



Computations of French Lifetables by Département

Florian Bonnet

► To cite this version:

| Florian Bonnet. Computations of French Lifetables by Département. 2018. halshs-01955515

HAL Id: halshs-01955515

<https://shs.hal.science/halshs-01955515>

Preprint submitted on 14 Dec 2018

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

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



WORKING PAPER N° 2018 – 57

Computations of French Lifetables by *Département*

Florian Bonnet

JEL Codes:
Keywords:



PARIS-JOURDAN SCIENCES ÉCONOMIQUES

48, Bd JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 80 52 16 00=
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
ÉCOLE DES PONTS PARISTECH – ÉCOLE NORMALE SUPÉRIEURE
INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE – UNIVERSITÉ PARIS 1

Computations of French Lifetables by *Département*, 1901–2014*

Florian Bonnet[†]

November 26, 2018

Abstract

Debates concerning the territorial divide in France are deep. To bring a contribution to this issue, I compute the departmental lifetables since 1901, for both men and women. In this paper, I present the raw data collected to do so, namely yearly births and deaths by age as well as population by age at each census carried out during the 20th century. I add statistics according to military mortality and mortality in deportation to cover the periods of the Two World Wars. I also present the methods I use to compute these lifetables, which come mainly from the Human Mortality Database protocol. I revise this protocol to take into account the specificities of French departmental data, mainly the few changes in French departmental boundaries, the underestimation of infant mortality and the lack of raw data homogeneity. This new database complements a still limited supply of long-term mortality statistics computed at local level.

*This study received financial support from the ERC Grant "Demographic Uncertainty" led by Hippolyte d'Albis. The author would especially like to thank Magali Barbieri, Carl Boe and Hippolyte d'Albis for their many advices.

[†]Université Paris 1 Panthéon-Sorbonne, Paris School of Economics, florian.bonnet@psemail.eu

Part I

Article

1 Introduction

Life expectancy has risen sharply in France since the beginning of the 20th century. The lifetables calculated by Vallin and Meslé (2001) for the 19th and 20th centuries show that men life expectancy at birth was 33 in 1806, 44.5 in 1898, 60 in 1946 and 74.5 in 1997. This increase at the national level does not say anything about increases at the local level. As such, significant differences exist between the French *départements*. Barbieri (2013) worked on departmental mortality and showed that the life expectancy at birth of men for the period 2006–2008 was 74.4 years in *Nord*, compared to 79.7 years in *Hauts-de-Seine*, a difference of more than 5 years. This situation explains why the debate on the territorial divide according to health is important in France. Indeed, departmental differences can not be explained from a public policy point of view: the State has to reduce these inequalities. In order to inform public decision-makers in their choices, it is important to know the history of these departmental differences.

Consequently, I compute in this paper the yearly departmental lifetables by sex for all French metropolitan *départements* between 1901 and 2014. The computation of these lifetables is based on the exhaustive collection of population flows (deaths by age and sex, births by sex) and population stocks at each census (population by age and sex). I exploit a French unique characteristic: since 1789, this country is divided into around 100 geographical units of similar size, namely *départements*. This division has changed very little during two centuries, and the statistical centralizations have been carried out at this geographical level. Moreover, in order to take into account the two World Wars that affected France between 1914–1918 and 1939–1945, I have collected in two original sources the military deaths by age during the two wars as well as the deaths in deportation by age and sex during the Second World War. With these lifetables, I get life expectancies and mortality rates at each age for more than 100 years. In addition, I get populations by age and sex at each January 1st.

These lifetables at the subnational level complete a still incomplete literature. Bonneuil (1997) worked on departmental mortality in the 19th century: he computed women lifetables by five-year period and for five-year age groups. He followed Van de Walle (1974) who computed similar lifetables with a different methodology. These two authors have not studied in the same way men's mortality, because of strong fluctuations due to the wars which afflicted France at this time. From 1954 to 1999, Daguet (2006) grouped lifetables established at the departmental level, but only for the census years. Barbieri (2013) used in her study departmental lifetables calculated by INSEE for the period 1975–2008. However, these data were provided exceptionally. Vallin and Meslé (2005) used departmental life expectancies for the period 1906–1954. However, both reconstruction methods and data have never been published. Lastly, various mortality indicators are available in official publications, namely *Statistique Annuelle du Mouvement de la Population*.¹ However, these indicators are relatively scarce: they relate only to infant mortality rates, or standardized mortality rates.

In addition, the lifetables I compute are based on a unified methodological protocol for the whole period 1901–2014, which is not the case of the papers previously cited. This methodological protocol is available

¹See, for example, page 74 of this publication for the year 1938.

in Wilmoth et al (2007). Many researchers are using this protocol to compute national lifetables for a large number of countries. It is also used to compute lifetables at the local level in two OECD countries. The results according to Canadian provinces for the period 1921–2011 are available in the Canadian Human Mortality Database², and those according to the Japanese provinces since 1975 are available in the Japan Mortality Database.³ This paper therefore complements a still limited supply of local mortality data freely available by adopting an internationally recognized protocol; this allows international comparisons without methodological bias.

The rest of this paper is organized as follows. In Section 2 I present the statistical sources used to compute departmental lifetables. The methods used are explained in Section 3 in which I distinguish the methods coming from the HMD protocol and the methods specific to this study. In Section 4 I illustrate some of the results available in this new database. Part 2 of this paper is the methodological appendix.

2 Sources

Computations of departmental lifetables requires two types of data: population movement (deaths and births domiciled), and population censuses. The deaths collected do not only concern civilian deaths: both military deaths during the two World Wars and deportation deaths between 1939 and 1945 have been included.

2.1 Deaths

Civilian deaths of each *département*, each sex and each year over the period 1901–2014 have been retrieved from the population movement statistics published by *Statistique Générale de la France* (SGF) and then by *Institut National de la Statistique et des Etudes Economiques* (INSEE). I have retrieved deaths by age group recorded in home *département*. Tables 7 and 8 in Appendix (Part 2, Section 6.9) provide sources in which raw statistics have been found. In addition, I have collected in Vallin and Meslé (2001) single-age and sex-specific civilian deaths at the national level for the same period.

I have retrieved deaths during the two World Wars from Defense Ministry’s website.⁴ They are available by year of birth at the departmental level, and by year of birth and year of death at the national level.

Individuals who died during deportation in the Second World War are not included in the civilian population movement. However, they were nearly 100,000. I have decided to include them in my statistics, using data from Memorialgenweb Website.⁵ This database records deportees who left France and died in deportation published in the *Journal Officiel*, by *département* of birth if they were born in France, and by country of birth otherwise. Table 1 presents figures of the foreign-born deportees by country of birth. One can see that the Poles were the most numerous. Although this database is not exhaustive, the large number of observations provides a sample close to the total of deaths in deportation.

²Computed by researchers in “Université de Montréal”, www.demo.umontreal.ca/chmd/.

³Computed by researchers at the National Institute of Population and Social Security Research, <http://www.ipss.go.jp/p-toukei/JMD/index-en.asp>.

⁴<http://www.memoiredeshommes.sga.defense.gouv.fr/>

⁵<http://www.memorialgenweb.org/memorial3/deportes/index.php>, forwarded on March 7th, 2016

Table 1: SUMMARY OF FOREIGN-BORN DEPORTEES BY NATIONALITY

Country	Deportees	In % of foreign-born deportees
Pologne	13,599	40.46%
Spain	5,075	15.10%
Russia	2,741	8.16%
Germany	2,425	7.21%
Romania	1,861	5.54%
Turkey	1,511	4.50%
Algeria	1,050	3.12%
Greece	939	2.79%
Italia	535	1.59%
Ukraine	534	1.59%

Notes: Deportees by country of birth in the Memorialgenweb’s database.

2.2 Births

I have retrieved births by year, sex and mother’s home *département* for the period 1901–2014. I have also recovered stillbirths by mother’s home *département* and year (both males and females). Finally, I have retrieved births by year, sex and mother’s home *département* for the period 1853-1900.⁶

2.3 Censuses

Finally, I have collected populations by birth year, home *département* and sex for each census of the period 1901–1962 from hard-copy publications of SGF and INSEE. For the period 1968–2014, these statistics have been found in on-line sources. These data are not available for each year because censuses were held at varying intervals. Between 1901 and 2014, censuses were made in 1901, 1906, 1911, 1921, 1926, 1931, 1936, 1946, 1954, 1962, 1968, 1975, 1982, 1990, 1999, 2008, 2013 and 2014.⁷

3 Methods

The protocol I use to compute departmental lifetables is largely inspired by the one of the Human Mortality Database (HMD). This database gathers all national lifetables computed using these methods. However, since my database is specific both for the small numbers in each *département* and the time period chosen (including the two World Wars), I have added specific methods.

3.1 HMD Protocol Methods

3.1.1 Raw Data Adjustments

Raw data adjustments according to deaths are the main issue since they are aggregated into five-year age groups until 1967 and by single age between 1968 and 2014. To get a 1×1 format (single age, year of death) for the deaths between 1901 and 1967, I distribute deaths at unknown age among age groups, and

⁶Tables 9, 10 and 11 in Appendix (Part 2, Section 6.9) give sources in which raw statistics have been found.

⁷Table 12 in Appendix (Part 2, Section 6.9) gives sources in which raw statistics have been found.

adjust the curve of cumulative deaths by cubic splines. Cubic Spline is a semi-parametric estimation method which joins the points of a cumulative distribution by third degree polynomials. Let $Y(x) = \sum_{u=0}^{x-1} D_u$ be the cumulative number of deaths up to age x . $Y(x)$ is known for a limited collection of ages including 1, 5, 10... etc from the raw data. I know $Y(x)$ for both the highest age in the distribution (80, 90 or 100) and the age above which no further deaths are observed, set at 105. Equation (1) fits a cubic spline by using these values (the indicator function $I(\cdot)$ equals one if the logical statement within parentheses is true and zero otherwise):

$$Y(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \beta_1 (x - k_1) I(x > k_1) + \dots + \beta_n (x - k_n) I(x > k_n). \quad (1)$$

I have to estimate the vector $(\alpha_0; \alpha_1; \alpha_2; \alpha_3; \beta_1; \dots; \beta_n)$ which contains $n + 4$ coefficients, but I only know $n + 2$ values of $Y(x)$, and therefore $n + 2$ constraints. Two further constraints must be introduced to identify the model. First I assume that there is no death at the upper bound, namely 105. Second I assume that deaths observed between 1 and 5-year-old occurred between 1 and 2 year-old. $\hat{Y}(x)$ are calculated for all ages, for each *département*, sex and year. Deaths at age x are found as follows:

$$\hat{D}(x) = \hat{Y}(x+1) - \hat{Y}(x).$$

Negative death counts may occur when the deaths in five-year age groups are extremely low.⁸ The method is to set zero-deaths in age groups where negative counts occur. To balance this, deaths in the adjacent age groups are reduced pro-rata their number of deaths. If D_{neg} is the sum of negative death counts for an observation, D_s^* the deaths at age s after allocation of negative death counts, D_s the estimated deaths at age s before allocation of negative death counts, x_1 and x_2 the lower and higher limits of the interval in which the negative death counts are observed, then:

$$\begin{cases} D_s^* = 0 & \text{for } s \in [x_1, x_2], \\ D_s^* = D_{neg} \times \frac{D_s}{\sum_{i \in \Omega_1} D_i} & \text{for } s \in [x_1 - 5, x_1] \cup [x_2, x_2 + 5], \\ D_s^* = D_s & \text{otherwise.} \end{cases} \quad (2)$$

Deaths estimated by cubic spline are too imprecise to be used at advanced ages: open-age interval of deaths is too low (see Tables 7 and 8, Column 5). These deaths are adjusted by means of the Kannisto model, which assumes a survival curve of logistic form, with a zero-asymptote for very old ages. I use this method for deaths beyond the open-age interval – different according to the periods, I keep a maximum of 95 so that estimates are not hindered by too small figures – and rely on the deaths observed for ages 10 years below this limit. Thus, if the open-age interval begins at age 90, I use the ages 80–89. Formally, I compute a fictitious survival curve $S(80+x)$:

$$S(80+x) = \frac{\sum_{u=80+x}^{105} D_u}{\sum_{u=80}^{105} D_u} \quad \text{for } x = 0, 1, 2, \dots, 9. \quad (3)$$

This survival function conditional on reaching age 80 may be seen as tracking a “synthetic extinct cohort”, since it is based on annual deaths and not on deaths in the cohort itself. Assuming that this fictitious

⁸ This usually happens at around age 30. I count only seven of these occurrences, but they need to be adjusted so as not to have negative mortality rates later.

cohort displays survival probabilities that can be fitted by the Kannisto model, the survival function $s(x)$ is:

$$s(x) = \left(\frac{1+a}{1+ae^{b(x-80)}} \right)^{1/b}. \quad (4)$$

with estimated values for a and b , I compute $\hat{s}(x)$ et $d(x) = \hat{s}(x) - \hat{s}(x+1)$. Finally, I obtain deaths at each age:

$$D(x) = \sum_{u=90}^{105} D_u \times \frac{d(x)}{\hat{s}(90)}. \quad (5)$$

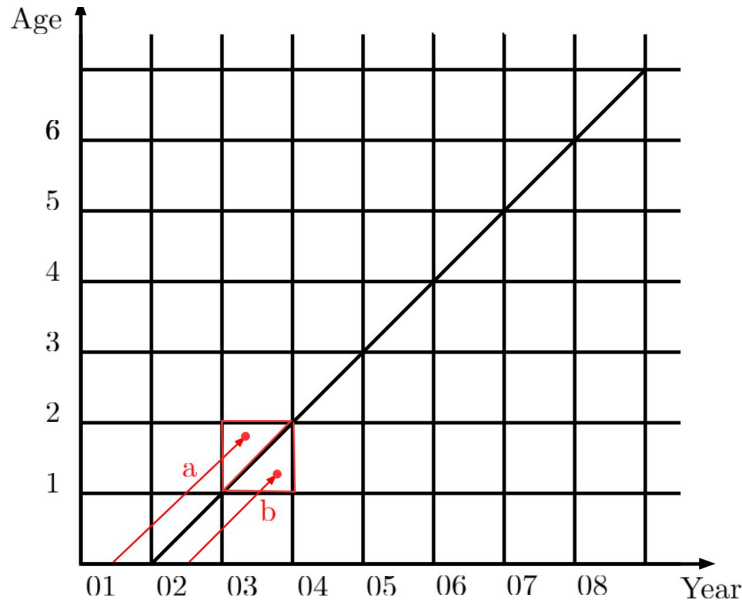
I finally proceed to a uniform adjustment so that the sum of the departmental deaths for each age, year, and sex corresponds to the national data.

For censuses, raw data are generally available for groups of five-year of births. I use the Cubic Splines method in the same way to estimate populations according to their year of birth.

3.1.2 Splitting Deaths into Lexis Triangles

Figure 1 presents deaths by year and age. They may be split into two triangles for a single year, known as Lexis triangles. For individuals who died between ages 1 and 2 in 1903, one may distinguish two kinds of deaths. The first who died between ages 1 and 2 in 1903, born in 1901 (“a” on Figure 1, upper triangle). The others who died between ages 1 and 2 in 1903, born in 1902 (“b” on Figure 1, lower triangle).

Figure 1: AN EXEMPLE OF LEXIS DIAGRAM



Overall, if the probability of death is equiprobable over time, one could think that the distribution of annual deaths by age for half in the lower triangle and the other half in the upper triangle would be sufficient. This is not, for two main reasons. The first is that infant mortality, when high, is observed largely in the first days after birth, and must therefore be integrated into the lower triangle. The second concerns the relative size of cohorts, which also influences the distribution between triangles. When the flow of births varies greatly from one year to the next (e.g. during the two World Wars), the half-death distribution in the lower triangle is strongly biased. The HMD protocol sets a sex-specific equation allowing the distribution

of deaths in Lexis triangles. This equation takes into account the relative size of two successive cohorts, age, some historical events (e.g Spanish influenza), and the infant mortality rate. If we call x the age and t the year, these sex-specific equations are as follows (Equation (6) for women, Equation (7) for men):

$$\begin{aligned}\hat{\pi}_d(x, t) = & 0.4710 + \hat{\alpha}_F + 0.7372 [\pi_b(x, t) - 0.5] \\ & + 0.1025 I_{t=1918} - 0.0237 I_{t=1919} \\ & - 0.0112 \log IMR(t) - 0.0688 \log IMR(t) I_{x=0} + 0.0268 \log IMR(t) I_{x=1} \\ & + 0.1526 [\log IMR(t) - \log(0.01)] I_{x=0} I_{IMR(t) < 0.01}\end{aligned}\quad ; \quad (6)$$

$$\begin{aligned}\hat{\pi}_d(x, t) = & 0.4836 + \hat{\alpha}_H + 0.6992 [\pi_b(x, t) - 0.5] \\ & + 0.0728 I_{t=1918} - 0.0352 I_{t=1919} \\ & - 0.0088 \log IMR(t) - 0.0745 \log IMR(t) I_{x=0} + 0.0259 \log IMR(t) I_{x=1} \\ & + 0.1673 [\log IMR(t) - \log(0.01)] I_{x=0} I_{IMR(t) < 0.01}\end{aligned}\quad . \quad (7)$$

$\hat{\pi}_d(x, t)$ is defined as the proportion of death of a given year and age allocated in the lower triangle. α_F and α_H are age-specific values coming from the HMD protocol.

$\pi_b(x, t)$ is defined as the ratio of births between two successive cohorts and calculated only once for both sexes:

$$\pi_b(x, t) = \frac{B(t-x)}{B(t-x) + B(t-x-1)} \quad (8)$$

Long historical series are required to calculate this ratio for all the cohorts tracked between 1901 and 2014. One can take individuals aged 80 in 1901 as an example. To calculate this ratio one needs birth in 1820 and 1821. I was unable to do so: my birth records only go back to 1853. For earlier years I assume that births before 1853 were equal to births in 1853 and use a birth ratio of 0.5.

$IMR(t)$, the same for both sexes, is calculated as follows:

$$IMR(t) = \frac{D(0, t)}{\frac{1}{3}B(t-1) + \frac{2}{3}B(t)} \quad (9)$$

If births are not available for one of the two years, $IMR(t)$ is calculated as follows⁹:

$$IMR(t) = \frac{D(0, t)}{B(t^*)} \quad (10)$$

with t^* the year for which births are available.¹⁰

3.1.3 Computations of Populations by Age at 1st January of each Year

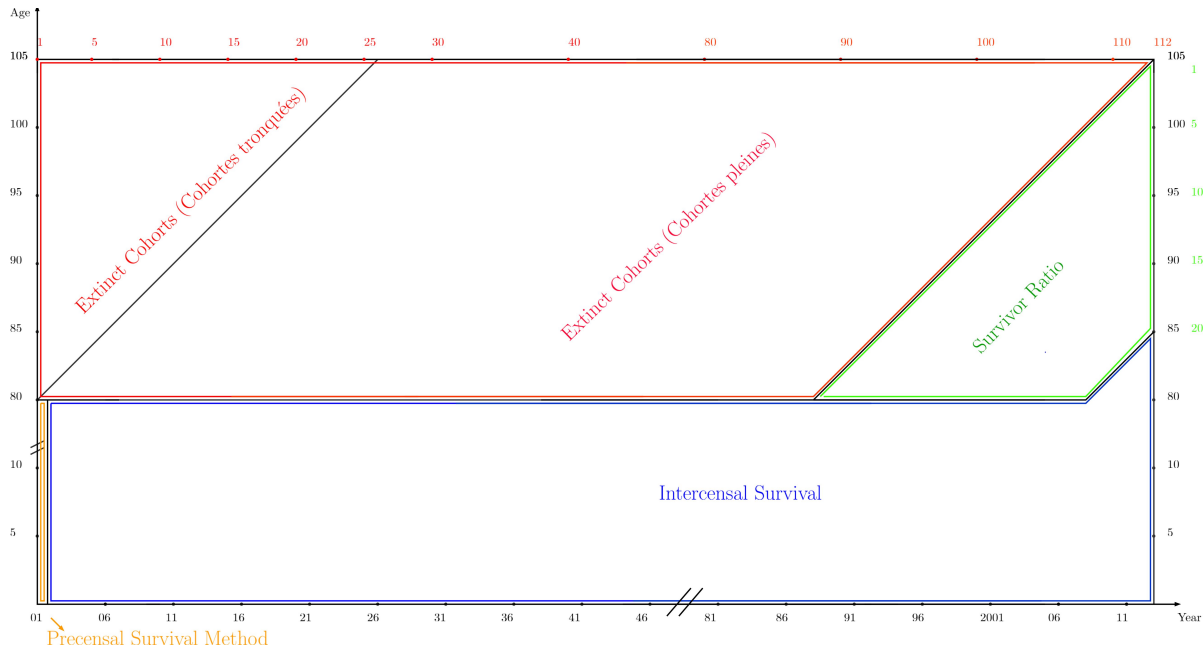
To calculate the mortality rates required for lifetables, I need populations by age at 1st January for each year from 1901 to 2014. I get populations by age in 2014 from official statistics so I may calculate populations

⁹When $IMR(t)$ is equal to zero because of no infant deaths, I assume a 0,00000001 IMR value so that $\log IMR(t)$ can be calculated.

¹⁰I obtain proportions of deaths in the lower triangle greater than 1 for 28 female observations and 30 for male observations, all in 1918 or 1919 and for deaths under age 1. This is due to the Spanish influenza epidemic, the high infant mortality rate and the size differences between the cohorts born in 1918 and 1919. To tackle this issue, the death proportions in the lower triangle are set at 1, leading to zero death in the upper triangle for these observations.

by age for the 1901–2013 period.¹¹ Figure 2 reveals the four methods used for various periods and ages. Section 6.1 of the Appendix precisely specifies each method used.

Figure 2: METHODS FOR COMPUTATIONS OF POPULATION AT 1st JANUARY



Notes: Methods used to compute populations by age at each 1st January. For more details, see Section 6.1 of Appendix.

The “Intercensal Survival” method is used to estimate the population under age 80 from 1902 to 2013. Starting from one census (say, 1901) the population by age at the following census (1906) is estimated by subtracting from the population by age in 1901 the deaths that occurred from 1901 to 1906. The difference between estimated and recorded populations in 1906, due to measurement errors and migrations, is then attributed to the intercensal population figures.

Second, the “Precensal Survival” method is used to estimate the population under age 80 in 1901. Since “Intercensal Survival” estimates the population under age 80 on 1st January of each intercensal year, I cannot compute population by age on 1st January the year of the first census. To do so, I use the population on the day of the 1901 census and add the deaths occurring between 1st January 1901 and the census day. Since there is no second census available as with “Intercensal Survival” method, I cannot correct for migrations and errors: I assumed that in so short a period these are minimal.

With the “Extinct Cohorts” method I can estimate the population aged 80 and over born in the cohorts that died between 1901 and 2013. I assume that migrations after age 80 are small; I compute the population of a cohort still alive by summing its future observed deaths.

Finally I estimate the population aged 85 and over in 2014 with the “Survivor Ratio” method. I assume that the survival ratio between two ages for the extinct cohorts can be applied to the still living cohorts in order to estimate their size at the last census. The 85-and-over estimated are then adjusted by the 85-and-over recorded in 2013. After this adjustment I compute the size of the intermediate populations located in the green quadrilateral by subtracting step by step the observed deaths.

¹¹I do not need to calculate populations by age in 2014 since these data are available in raw statistics provided by INSEE

3.1.4 Adjustment of Computed Mortality Rates

I can compute departmental mortality rates by age and sex with deaths in Lexis triangles and populations at each 1st January. Mortality rates are the ratio between the number of deaths and the number of individuals exposed to the risk¹²:

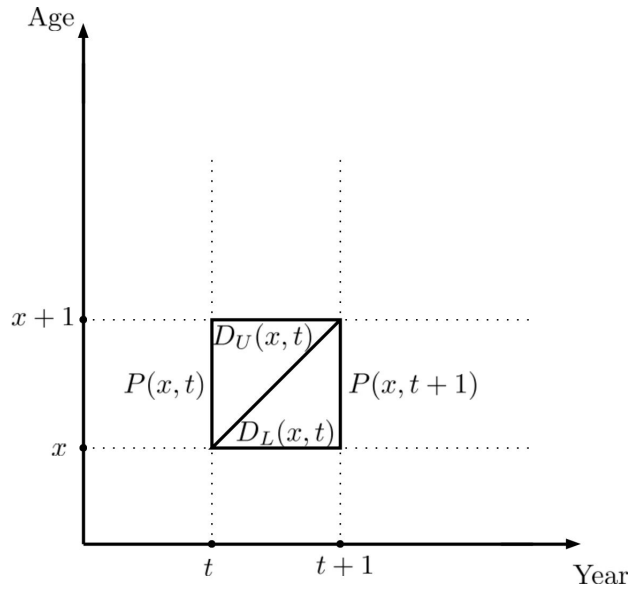
$$M_{xt} = \frac{D_{xt}}{E_{xt}} = \frac{D_L(x, t) + D_U(x, t)}{\frac{1}{2} [P(x, t) + P(x, t + 1)] + \frac{1}{6} [D_L(x, t) - D_U(x, t)]}. \quad (11)$$

Note that I do not calculate populations for 2015, although these are needed for 2014. To estimate mortality rates for that year, I assume that the population at each age in 2015 is equal to that in 2014, and the formula becomes:

$$M_{x2014} = \frac{D_{x2014}}{E_{x2014}} = \frac{D_L(x, 2014) + D_U(x, 2014)}{P(x, 2014) + \frac{1}{6} [D_L(x, 2014) - D_U(x, 2014)]}. \quad (12)$$

Figure 3 presents the set of data needed to compute mortality rates.

Figure 3: MORTALITY RATES COMPUTATIONS



These rates are not used directly to calculate lifetables. I smooth mortality rates beyond age 90 in order to avoid erratic fluctuations due to small numbers of deaths and population at risk. The instantaneous probability of dying over age 80 in the Kannisto model can be expressed as follows (with a and $b \geq 0$):

$$\mu_x(a, b) = \frac{ae^{b(x-80)}}{1 + ae^{b(x-80)}}. \quad (13)$$

Mortality rates estimated with the Kannisto model $M_x(a, b)$ are:

$$M_x(a, b) = \mu_{x+0.5}(a, b). \quad (14)$$

If $D_x \sim \text{Poisson}(E_x \mu_{x+0.5}(a, b))$, then parameters a and b may be calculated by minimizing the following function:

¹²For the explanation of the presence of the difference between the two Lexis triangles at the denominator, please see HMD Protocol, Appendix E.

$$-\log L(a, b) = \sum_{x=80}^{105} [D_x \log \mu_{x+0,5}(a, b) - E_x \mu_{x+0,5}(a, b)]. \quad (15)$$

I can calculate $\hat{M}_x(\hat{a}, \hat{b})$ for all ages above 90, with estimated parameters (\hat{a}, \hat{b}) . I assume that the population's mortality rates are equal to the mortality rates in the survival tables (m_x):

$$\begin{cases} m_x = M_x & x \in [0, 89] \\ m_x = \hat{M}_x & x \in [90, 105] \end{cases}. \quad (16)$$

To convert the survival table mortality rates into probabilities of dying, one must define a_x , the mean number of years lived by people dying between ages x and $x + 1$. I assume that deaths are uniformly distributed at each age:

$$\begin{cases} a_x = 1/2 & x \in [1, 104] \\ a_x = \frac{1}{m_{105}^{\infty}} & x = 105+ \end{cases}. \quad (17)$$

For age 0, I follow Preston (2001), who refers on Coale and Demeny (1983)'s lifetables. Thus:

$$\begin{cases} m_0 \geq 0.107 & \begin{cases} a_0 = 0,350 & \text{for women,} \\ a_0 = 0,330 & \text{for men,} \end{cases} \\ m_0 < 0.107 & \begin{cases} a_0 = 0,053 + 2.800 & \text{for women,} \\ a_0 = 0,045 + 2.684 & \text{for men.} \end{cases} \end{cases} \quad (18)$$

The probabilities of death may be calculated as follows:

$$\begin{cases} q_x = \frac{m_x}{1 + (1 + a_x)m_x} & x \in [0, 104] \\ q_x = 1 & x = 105+ \end{cases}. \quad (19)$$

With values of q_x , I can compute each of the lifetable values, for each age: the number of survivors (l_x), the number of deaths (d_x), and the life expectancies (e_x). Two lifetables are estimated: complete in format (1×1) i. e. for each age and each year, and in the format (1×5) i. e. for each age and each group of 5 years. For the sake of readability, lifetables in the (1×5) and (5×5) formats are also estimated. So I get values for age groups $[0, 1[$, $[1, 5[$, $[5, 10[$, $[10, 15[$... etc until ages 105 and over. Section 6.2 of the Appendix reviews the computations made to estimate each of the outstanding lifetable values in each specification.

3.2 Specific Departmental Methods for the Period 1901–2014

The methods presented previously come from the Human Mortality Database protocol. However, they are too general to be applied without correction to the case of French *départements* during the 20th century. These corrections are due to three main issues: the quality of the raw data, the two World Wars, and the territorial changes in my departmental classification.

3.2.1 Specific Methods Due to Data Quality

I include false stillbirths in births and deaths before first birthday as Vallin and Meslé (2001) did for the national lifetables. In their work they explained that before 1993, a child born alive who died before the official statement of birth was considered to be stillborn, which distorts both deaths before first birthday and births. To reduce this bias I have retrieved from official publications the false stillbirths by sex at the national level (Vallin and Meslé, 2001) and I have distributed them among *départements* pro rata of stillbirths. I added them to deaths before first birthday and births.

Moreover, the data retrieved from censuses are not of identical quality so I make some adjustments. The first is to distribute individuals of unknown year of birth pro rata of the numbers in known year of birth age groups. Although this do not present any problem for most censuses, this is not true for the 1901 one, when these numbers were included in those of the open-age interval. The second is to split the open-age interval 80-year-old and over in the 1906, 1921, 1926, 1931, 1936 and 1946 censuses. This open-age interval occurs too soon and generated some negative population figures. I split it in two age groupings: ages 80 to 84 and 85 and over. For these first two adjustments I use the 1911 census particularly detailed. Moreover, younger age groups did not always use the same variables: sometimes year of birth, sometimes age. I use a linear interpolation to compute figures per year of birth. Section 6.3 of the Appendix presents in a more detailed manner these three adjustments.

3.2.2 Specific Methods due to the Two World Wars

The two World Wars had significant demographic effects both at national and departmental level. The first is due to internal migrations caused by the conflict and the France's division into occupied and unoccupied zones in 1940. The raw statistics give no direct indication for this question. The second concerns the heavy military losses, which had to be included in death statistics. On this particular point, this study is the first to integrate military and deportation deaths into lifetables at subnational level.

Ideally, the statistics of military deaths should be available according to the age and the year of the soldier's death, as well as his home *département* before the war. Since the sources used are incomplete, I couple two different matrices. The first provides the total of deaths by *département* and year of birth. It comes from the Defense Ministry's database, which lists all the "*Morts pour la France*" (MPLF) of the two wars. The second provides the total of deaths at the national level by year of death and year of birth. It mobilizes the crowd-based indexing on the *Mémoire des Hommes* website: each individual, using his personal research on a specific soldier, inform both his year of death and his year of birth. This work has been done for just over 20% of total deaths. I wonder if this sample is representative of the distribution by year of death. For that, I use Pedroncini (1992)'s work: it gives total military deaths by year of death. Table 2 shows these distributions according to both sources. Even if discrepancies exist, I can use the sample coming from *Mémoire des Hommes*. Data by year of birth and year of death are therefore extracted from the Defense Ministry's database.

By cross-referencing these two matrices, I get a matrix giving total deaths by *département*, year of birth and year of death. I assume that there is little variation between *départements* in the year of death according to the cohort.

This distribution of deaths is then adjusted by the total of deaths as estimated by researchers at national level, so as to verify the overall consistency of the various sources. Prost (2008) makes an inventory of the

Table 2: DISTRIBUTION BY YEAR OF DEATH OF SOLDIERS

	Year	1914	1915	1916	1917	1918	Total
<i>Mémoire des hommes</i>	Deaths	75,403	82,878	50,933	34,436	52,459	296,109
	% of the total	25.46%	27.99%	17.20%	11.63%	17.72%	100%
Pedroncini (1992)	Deaths	301,000	349,000	252,000	164,000	235,000	1,301,000
	% of the total	23.14%	26.83%	19.37%	12.61%	18.06%	100%

statistical estimates of deaths during the First World War. He used the Marin’s report, followed by Hubert (1931) and Dupaquier (1988). Roure’s report cited by Prost (2008) revealed 1,357,800 military casualties, taking into account deaths of foreigners. Hubert (1931) added 40,000 soldiers dead during the 6 months after the armistice as well as sailors. Table 3 summarizes these numbers. Regarding the 28,600 deaths that occurred 6 months after the armistice, I assume that they had been included in the 1919 deaths of the population movement and do not take them into account. With regard to the 75,700 deaths of soldiers coming from settlements and abroad, since these populations were not registered in 1911 in the French *départements* and were surely recorded in the civilian deaths of their home country, I do not keep them in the total. Finally, I obtain 1,304,400 deaths.

Table 3: MILITARY DEATHS DURING THE FIRST WORLD WAR

Source	Variable	Deaths
Roure	Total of French military deaths	1,282,100
	Total foreign-born and settlements	75,700
	Total Roure	1,357,800
Hubert	Deaths 6 months after armistice	28,600
	Sailors	11,400
	Final total	1,397,800

The principle is the same for the Second World War. The two matrices combined come from the Defense Ministry’s database. The total of deaths I use is 200,000, in line with Lagrou et al. (2002). Section 6.4 of the Appendix reviews the departmental classification problems in the Defense Ministry website, as well as the cubic splines used to distribute departmental deaths by single year of birth for the two World Wars.

According to deportation during the Second World War, deportees are classified by birth place in the database, which is different from home place. I build cross-matrices between birth place and home place for the deportees born in France and those born abroad. For that purpose I use two raw materials. The first is the 1936 census for the foreign-born, which provides their distribution among *départements* in France. The second is the 1946 census for the French-born, which provides their distribution by birth place and home place at departmental level. Finally, I adjust these figures by the total of deportees estimated by researchers, namely 110,000, in line with Dupaquier (1988). Section 6.5 of the Appendix presents the computations of deportees by age, sex and home-*département*.

3.2.3 Specific Methods Due to Territorial Changes

The main advantage of the French *départements* is their stability since the beginning of the 19th century. However, there were some changes during the two last centuries, especially with regard to the eastern borders and the Paris region. To take this into account, some adjustments are necessary. In this study, I use a

departmental classification with 97 *départements*: the 95 *départements* of the current metropolitan France (*Corse* counting as one), as well as the *Seine* and *Seine-et-Oise* in their pre-1968 boundaries. Territorial breakdowns are twofold in this study: either departmental boundaries changed because of a territorial reorganization, or the data are missing within the unified departmental classification that I use.

The departmental boundary changes are of two types for the period 1853–2014. The first concerns the pre-1901 period. *Savoie* and Nice’s *Comté* were attached to France following the April, 22th and 23th, 1860 *plébiscite*. *Savoie* and *Haute-Savoie* were created ex nihilo on June 14th, 1860 while *Alpes-Maritimes* was created by aggregating a part of *Var* (*Grasse*’s *canton*) to the *Comté*. Moreover, following the war against Prussia in 1870, *Meurthe* and *Moselle* in their old form disappeared to form *Moselle* and *Meurthe-et-Moselle*.¹³ In addition, the *départements* boundaries of *Haut-Rhin*¹⁴, *Bas-Rhin* and *Vosges*¹⁵ changed. For this period, I distributed births of the old-classification *départements* between the unified-classification *départements*. The second change concerns the 1901–2014 period. It follows the *Ile-de-France* reorganization in 1964, effective in 1968. This reorganization led to the dissolution of *Seine* and *Seine-et-Oise*. These *départements* were divided between *Paris*, *Yvelines*, *Essonne*, *Hauts-de-Seine*, *Seine-Saint-Denis*, *Val-de-Marne* and *Val d’Oise*.

The missing data in the unified departmental classification are also of two types. The first concerns the missing data due to the two World wars: *Aisne*, *Ardennes*, *Marne*, *Meurthe-et-Moselle*, *Meuse*, *Nord*, *Oise*, *Pas-de-Calais*, *Somme* and *Vosges* for the 1914–1918 period, and *Moselle*, *Bas-Rhin* and *Haut-Rhin* for the period 1939–1945. *Corse* is also concerned in 1943 and 1944. The second category is *départements* temporarily under German control: this is the case of *Bas-Rhin*, *Haut-Rhin* and *Moselle* before 1919.

3.2.4 Specific Methods Due to Missing Data

Births of the missing *départements* during the period 1853–1900 are first estimated. Recall that these births allow the distribution of deaths according to Lexis triangles. I consider that the changes were synchronized between missing *départements* and a neighboring *département*. For *Var* and *Alpes-Maritimes*, whose limits are stable since 1861, I use the ratio between births in 1861 and births in *Bouches-du-Rhône* to deduce births between 1853 and 1860. I proceed in the same way for *Savoie* and *Haute-Savoie*, for which I use *Ain* as reference. Regarding *Vosges*, *Territoire de Belfort* and *Meurthe-et-Moselle*, I used *Haute-Saône* as reference for the 1853–1869 period. As I know values for *Meurthe*, *Moselle*, *Haut-Rhin* and *Vosges* (former *départements*), it was easy to deduce values for *Moselle* and *Haut-Rhin* in their current boundaries. For the 1870–1900 period, births in *Moselle*, *Bas-Rhin* and *Haut-Rhin* were estimated using *Haute-Saône* as reference.

Data from the population movement for missing *départements* during the two World wars are also estimated. Even if the lifetables of these *départements* should be analyzed with caution, this allows an approximation of their current mortality conditions. For that, I go further than the method used for births

¹³Until 1870, two departments existed, namely *Meurthe* and *Moselle*. Their gathering fell within the same limits as *Meurthe-et-Moselle* and the new *Moselle*. The new *Moselle* includes the territories under German control in 1870, namely the districts of *Château-Salins* and *Sarrebourg* for the old *Meurthe* and *Thionville*, *Metz*, *Forbach-Boulay Moselle* and *Sarreguemines* for the old *Moselle*. In contrast, the new *Meurthe-et-Moselle* includes the territories remained French at that time, i.e. the districts of *Luneville*, *Nancy* and *Toul* for the old *Meurthe* and the *canton* of *Briey* for the old *Moselle*.

¹⁴In 1870, *Haut-Rhin* in its former boundaries is divided between *Haut-Rhin* as we know today – which passes under German control until the end of the Second World War – and *Territoire de Belfort*, which remains under French control.

¹⁵In 1870, the former cantons of *Schirmeck* and *Saales* (in *Vosges*) are attached to *Bas-Rhin*, which passes under German control. The new boundaries of these two *départements* are those that we know nowadays.

by endogenizing the choice of the reference *département*. For each couple of *département* and missing period, I choose a panel of geographically close *départements* whose data are available. Table 4 gives these candidates for each set of missing *départements*. I then calculate a score based on the synchronicity of demographic variations over the period surrounding the missing period. From this score, a reference *département* is defined for each *département* with missing data and used to estimate these values. This method is used to both total births and deaths by age (sum of civilian, military and in deportation deaths). Section 6.6 of the Appendix goes into detail about the choice of reference *département* and the method used.

Table 4: PANEL OF CANDIDATE REFERENCE DÉPARTEMENTS

Period	Missing départements	Panel of reference départements
1914–1919	Aisne, Ardennes, Marne, Meurthe-et-Moselle, Meuse, Nord, Oise, Pas-de-Calais, Somme, Vosges	Aube, Eure, Haute-Marne, Haute-Saône, Seine-Inférieure, Seine-et-Marne, Seine-et-Oise
1939–1945	Moselle, Bas-Rhin, Haut-Rhin	Doubs, Meurthe-et-Moselle, Haute-Saône, Vosges
1943–1944	Corse	Alpes-Maritimes, Bouches-du-Rhône, Gard, Hérault, Var

With the reorganization of *Ile-de-France* in 1968 I must differentiate the *départements* belonging to the old classification from those belonging to the new ones. The former *départements* are followed over the 1901–1968 period, and the new ones between 1968 and 2014. As such, I make several adjustments. The first concerns the distribution of births before 1968 among the *départements* of the new classification, in order to distribute deaths in Lexis triangles. It is done pro-rata 1968’s births. Then I estimate 1968’s age-populations for *départements* of the old classification by using the “Intercensal Survival” method: I assume that *Ile-de-France* migratory profile was the same for *Seine* and *Seine-et-Oise*. Section 6.7 of the Appendix discusses these two adjustments.

Finally, computation periods vary by *département*. I distinguish them according to four classes. Class 1 (C_1) concerns all *départements* outside *Moselle*, *Bas-Rhin*, *Haut-Rhin* and *Ile-de-France* (except *Seine-et-Marne*). These 85 *départements* are tracked over the period 1901–2014. Computations of population at each 1st January is done as shown in Figure 2. *Départements* in class 2 (C_2) are the former *Ile-de-France* *départements*, namely *Seine* (75) and *Seine-et-Oise* (78). Lifetables were estimated over the period 1901–1968. Class 3 (C_3) concerns the new *Ile-de-France* *départements*: *Essonne* (91), *Hauts-de-Seine* (92), *Seine-Saint-Denis* (93), *Val-de-Marne* (94), *Val d’Oise* (95), *Paris* (96), *Yvelines* (97). Lifetables are available for the period 1968–2014. *Bas-Rhin*, *Haut-Rhin* and *Moselle* are in class 4 (C_4): lifetables are estimated between 1921 and 2014. Figures in Section 6.8 of the Appendix draw the methods used to estimate the January 1st populations for each of these four classes. These are variants of Figure 2.

3.3 Reliability of the Data and Comparison with Other Studies

The raw data used in this study come from old statistical sources. I therefore verified that their use could be done without introducing bias in future analyzes.

Firstly, I was interested in the consistency of departmental and national data. Vallin and Meslé (2001) calculated the national lifetables for the 19th and 20th centuries. Consequently, I verified that the departmental sums of deaths, births, false stillbirths and populations are equal to national values. These expectations were true, which testify to the quality of the raw data. My results are therefore consistent with the

results established at the national level.

Second, I was interested in the coherence of my results with the works already done at the departmental level. To do so, I calculated the differences between the departmental life expectancies of my paper and those of Bonneuil (1997) and Daguet (2006). Results are presented in Table 5.

Bonneuil (1997) calculated the life expectancies of women in 1901–1905. I have calculated life expectancies for the same period as well. The comparison between these estimates shows that mine are on average higher: the median of the difference is 3.34%. In addition, 50% of *départements* have a difference between 0.49% and 6.05%, and 25% of them have a difference of more than 6.05%. The in-depth study of age-specific mortality rates reveals that these differences are largely explained by lower infant mortality rates (deaths under age 5). Nevertheless, since I cannot retrieve the death and population statistics of Bonneuil (1997), I do not know if this difference comes from an underestimation of the number of deaths or an overestimation of the population at risk.

Daguet (2006) also revealed the departmental life expectancies at birth at the date of each census between 1954 and 1999. I compute the differences for both men and women. Overall, differences are much smaller. The median is around 0.2%, with no distinction for men and women and no temporal trend. The differences for 50% of the *départements* fall between 0% and 0.7% in 1962. These differences in 1999 for men are 0.22% and 0.73%, respectively. Although slight differences remain, one can conclude that life expectancies are reliable, even if slightly overestimated.

Table 5: DIFFERENCES OF DEPARTMENTAL LIFE EXPECTANCIES AT BIRTH WITH OTHER STUDIES

	Men			Women		
	1 st Quart.	Med.	3 rd Quart.	1 st Quart.	Med.	3 rd Quart.
1901–1905				0.49	3.34	6.05
1954	0.18	0.65	1	0.54	0.84	1.34
1962	0	0.4	0.72	-0.01	0.37	0.68
1968	0.17	0.38	0.73	-0.02	0.33	0.78
1975	-0.17	0.15	0.5	-0.11	0.19	0.47
1982	0.01	0.27	0.59	0.04	0.21	0.5
1990	0.09	0.31	0.55	0.21	0.4	0.62
1999	0.22	0.49	0.73	0.47	0.66	0.99

Notes: Differences in % of my computations. Distribution of 90 or 95 departmental differences, according to the classification of the year.

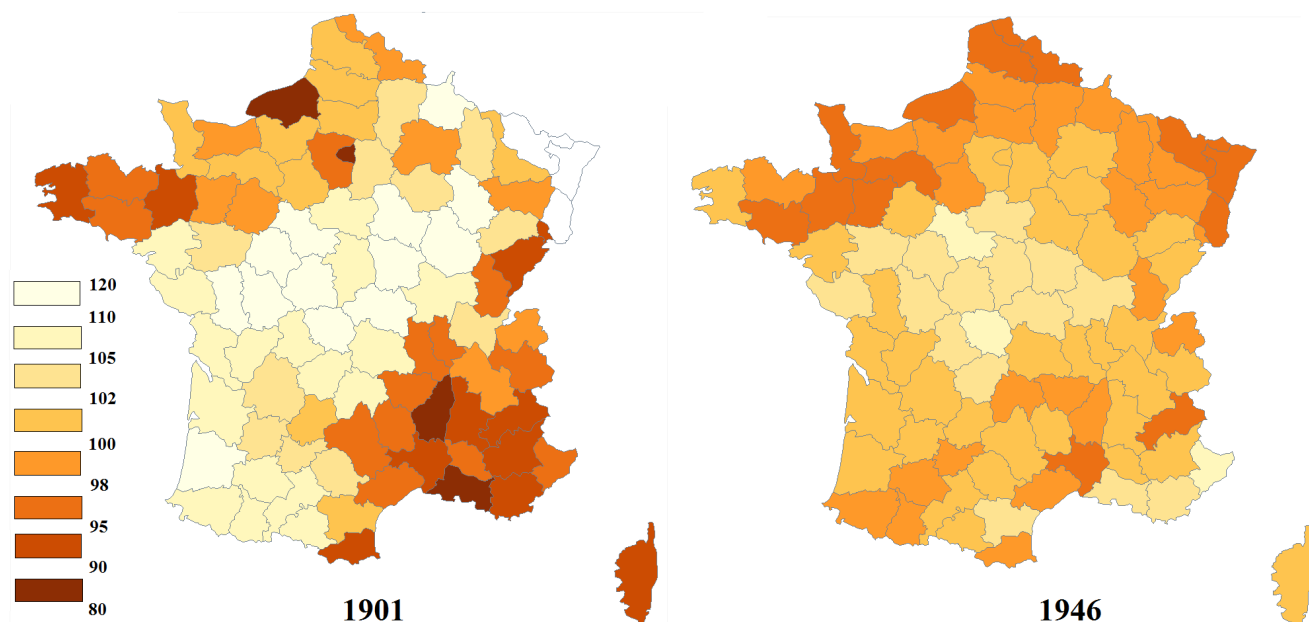
4 Available Results and Discussion

4.1 Available Results

Results are available for the 97 metropolitan *départements* monitored over the period 1901–2014, namely the *départements* of the current classification (*Corse* counting as one) as well as the old *Seine* and *Seine-et-Oise*. Due to their additivity, results are also available at the regional level in the classification prior to January 2016 (22 regions). The variables available are the life expectancies at each age (e_x) as well as a set of lifetable variables between ages 0 to 105 and over (number of survivors, mortality rates, proportions of deaths). Yearly births and populations by age are also available.

Figure 4 reveals the departmental life expectancies at birth relative to the metropolitan average, for women. I chose to present the results for women, but these results are available for men too. The first map shows the results for 1901. One can see that the highest life expectancies were located on an axis connecting the South-West to the North-East, from *Ardennes* to *Landes*. Maximums were reached in *Ardennes* but also in *Pays de la Loire* (*Loir-et-Cher*, *Indre*, *Indre-et-Loire*, *Deux-Sèvres*, ... etc.) and *Bourgogne* (*Côte d'Or*, *Yonne*, *Nièvre*, ... etc.) with values 10 to 20% higher than the metropolitan average. In contrast, life expectancies at birth in the South-East, *Seine* and *Bretagne* are significantly lower than the metropolitan average (between 5 and 20% according to the *département*). The second map presents these life expectancies at birth in the aftermath of the Second World War. At that time, maximums were reached in *Loir-et-Cher*, *Creuse* and *Alpes-Maritimes* with life expectancies 5 to 10% higher than the metropolitan average : Central-West was still a leader region, while the regions of *Bretagne* and *Normandie* were still lagging behind.

Figure 4: LIFE EXPECTANCY AT BIRTH FOR WOMEN (IN % OF THE METROPOLITAN MEAN): 1901 AND 1946

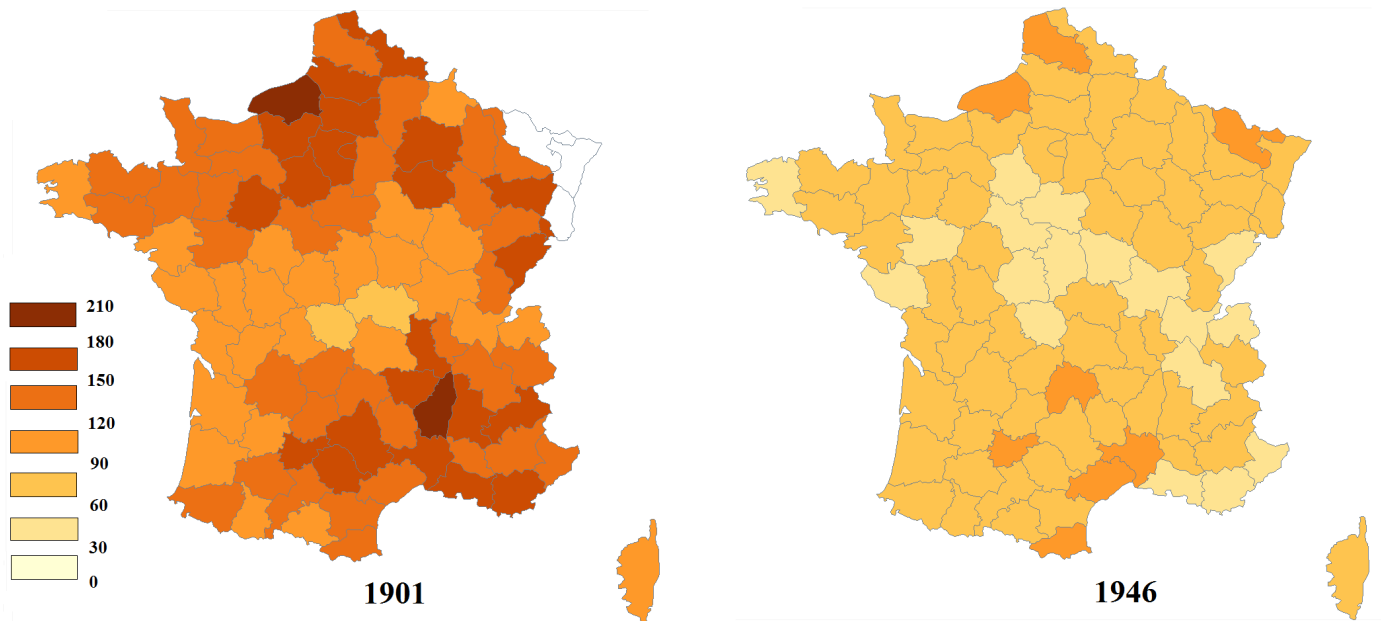


Notes: Sample includes 90 *départements*. *Moselle*, *Bas-Rhin* and *Haut-Rhin* values are non available in 1901 (*départements* under German administration).

Rather than analyzing synthetic indicators such as life expectancy, one can look at age-specific indicators. Since they impacted strongly life expectancies at birth, Figure 5 presents infant mortality rates for women. One more time I chose to present the results for women, but these results are available for men too. I represent the rates per thousand, and no longer relative to the metropolitan average. The landscape in 1901 was relatively similar to the map of life expectancy, since infant mortality rates were in 1901 very high. One can see that in extreme cases (*Seine-Inferieure*, *Ardèche*), for a thousand children under one year, between 180 and 210 died before their first birthday. Rates were generally high in the North and the South-East (between 120 and 150), while they were lower in a broad central band connecting the *Saône-et-Loire* to the *Charente-Maritime* and the Atlantic coast. Minimums (between 60 and 90) were reached in *Creuse* and *Allier*. The second map shows the same values in 1946. Infant mortality rates decreased between the two years since they were globally around 60 per thousand in 1946. An under-mortality zone was visible, from *Eure-et-Loir* to *Isère* via *Nièvre*. The Mediterranean coast presented diverse situations: early mortality was

low in the East (*Var, Alpes-Maritimes*) and strong in the West (*Hérault, Gard, Pyrénées-Orientales*).

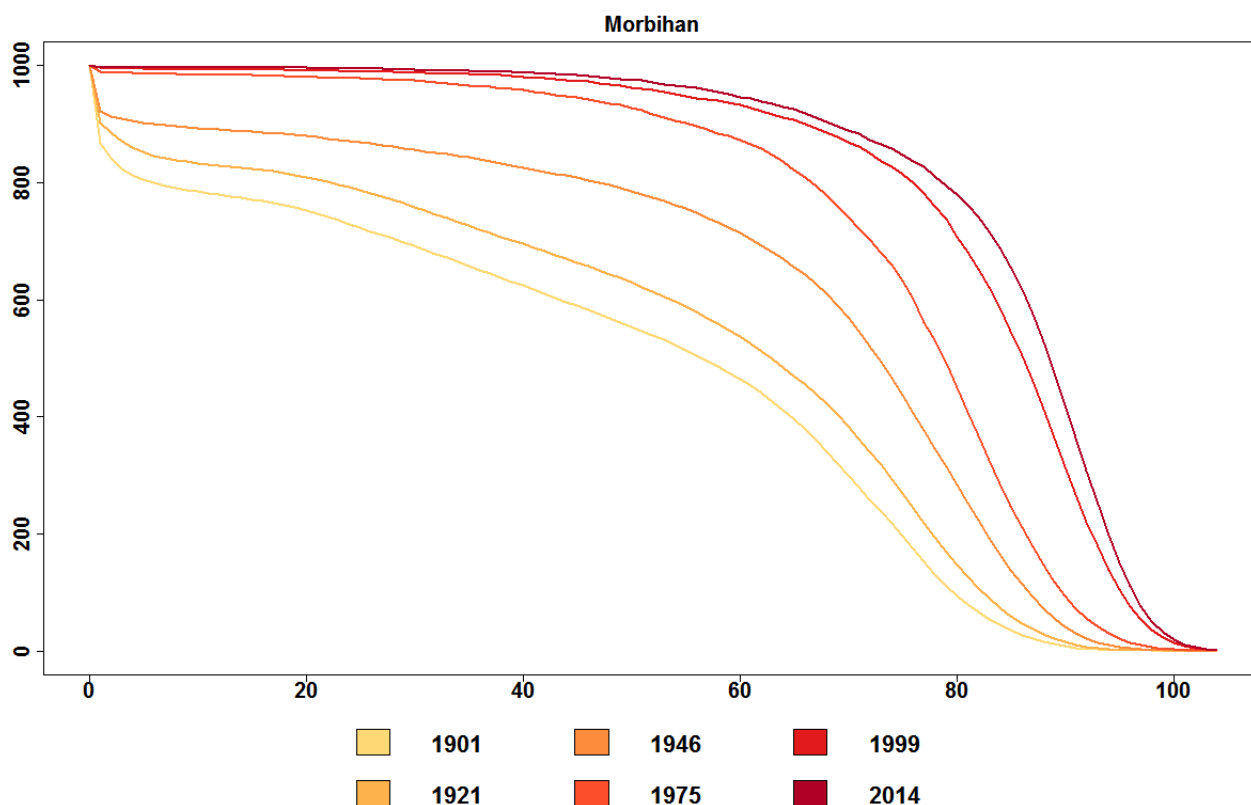
Figure 5: INFANT MORTALITY RATES FOR WOMEN (PER THOUSAND): 1901 AND 1946



Notes: Sample includes 90 *départements*. *Moselle, Bas-Rhin and Haut-Rhin* values are non available in 1901 (*départements* under German administration).

Finally, one can analyze evolvments of a single *département* over the 1901–2014 period. Figure 6 shows female survivors at each age for different dates in *Morbihan*. I have chosen this *département* since it was a place of high mortality in 1901. Indeed, there was high infant mortality at that time: there were only 850 survivors in the fictitious cohort. This infant mortality almost completely disappeared in 1975. The survival curve shifted to the upper-right corner as mortality rates were globally declining. This displacement was important until 1975, mainly because of the drop in infant mortality. Subsequently, the curve moved mainly because of the decrease in mortality between 60 and 80 years, then beyond 80 years for the 1999–2014 period. This is in line with the literature about rectangularization of the survival curve (see Wilmoth and Horiuchi (1999), Fries (2002), Cheung et al. (2005) for example): this curve was in 2014 very flat until age 60 (there is almost no death below this age). Beyond this age the curve decreases dramatically, especially beyond age 80.

Figure 6: EVOLUTION OF SURVIVORS AT EACH AGE IN MORBIHAN



4.2 Discussion

4.2.1 Censuses Reliability

With population censuses one knows the spatial distribution of the population by age and sex between the French *départements* along the 20th century. During this period, censuses served as a support for some public choices. The first concerns local budgets: allocations coming from central administration were based on the population of each territory. These censuses therefore affected the spatial distribution of public finance. The second concerns the electoral divisions: in order to obtain a fair representation in local or national assemblies, electoral divisions are divided so that each of them represents roughly the same population percentage. Censuses therefore had a very strong political impact. As a result, some regions have sought to inflate their census populations in order to get greater financial or electoral weight. Historians and statisticians have shown that *Marseille's* population was overestimated in the 1930s.¹⁶ This was also true in *Corse* in 1962: results of the exhaustive counting were not published because of inconsistencies. These censuses are, however, the basis of age-population computations. Even though ambiguous cases remain marginal over the period, they nevertheless existed.

4.2.2 Interdepartmental Migrations

Methods used in this study partly take into account the issue of migrations. At each census date, the difference between estimated and recorded population can be seen as an approximation of net migration

¹⁶See *Statistique Annuelle du Mouvement de la Population*, 1939–1942, page 4

flows at each age. These flows are then distributed in proportion to the time elapsed between the first census and January 1st of each year of the intercensal period. This approximation does not affect our results when the flows are weak or if they follow the approximation used. This is not the case in war periods. The May-June 1940 Exodus is an emblematic example. To escape the advance of German troops on French territory, the populations of the North-East migrate in mass towards the South and the West. I cannot take into account this exodus with the methodology used: the population of *Ardennes* on January 1st, 1941 is for example largely overestimated. This issue is presented on several occasions in the *Statistique Annuelle du Mouvement de la Population* between 1939 and 1942¹⁷; this publication has suggested to estimate the present population with ration tickets dispensed to the population. However, Alary et al. (2006) showed that these tickets were circumvented during the war, questioning their reliability in counting the present population. Bonnet (2018) try to estimate these departmental populations, but only for females and for the total population.

Another issue relating to interdepartmental migrations concerns the nursery of children born in urban *départements*. Newborns were sent to rural *départements* close to major urban centers. Thus, *Seine* has a lower infant mortality rate than it should be because some of the infants are sent to suburbs. To overcome this issue, official publications suggests¹⁸ to divide deaths of children under age 1 born in a *département* and living anywhere on the national territory, by the total of births in this *département*. I cannot do this because I do not find these raw data in official publications; this suggests that my infant mortality rates are slightly underestimated in urban *départements*.

4.2.3 Domiciliation of Deaths during the Two World Wars

The sources I use to estimate life expectancies during the two World Wars are incomplete: military and deportee deaths were recorded by birth *département* and not by home *département*. I build matrices linking birth *département* and home *département* before the deportation; nevertheless, they rely on strong assumptions about the representativity of pre- and post-war situations concerning the phenomena that took place during the war. The few statistics kept for this period limit the possibilities to go further. Regarding military deaths, I assume that the home *département* was similar to the birth *département* concerning the “*Morts pour la France*”. If this hypothesis seems weaker than those assumed for deportees, it is not entirely satisfactory. Again, I miss reliable and available data to overcome this issue.

4.2.4 Small *Département* Figures

Estimating fertility or mortality rates is difficult when figures are small (namely around 0). Papers tackle this issue by using bayesian estimation process (Asunção et al. (2005), Schmertmann et al. (2014) for fertility rates, Alexander et al. (2017) for mortality rates). The question arose of using these methods to supplement the HMD Protocol. However, the French *départements* figures are not as small as geographical units used in these studies: for example, the minimum according to population was reached in *Territoire de Belfort* in 1901 with 50,000 women, compared to 2,000 for some counties. However, these estimation models may be applied in the future, particularly to compute confidence intervals around departmental life expectancies.

¹⁷See *Statistique Annuelle du Mouvement de la Population*, 1939–1942, pages 3-4, 47 and 55

¹⁸See *Statistique Annuelle du Mouvement de la Population*, 1939–1942, pages 55

5 Conclusion

In this paper, I have presented the sources and methods used to estimate lifetables by sex for all French metropolitan *départements* from 1901 to 2014. To do so, I have collected vital records and census statistics at the departmental level since the beginning of the 20th century. Since the two World Wars afflicted France between 1914–1918 and 1939–1945, military deaths and deaths in deportation were of great importance in the lifetables estimates; these statistics have been collected at the departmental level in original sources, namely the “*Mémoire des Hommes*” and “*MemorialGenWeb*” databases.

To estimate departmental lifetables, I have referred to the methods used in a large number of countries by the researchers of the Human Mortality Database. These methods transform the collected raw data into homogeneous data. They include the use of Cubics Splines to estimate deaths by age groups, the Kannisto model to extrapolate deaths at older ages, and a panel of methods to estimate populations at 1st January of each year. The HMD protocol has been amended to take into account the French data specificities. This concerns false stillbirths which are reintroduced in the statistics of births and infant deaths, and territorial breaks such as those which affected the Paris region in 1968.

This work provides a new database on departmental mortality for the entire 20th century. Coupled with Bonneuil (1997)’s estimations for the 19th century, it provides an overview of the local trends in mortality since the French Revolution. As they have been calculated for each sex, these data shed new insights on the reasons explaining the differences in life expectancy between men and women. Moreover, beyond mortality statistics, this new database can be used to analyze all demographic fields at local level: birth rates since it includes annual births, the spatial distribution of population since it provides yearly populations by age, and finally internal migrations. These fields of research are on my future agenda.

References

- Alary, E., B. Vergez-Chaignon, and G. Gauvin (2006). *Les Français au quotidien, 1939–1949*. Perrin.
- Barbieri, M. (2013). La mortalité départementale en France. *Population* 68(3), 433–479.
- Bonnet, F. (2018). Beyond the Exodus of May-June 1940: Internal Migrations in France during the Second World War. *mimeo*.
- Cheung, S. L. K., J.-M. Robine, E. J.-C. Tu, and G. Caselli (2005). Three Dimensions of the Survival Curve: Horizontalization, Verticalization, and Longevity Extension. *Demography* 42(2), 243–258.
- Coale, A. J., P. Demeny, and B. Vaughan (2013). *Regional Model Life Tables and Stable Populations: Studies in Population*. Elsevier.
- Daguet, F. (2006). *Données de démographie régionale de 1954 à 1999*. INSEE.
- Dupâquier, J. and J.-P. Bardet (1988). *Histoire de la population française*, Volume 4. Presses universitaires de France.
- Fries, J. F. (2002). Aging, Natural Death, and the Compression of Morbidity. *Bulletin of the World Health Organization* 80(3), 245–250.
- Huber, M. (1931). *La population de la France pendant la guerre*. Presses Universitaires de France.
- Pedroncini, G., A. Corvisier, and A. Blanchard (1992). *Histoire militaire de la France. 3. De 1871 à 1940*. Presses Universitaires de France.
- Preston, S., P. Heuveline, and M. Guillot (2000). *Demography: Measuring and Modeling Population Processes*. Wiley-Blackwell.
- Prost, A. (2008). Compter les vivants et les morts: l'évaluation des pertes françaises de 1914–1918. *Le Mouvement Social* (1), 41–60.
- Vallin, J. and F. Meslé (2001). *Tables de mortalité françaises pour les XIXe et XXe siècles et projections pour le XXIe siècle*. Éditions de l'Institut National d'Études Démographiques.
- Vallin, J. and F. Meslé (2005). Convergences and Divergences: an Analytical Framework of National and Sub-National Trends in Life Expectancy. *Genus* 61(1), 83–124.
- Van de Walle, É. (1974). *The Female Population of France in the 19th Century: A Reconstruction of 82 Départements*. University Press.
- Wilmoth, J. R., K. Andreev, D. Jdanov, D. A. Gleijer, C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, and P. Vachon (2007). Methods Protocol for the Human Mortality Database. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock*. URL: <http://mortality.org> [version 31/05/2007] 9, 10–11.
- Wilmoth, J. R. and S. Horiuchi (1999). Rectangularization Revisited: Variability of Age at Death Within Human Populations. *Demography* 36(4), 475–495.

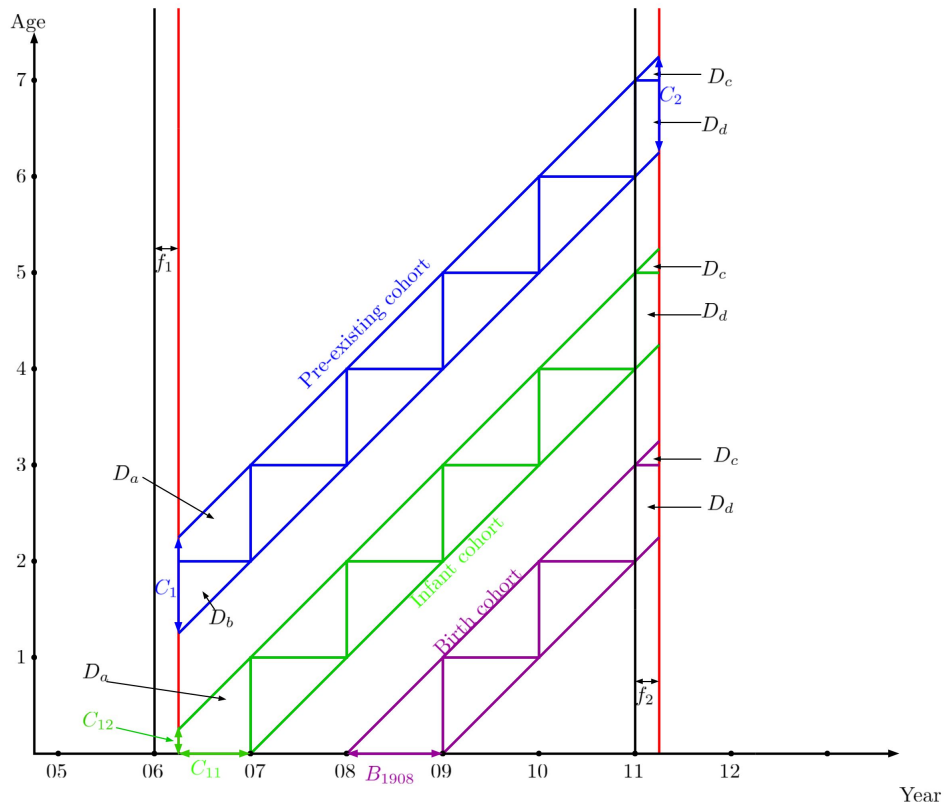
6 Appendices

6.1 Computations of Population on 1st January

6.1.1 Intercensal Survival

The first method used to compute populations on 1st January of each year is “Intercensal Survival”. With this method I can estimate population by age for each intercensal period. Populations at the second census (e.g. 1906 for 1901–1906) are not estimated in the same way for all cohorts. Figure 7 presents the three types of cohorts which exist in this method. There are “Pre-existing cohorts” (born before the census year), “Infant cohort” (born during the census year) and “Birth cohorts” (born after the census year). The gaps between the census date and 1st January of the census year are crucial. This gap is called f_1 for the first census and f_2 for the second.

Figure 7: CLASSIFICATION OF DIFFERENT COHORTS FOR INTERCENSAL SURVIVAL METHOD



I begin with “Pre-existing cohorts”. I estimate age-population at date of the second census. Let t and $t + N$ be the first and last 1st January in the intercensal period. N is the number of full calendar years between censuses. The dates of the two censuses are:

$$t_1 = t - 1 + f_1,$$

$$t_2 = t + N + f_2.$$

The elapsed time between the censuses is thus:

$$t_2 - t_1 = N + 1 - f_1 + f_2.$$

The cohort tracked (Figure 7, in blue) was 1- or 2-years-old at the time of the 1906 census and was born in 1904. Data are by year of birth and not by age, which simplifies computations. I assume a uniform distribution of deaths in each Lexis triangle, so that for the cohort aged x on 1st January of the year of the first census,

$$D_a = (1 - f_1^2) \times D_L(x, t - 1),$$

$$D_b = (1 - f_1)^2 \times D_U(x - 1, t - 1),$$

$$D_c = f_2^2 \times D_L(x + N + 1, t + N),$$

$$D_d = (2f_2 - f_2^2) \times D_U(x + N, t + N).$$

This cohort's estimated population at the second census may be called \hat{C}_2 and is calculated as follows:

$$\hat{C}_2 = C_1 - (D_a + D_b) - \sum_{i=0}^{N-1} [D_U(x + i, t + i) + D_L(x + i + 1, t + i)] - (D_c + D_d), \quad (20)$$

where $\Delta_x = C_2 - \hat{C}_2$ (the difference between the estimated population and that recorded at the date of the second census) comprises estimation errors and intercensal migrations within the cohort. In order to compute age-population at 1st January of each intercensal year, the Δ_x error must be split between the age-populations in each intercensal year. I assume that these rough migrations are uniformly distributed over time. Population by age is calculated as follows:

$$P(x + n, t + n) = C_1 - (D_a + D_b) - \sum_{i=0}^{n-1} [D_U(x + i, t + i) + D_L(x + i + 1, t + i)] + \frac{1 - f_1 + n}{N + 1 - f_1 + f_2} \Delta_x. \quad (21)$$

There is only one “Infant cohort” to track for each intercensal period (in Figure 7, the cohort born in 1906). Thus, $C_1 = C_{11} + C_{12}$, with $C_{11} = (1 - f_1) \times B_{t-1}$ and C_{12} the population recorded as born during the year of the census. Thus,

$$\hat{C}_2 = C_1 - D_a - \sum_{i=0}^{N-1} [D_U(i, t + i) + D_L(i + 1, t + i)] - (D_c + D_d), \quad (22)$$

and

$$P(n, t + n) = C_1 - (D_a + D_b) - \sum_{i=0}^{n-1} [D_U(i, t + i) + D_L(i + 1, t + i)] + \frac{\frac{1}{2}(1 - f_1^2) + n}{N + \frac{1}{2}(1 - f_1^2) + f_2} \Delta_0. \quad (23)$$

Finally, since N is the number of full calendar years during the intercensal interval, I track N birth cohorts. A cohort born in year $t + j$ is aged $K = N - j - 1$ on 01/01/ $t + N$. The estimated population of

this cohort may be expressed as:

$$\hat{C}_2 = B_{t+j} - D_L(0, t+j) - \sum_{i=1}^{N-1} [D_U(i-1, t+j+i) + D_L(i, t+j+i)] - (D_c + D_d). \quad (24)$$

Note that the number of intermediate populations produced by the various cohorts depends on K . For $k = 0, \dots, K$, the intermediate populations of each cohort are computed as follows:

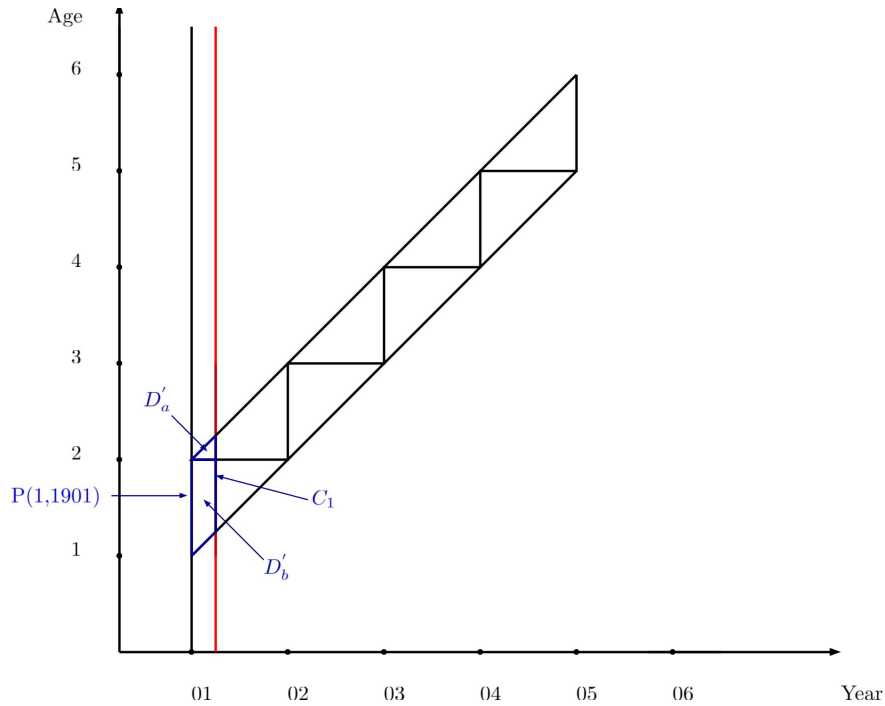
$$P(k, t+j+k+1) = B_{t+j} - D_L(0, t+j) - \sum_{i=1}^k [D_U(i-1, t+j+i) + D_L(i, t+j+i)] + \frac{2k+1}{2K+1+2f_2} \Delta_{t+j}. \quad (25)$$

6.1.2 Precensal Survival Method

The second method I use is “Precensal Survival”, to compute populations for the first 1st January of the whole period. Figure 8 presents the computations for population of age 1 in 1901. To do so, I must add D'_a et D'_b to the population born in 1901 and recorded on March 6th, 1901. If t_1 is the first 1st January of the intercensal period, then:

$$P(x-1, t_1-1) = C_1 + D'_a + D'_b. \quad (26)$$

Figure 8: PRECENSAL SURVIVAL METHOD



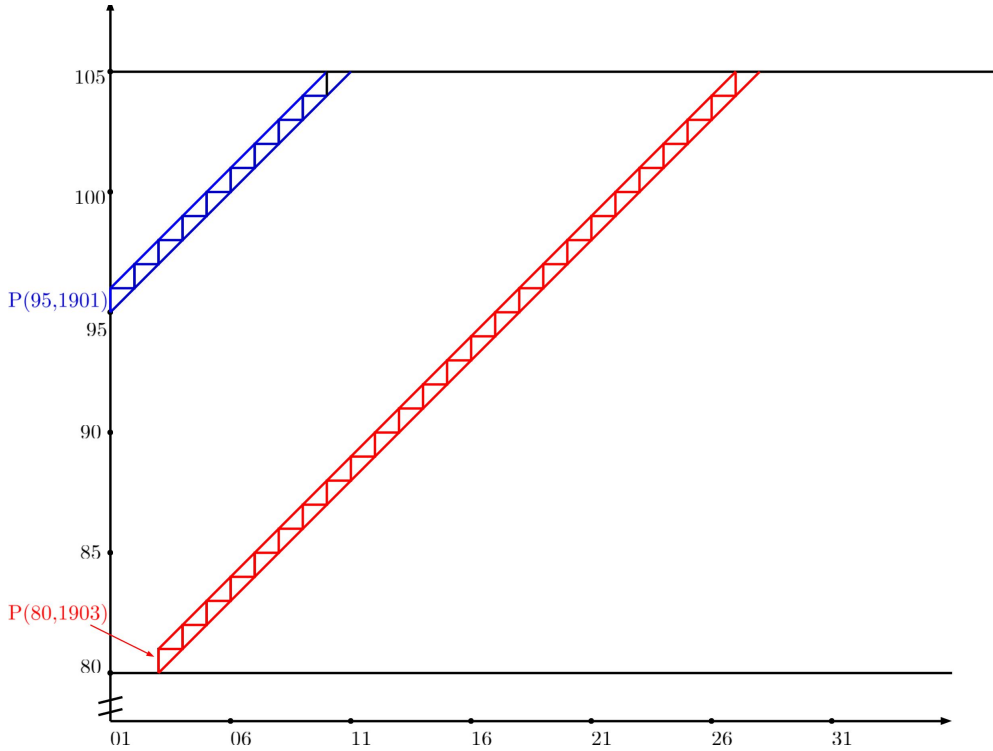
6.1.3 Extinct Cohorts Method

The third method I use is “Extinct Cohorts”, to calculate age-population for the cohorts extinct in 2013. Since the maximum age in my database is 105, a cohort is considered to be extinct if it reached 105 or over in 2013. Figure 9 reveals that my data comprise two kinds of extinct cohorts. The first are “Full

cohorts” (Figure 9, in red), which can be tracked from ages 80 to 105 in 1901–2013. Thus, the 80-year-old population in 1903 equals the sum of the cohort’s Lexis triangles between ages 80 and 105. The others are “Truncated cohorts” (Figure 9, in blue), those over age 80 in 1901. Thus, the 95-year-old population in 1901 equals the sum of the cohort’s Lexis triangles between 95 and 105. More generally, the population of age x in year t can be calculated as follows:

$$P(x, t) = \sum_{i=0}^{\infty} [D_U(x + i, t + i) + D_L(x + i, t + i)].$$

Figure 9: EXTINCT COHORTS METHOD



6.1.4 Survivor Ratio Method

The last method I use is “Survivor ratio”, to calculate non-extinct cohorts of age 85 and over in 2013. Figure 10 presents the computations for the cohort aged 104 in 2013. The survivor ratio R may be defined as the number of individuals alive at age x on 1st January t , divided by the number of individuals in the same cohort alive k years previously. Formally:

$$R = \frac{P(x, t)}{P(x - k, t - k)}.$$

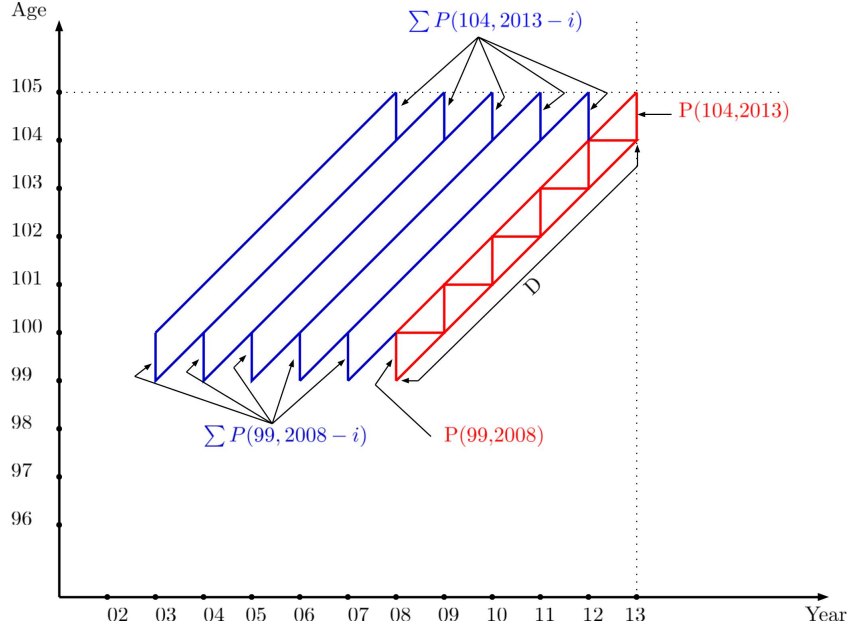
I assume that there is no migration at these ages. R may also be expressed:

$$R = \frac{P(x, t)}{P(x, t) + \dot{D}}.$$

where $\dot{D} = \sum_{i=1}^k [D_U(x - i, t - i) + D_L(x - i + 1, t - i)]$. Finally, $P(x, t)$ may be expressed as a function of R :

$$P(x, t) = \frac{R}{1 - R} \dot{D}. \quad (27)$$

Figure 10: SURVIVOR RATIO METHOD



Since the survivor ratio cannot be directly observed for a cohort, I use preceding cohorts whose age-populations have been calculated by the “Extinct Cohorts” method. I assume that the survival ratio has roughly the same value in the studied cohort and in the preceding ones. As such, the mean ratio R^* of the preceding m cohorts may be calculated as follows:

$$R^*(x, 2013, k, m) = \frac{\sum_{i=1}^m P(x, 2013 - i)}{\sum_{i=1}^m P(x - k, 2013 - k - i)}.$$

I may then estimate $\tilde{P}(x, 2013)$:

$$\tilde{P}(x, 2013) = \frac{R^*}{1 - R^*} \dot{D}.$$

Subsequently, I may track the cohort back in time and estimate $\tilde{P}(x - 1, 2012)$, $\tilde{P}(x - 2, 2011)$, ... by adding step by step the cohort’s deaths. I apply this method for any non-extinct cohort in 2013. For my estimations I follow the guidelines of the HMD Protocol, with $k = m = 5$.

The assumption of a constant survivor ratio over time is strong, and I may control by the recorded population on 1st January 2013. I compare the 85-and-over population on 1st January 2013 – retrieved from the census of that year (called P_{85+}^{Rec}) – with the 85-and-over population on 1st January 2013 as calculated by the Survivor Ratio method (called P_{85+}^{SR}). Thus, populations at each age in 2013 can be computed as follows:

$$\hat{P}(x, 2013) = c \tilde{P}(x, 2013) = c \frac{R^*}{1 - R^*} \dot{D},$$

where $c = \frac{P_{85+}^{Rec}}{P_{85+}^{SR}}$. As before, each cohort is back-followed: I make estimates for $\hat{P}(x - 1, 2012)$, $\hat{P}(x - 2, 2011)$, ...

6.2 Set of Different Lifetables

Concerning computations of (1×1) and (1×5) lifetables, I start from values of q_x . With these values I compute p_x the probability of staying alive between x and $x + 1$. Then I compute the number of survivors at each age per 100,000 births

$$l_x = l_0 \prod_{i=0}^{x-1} p_i \text{ with } (l_0 = 100,000),$$

the deaths at each age (d_x)

$$\begin{cases} d_x = l_x q_x & x \in (0, 104) \\ d_x = l_x & x = 105 \end{cases},$$

the number of years lived between x and $x + 1$

$$\begin{cases} L_x = l_x - (1 - a_x)d_x & x \in (0, 104) \\ L_{105}^\infty = l_x a_x & x = 105 \end{cases},$$

the number of life years remaining to live

$$\begin{cases} T_x = \sum_{i=x}^{104} L_i + L_{105}^\infty & x \in (0, 104) \\ T_x = L_{105}^\infty & x = 105 \end{cases}.$$

Finally, life expectancy at age x is computed as follows:

$$e_x = \frac{T_x}{l_x}.$$

Methods are quite the same for (1×5) lifetables. I therefore get lifetables for quinquennial periods: 1901–1905, 1906–1910, 1911–1915... etc. Values in abridged (5×1) and (5×5) lifetables are computed with previous variables. ${}_5e_x$, ${}_5l_x$ and ${}_5T_x$ are directly retrieved from the complete lifetables. Finally, ${}_5d_x = l_x - l_{x+5}$, ${}_5q_x = \frac{d_x}{l_x}$ and ${}_5L_x = T_x - T_{x+5}$. One can also find ${}_5a_x$ and ${}_5m_x$ from the basic formula linking all these variables.

6.3 Census Adjustments

For my purposes it is simpler to compute population figures by birth year. Census data are given by single age after 1968. I gather populations by five-year age groups between ages 15 and 89, before taking the open-age interval 90 and over. The cubic splines adjustment takes into account that populations were given by age and not by birth year. Thus, taking the 1968 census as an example, I isolated the populations born between 01/01/1968 and the date of the census.¹⁹ Before 1968, data are given by birth year. Nevertheless some specific adjustments are needed.

¹⁹Note that the estimates of the population born in the census year are important because they are used to calculate populations by 1st January of each year.

6.3.1 Distribution of Deaths of Unknown Age in 1901 Census

For the 1901 census, individuals whose birth year is unknown are put together in the open-age interval. To allocate them I use the 1911 census, which has a useful degree of detail. The process follows three steps. The first is based on the calculation of the quotient of individuals aged 95 and over by individuals aged 80 and over for each *département* i and each sex j in 1911:

$$R_{95ij}^{1911} = \frac{\sum_{s=95}^{105} P_{sij}^{1911}}{\sum_{s=80}^{105} P_{sij}^{1911}}. \quad (28)$$

These quotients are then applied to the 1901 census to compute the proportion of individuals aged 95 and over among individuals aged 80 and over:

$$\sum_{s=95}^{105} P_{sij}^{1901} = R_{95ij}^{1911} \times \sum_{s=80}^{105} P_{sij}^{1901}. \quad (29)$$

By subtraction, I finally deduce death of unknown year of birth for each *département* and sex.

6.3.2 Addition of Age Group for Pre-1946 Censuses

The 1906, 1921, 1926, 1931, 1936 and 1946 censuses did not use the same methodology for populations in the first three age groups. Some groups have to be combined or splitted (Table 6, in italics). For that purpose I assume that births were spread uniformly over time.

TABLE 6: CLASSIFICATION AND AVAILABILITY OF POPULATIONS BORN TWO YEARS BEFORE THE CENSUS

Census	1 st class	2 nd class	3 rd class
1901	Born from 01/01/01 to 04/03/01	Born in 1900	Born in 1899
1906	Born from 01/01/06 to 03/06/06	Born in 1905	Born in 1904
1911	Born from 01/01/11 to 03/05/11	Born in 1910	Born in 1909
1921	Born from 01/01/21 to 03/05/21	<i>Born from 03/06/20 to 12/31/20</i>	<i>Born from 01/01/20 to 03/05/20</i>
1926	Born from 01/01/26 to 03/07/26	<i>Born from 03/08/25 to 12/31/25</i>	<i>Born from 01/01/25 to 03/07/25</i>
1931	Born from 01/01/31 to 03/07/31	<i>Born from 03/08/30 to 12/31/30</i>	<i>Born from 01/01/30 to 03/07/30</i>
1936	Born from 01/01/36 to 03/07/36	<i>Born from 08/03/35 to 31/12/35</i>	<i>Born from 01/01/35 to 7/03/35</i>
1946	<i>Born from 03/10/45 to 03/09/46</i>	<i>Born from 01/01/44 to 03/09/45</i>	Born in 1943

Notes: Periods in italics in the table have to be combined or splitted to get populations by year of birth. 01/01/01 means 01/01/1901.

Finally, the 1911 census is rather different because it provides data for each year of birth and not per five-year groups. However, these numbers fluctuate considerably. There were two possible methods: either use the numbers given, or combine the numbers in five-year groups as for the other censuses and apply cubic splines. Although the first method provides more information, it includes inconsistent fluctuations at adult ages. Since I need to maintain consistency, I choose the second method. Raw data in 1911 have to be thoroughly reprocessed: I keep the first fifteen birth year groups, and then combine them by five-year groups (1891–1895, 1886–1890, etc.) plus the open-age interval “1820 and earlier”.

6.3.3 Adjustment of Censuses by Cubic Splines

To get populations by single year of birth and not five-year groups, I adjust census populations by cubic splines, as I do for civilian and military deaths. The cubic splines are fitted to the cumulative curve of population born before 1st January of the census year. For example, according to the 1901 census, I consider the population born before 1st January 1901. The population born between 1st January 1901 and the day of the census provide no further information and would involve fractional knots.

6.4 Estimates of Military Deaths during the Two World Wars

The classification of *départements* from the “*Mémoire des Hommes*” website is modified to fit the classification for civilian deaths. Problems concern *Corse* (two *départements* counting as one) and the old *départements* of *Seine* and *Seine-et-Oise*. For these last two, deaths are given for the new *départements*. To allocate deaths between *Seine* and *Seine-et-Oise* I first sum all deaths in *Ile-de-France* (without *Seine-et-Marne*), then I allocate these military deaths pro rata of population in the cohorts born from 1880 to 1896. These cohorts account for 83% of total military deaths in the First World War. Concerning the distribution of deaths in the Parisian *départements* between *Seine* and *Seine-et-Oise* for the Second World War, I allocate them pro rata of populations born between 1905 and 1921 (70% of total deaths during the Second World War). *Seine*’s deaths are equal to 78.6% of the total.

Moreover, to ease the collection of data from the website, military deaths have been retrieved by year of birth for the youngest (born after 1889), then by five-year group for those born in 1889 and earlier. These deaths must be split by year of birth, which is done by cubic splines. The two assumptions made are (1) no deaths under age 16 and (2) no deaths over age 60.

6.5 Estimates of Deportees

The deportee database is nominative (1 line for each deportee). Sex, birth *département* (or country of birth if born abroad), day-month-year of birth, day-month-year of death were extracted. The age of death in days-months-years follows. For dates of birth and death, data are kept since the year was available. Thus, if only the year was available, the date chosen was January 1st. Likewise, if only the month and year of birth were available, the full date of birth was set to the first day of the month. If the date was considered irrelevant (namely if date of birth after date of death), the date is erased. For individuals whose year of death was after 1946 (for about forty individuals), I consider that those are unknown. 93% of the deceased have well-informed data for the four variables (sex, date of death, age, place of birth). For those with two or three variables missing, data were not used. This corresponds to 6.5% of the database. I did not use deportees with one variable missing too since they represented only 0.5% of the total. From these nominative data, I thus extract matrices crossing the age of death, the year of death (1940–1946), the place of birth and the sex.

One of the variables available in the deportee database is the place of birth. One has to differentiate this variable from the home place before deportation, that is where the deceased would have to be located in my lifetables. Since a 40-year-old have a non-zero probability to migrate in a different *département* from where he is born, I may infer the home-*département* before deportation. Similarly, deportees born abroad must be located in a French *département*.

6.5.1 Born-abroad Deportees

There are 33,609 deaths of born-abroad deportees, some 44% of the database. Those born outside France need to be allocated across France on the assumption that they immigrated before they were arrested and deported. One may suppose that these deportees born outside France fled Nazi persecution and settled in France before the start of the war. I make the assumption that the probability of being in each *département* can be inferred by the spatial distribution of foreigners in 1936. Moreover, I assume that this distribution does not vary by age, and also that the 1936 distribution is representative of the war-time one. I can construct the following matrices:

1. N : $PaysN \times Age$ (90×105) (taken from the MemGenWeb base),
2. P : $DeptR \times PaysN$ (91×48) (taken from the 1936 census),
3. R : $DeptR \times Age$ (90×105).

The first modification concerns *Seine*. Matrix P comprises 91 *départements* and not 90 because of the distinction we make between the city of Paris and the inner suburbs, so these two lines are summed to get the same administrative boundaries in the two matrices. Next P must be transformed so that the matrix gives us the probability that an individual born in country i lives in *département* j . Each element in the matrix equals:

$$P_{ij}^* = \frac{T_{ij}}{\sum_i T_{ij}} = \frac{T_{ij}}{T_{.j}},$$

with T the total of deportees.

Third, the names of countries of birth for Matrices P and N must be linked: there are 48 countries or regions in Matrix P and 90 countries in Matrix N .²⁰ I need to reclassify them to calculate the product of Matrices N and P . Thus I get a Matrix P^* (90×90) and calculate the R Matrix:

$$R' = N'P^*$$

Ultimately, each element R_{sj} in Matrix R corresponds to the sum of individuals aged s born in each of the countries i who emigrated to *département* j before being arrested and deported.

6.5.2 French Deportees

There are 43,055 deaths of French-born deportees in the database. I cannot assume that any deportee born in a *département* stayed in that *département*. A transfer matrix must therefore be constructed linking *département* of birth and *département* of residence before deportation. I use the matrix cross-referencing *département* of residence and *département* of birth in the 1946 census.²¹ I assume both this matrix is representative of the pre-war situation and of deportee migrations, and that the probability of migration is equal for all ages.

²⁰The level of detail in the MemGenWeb database is quite high, whereas the one in the census is lower (many Asian, South American and African countries are not directly specified, and colonies are often included in the generic term “French possessions in Africa”).

²¹This matrix distinguishes males and females, enabling us to refine the estimates.

I make a few preliminary modifications. The main is to allocate the deportees according to the post-1968 *départements* between *Seine* and *Seine-et-Oise*. The allocation key is the same as the one used for military deaths in the Second World War. I construct the following matrices:

1. N : $DeptN \times Age$ (90×105),
2. P : $DeptN \times DeptR$ (90×90),
3. R : $DeptR \times Age$ (90×105).

P is transformed so that the matrix gives the probability that an individual born in *département* i lives in *département* j . Each element in the matrix equals:

$$P_{ij}^* = \frac{T_{ij}}{\sum_i T_{ij}} = \frac{T_{ij}}{T_{.j}}.$$

Thus I deduct:

$$R' = N'P^*.$$

The matrices of French and foreign-born deportees are finally added. This final matrix is the sum for each *département*, each age, each sex and each year, of the deportees born in a French *département* and deportees born outside France but living in France when they were arrested. For the total number of deportees, I based my computations on Dupaquier (1988). He followed Sauvy, who reports 27,000 resisters who died in deportation and 83,000 jewish and other deportees. Consequently, I consider that 110,000 individuals died in the camps.

6.6 Missing Data During the Two World Wars

There are ten *départements* (*Aisne*, *Ardennes*, *Marne*, *Meurthe-et-Moselle*, *Meuse*, *Nord*, *Oise*, *Pas-de-Calais*, *Somme*, *Vosges*) with missing data during the First World War, and four during the Second World War (*Corse* between 1943 and 1944, *Moselle*, *Bas-Rhin*, *Haut-Rhin* between 1939 and 1945). These missing data are of two types: births and stillbirths, as well as deaths. The general assumption used for the estimations of these missing data is that the neighbours *départements* have similar evolutions concerning their demographic variables, because of their culture and their shared living conditions. As a result, I can estimate the evolution of the missing demographic variables using a reference *département*.

6.6.1 Births and stillbirths

The choice of the reference *département* for each of the missing *départements* and each sub-period must consider how their demographic variables were synchronized. For that purpose, I may define a support interval and then track changes in the ratio between the variable in the missing *département* and in the reference *département* during that interval. Let t_1 and t_2 be the first and last years of the subperiod for which there are missing data, $\Omega_\Delta = [t_1, t_2]$ the subperiod for which there are missing data, $\Omega_t = [t_1 - h, t_1] \cup [t_2, t_2 + h]$ the support interval with $h = 4$, i the missing *département*, j the potential reference *département*. The ratio R_{ij}^t is calculated for a demographic variable V :

$$R_{ij}^t = \frac{V_j^t}{V_i^t}, t \in \Omega_t \quad .$$

Then mean (\bar{x}_{ij}) and standard deviation (σ_{ij}) of R_{ij}^t are calculated over the interval Ω_t . The stability of the ratio is measured as the coefficient of variation of R_{ij}^t over the interval Ω_t :

$$CV_{ij} = \frac{\sigma_{ij}}{\bar{x}_{ij}}.$$

The reference *département* j^* chosen is the one with the lowest coefficient of variation among all the possible reference *départements*. This criterion is used for both stillbirths and births. Since births moves in a similar way for both sexes, the reference *département* is chosen by examining female births. The same choice is then applied to males and females. After choosing the reference *département* for each missing *département*, the missing data for *département* i and variable V is estimated as follows:

$$V_i^t = V_{j^*}^t \times \bar{x}_{ij^*}.$$

6.6.2 Deaths

The method used to estimate missing deaths is similar to the one used for stillbirths and births. Note that computations are made for total deaths (including military deaths and deportees). Let t_1 and t_2 be the first and last years of the subperiod for which there are missing data, $\Omega_\Delta = [t_1, t_2]$ the subperiod for which there are missing data, $\Omega_t = [t_1 - h, t_1] \cup [t_2, t_2 + h]$ the support interval with $h = 4$, i the missing *département*, j the potential reference *département*. The ratio R_{xij}^t is calculated for deaths D at age x :

$$R_{xij}^t = \frac{D_{xj}^t}{D_{xi}^t} (t \in \Omega_t).$$

Then mean (\bar{x}_{xij}) and standard deviation (σ_{xij}) of R_{xij}^t are calculated over the interval Ω_t . The stability of the ratio is measured as the coefficient of variation of R_{xij}^t over the interval Ω_t :

$$CV_{xij} = \frac{\sigma_{xij}}{\bar{x}_{xij}}.$$

The fit between missing *département* and reference *département* needs to take the lowest value of the coefficient of variation over a number of ages Ω_x and not a single point. I calculate a score S_{ij} :

$$S_{ij} = \frac{1}{\Omega_x} \sum_{x \in \Omega_x} CV_{xij},$$

where Ω_x is defined as ages 0–4 and 50–89 in order to avoid erratic results due to small number of deaths.

The reference *département* j^* chosen is the one with the lowest score among all the possible reference *départements*. After choosing the reference *département* for each missing *département* and subperiod, deaths at age x for the *département* i are estimated as follows:

$$D_{xi}^t = D_{xj^*}^t \times \bar{x}_{xij^*}.$$

6.7 Reorganization of *Ile-de-France* in 1968

By changing the three *départements* of *Ile-de-France* (*Seine*, *Seine-et-Marne*, *Seine-et-Oise*) in eight new ones (*Paris*, *Seine-et-Marne*, *Yvelines*, *Essonne*, *Hauts-de-Seine*, *Seine-Saint-Denis*, *Val-de-Marne*, *Val-d'Oise*), the reorganization of this region in 1968 creates a discontinuity in data. I change my methodology so as to track each of these *départements* over the most appropriate period. Note that *Seine-et-Marne* was not affected by these changes. When I talk about *Ile-de-France* hereafter, I mean the *Ile-de-France* region less *Seine-et-Marne*.

For the intercensal period 1901–1962, I can track the old *départements*: I have all the censuses between these years and population flows (births and deaths). For the intercensal period 1968–2014, I can track the new *départements*: I have all the censuses and population flows between those dates. For the intercensal period 1962–1968, I have 1962 and 1968 censuses for the new *départements*, but no population flows. For the same intercensal period, I have population flows and the 1962 census for the old *départements*, but no data according to the 1968 census. I choose to track the old *départements* until 1968, and the new ones from 1968 onwards. To do so, I make two adjustments. The first is about pre-1968 births for the new *départements*, useful to split deaths in Lexis triangles. The second is about populations of the old *départements* in 1968, to estimate the 1st January population of these *départements* between 1962 and 1968.

To estimate births of the new *départements* before 1968, I use the 1968 distribution. I assume that the weight of each *département* remains constant. Although this is a strong assumption if one want to know the accurate number of births, it is less strong for the relative size of two successive cohorts.

I am not able to calculate 1st January populations of the 1962–1968 intercensal period for *Seine* and *Seine-et-Oise*. Indeed, the turning census available for both old and new *départements* is the 1962 one. In order to estimate pre-1968 population, one need population aged 85 and over to implement the “Survivor Ratio” method, and populations aged 0 to 84 to implement the “Intercensal Survival” one. To estimate the population aged 85-and-over for *Seine* and *Seine-et-Oise*, I assume that the weight of the two *départements* in the *Ile-de-France* 85-and-over population did not vary between 1962 and 1968.

It is more difficult concerning the population aged 0 to 84. To do so, I draw on the Intercensal Survival method. First, I calculate the estimated population in 1968 for *Seine* and *Seine-et-Oise* and the sum of these two *départements* ($\hat{P}_{IdF}^{68}(x)$), by subtracting from each cohort counted in 1962 deaths occurring during the intercensal period. I also know the population estimated for these two *départements* in 1968 ($P_{IdF}^{68}(x)$) by summing the new *départements*. I can therefore deduce the migratory profile for *Ile-de-France*:

$$R_{IdF}^{68}(x) = \frac{\hat{P}_{IdF}^{68}(x)}{P_{IdF}^{68}(x)}.$$

I assume this profile was similar for each of the old *départements* j and use this migratory profile to compute 1968 census populations:

$$P_j^{68}(x) = R_{IdF}^{68}(x) \times \hat{P}_j^{68}(x).$$

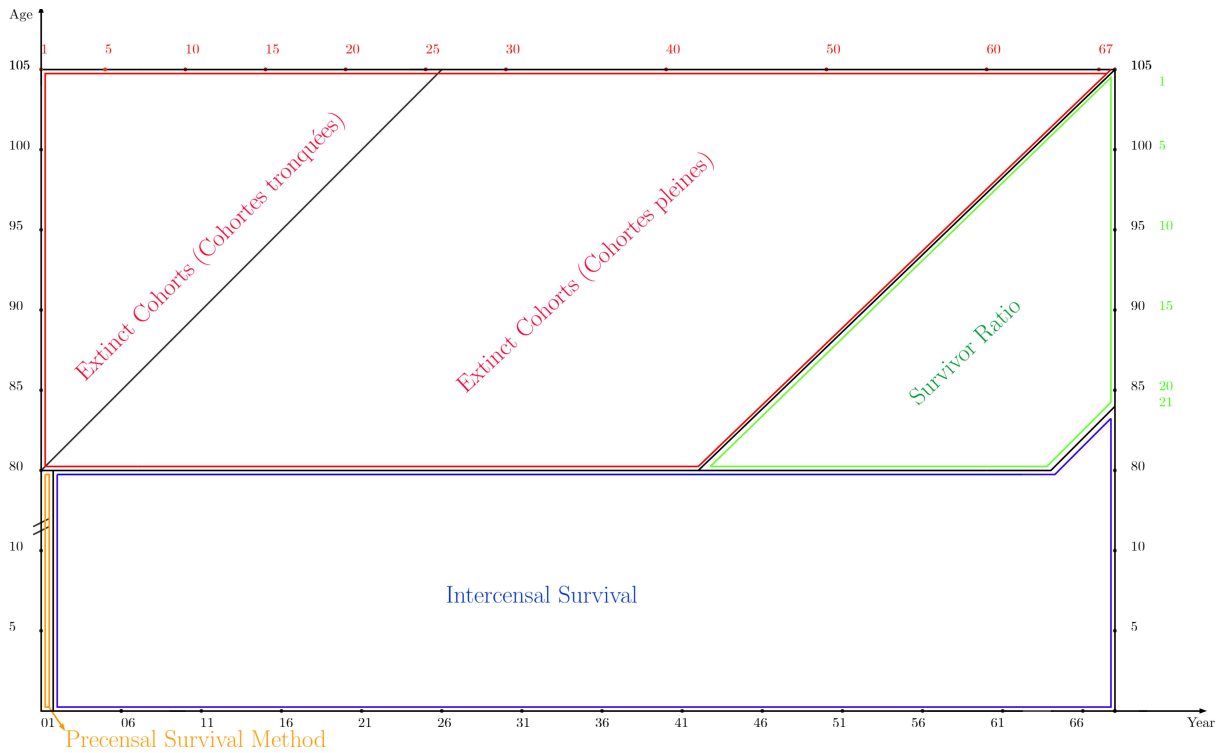
6.8 Computations of 1st January Populations by Class of *Départements*

I miss data for some *départements* between 1901 and 2014 in order to compute lifetables. Consequently, I divided my panel into four classes:

C₁ All départements except *Moselle*, *Bas-Rhin*, *Haut-Rhin*, *Seine-et-Oise* and *Ile-de-France* (except *Seine-et-Marne*). These *départements* are tracked between 1901 and 2014. Figure 2 presents how I calculate populations at each 1st January.

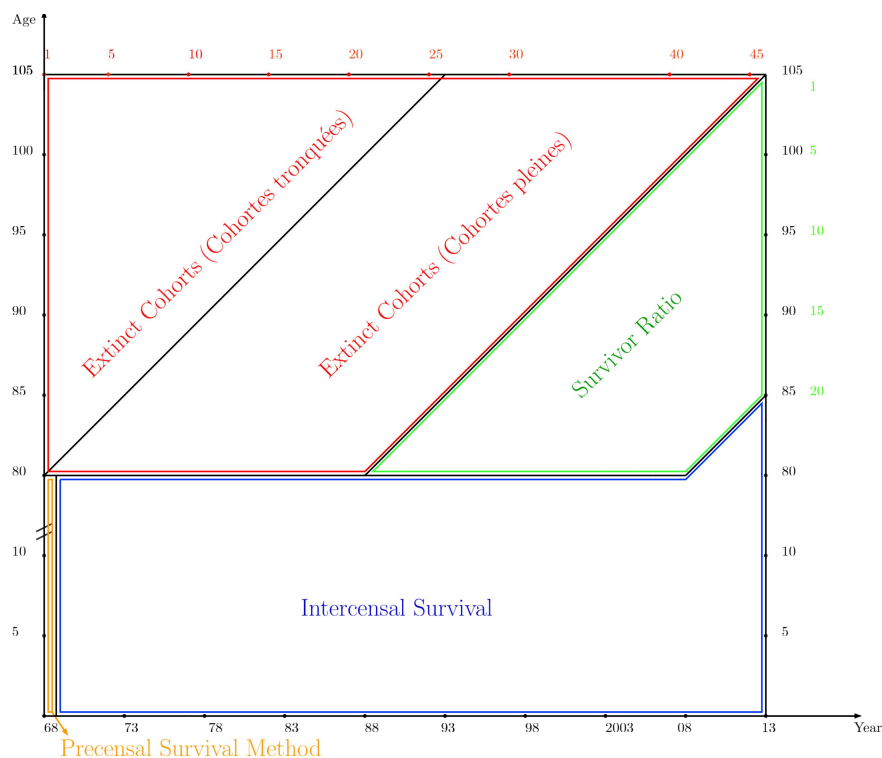
C₂ *Seine* (75) and *Seine-et-Oise* (78). The lifetables for these *départements* are estimated for the period 1901–1968. Figure 11 presents the methods used to compute populations at each 1st January. One can see that the Survivor Ratio method is applied to the 1968 census and not the 2013 census.

Figure 11: ESTIMATIONS OF POPULATIONS FOR DÉPARTEMENTS OF CLASS 2



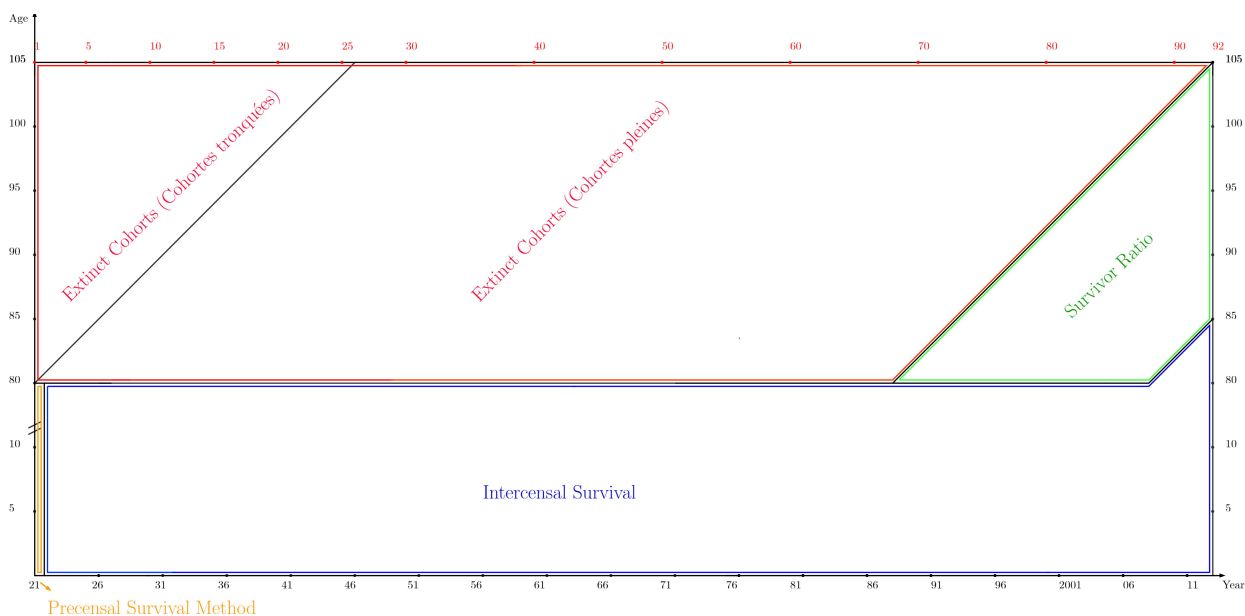
C₃ The new *départements* in *Ile-de-France*: *Essonne* (91), *Hauts-de-Seine* (92), *Seine-Saint-Denis* (93), *Val-de-Marne* (94), *Val d'Oise* (95), *Paris* (96), *Yvelines* (97). These lifetables are estimated between 1968 and 2014. Figure 12 presents the methods used to compute populations at each 1st January.

Figure 12: ESTIMATIONS OF POPULATIONS FOR DÉPARTEMENTS OF CLASS 3



C_4 *Moselle, Bas-Rhin and Haut-Rhin*. From 1870 to 1918 these three *départements* were under German administration. Consequently, the public records were not kept by the French authorities. I have not been able to do research in Germany to find data for this territory, so my estimates begin at the first available census, namely 1921, as shown in Figure 13.

Figure 13: COMPUTATIONS OF POPULATIONS FOR DÉPARTEMENTS OF CLASS 4



6.9 Sources of Raw Data

Table 7: SOURCES FOR CIVILIAN DEATHS, 1901–1929

Year	Départements		Classes		Publication	Book	Page
	Total	Missing	Step	Ceiling			
1901	87	(2)	5	100+	SAMP (Year 1901)	31	62–73
1902	87	(2)	5	100+	SAMP (Year 1902)	32	62–73
1903	87	(2)	5	100+	SAMP (Year 1903)	33	72–83
1904	87	(2)	5	100+	SAMP (Year 1904)	34	62–73
1905	87	(2)	5	100+	SAMP (Years 1905 et 1906)	35-36	62–73
1906	87	(2)	5	100+	SAMP (Years 1905 et 1906)	35-36	140–144
1907	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	190–193
1908	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	194–197
1909	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	198–201
1910	87	(2)	5	100+	SAMP (Years 1907-1908-1909-1910)	1	202–205
1911	87	(2)	5	100+	SAMP (Years 1911-1912-1913)	2	152–155
1912	87	(2)	5	100+	SAMP (Years 1911-1912-1913)	2	156–159
1913	87	(2)	5	100+	SAMP (Years 1911-1912-1913)	2	160–163
1914	77	(1)	5	100+	SAMP (Years 1914–1919)	3	100–103
1915	77	(1)	5	100+	SAMP (Years 1914–1919)	3	104–107
1916	77	(1)	5	100+	SAMP (Years 1914–1919)	3	108–111
1917	77	(1)	5	100+	SAMP (Years 1914–1919)	3	112–115
1918	77	(1)	5	100+	SAMP (Years 1914–1919)	3	116–119
1919	77	(1)	5	100+	SAMP (Years 1914–1919)	3	120–123
1920	90		5	100+	SAMP (Years 1920–1924)	4	82–85
1921	90		5	100+	SAMP (Years 1920–1924)	4	86–89
1922	90		5	100+	SAMP (Years 1920–1924)	4	90–93
1923	90		5	100+	SAMP (Years 1920–1924)	4	94–97
1924	90		5	100+	SAMP (Years 1920–1924)	4	98–101
1925	85		5	100+	SAMP (Year 1925) - CD	5	2–183
1926	90		5	100+	SAMP (Year 1926) - CD	6	2–183
1927	90		5	100+	SAMP (Year 1927) - CD	7	2–183
1928	90		5	100+	SAMP (Year 1928) - CD	8	2–183
1929	90		5	100+	SAMP (Year 1929) - CD	9	2–183

Notes: “SAMP” refers to *Statistique Annuelle du Mouvement de la Population*; “CD” refers to *Causes de Décès*.

(1) Aisne - Ardennes - Marne - Meurthe et Moselle - Meuse - Moselle - Nord-Oise - Pas de Calais - Bas Rhin - Haut Rhin - Somme - Vosges

(2) Moselle - Bas Rhin - Haut Rhin

Table 8: SOURCES FOR CIVILIAN DEATHS, 1930–2014

Year	Départements		Classes		Publication	Book	Page
	Total	Missing	Step	Ceiling			
1930	90		5	80+	SAMP (Year 1930) - CD	10	16–195
1931	90		5	80+	SAMP (Year 1931) - CD	11	16–195
1932	90		5	80+	SAMP (Year 1932) - CD	12	16–195
1933	90		5	80+	SAMP (Year 1933) - CD	13	16–195
1934	90		5	80+	SAMP (Year 1934) - CD	14	16–195
1935	90		5	80+	SAMP (Year 1935) - CD	15	16–195
1936	90		5	80+	SAMP (Year 1936) - CD	16	16–195
1937	90		5	100+	SAMP (Year 1937)	17	54–57
1938	90		5	100+	SAMP (Year 1938)	18	154–157
1939	87	(2)	5	100+	SAMP (Years 1939–1942)	19	118–125
1940	87	(2)	5	100+	SAMP (Years 1939–1942)	19	178–185
1941	87	(2)	5	100+	SAMP (Years 1939–1942)	19	238–245
1942	87	(2)	5	100+	SAMP (Years 1939–1942)	19	298–245
1943	86	Corse + (2)	5	100+	SAMP (Year 1943)	20	58–65
1944	86	Corse + (2)	5	100+	SAMP (Year 1944)	21	58–65
1945	87	(2)	5	100+	SAMP (Year 1945)	22	60–67
1946	90		5	100+	SAMP (Years 1946–1947)	23	110–117
1947	90		5	100+	SAMP (Years 1946–1947)	23	170–177
1948	90		5	100+	SAMP (Years 1948–1949)	24	242–249
1949	90		5	100+	SAMP (Years 1948–1949)	24	308–315
1950	90		5	100+	SAMP (Years 1950–1951)	25	240–247
1951	90		5	100+	SAMP (Years 1950–1951)	25	314–321
1952	90		5	85+	SAMP (Year 1952)	26	196–203
1953	90		5	90+	SAMP (Years 1953–1955)		291–294
1954	90		5	90+	SAMP (Years 1953–1955)		360–363
1955	90		5	90+	SAMP (Years 1953–1955)		434–437
1956	90		5	90+	SAMP (Years 1956–1959)	II	104–115
1957	90		5	90+	SAMP (Years 1956–1959)	II	272–283
1958	90		5	90+	SAMP (Years 1956–1959)	II	438–449
1959	90		5	90+	SAMP (Years 1956–1959)	II	608–619
1960	90		5	90+	SAMP (Years 1960–1962)	II	134–145
1961	90		5	90+	SAMP (Years 1960–1962)	II	364–375
1962	90		5	90+	SAMP (Years 1960–1962)	II	594–605
1963	90		5	90+	SAMP (Years 1963–1964)	II	140–145
1964	90		5	90+	SAMP (Years 1963–1964)	II	312–317
1965	90		5	90+	SAMP (Years 1965–1966)	II	156–165
1966	90		5	90+	SAMP (Years 1965–1966)	II	360–369
1967	90		10	75+	SCD (Years 1966–1967)		210–211
1968–1997	95		1	125+	Detailed Files INSEE (*)		
1998–2014	95		1	125+	www.insee.fr (Detailed Files)		

Notes: “SAMP” refers to *Statistique Annuelle du Mouvement de la Population*; “SCD” refers to *Statistiques des Causes de Décès*.

(*) Detailed Files obtained with ADISP

(2) Moselle - Bas Rhin - Haut Rhin

INSEE.fr: <https://www.insee.fr/fr/statistiques/2408054?sommaire=2117120>

Table 9: SOURCES FOR BIRTHS, 1901–1935

Year	Départements		Publication	Book	Page
	Total	Missing			
1901	87	(2)	SAMP (Year 1901)	31	32
1902	87	(2)	SAMP (Year 1902)	32	31
1903	87	(2)	SAMP (Year 1903)	33	32
1904	87	(2)	SAMP (Year 1904)	34	32
1905	87	(2)	SAMP (Years 1905 et 1906)	35-36	32
1906	87	(2)	SAMP (Years 1905 et 1906)	35-36	113
1907	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	128–131
1908	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	132–135
1909	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	136–139
1910	87	(2)	SAMP (Years 1907-1908-1909-1910)	1	140–143
1911	87	(2)	SAMP (Years 1911-1912-1913)	2	104–107
1912	87	(2)	SAMP (Years 1911-1912-1913)	2	108–111
1913	87	(2)	SAMP (Years 1911-1912-1913)	2	112–115
1914	77	(1)	SAMP (Years 1914–1919)	3	44–47
1915	77	(1)	SAMP (Years 1914–1919)	3	48–51
1916	77	(1)	SAMP (Years 1914–1919)	3	52–55
1917	77	(1)	SAMP (Years 1914–1919)	3	56–59
1918	77	(1)	SAMP (Years 1914–1919)	3	60–63
1919	90		SAMP (Years 1914–1919)	3	64–67
1920	90		SAMP (Years 1920–1924)	4	34–37
1921	90		SAMP (Years 1920–1924)	4	38–41
1922	90		SAMP (Years 1920–1924)	4	42–45
1923	90		SAMP (Years 1920–1924)	4	46–49
1924	90		SAMP (Years 1920–1924)	4	50–53
1925	90		SAMP (Year 1925) - CD	5	12–15
1926	90		SAMP (Year 1926) - CD	6	12–15
1927	90		SAMP (Year 1927) - CD	7	14–17
1928	90		SAMP (Year 1928) - CD	8	14–17
1929	90		SAMP (Year 1929) - CD	9	16–19
1930	90		SAMP (Year 1930) - CD	10	16–19
1931	90		SAMP (Year 1931) - CD	11	16–19
1932	90		SAMP (Year 1932) - CD	12	14–17
1933	90		SAMP (Year 1933) - CD	13	14–17
1934	90		SAMP (Year 1934) - CD	14	14–17
1935	90		SAMP (Year 1935) - CD	15	14–17

Notes: “SAMP” refers to *Statistique Annuelle du Mouvement de la Population*; “CD” refers to *Causes de Décès*.

(1) Aisne - Ardennes - Marne - Meurthe et Moselle - Meuse - Moselle - Nord-Oise - Pas de Calais - Bas Rhin - Haut Rhin - Somme - Vosges

(2) Moselle - Bas Rhin - Haut Rhin

Table 10: SOURCES FOR BIRTHS, 1936–1971

Year	Départements		Publication	Book	Page
	Total	Missing			
1936	90		SAMP (Year 1936) - CD	16	14–17
1937	90		SAMP (Year 1937)	17	14–17
1938	90		SAMP (Year 1938)	18	114–117
1939	87	(2)	SAMP (Years 1939–1942)	19	78–81
1940	87	(2)	SAMP (Years 1939–1942)	19	138–141
1941	87	(2)	SAMP (Years 1939–1942)	19	200–203
1942	87	(2)	SAMP (Years 1939–1942)	19	260–263
1943	86	Corse + (2)	SAMP (Year 1943)	20	18–21
1944	86	Corse + (2)	SAMP (Year 1944)	21	18–21
1945	87	(2)	SAMP (Year 1945)	22	20–23
1946	90		SAMP (Years 1946–1947)	23	74–77
1947	90		SAMP (Years 1946–1947)	23	132–135
1948	90		SAMP (Years 1948–1949)	24	198–201
1949	90		SAMP (Years 1948–1949)	24	266–269
1950	90		SAMP (Years 1950–1951)	25	196–199
1951	90		SAMP (Years 1950–1951)	25	268–271
1952	90		SAMP (Year 1952)	26	152–155
1953	90		SAMP (Years 1953–1955)		274
1954	90		SAMP (Years 1953–1955)		334
1955	90		SAMP (Years 1953–1955)		408
1956	90		SAMP (Years 1956–1959)	II	53–54
1957	90		SAMP (Years 1956–1959)	II	203–204
1958	90		SAMP (Years 1956–1959)	II	371–372
1959	90		SAMP (Years 1956–1959)	II	541–542
1960	90		SAMP (Years 1960–1962)	II	56–57
1961	90		SAMP (Years 1960–1962)	II	252–253
1962	90		SAMP (Years 1960–1962)	II	494–495
1963	90		SAMP (Years 1963–1964)	II	70–72
1964	90		SAMP (Years 1963–1964)	II	240–243
1965	90		SAMP (Years 1965–1966)	II	69–71
1966	90		SAMP (Years 1965–1966)	II	267–269
1967	90		AS 1968 Tableau XVIII (*)		50
1968	95		SAMP (Year 1968)		136–137; 144–145
1969	95		SAMP (Year 1969)		136–137; 144–145
1970	95		SAMP (Year 1970)		138–139; 146–147
1971	95		SAMP (Year 1971)		140–141; 146–147

Notes: “SAMP” refers to *Statistique Annuelle du Mouvement de la Population*; “CD” refers to *Causes de Décès*; “AS” refers to *Annuaire Statistique*

(2) *Moselle - Bas Rhin - Haut Rhin*

(*) Since SAMP in 1967 does not exist, I collect the births for the two sexes and distribute them between boys and girls pro rata births in 1966.

Table 11: SOURCES FOR BIRTHS, 1972–2014

Year	Départements		Missing	Book	Page
	Total	Missing			
1972	95		SAMP (Year 1972)		138–139; 148–149
1973	95		SAMP (Year 1973)		138–139; 144–145
1974	95		SAMP (Year 1974)		136–137; 144–145
1975	95		SAMP (Year 1975)		148–151
1976	95		SAMP (Year 1976)		148–151
1977	95		SAMP (Year 1977)		148–151
1978	95		SCD (1978)	II	29–32
1979	95		SAMP (Year 1979)		146–149
1980	95		SAMP (Year 1980)		146–149
1981	95		SAMP (Year 1981)		146–149
1982	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1982)		171–174
1983	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1983)		171–174
1984	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1984)		181–184
1985	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1985)		172–175
1986	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1986)		172–175
1987	95		Collec. de l'INSEE Série D - La Sit. Dem. (Year 1987)		150–153
1988	95		IR-DS n° 3–4		176–179
1989	95		IR-DS n° 10		174–177
1990	95		IR-DS n° 16–17		212–215
1991	95		IR-DS n° 26–27		186–189
1992	95		IR-DS n° 42–43		188–191
1993	95		IR-DS n° 49–50		188–191
1994	95		IR-DS n° 51–52		188–191
1995	95		IR-DS n° 65–66		188–191
1996	95		IR-DS n° 70–71		217–220
1997	95		IR-DS n° 75–76		194–197
1998–2014	95		www.insee.fr		

Notes: “SAMP” refers to *Statistique Annuelle du Mouvement de la Population*; “CD” refers to *Causes de Décès*; “AS” refers to *Annuaire Statistique*; “IR-DS” refers to *Insee Résultats-Démographie et Société*

www.insee.fr: <https://www.insee.fr/fr/statistiques/2408051?sommaire=2117120>

Table 12: SOURCES FOR CENSUSES, 1901–2014

Date	Départements		Publication	Book	Table	Variable	Ceiling age
	Total	Missing					
March 4th, 1901	87	(1)	Stat. du RGP Stat.	I to III	I et III	Y. of birth	95
March 6th, 1906	87	(1)	Stat. du RGP Stat.	II and III	II	Y. of birth	80
March 5th, 1911	87	(1)	Stat. du RGP Stat.	II	III	Y. of birth	105
March 6th, 1921	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 7th, 1926	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 8th, 1931	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 8th, 1936	90		Stat. du RGP Stat.	II and III	I	Y. of birth	80
March 10th, 1946	90		RGP - Results by dept	I to VI	I	Y. of birth	80
May 8th, 1954	90		RGP - Results by dept	I to VI	D1	Y. of birth	89
March 7th, 1962	94	(2)	DE - Results by dept	I to VI	D1	Y. of birth	84
March 1st, 1968	95		www.insee.fr			Age	120
Feb. 20th, 1975	95		www.insee.fr			Age	120
March 4th, 1982	95		www.insee.fr			Age	120
March 5th, 1990	95		www.insee.fr			Age	120
March 8th, 1999	95		www.insee.fr			Age	120
January 1st, 2008	95		www.insee.fr			Age	120
January 1st, 2013	95		www.insee.fr			Age	120
January 1st, 2014	95		www.insee.fr			Âge	120

Notes: “RGP” refers to *Recensement Général de la Population*; “DE” refers to *Dépouillement Exhaustif*

(1) *Moselle - Bas Rhin - Haut Rhin*

(2) In 1962, the census made in *Corse* was irrelevant (cf p. 5 of the book). Only the 1/20th exploitation available in the regional *Provence-Alpes-Côte d’Azur* book was used. This one provided population by quinquennial group of birth years while the last class provided the 74 year-old and over, not the 84 year-old and over. To compute these age classes and get an homogeneous census, I use the distribution of the other *départements*. As an exemple, for ladies born between 1958 and 1962, 23.95% were born in 1961 elsewhere. So I apply this percentage on the sum of ladies born between 1958 and 1962 in *Corse* (4,860) and I deduct that 1,164 were born in 1961.

www.insee.fr : <https://www.insee.fr/fr/statistiques/2414232>