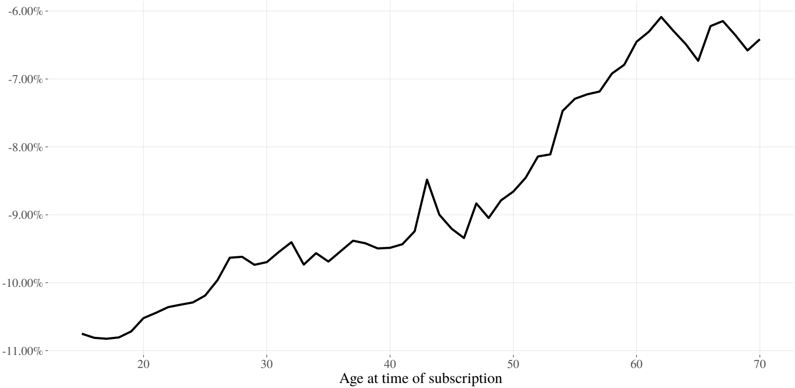


Figure 6: Widow's pension, $a_{\mathsf{m}|\mathsf{f}}$ (relative to independent case $a^{\perp}_{\mathsf{m}|\mathsf{f}}$).

Husband-Wife dependencies

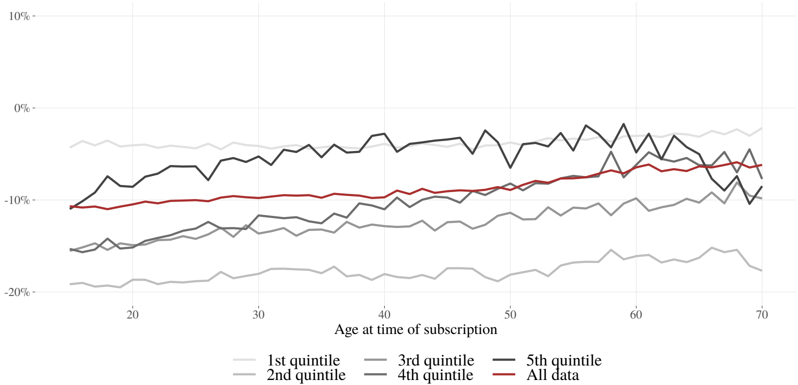


Figure 7: Widow's pension, $a_{m|f}$ (relative to independent case $a^{\perp}_{m|f}$), as a function of x_m, depending on the age difference between the annuitant and her late husband

Husband-Wife dependencies

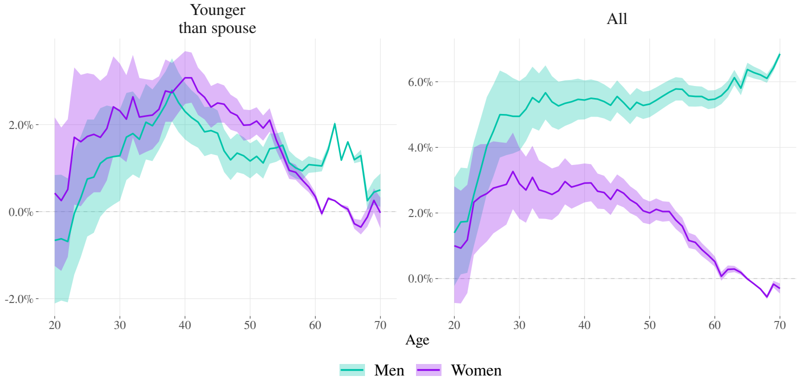


Figure 8: Relative change in residual life expectancy depending on the death status of the spouse and who is older in the couple.

Husband-Wife dependencies

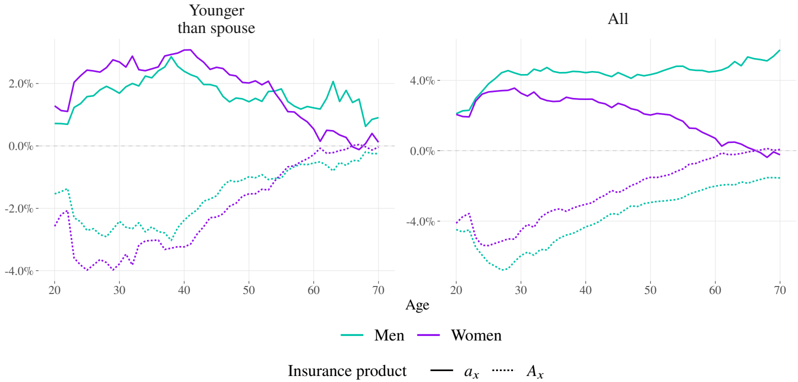


Figure 9: Relative change in present value of insurance products depending on the death status of the spouse at the time of signing the contract.

Children-Parents

"*inheritance of longevity*"
coined in Pearl (1931)

"*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*"
Vaupel (1988)

"*the chance of reaching a high age is transmitted from parents to children in a modest, but robust way*"
Vågerö et al. (2018)

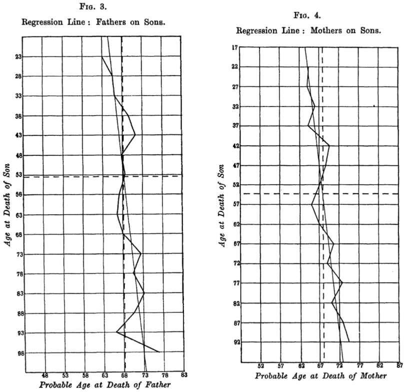


Figure 10: Son vs. parents Beeton and Pearson (1901).

Children-Parents

Beeton and Pearson (1901), regression of T_{x_c} given T_{x_f} or T_{x_m}

slope :
Daughter–mother
0.1968 [0.1910,0.20260]
Son–mother
0.1791 [0.1737,0.18443]
Daughter–father
0.1186 [0.1122,0.12507]
Son–father
0.1197 [0.1138,0.12567]

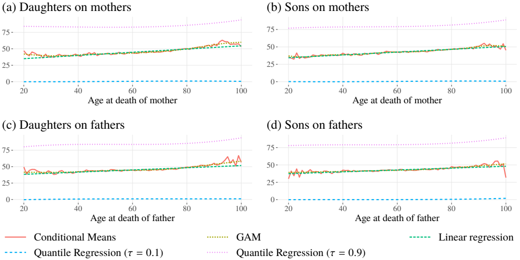


Figure 11: Age of the children given information relative to the parents.

Children-Parents

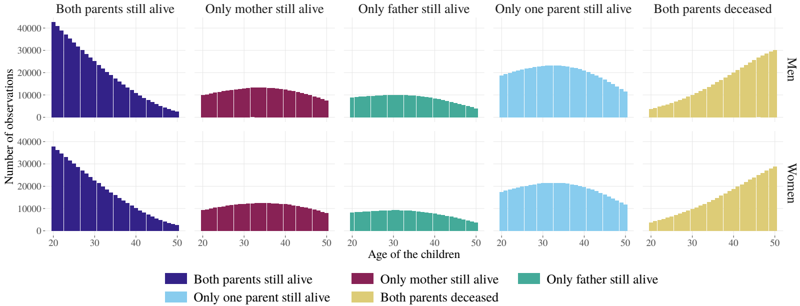


Figure 12: Number of observations for each subset.

Children-Parents

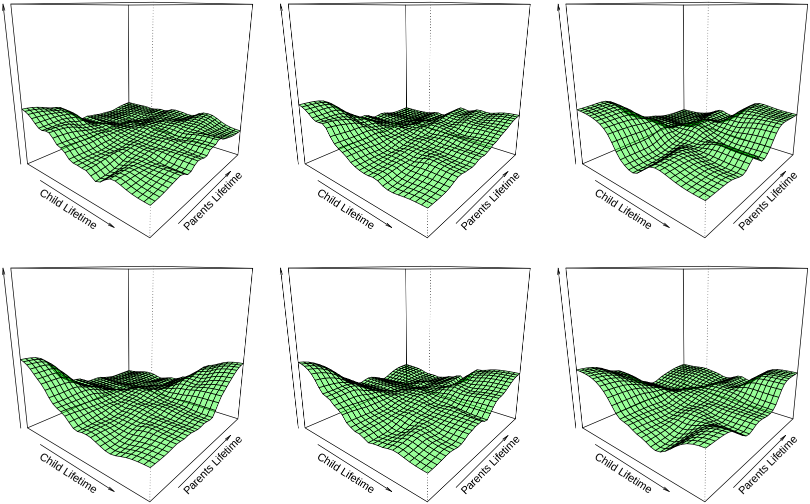


Figure 13: Copula density, children and father/mother/min/max.

Children-Parents, life expectancy

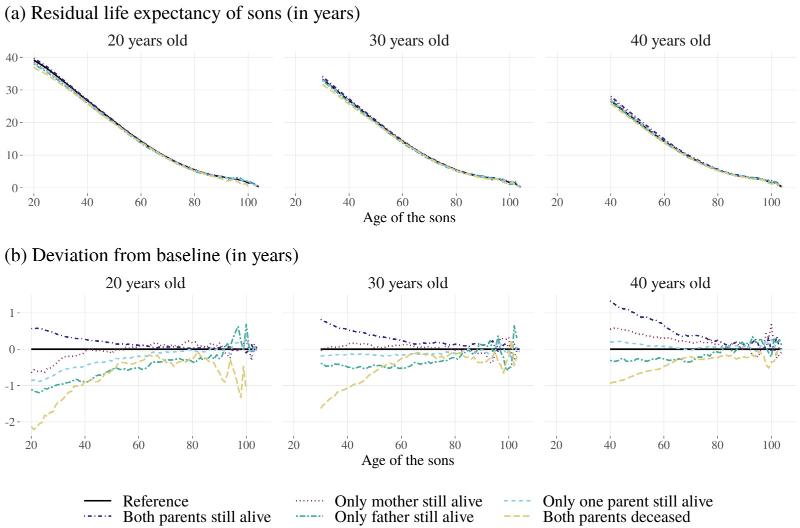


Figure 14: Residual life expectancy e_x with information about parents at age 20, 30 or 40.

Children-Parents, life expectancy

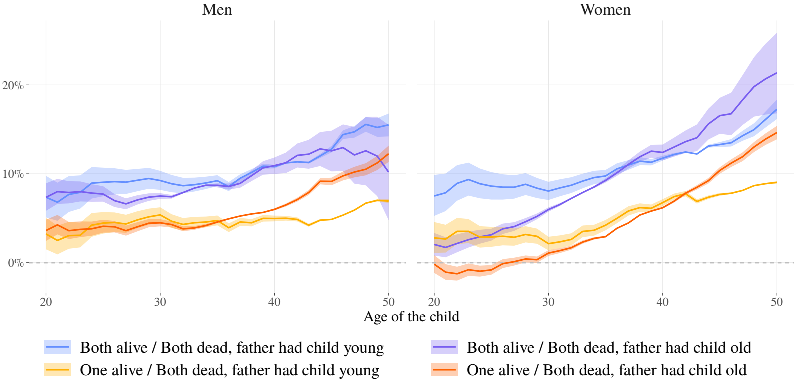


Figure 15: Gain in residual life expectancy depending on the death status of the parents and the age at which the father had the child.

Children-Parents, life expectancy

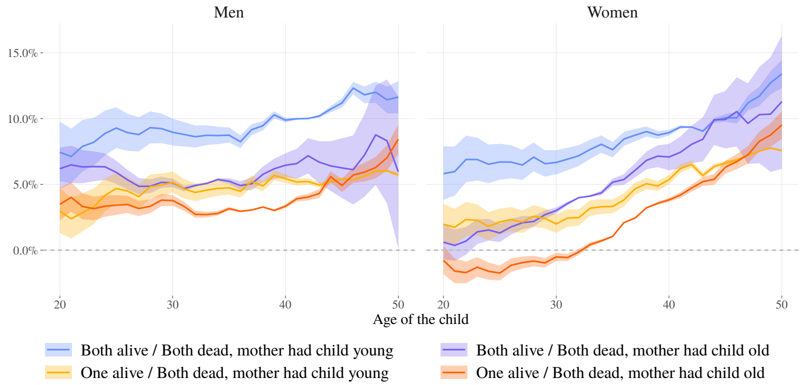


Figure 16: Gain in residual life expectancy depending on the death status of the parents and the age at which the mother had the child.

Children-Parents, annuities and insurance

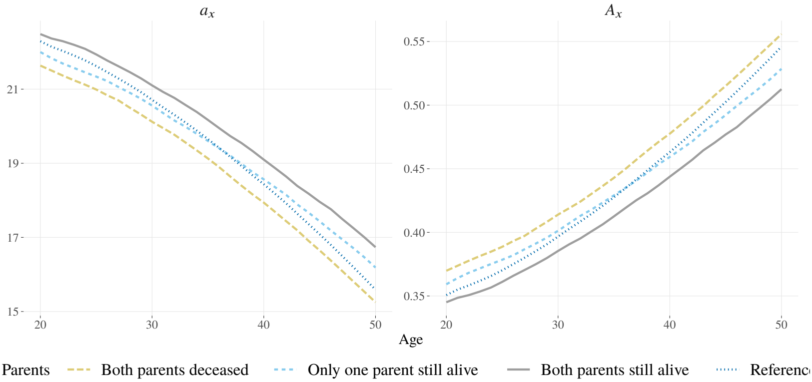


Figure 17: Annuity a_x and whole life insurance A_x, given information about the number of parents still alive, when child has age x.

Children-Parents, annuities and insurance

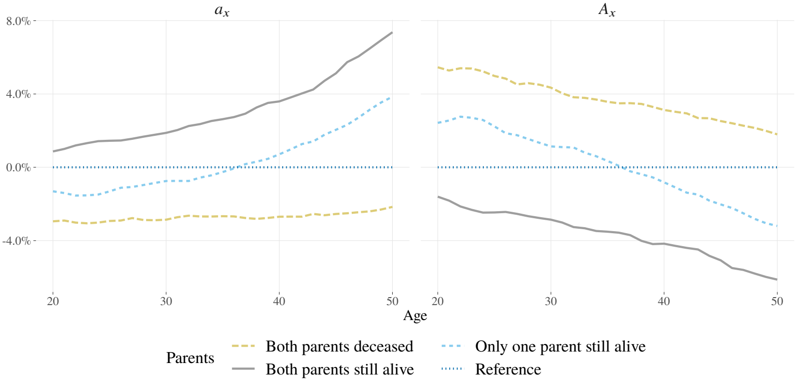


Figure 18: Annuity a_x and whole life insurance A_x, given information about the number of parents still alive, when child has age x (relative difference).

Children-Grandparents

Choi (2020), "*little is known about whether and how intergenerational relationships influence older adult mortality*"

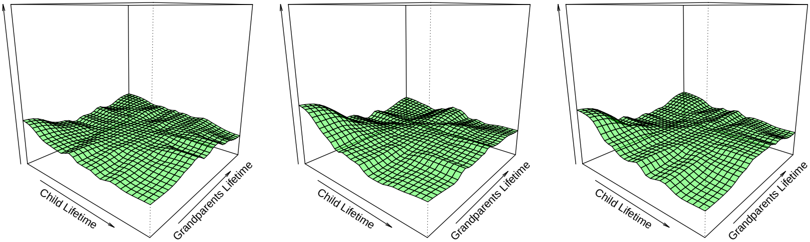


Figure 19: Copula density, children and grandparents min/max/mean.

Children-Grandparents, life expectancy

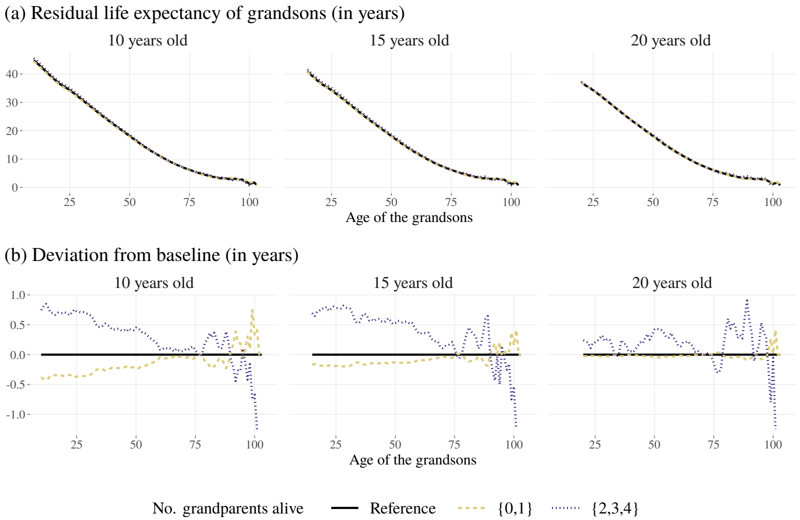


Figure 20: Residual life expectancy e_x with information about grandparents, at age 10, 15 or 20.

Children-Grandparents, annuities and insurance

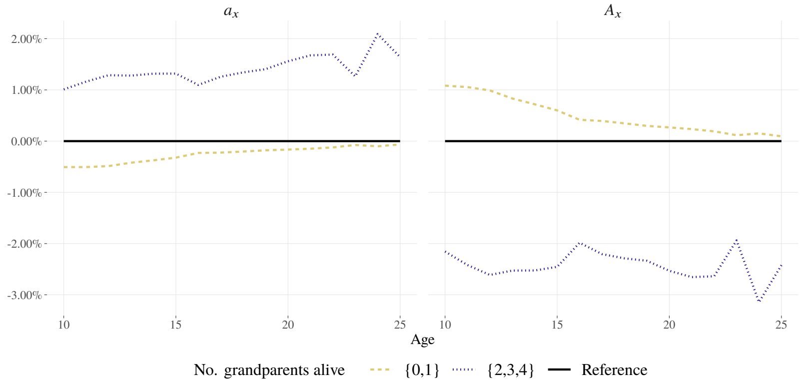


Figure 21: Annuity a_x and whole life insurance A_x, given information about the number of grandparents still alive, when child has age x.

References I

Bean, L. L., May, D. L., and Skolnick, M. (1978). The Mormon historical demography project. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 11(1):45–53. doi:10.1080/01615440.1978.9955216.

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.

Bouchard, G., Roy, R., Casgrain, B., and Hubert, M. (1989). Fichier de population et structures de gestion de base de données : le fichier-réseau BALSAC et le système INGRES/INGRID. *Histoire & Mesure*, 4(1):39–57. doi:10.3406/hism.1989.874.

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.

Carriere, J. (1997). Bivariate survival models for coupled lives. *Scandinavian Actuarial Journal*, pages 17–32.

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*, 61(1):96–112.

Cummins, N. (2017). Lifespans of the European elite, 800–1800. *The Journal of Economic History*, 77(02):406–439. doi:10.1017/s0022050717000468.

References II

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.

Fire, M. and Elovici, Y. (2015). Data mining of online genealogy datasets for revealing lifespan patterns in human population. *ACM Transactions on Intelligent Systems and Technology*, 6(2):1–22. doi:10.1145/2700464.

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

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

Gergaud, O., Laouenan, M., and Wasmer, E. (2016). A Brief History of Human Time. Exploring a database of" notable people".

Kaplanis, J., Gordon, A., Shor, T., Weissbrod, O., Geiger, D., Wahl, M., Gershovits, M., Markus, B., Sheikh, M., Gymrek, M., Bhatia, G., MacArthur, D. G., Price, A. L., and Erlich, Y. (2018). Quantitative analysis of population-scale family trees with millions of relatives. *Science*. doi:10.1126/science.aam9309.

Luciano, E., Spreeuw, J., and Vigna, E. (2008). Modelling stochastic mortality for dependent lives. *Insurance: Mathematics and Economics*, 43(2):234 – 244.

Mandemakers, K. (2000). Historical sample of the Netherlands. *Handbook of international historical microdata for population research*, pages 149–177.

References III

Matthijs, K. and Moreels, S. (2010). The antwerp COR*-database: A unique Flemish source for historical-demographic research. *The History of the Family*, 15(1):109–115. doi:10.1016/j.hisfam.2010.01.002.

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.

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.

Shemyakin, A. and Youn, H. (2001). Bayesian estimation of joint survival functions in life insurance. *Monographs of Official Statistics. Bayesian Methods with applications to science, policy and official statistics, European Communities*, 4891496.

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

Vågerö, D., Aronsson, V., and Modin, B. (2018). Why is parental lifespan linked to children's chances of reaching a high age? a transgenerational hypothesis. *SSM - Population Health*, 4:45 – 54.

Youn, H. and Shemyakin, A. (1999). Statistical aspects of joint life insurance pricing. *1999 Proceedings of the Business and Statistics Section of the American Statistical Association*, 34138.