

Computations of French Lifetables by Department,
1901–2014. On-line appendix.

February 14, 2020

A1 From Deaths by Age Group to Deaths by Lexis Triangle

Use of 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, the highest age in the distribution and the age above which no further deaths are observed, set at 105. Equation (1) fits a cubic spline by using these values ($I(.)$ 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 department, 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. 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. 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:

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

Use of Kannisto Model

Deaths obtained using cubic splines 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 $S(80+x)$:

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

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 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 $d(x) = \hat{s}(x) - \hat{s}(x+1)$. Then, I obtain deaths at each age:

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

Splitting deaths into Lexis triangles

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 since my birth records only go back to 1853. For earlier years I assume 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 follows¹:

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

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

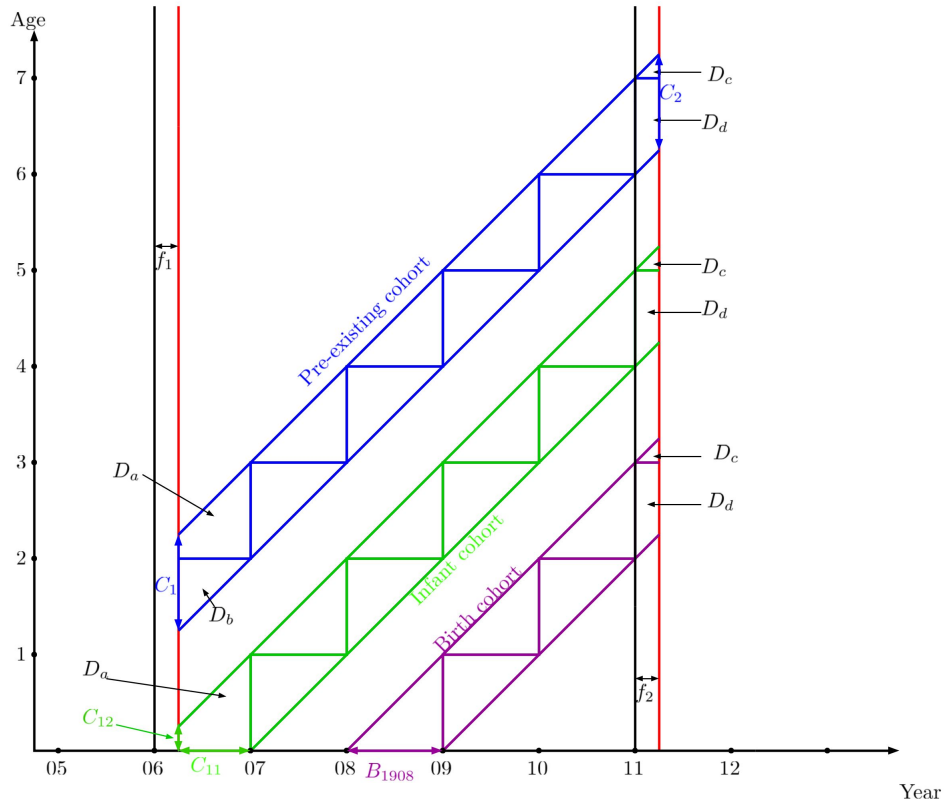
A2 Computations of Population on 1st January

Intercensal Survival

The first method used to compute yearly populations on 1st January is called “Intercensal Survival”. With this method I can estimate population by age (for those aged under 80) for each intercensal period. Populations at the second census (e.g. 1911 for 1906–1911) are not estimated in the same way for all cohorts. Figure 1 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.

¹When $IMR(t)$ is equal to zero because of no infant deaths, I assume a 0,00000001 value so that $\log IMR(t)$ can be calculated.

Figure 1: 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 1, in blue) was 1- or 2-years-old at the time of the 1906 census and was born in 1904. The data is by birth year 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:

$$\hat{C}_2 = C_1 - (D_a + D_b) - \sum_{i=0}^{N-1} [D_U(x + i, t + i) + D_L(x + i + 1, t + i)] - (D_c + D_d), \quad (11)$$

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:

$$P(x + n, t + n) = C_1 - (D_a + D_b) - \sum_{i=0}^{n-1} [D_U(x + i, t + i) + D_L(x + i + 1, t + i)] + \frac{1 - f_1 + n}{N + 1 - f_1 + f_2} \Delta_x. \quad (12)$$

There is only one “Infant cohort” to track for each intercensal period: in Figure 1, it is 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,

$$\hat{C}_2 = C_1 - D_a - \sum_{i=0}^{N-1} [D_U(i, t + i) + D_L(i + 1, t + i)] - (D_c + D_d), \quad (13)$$

and

$$P(n, t + n) = C_1 - (D_a + D_b) - \sum_{i=0}^{n-1} [D_U(i, t + i) + D_L(i + 1, t + i)] + \frac{\frac{1}{2}(1 - f_1^2) + n}{N + \frac{1}{2}(1 - f_1^2) + f_2} \Delta_0. \quad (14)$$

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:

$$\hat{C}_2 = B_{t+j} - D_L(0, t+j) - \sum_{i=1}^{N-1} [D_U(i-1, t+j+i) + D_L(i, t+j+i)] - (D_c + D_d). \quad (15)$$

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:

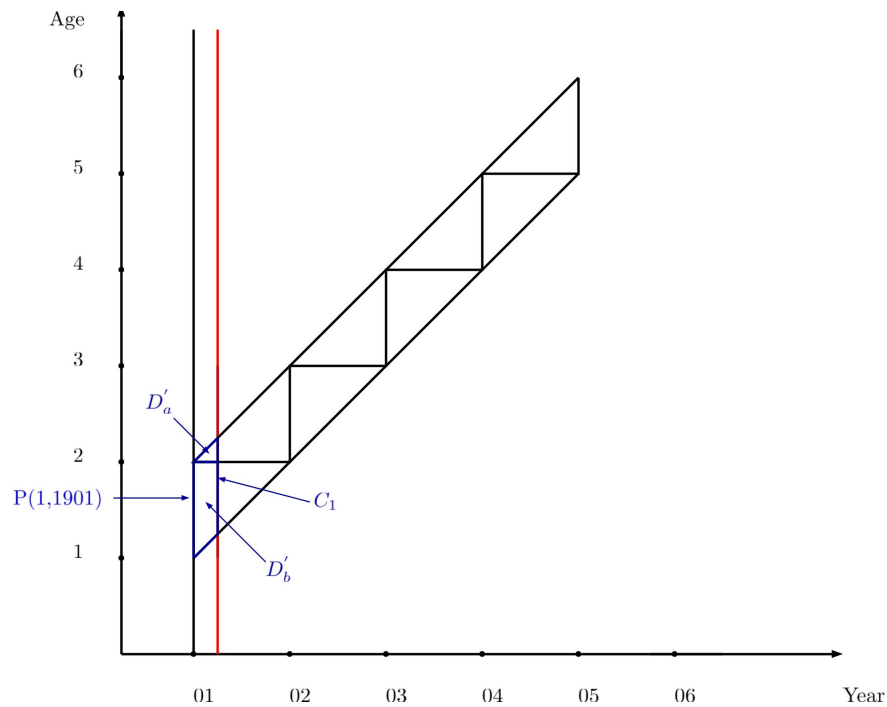
$$P(k, t+j+k+1) = B_{t+j} - D_L(0, t+j) - \sum_{i=1}^k [D_U(i-1, t+j+i) + D_L(i, t+j+i)] + \frac{2k+1}{2K+1+2f_2} \Delta_{t+j}. \quad (16)$$

Precensal Survival Method

The second method I use is “Precensal Survival”, to compute populations for the first 1st January of the whole period. Figure 2 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:

$$P(x-1, t_1-1) = C_1 + D'_a + D'_b. \quad (17)$$

Figure 2: PRECENSAL SURVIVAL METHOD

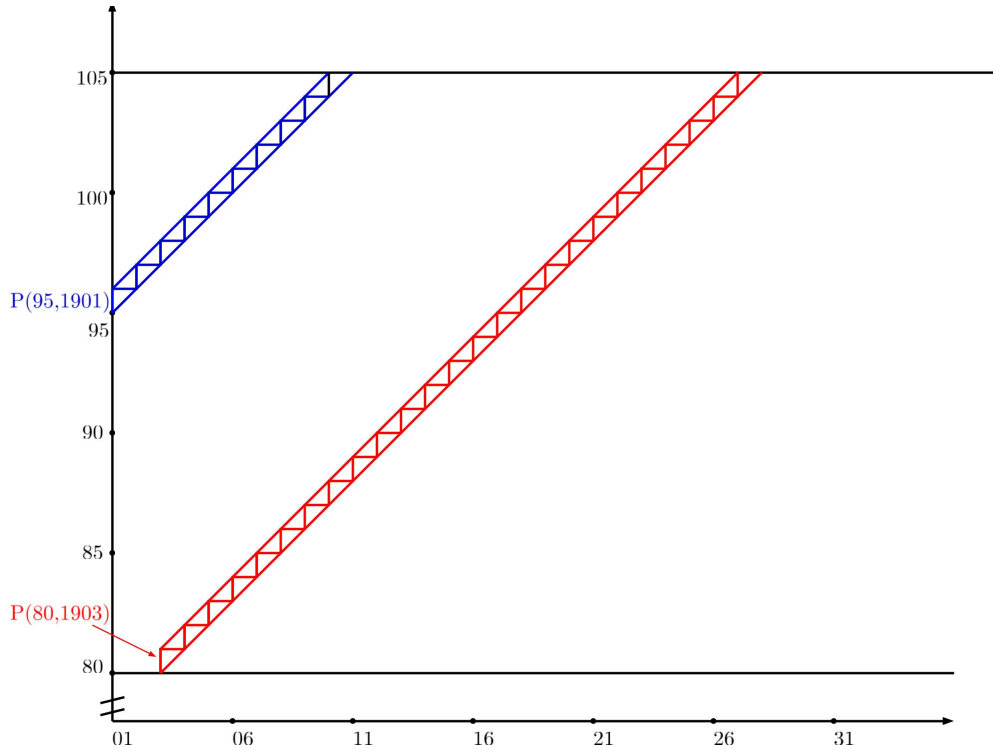


Extinct Cohorts Method

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 3 reveals that my data comprises two kinds of extinct cohorts. The first are “Full cohorts” (Figure 3, 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 3, in blue), aged 80 and over 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 3: EXTINCT COHORTS METHOD



Survivor Ratio Method

The last method I use is “Survivor ratio”, to calculate non-extinct cohorts of age 85 and over in 2013. Figure 4 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:

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

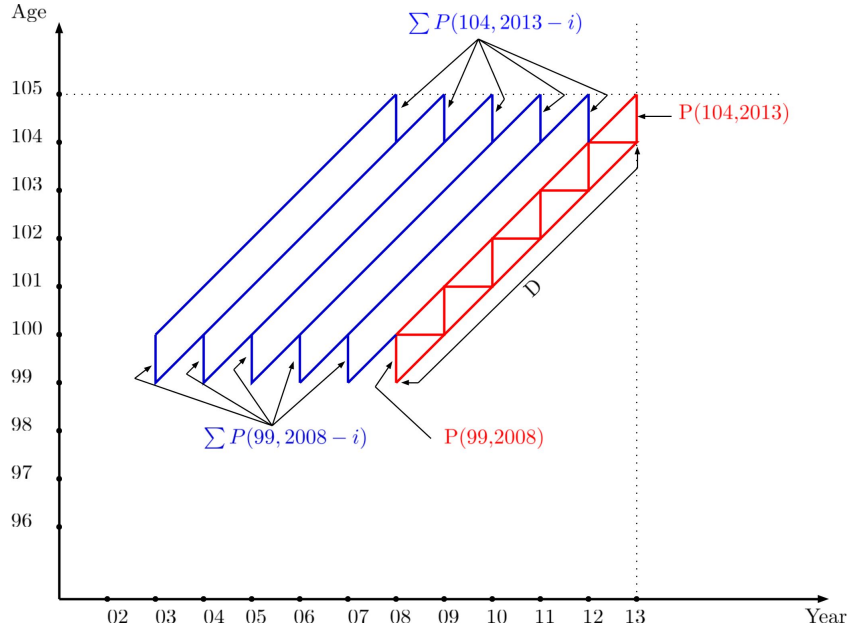
I assume that there is no migration at these ages. R may also be expressed:

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

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 :

$$P(x,t) = \frac{R}{1-R} \dot{D}. \quad (18)$$

Figure 4: 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:

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

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 and called P_{85+}^{Rec} – with the 85-and-over population on 1st January 2013, as calculated by the Survivor Ratio method and 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 $\hat{P}(x-1, 2012), \hat{P}(x-2, 2011), \dots$

A3 Census Adjustments

For my purposes it is simpler to compute population figures by birth year. The cubic splines adjustment takes into account that populations were given by age and not by birth year from 1968 onwards. Before 1968, census data is given by birth year.

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 department i and each sex j in 1911:

$$R_{95ij}^{1911} = \frac{\sum_{s=95}^{105} P_{sij}^{1911}}{\sum_{s=80}^{105} P_{sij}^{1911}}. \quad (19)$$

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:

$$\sum_{s=95}^{105} P_{sij}^{1901} = R_{95ij}^{1911} \times \sum_{s=80}^{105} P_{sij}^{1901}. \quad (20)$$

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

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, as shown in Table 1, in italics. For that purpose I assume that births were spread uniformly over time.

Finally, the 1911 census is rather different because it provides data for each birth year 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

TABLE 1: 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

Note: 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.

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”.

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 provides no further information and would involve fractional knots.

A4 Military deaths and deaths in deportation

Estimates of Military Deaths during the Two World Wars

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 department before the war. Since the sources used are incomplete, I couple two different matrices. The first provides the total of deaths by department and birth year.² It comes from the

²The classification of departments from the “*Mémoire des Hommes*” website is modified to fit the classification for civilian deaths. Problems concern *Corse* (two departments counting as one) and the old departments of *Seine* and *Seine-et-Oise*. For these last two, deaths are given according to the new departments. 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 departments 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%

Defense Ministry’s database, which lists all the “*Morts pour la France*” (MPLF) of the two World Wars. The second provides the total of deaths at the national level by birth year and death year. It mobilizes the crowd-based indexing on the *Mémoire des Hommes* website: Anyone, using his personal research on a specific soldier, can inform both his birth year and death year. This work has been done for just over 20% of total deaths. I wonder if this sample is representative of the distribution by death year. To do so, I use Pedroncini (1992)’s work: It gives total military deaths by death year. 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 birth year and death year are therefore extracted from the Defense Ministry’s database.

Table 2: DISTRIBUTION BY YEAR OF DEATH OF SOLDIERS

Source		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%

By cross-referencing these two matrices, I get a matrix giving total deaths by department, birth year and death year. I assume that there is little variation between departments in the death year 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 statistical estimates of deaths during the First World War. He uses the Marin’s report, followed by Hubert (1931) and Dupaquier (1988). Roure’s report cited by Prost (2008) reveals 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, I do not keep them in the total since these populations were not registered in 1911 in the French *départements* and were surely recorded in the civilian deaths of their home country. Finally, I obtain 1,304,400 deaths.

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

Table 3: MILITARY DEATHS DURING THE FIRST WORLD WAR

Source	Variable	Deaths
	Total of French military deaths	1,282,100
Roure	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).

Estimates of Deportees

The deportee database is nominative (1 line for each deportee). Sex, birth department (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 the date of birth follows date of death), the date is erased. For individuals whose year of death was after 1946 (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 department from where he is born, I may infer the home-department before deportation. Similarly, deportees born abroad must be located in a French department.

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 department 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 : Birth country \times Age (90×105), retrieved from the “*MemorialGenWeb*” database,
2. P : Department of residence \times Birth country (91×48), retrieved from the 1936 census,
3. R : Department of residence \times Age (90×105).

The first modification concerns *Seine*. Matrix P comprises 91 departments 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 department 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 . The level of detail in the “*MemorialGenWeb*” 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”. 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.

French Deportees

There are 43,055 deaths of French-born deportees in the database. I cannot assume that any deportee born in a department stayed in that department. A transfer matrix must therefore be constructed linking department

of birth and department of residence before deportation. I use the matrix cross-referencing department of residence and department of birth in the 1946 census. This matrix distinguishes males and females. 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.

I make a few preliminary modifications. The main is to allocate the deportees according to the post-1968 departments 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 : Birth department \times Age (90×105), retrieved from “*Mémoire des Hommes*” database
2. P : Department of residence \times Department of birth (90×90), retrieved from the 1936 census,
3. R : Department of residence \times Age (90×105).

P is transformed so that the matrix gives the probability that an individual born in *département* i lives in department 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 department, 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 reports 27,000 resisters dead in deportation and 83,000 jewish and other deportees. Consequently, I consider that 110,000 individuals died in the camps.

A5 Missing Data During the Two World Wars

There are 10 departments (*Aisne, Ardennes, Marne, Meurthe-et-Moselle, Meuse, Nord, Oise, Pas-de-Calais, Somme, Vosges*) with missing data during the First World War, and 4 during the Second World War (*Corse* between 1943 and 1944, *Moselle, Bas-Rhin, Haut-Rhin* between 1939 and 1945). These missing data are of 2 types: births and stillbirths, as well as deaths. The general assumption used for the estimations of these missing data is that the neighbours departments 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 department.

Births and stillbirths

The choice of the reference department for each of the missing departments 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 department and the variable 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 department, j the potential reference department. The ratio R_{ij}^t is calculated for a demographic variable V :

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

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

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

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 department, j

the potential reference department. The ratio R_{xij}^t is calculated for deaths D at age x :

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

The mean \bar{x}_{xij} and the 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 department and reference department 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 department j^* chosen is the one with the lowest score among all the possible reference departments. After choosing the reference department for each missing department and subperiod, deaths at age x for the department i are estimated as follows:

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

A6 Reorganization of *Ile-de-France* in 1968

By changing the three departments 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 the data. I change my methodology so as to track each of these departments 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 departments since 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 departments since 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 departments, but no population

flows. For the same intercensal period, I have population flows and the 1962 census for the old departments, but no data according to the 1968 census. I choose to track the old departments 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 departments, useful to split deaths in Lexis triangles. The second is about populations of the old departments in 1968, to estimate the 1st January population of these departments between 1962 and 1968.

To estimate births of the new departments before 1968, I use the 1968 distribution. I assume that the weight of each department 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 departments 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 departments 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 departments $\hat{P}_{IdF}^{68}(x)$, 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 departments. 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 departments j and use this migratory profile to compute 1968 census populations:

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

A7 Computations of 1st January Populations by Class of Departments

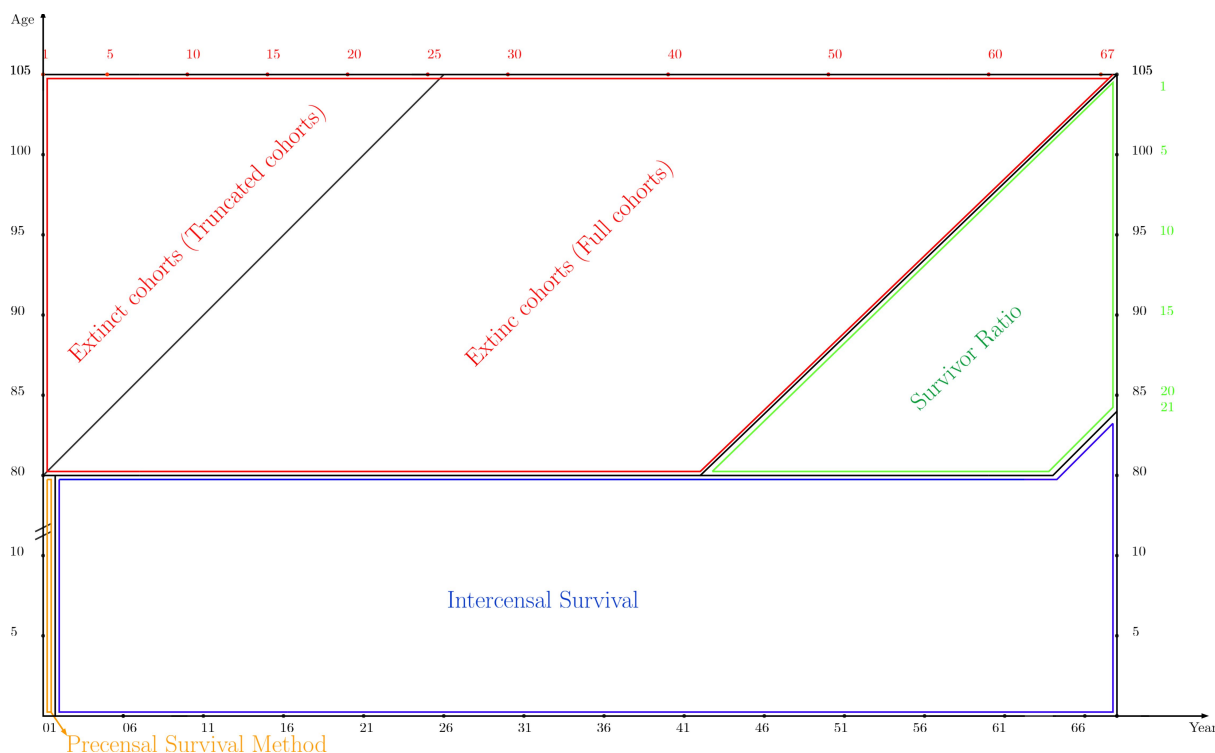
I miss data for some departments between 1901 and 2014 in order to compute lifetables. Consequently, I divided my panel into four classes:

C_1 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.

C_2 *Seine* (75) and *Seine-et-Oise* (78). The lifetables for these departments are estimated for the period 1901–1968. Figure 5 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 5: ESTIMATIONS OF POPULATIONS FOR DEPARTMENTS OF CLASS 2



C_3 The new departments 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 6 presents the methods used to compute populations at each 1st January.

A8 Sources of Raw Data

Table 4: SOURCES FOR CIVILIAN DEATHS, 1901–1929

Year	Departments		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

Note: “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 5: SOURCES FOR CIVILIAN DEATHS, 1930–2014

Year	Departments		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	<i>Corse</i> + (2)	5	100+	SAMP (Year 1943)	20	58–65
1944	86	<i>Corse</i> + (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+	Detailed Files, www.insee.fr		

Note: “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>, data downloaded on March, 2016.

Table 6: SOURCES FOR BIRTHS, 1901–1935

Year	Departments		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

Note: “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 7: SOURCES FOR BIRTHS, 1936–1971

Year	Departments		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	<i>Corse</i> + (2)	SAMP (Year 1943)	20	18–21
1944	86	<i>Corse</i> + (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 Table 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

Note: “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 8: SOURCES FOR BIRTHS, 1972–2014

Year	Departments		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		<i>Collec. de l'INSEE Série D - La Sit. Dem.</i> (Year 1982)		171–174
1983	95		<i>Collec. de l'INSEE Série D - La Sit. Dem.</i> (Year 1983)		171–174
1984	95		<i>Collec. de l'INSEE Série D - La Sit. Dem.</i> (Year 1984)		181–184
1985	95		<i>Collec. de l'INSEE Série D - La Sit. Dem.</i> (Year 1985)		172–175
1986	95		<i>Collec. de l'INSEE Série D - La Sit. Dem.</i> (Year 1986)		172–175
1987	95		<i>Collec. de l'INSEE Série D - La Sit. Dem.</i> (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		Detailed Files, www.insee.fr		

Note: “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>, data downloaded on March, 2016.

Table 9: SOURCES FOR CENSUSES, 1901–2014

Date	Departments		Publication	Book	Table	Variable	Ceiling age
	Total	Missing					
March 4th, 1901	87	(1)	RGP Stat.	I to III	I et III	Y. of birth	95
March 6th, 1906	87	(1)	RGP Stat.	II and III	II	Y. of birth	80
March 5th, 1911	87	(1)	RGP Stat.	II	III	Y. of birth	105
March 6th, 1921	90		RGP Stat.	II and III	I	Y. of birth	80
March 7th, 1926	90		RGP Stat.	II and III	I	Y. of birth	80
March 8th, 1931	90		RGP Stat.	II and III	I	Y. of birth	80
March 8th, 1936	90		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

Note: “RGP” refers to *Recensement Général de la Population*; “DE” refers to *Dépoulement 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>, data downloaded on March, 2016.