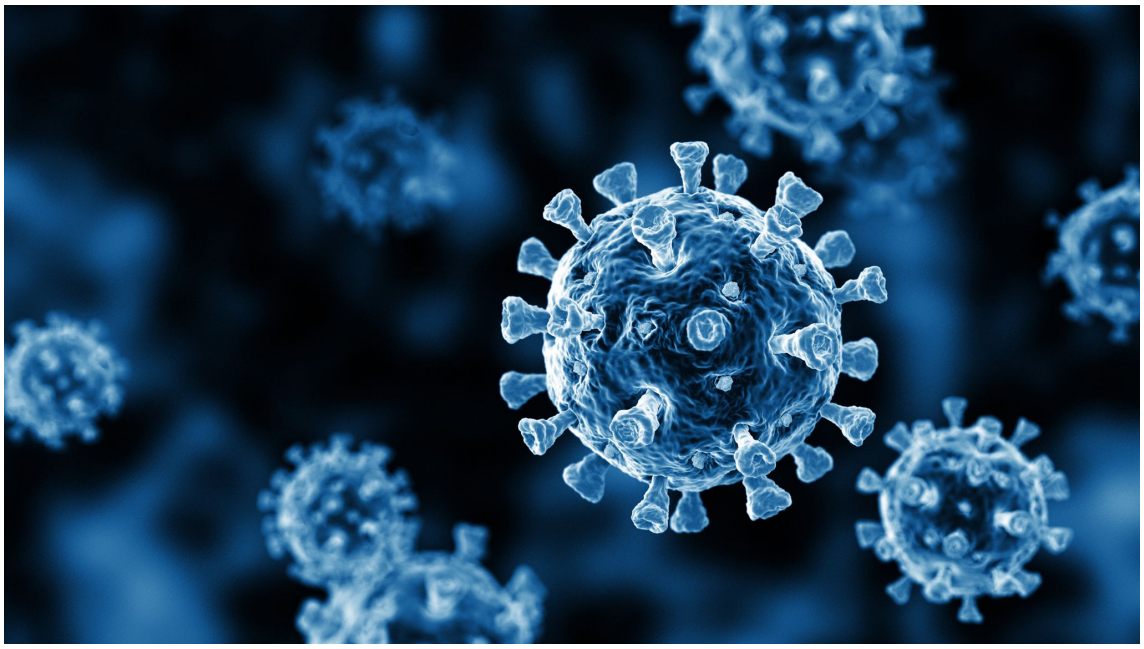
How Many People Did COVID-19 Kill in France in 2020?

Côme Pollet - Jules Chapon - Victor Parent - Diego Renaud

May 17, 2023



Contents

1 Introduction

This document presents our results and analysis to meet the requirements of our 2nd-year StatApp project at ENSAE Paris, which is to answer the question: **"How many people did COVID-19 kill in France in 2020?"**

The objective of this StatApp project is to replicate and criticize the article "The COVID-19 Epidemic Has Had a Relatively Small Impact on Mortality in France" published by Laurent Toubiana, Laurent Mucchielli, Pierre Chaillot, and Jacques Bouaud. In this article, the four authors conducted a statistical study leading to results that contradict those obtained by INSEE (National Institute of Statistics and Economic Studies). According to the authors, the COVID-19 epidemic did not lead to excess mortality in the 0-65 age group, and only caused a slight excess mortality in those over 65. They reported a much lower excess mortality (around 4%) compared to the INSEE's estimate (nearly 8%). It seems reasonable to replicate their analysis to verify the accuracy of these results, which contradict the findings presented by the scientific community, government, and independent statistical institutes (INSEE, Inserm, etc.). We will attempt to explain how these four authors arrived at such results and, at the same time, try to quantify the impact of COVID-19 on mortality in France in 2020.

To conduct this fact-checking work, we will start by closely studying the article in question to identify potential flaws in the authors' analysis. Following that, we will compare their methodology with the approaches typically used in the field of demography and excess mortality studies. Finally, we will present our own statistical model to independently estimate excess mortality in France in 2020.

2 Study of the Article

The article published by Laurent Toubiana, Laurent Mucchielli, Pierre Chaillot, and Jacques Bouaud is available here.

2.1 Summary of the Article

In this article, the authors examine the impact of the COVID-19 pandemic on mortality in France in 2020, suggesting that lockdown measures might have been more harmful than beneficial. They begin with a historical overview, noting that the SARS-CoV-2 virus emerged in late 2019 in Wuhan, China, and that the WHO declared a pandemic on March 11, 2020. They mention that the arrival of the epidemic in France around February 15, 2020, surprised health authorities due to the high number of hospitalizations and deaths. In response to the crisis, a nationwide lockdown was implemented on March 17, 2020.

In epidemiology, excess mortality is used to measure the impact of a health event. The authors define excess mortality as **the difference between the observed number**

of deaths during a health event and the expected number of deaths for the same period. They use data from the National Directory of Identification of Physical Persons (RNIPP), managed by the National Institute of Statistics and Economic Studies (INSEE), to obtain accurate information about deaths. They note that the cause of death is not always clearly indicated, which could lead to an overestimation of COVID-19-related mortality.

They emphasize **the importance of the age structure of the population** in analyzing mortality. Due to the aging population in France, the number of deaths increases steadily. They conclude that the impact of the pandemic on overall mortality in 2020 must be analyzed while considering this growing trend. The authors examine the impact of health events, such as epidemics, on the mortality of an aging population. Their analysis is primarily based on the following dynamics:

1. Medical advancements and improvements in living conditions have led to an increase in life expectancy, but this trend is starting to slow down, especially in Western countries.

2. Elderly individuals are particularly vulnerable to infectious diseases like the flu, which can cause spikes in mortality, especially during the winter months.

3. The article highlights the existence of seasonal variations in mortality, with winter peaks linked to infectious diseases and summer peaks linked to heatwaves. For example, the highest monthly peak over a fifteen-year period was recorded during the winter 2016-2017 flu epidemic with 68,969 deaths, compared to 67,537 in April 2020 during the COVID-19 epidemic.

4. The "harvesting" effect is also discussed. This phenomenon involves elderly and frail individuals dying "synchronously" during a strong health event (epidemic, heatwave), leading to a temporary "depletion" of this population category. For example, the high mortality in 2017 resulted in low mortality in 2018 and 2019.

5. In 2020, 90% of people admitted to intensive care and 65% of people who died in connection with the COVID-19 epidemic had at least one comorbidity. The median age at death in the hospital for COVID-19 patients was 85 years, and nearly 92.5% of individuals were aged 65 and older.

6. The demographic evolution of the French population is also analyzed. Individuals aged 65 and older, although representing only 19.45% of the population in 2020, accounted for 84.3% of deaths.

7. The average annual mortality rate for the entire French population in 2020 was 1.004%. For individuals under 65, it was 0.17%, but it reached 4.18% for those over 65. The mortality rate increases exponentially with age.

To estimate excess mortality in 2020, the authors claim to have used a demographic method endorsed by the National Institute of Health and Medical Research (Inserm) and detailed in this document. This method involves taking the empirical average of mortality ratios from the previous 3 years (2017, 2018, and 2019). Two estimates are made based on mortality rates for the reference years. The first is based on age groups of 5 years, while the second groups all ages into two categories: under 65 and 65 and older. For each period, confidence intervals are calculated to account for data variability. Any observed value within the confidence interval of the predicted value is considered expected, regardless of observed differences.

Tableau 2 : Evaluation de la surmortalité en 2020 (3 années de référence : 2017-19) selon des classes d'âges de 5 ans

Classe d'âge	Population en million d'habitants	Décès attendus	Décès observés	Décès en excès sur classe d'âge	Décès en excès par âge
0-4 ans	3,52 (5,41 %)	2 940 [2 925 ; 2 955]	2 691	- 234 (-7,97 % [-7,93 % ; -8,01 %])	- 206 (-6,99 % [-6,96 % ; -7,03 %])
5-9 ans	3,94 (6,05 %)	289 [275 ; 303]	272	- 3 (-0,97 % [-0,92 % ; -1,02 %])	- 25 (-8,79 % [-8,38 % ; -9,24 %])
10-14 ans	4,04 (6,21 %)	311 [293 ; 329]	321	0 %	8 (2,49 % [2,35 % ; 2,65 %])
15-19 ans	4 (6,15 %)	853 [812 ; 895]	813	0 %	- 25 (-2,92 % [-2,79 % ; -3,07 %])
20-24 ans	3,65 (5,6 %)	1 425 [1 376 ; 1 474]	1 375	- 1 (-0,07 % [-0,06 % ; -0,07 %])	- 30 (-2,14 % [-2,07 % ; -2,21 %])
25-29 ans	3,63 (5,57 %)	1 622 [1 572 ; 1 672]	1 565	- 7 (-0,44 % [-0,42 % ; -0,45 %])	- 2 (-0,12 % [-0,11 % ; -0,12 %])
30-34 ans	3,95 (6,07 %)	2 259 [2 233 ; 2 286]	2 255	0 %	- 13 (-0,59 % [-0,59 % ; -0,6 %])
35-39 ans	4,12 (6,32 %)	3 421 [3 383 ; 3 460]	3 276	- 107 (-3,12 % [-3,08 % ; -3,16 %])	- 134 (-3,91 % [-3,87 % ; -3,96 %])
40-44 ans	3,95 (6,07 %)	4 758 [4 653 ; 4 864]	4 852	0 %	89 (1,87 % [1,83 % ; 1,92 %])
45-49 ans	4,38 (6,73 %)	8 925 [8 674 ; 9 177]	8 679	0 %	- 168 (-1,88 % [-1,83 % ; -1,94 %])
50-54 ans	4,29 (6,58 %)	14 002 [13 706 ; 14 298]	13 958	0 %	2 (0,02 % [0,02 % ; 0,02 %])
55-59 ans	4,23 (6,49 %)	21 585 [20 857 ; 22 313]	21 305	0 %	- 131 (-0,61 % [-0,59 % ; -0,63 %])
60-64 ans	3,98 (6,12 %)	30 819 [30 185 ; 31 454]	31 116	0 %	- 69 (-0,22 % [-0,22 % ; -0,23 %])
65-69 ans	3,8 (5,84 %)	40 520 [40 404 ; 40 635]	42 476	1 841 (4,54 % [4,53 % ; 4,56 %])	988 (2,44 % [2,43 % ; 2,45 %])
70-74 ans	3,4 (5,22 %)	55 757 [54 681 ; 56 834]	58 327	1 493 (2,68 % [2,63 % ; 2,73 %])	3 109 (5,58 % [5,47 % ; 5,69 %])
75-79 ans	2,16 (3,32 %)	51 767 [51 303 ; 52 232]	57 516	5 284 (10,21 % [10,12 % ; 10,3 %])	2 606 (5,03 % [4,99 % ; 5,08 %])
80-84 ans	1,83 (2,82 %)	77 960 [76 645 ; 79 276]	83 367	4 091 (5,25 % [5,16 % ; 5,34 %])	3 104 (3,98 % [3,92 % ; 4,05 %])
85-89 ans	1,35 (2,08 %)	116 220 [112 979 ; 119 462]	121 025	1 563 (1,34 % [1,31 % ; 1,38 %])	1 124 (0,97 % [0,94 % ; 0,99 %])
90-94 ans	0,66 (1,02 %)	112 700 [110 672 ; 114 729]	120 443	5 714 (5,07 % [4,98 % ; 5,16 %])	5 035 (4,47 % [4,39 % ; 4,55 %])
95-99 ans	0,22 (0,34 %)	81 354 [78 771 ; 83 937]	78 384	- 387 (-0,48 % [-0,46 % ; -0,49 %])	1 047 (1,29 % [1,25 % ; 1,33 %])
Somme	65,12 (100 %)	629 490 [628 011 ; 630 970]	654 016	23 046 (3,66 % [3,65 % ; 3,67 %])	16 309 (2,59 % [2,58 % ; 2,6 %])

Source : Insee, Indicateurs démographiques, calculs des auteurs.
*NB : L'excès (ou le défaut) de décès pour les 2 grands groupes d'âge (supérieur et inférieur à 65 ans) est la somme des excès (ou défaut) estimés au niveau de chaque âge pour prendre en considération l'évolution démographique de la manière la plus fine possible. Ceci explique la légère différence de résultat entre cette estimation et celle obtenue avec un calcul direct sur les 2 grands groupes d'âge (supérieur et inférieur à 65 ans).

Figure 1: Detailed Results of the Authors' Analysis

Overall, excess mortality in 2020 compared to the previous three years (2017-2019) ranged from 23,046 to 16,309 additional deaths for the entire population. **The increase in mortality due to the COVID-19 epidemic in 2020 was estimated between 3.66% and 2.59%**, depending on the calculation method used by the authors. It's also noteworthy that excess mortality was particularly high in individuals over 65, with an excess of 17,920 deaths, representing **a 3.34% increase**. However, in individuals under 65, there was no significant excess mortality in 2020. These numbers particularly caught our attention as they are significantly lower than those provided by INSEE, which were around 9%.

The authors conclude their article by arguing that the impact of the COVID-19 epidemic on mortality is less significant than often presented in the public debate, particularly in comparison with past diseases like the Spanish flu. They note that mortality in 2019 was exceptionally low, and the increase observed in 2020 could partly be a catch-up from this under-mortality. Regarding mortality predictions due to COVID-19, the authors criticize the use of predictive models that anticipated a very high number of deaths, arguing

that these predictions were not supported by actual figures. Finally, the authors caution against a misinterpretation of the increase in mortality in 2020 as entirely attributed to COVID-19. They argue that this increase is partly due to population aging and the end of life for the baby boomer generations. In summary, they assert that COVID-19 should be seen as another of the many viral illnesses causing severe respiratory pathologies, despite what reputable scientific institutions advocate to the contrary.

2.2 Inconsistencies and Potential Questions

We have compiled here all the clues and evidence that could explain potential errors or oversights by the authors.

Manipulation of Confidence Intervals: First and foremost, the authors manipulate the bounds of confidence intervals to minimize the impact of COVID-19. In the Table 1A (Figure 2), the authors arrive at 654,016 observed deaths and 629,490 expected deaths. If you subtract them, you get an excess of 24,526 deaths, not 23,046. The latter number was obtained by taking the upper bound of expected deaths (630,970), which allows them to reach their result of 23,046. This method helps to reduce the estimated excess mortality in 2020.

Tableau 1a : Evaluation de la surmortalité en 2020 (3 années de référence : 2017-19)

Classe d'âge	Population en million d'habitants	Décès attendus	Décès observés	Décès en excès sur classe d'âge	Décès en excès par âge
0-64 ans	51,68 (79,36 %)	93 211 [91 612 ; 94 810]	92 478	(0 % [0 % ; 0 %])	- 705 (-0,76 % [-0,74 % ; -0,77 %])
65-99 ans	13,44 (20,64 %)	536 279 [528 940 ; 543 618]	561 538	17 920 (3,34 % [3,3 % ; 3,39 %])	17 013 (3,17 % [3,13 % ; 3,22 %])
Somme	65,12 (100 %)	629 490 [628 011 ; 630 970]	654 016	23 046 (3,66 % [3,65 % ; 3,67 %])	16 309 (2,59 % [2,58 % ; 2,6 %])

Figure 2: Aggregated Results of the Authors' Analysis

Confidence Interval Manipulation: The bounds of the confidence intervals for expected deaths surprised us. In fact, the bounds for expected deaths in 2020 are [628,011; 630,970], whereas in 2019, it's [617,657; 632,850]. The difference in bounds is much larger in 2019 than in 2020. Hence, we need to delve into the confidence intervals for expected deaths in 2020 to verify the authors' work.

Opacity of Results: Additionally, there's a certain opacity in their results. On one hand, they do not provide their databases, and on the other, they do not explicitly explain the statistical model used to estimate their mortality ratios.

2.3 Key Takeaways

The article under study presents an analysis that, despite claiming to be precise and impartial, leaves several significant gaps. The authors focus on general statistical concepts known to all, but they neglect to provide essential details about the more technical aspects

of their methodology. This includes their manipulation of confidence intervals as well as the opacity of their statistical model and the used databases, which are crucial elements of any rigorous statistical analysis.

Surprising Results: The authors reach conclusions that are, at best, surprising. They find a significantly lower excess mortality than what official figures have reported. This raises legitimate questions about the validity of their results.

Questionable Methodological Approach: Their approach is hasty and, in many respects, precarious. They skim over key concepts without giving them the time or importance they deserve. Their treatment of confidence intervals, for instance, is particularly problematic, as it appears they are using these tools to downplay the impact of COVID-19 on mortality in 2020.

Opacity of Data and Methodology: They exhibit a certain opacity both in their handling of processed data and in their methodology. They don't provide their databases and don't clearly explain the demographic method they use to estimate their mortality ratios. This makes it challenging for readers to evaluate the accuracy of their conclusions.

Objective of Our Analysis: Our objective through this analysis is to untangle the threads of their argumentation, understand their methods, and see if their conclusions hold up. Ultimately, we want to determine whether the authors reasonably downplayed the impact of COVID-19 on mortality in France in 2020, or if, as we suspect, they underestimated the true severity of the epidemic.

This study is important not only for understanding the actual impact of COVID-19 on our society, but also for emphasizing the importance of rigorous and transparent statistical analysis. In an era where information is omnipresent, it is more crucial than ever to distinguish between serious analyses and hasty or misleading interpretations.

3 Statistical Model

3.1 Definition of Variables

To establish a prediction of the expected number of deaths in France in 2020, we need to predict the mortality ratios in France in 2020 if COVID-19 had never appeared. These ratios represent the probability that an individual will die during the year based on their age and sex. Hence, we need to define the set of variables at play in our experiment.

3.1.1 Population Counts

We define:

• n_t as the total population count for a given year t.

3.1.2 Random Variables

For any year t, we have a sample of n_t individuals that are i.i.d but heterogeneous. For any individual i living in year t, we define the following random variables:

• $A_{i,t}$ representing the age of individual i on December 31st of year t.

• $S_{i,t}$ representing the sex of individual i during year t.

• $N_{t,a,s} = \sum_{i=1}^{n_t}(\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\})$ representing the number of individuals of age a and sex s during year t.

Applying the **law of large numbers**, we have: $\forall\, t \in [\![t_{min}; t_{max}]\!], \forall\, a \in [\![a_{min}; a_{max}]\!], \forall\, s \in \{0, 1\}$

$$\frac{N_{t,a}}{n_t} \xrightarrow[n_t \to \infty]{p.s} \mathbb{P}(A_{i,t} = a)$$

$$\frac{N_{t,s}}{n_t} \xrightarrow[n_t \to \infty]{p.s} \mathbb{P}(S_{i,t} = s)$$

$$\frac{N_{t,a,s}}{n_t} \xrightarrow[n_t \to \infty]{p.s} \mathbb{P}(A_{i,t} = a \cap S_{i,t} = s) = p_{t,a,s}$$

• $X_{i,t}$ representing the probability for individual i to die during year t.
We have:

$$X_{i,t} = \begin{cases} 1 & \text{if individual } i \text{ died during year } t \\ 0 & \text{otherwise} \end{cases}$$

We then define

$$q_{i,t,a,s} = \mathbb{E}[X_{i,t} \mid A_{i,t} = a, S_{i,t} = s]$$

We have $((X_{i,t} \mid A_{i,t} = a, S_{i,t} = s))_{i=1,\dots,n_t} \overset{iid}{\sim} \mathcal{B}(q_{t,a,s})$ a sequence of independent and identically distributed random variables following a common Bernoulli distribution with parameter $q_{t,a,s}$. Intuitively, we model our experiment as a Bernoulli urn with a conditional probability of success.

Example:
$q_{2020,30,1} = \mathbb{E}[X_{1,2020} \mid A_{1,2020} = 30,\ S_{1,2020} = 1]$ represents the probability for an individual to die during the year 2020, given that they are a 30-year-old male.

3.1.3 Observed Mortality Ratios

To predict our mortality ratios in 2020, we will use those from previous years. These ratios, which can be observed in INSEE tables, are realizations of the previously introduced random variables. We can define these observed mortality rates $\hat{q}_{t,a,s}$ at year t, age a, and sex s:

$$\hat{q}_{t,a,s} = \frac{1}{N_{t,a,s}} . \sum_{i=1}^{n_t} (X_{i,t} . \mathbb{1}\{A_{i,t} = a\} . \mathbb{1}\{S_{i,t} = s\})$$

with $\mathbb{E}[X_{i,t} \mid A_{i,t} = a, S_{i,t} = s] = q_{t,a,s}$ and $\mathbb{V}[X_{i,t} \mid A_{i,t} = a, S_{i,t} = s] = q_{t,a,s} . (1 - q_{t,a,s})$. For any period $\mathcal{T} = [\![\underline{t}; \overline{t}]\!]$ (the authors focus on years 2017, 2018, and 2019 in their article to predict the ratios of 2020), we define the vector $\hat{q}_{\mathcal{T},a,s}$ of mortality ratios at period $\mathcal{T}$, age a, and sex s:

$$\hat{q}_{\mathcal{T},a,s} = \begin{pmatrix} \hat{q}_{\underline{t},a,s} \\ ... \\ \hat{q}_{\overline{t},a,s} \end{pmatrix} = \begin{pmatrix} \frac{1}{N_{\underline{t},a,s}} \sum_{i=1}^{n_{\underline{t}}} (X_{i,\underline{t}} . \mathbb{1}\{A_{i,\underline{t}} = a\} . \mathbb{1}\{S_{i,\underline{t}} = s\}) \\ ... \\ \frac{1}{N_{\overline{t},a,s}} \sum_{i=1}^{n_{\overline{t}}} (X_{i,\overline{t}} . \mathbb{1}\{A_{i,\overline{t}} = a\} . \mathbb{1}\{S_{i,\overline{t}} = s\}) \end{pmatrix}$$

Example:
Considering $\mathcal{T} = [\![2017; 2019]\!]$, $a = 30$, and $s = 1$, we have:

$$\hat{q}_{[\![2017;2019]\!],30,1} = \begin{pmatrix} \hat{q}_{2017,30,1} \\ \vdots \\ \hat{q}_{2019,30,1} \end{pmatrix} = \begin{pmatrix} \frac{1}{N_{2017,30,1}} \sum_{i=1}^{n_{2017}} (X_{i,2017} . \mathbb{1}\{A_{i,2017} = 30\} . \mathbb{1}\{S_{i,2017} = 1\}) \\ \vdots \\ \frac{1}{N_{2019,30,1}} \sum_{i=1}^{n_{2019}} (X_{i,2019} . \mathbb{1}\{A_{i,2019} = 30\} . \mathbb{1}\{S_{i,2019} = 1\}) \end{pmatrix}$$

which represents the vector of mortality ratios from 2017 to 2019 for 30-year-old males, and will subsequently allow us to estimate the mortality rate for 2020 for 30-year-old males.

3.1.4 Predicted Mortality Ratios

Once the past mortality ratios are defined, we need to create the estimator for our parameters of interest, namely the mortality ratios in 2020. For this purpose, we define f, a differentiable function:

$$\begin{array}{rccc} f & : & \mathbb{R}^d & \to & \mathbb{R}^p \\ & & (x_1, ..., x_d) & \mapsto & f(x_1, ..., x_d) \end{array}$$

We can then define $\tilde{q}_{\tilde{t},a,s}$ as a new variable aiming to predict the mortality ratios for year $\tilde{t}$, age a, and sex s given in terms of past mortality ratios:

$$\tilde{q}_{\tilde{t},a,s} = f(\hat{q}_{\mathcal{T},a,s})$$

This function f aims to predict a mortality ratio using past mortality ratios. We can mention some examples:
• defining the mortality rate as that of the previous year
• defining the mortality rate as the average of the rates of the previous three years. This is the method chosen by our authors.

We also set $q^P_{\tilde{t},a,s} = f(q_{\mathcal{T},a,s})$. This theoretical variable represents the theoretical prediction using our function f. It differs from $\tilde{q}_{\tilde{t},a,s}$ in that $q^P_{\tilde{t},a,s}$ does not depend on the realization of random variables. We will use this theoretical prediction in creating our confidence intervals.

Example:
Considering $\mathcal{T} = [\![2017; 2019]\!]$, $a = 30$, $s = 1$, and $\tilde{t} = 2020$, we have:

$$\tilde{q}_{2020,30,1} = f(\hat{q}_{[\![2017;2019]\!],30,1})$$

which represents the vector of predicted mortality ratios for 30-year-old males in 2020, based on the observed mortality ratios for 30-year-old males in 2017, 2018, and 2019.

3.2 Observed Mortality Ratios: Limit Theorems

In order to construct confidence intervals or conduct tests on our prediction, we need to study the convergence of our estimators for observed mortality ratios. We recall that we have:

$$N_{t,a,s} = \sum_{i=1}^{n_t} (\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\})$$

and

$$\hat{q}_{t,a,s} = \frac{1}{N_{t,a,s}} \sum_{i=1}^{n_t} (X_{i,t}.\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\}) = \left(\frac{n_t}{N_{t,a,s}}\right) \frac{1}{n_t} \sum_{i=1}^{n_t} (X_{i,t}.\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\})$$

for all i (recalling that we have $p_{t,a,s} = \mathbb{P}(A_t = a \cap S_t = s)$):

• $\mathbb{E}[X_{i,t} \mid A_{i,t} = a, S_{i,t} = s] = q_{t,a,s}$

• $\mathbb{V}[X_{i,t} \mid A_{i,t} = a, S_{i,t} = s] = q_{t,a,s}.(1 - q_{t,a,s})$

• $\mathbb{E}\left[X_{i,t}.\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\}\right] = p_{t,a,s}.q_{t,a,s}$

• $\mathbb{V}\left[X_{i,t}.\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\}\right] = p_{t,a,s}.q_{t,a,s}.(1 - p_{t,a,s}.q_{t,a,s})$

Hence, according to the **law of large numbers**, we have:

$$\frac{N_{t,a,s}}{n_t} \xrightarrow[n_t \to \infty]{a.s} p_{t,a,s}$$

and

$$\frac{1}{n_t} \sum_{i=1}^{n_t} (X_{i,t}.\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\}) \xrightarrow[n_t \to \infty]{a.s} p_{t,a,s}.q_{t,a,s}$$

By applying the **first theorem of continuity** to the function $(x, y) \mapsto \frac{x}{y}$ with $y > 0$, we have:

$$\hat{q}_{t,a,s} \xrightarrow[n_t \to \infty]{a.s} q_{t,a,s}$$

We can also apply the **central limit theorem**, yielding:

$$\sqrt{n_t} \begin{pmatrix} \frac{1}{n_t} \sum_{i=1}^{n_t} (X_{i,t}.\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\}) - p_{t,a,s}.q_{t,a,s} \\ \frac{1}{n_t} \sum_{i=1}^{n_t} (\mathbb{1}\{A_{i,t} = a\}.\mathbb{1}\{S_{i,t} = s\}) - p_{t,a,s} \end{pmatrix} \xrightarrow[n_t \to \infty]{d} \mathcal{N}(0, \Sigma_{t,a,s})$$

where

$$\Sigma_{t,a,s} = \begin{pmatrix} \mathbb{V}\left[X_t.\mathbb{1}\{A_t = a\}.\mathbb{1}\{S_t = s\}\right] & c \\ c & \mathbb{V}\left[\mathbb{1}\{A_t = a\}.\mathbb{1}\{S_t = s\}\right] \end{pmatrix}$$

with

$$c = \mathrm{Cov}\left(X_t.\mathbb{1}\{A_t = a\}.\mathbb{1}\{S_t = s\}, \mathbb{1}\{A_t = a\}.\mathbb{1}\{S_t = s\}\right)$$

We can then apply the **delta method** to the following function:

$$\begin{array}{rccc} h & : & \mathbb{R} \times \mathbb{R}^* & \to & \mathbb{R} \\ & & (x, y) & \mapsto & \frac{x}{y} \end{array}$$

which is differentiable for $y > 0$ and has $\nabla h(x, y) = (\frac{1}{y}, \frac{-x}{y^2})$

Thus, after calculations, we have:

$$\sqrt{n_t}\left(\hat{q}_{t,a,s} - q_{t,a,s}\right) \xrightarrow[n_t \to \infty]{d} \mathcal{N}\left(0, \frac{q_{t,a,s}.(1 - q_{t,a,s})}{p_{t,a,s}}\right)$$

i.e.

$$\sqrt{n_t.p_{t,a,s}} \frac{\hat{q}_{t,a,s} - q_{t,a,s}}{\sqrt{q_{t,a,s}.(1 - q_{t,a,s})}} \xrightarrow[n_t \to \infty]{d} \mathcal{N}(0, 1)$$

And thus, according to the **Slutsky's theorem**, we have:

$$\sqrt{N_{t,a,s}}\frac{\hat{q}_{t,a,s}-q_{t,a,s}}{\sqrt{\hat{q}_{t,a,s}.(1-\hat{q}_{t,a,s})}} \xrightarrow[n_t\to\infty]{d} \mathcal{N}(0,1)$$

To extend to the multivariate analysis, we make the following assumption:

$$\forall\ t\in\mathcal{T},\ \frac{n_{\underline{t}}}{n_t} \xrightarrow[n_t,n_{\underline{t}}\to+\infty]{} \lambda_t\in\mathbb{R}$$

This implies that sample sizes grow at the same rate.

Similar to before, we can extend our reasoning to a vector of mortality ratios.

Indeed, according to the **law of large numbers** and the **central limit theorem**, we have:

$$\hat{q}_{\mathcal{T},a,s} \xrightarrow[n_t\to\infty]{a.s} q_{\mathcal{T},a,s}=\begin{pmatrix} q_{\underline{t},a,s}\\ \vdots \\ q_{\bar{t},a,s}\end{pmatrix}$$

and

$$\sqrt{n_{\underline{t}}}.\left(\hat{q}_{\mathcal{T},a,s}-q_{\mathcal{T},a,s}\right) \xrightarrow[n_t\to\infty]{d} \mathcal{N}(0,\Sigma_{\mathcal{T},a,s})$$

where

$$\Sigma_{\mathcal{T},a,s}=\begin{bmatrix} \frac{q_{\underline{t},a,s}.(1-q_{\underline{t},a,s}).\lambda_{\underline{t}}}{p_{t,a,s}} & & (0)\\ & \ddots & \\ (0) & & \frac{q_{\bar{t},a,s}.(1-q_{\bar{t},a,s}).\lambda_{\bar{t}}}{p_{t,a,s}}\end{bmatrix}$$

We then define: $\hat{\Sigma}_{\mathcal{T},a,s}=\begin{bmatrix} \frac{\hat{q}_{\underline{t},a,s}.(1-\hat{q}_{\underline{t},a,s}).n_{\underline{t}}}{N_{\underline{t},a,s}} & & (0)\\ & \ddots & \\ (0) & & \frac{\hat{q}_{\bar{t},a,s}.(1-\hat{q}_{\bar{t},a,s}).n_{\underline{t}}}{N_{\bar{t},a,s}}\end{bmatrix}$

Then, according to the **first theorem of continuity**:

$$\hat{\Sigma}_{\mathcal{T},a,s} \xrightarrow[n_t\to\infty]{a.s} \Sigma_{\mathcal{T},a,s}$$

Thus, according to the **Slutsky's theorem**:

$$\sqrt{n_{\underline{t}}}\frac{\hat{q}_{\mathcal{T},a,s} - q_{\mathcal{T},a,s}}{\sqrt{\hat{\Sigma}_{\mathcal{T},a,s}}} \xrightarrow[n_t \to \infty]{d} \mathcal{N}(0,1)$$

3.3 Predicted Mortality Ratios: Limit Theorems

Now that we have established the convergence of our observed ratios, we can proceed to study the convergence of our predicted ratios. These latter are functions of the observed ratios, allowing us to establish the necessary convergences for our future analyses. We recall that we have:

$$\hat{q}_{\mathcal{T},a,s} \xrightarrow[n_t \to \infty]{p.s} q_{\mathcal{T},a,s} = \begin{pmatrix} q_{\underline{t},a,s} \\ ... \\ q_{\overline{t},a,s} \end{pmatrix}$$

and

$$\sqrt{n_{\underline{t}}}.\left(\hat{q}_{\mathcal{T},a,s} - q_{\mathcal{T},a,s}\right) \xrightarrow[n_t \to \infty]{d} \mathcal{N}(0, \Sigma_{\mathcal{T},a,s})$$

where

$$\Sigma_{\mathcal{T},a,s} = \begin{bmatrix} \frac{q_{\underline{t},a,s}.(1-q_{\underline{t},a,s}).\lambda_{\underline{t}}}{p_{t,a,s}} & & (0) \\ & \ddots & \\ (0) & & \frac{q_{\overline{t},a,s}.(1-q_{\overline{t},a,s}).\lambda_{\overline{t}}}{p_{t,a,s}} \end{bmatrix}$$

Thus, leveraging the **First Continuity Theorem** and the **delta-method** applied to the function f, which is both continuous and differentiable (by assumption), we have:

$$f(\hat{q}_{\mathcal{T},a,s}) \xrightarrow[n_t \to \infty]{p.s} f(q_{\mathcal{T},a,s})$$

and

$$\sqrt{n_{\underline{t}}}\left(f(\hat{q}_{\mathcal{T},a,s}) - f(q_{\mathcal{T},a,s})\right) \xrightarrow[n_t \to \infty]{d} \mathcal{N}\left(0, \nabla f(q_{\mathcal{T},a,s})^T \Sigma_{\mathcal{T},a,s} \nabla f(q_{\mathcal{T},a,s})\right)$$

i.e.

$$\tilde{q}_{\tilde{t},a,s} \xrightarrow[n_t \to \infty]{p.s} q^P_{\tilde{t},a,s} \text{ } \textit{(convergence to a predicted ratio)}$$

and

$$\sqrt{n_{\underline{t}}}\left(\tilde{q}_{\tilde{t},a,s} - q^P_{\tilde{t},a,s}\right) \xrightarrow[n_t \to \infty]{d} \mathcal{N}\left(0, \nabla f(q_{\mathcal{T},a,s})^T \Sigma_{\mathcal{T},a,s} \nabla f(q_{\mathcal{T},a,s})\right)$$

According to the **First Continuity Theorem** and the **Slutsky's Theorem**, we have:

$$\sqrt{n_{\tilde{t}}}\frac{\tilde{q}_{\tilde{t},a,s} - q^P_{\tilde{t},a,s}}{\sqrt{\frac{1}{\lambda_{\tilde{t}}}\nabla f(\hat{q}_{\mathcal{T},a,s})^T \hat{\Sigma}_{\mathcal{T},a,s} \nabla f(\hat{q}_{\mathcal{T},a,s})}} \xrightarrow[n_t \to \infty]{d} \mathcal{N}(0,1)$$

3.4 Confidence Intervals

In this section, we will determine confidence intervals at level α for observed and predicted mortality rates during year t.

3.4.1 Observed Mortality Rates

We begin by establishing confidence intervals for our observed mortality rates. We remind that we have:

$$\sqrt{n_{\tilde{t}}}\frac{\hat{q}_{\tilde{t},a,s} - q_{\tilde{t},a,s}}{\sqrt{\hat{\Sigma}_{\tilde{t},a,s}}} \xrightarrow[n_{\tilde{t}} \to \infty]{d} \mathcal{N}(0,1)$$

Let $(b,c) \in [\![0;1]\!]^2$, we have:

$$\mathbb{P}\left(\sqrt{n_{\tilde{t}}}\frac{\hat{q}_{\tilde{t},a,s} - q_{\tilde{t},a,s}}{\sqrt{\hat{\Sigma}_{\tilde{t},a,s}}} \in [b,c]\right) \xrightarrow[n_t \to \infty]{p.s} \mathbb{P}\left(\mathcal{N}(0,1) \in [b,c]\right)$$

Hence

$$\mathbb{P}\left(\sqrt{n_{\tilde{t}}}\frac{\hat{q}_{\tilde{t},a,s} - q_{\tilde{t},a,s}}{\sqrt{\hat{\Sigma}_{\tilde{t},a,s}}} \in [-z_{1-\alpha/2}; z_{1-\alpha/2}]\right) \xrightarrow[n_t \to \infty]{p.s} 1-\alpha$$

where z_r represents the r-th quantile of a standard centered normal distribution.

Thus, we have:

$$\mathbb{P}\left(q_{\tilde{t},a,s} \in \left[\hat{q}_{\tilde{t},a,s} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\Sigma}_{\tilde{t},a,s}}{n_{\tilde{t}}}}\right]\right) \xrightarrow[n_t \to \infty]{p.s} 1-\alpha$$

We can then multiply the bounds of the confidence interval by the associated population size to determine a level α confidence interval for the number of observed deaths during year $\tilde{t}$ at given age and sex.

Now, we wish to determine a level α confidence interval for the average mortality rate

across the entire population during year $\tilde{t}$.

We then set:

$$\hat{q}_{\tilde{t}} = \sum_{a,s} \frac{N_{\tilde{t},a,s}}{n_{\tilde{t}}} \hat{q}_{\tilde{t},a,s}$$

We have:

$$\hat{q}_{\tilde{t}} \xrightarrow[n_t \to \infty]{p.s} \sum_{a,s} p_{t,a,s} q_{\tilde{t},a,s} = q_{\tilde{t}}$$

and thus, due to the independence of $(\hat{q}_{\tilde{t},a,s})_{a,s}$,

$$\mathbb{P}\left(q_{\tilde{t}} \in \left[\hat{q}_{\tilde{t}} \pm z_{1-\alpha/2} \sum_{a,s} \frac{N_{\tilde{t},a,s}}{n_{\tilde{t}}} \sqrt{\frac{\hat{\Sigma}_{\tilde{t},a,s}}{n_{\tilde{t}}}} \right] \right) \xrightarrow[n_t \to \infty]{p.s} 1 - \alpha$$

We can then multiply the bounds of the confidence interval by the associated population size to determine a level α confidence interval for the number of observed deaths during year $\tilde{t}$.

3.4.2 Predicted Mortality Ratios

Now, we can move on to constructing confidence intervals for the predicted ratios. We remind that we have:

$$\sqrt{n_{\tilde{t}}} \frac{\tilde{q}_{\tilde{t},a,s} - q^P_{\tilde{t},a,s}}{\sqrt{\frac{1}{\lambda_{\tilde{t}}} \nabla f(\hat{q}_{\mathcal{T},a,s})^T \hat{\Sigma}_{\mathcal{T},a,s} \nabla f(\hat{q}_{\mathcal{T},a,s})}} \xrightarrow[n_t \to \infty]{d} \mathcal{N}(0,1)$$

Thus, similarly to before, we have:

$$\mathbb{P}\left(q^P_{\tilde{t},a,s} \in \left[\tilde{q}_{\tilde{t},a,s} \pm z_{1-\alpha/2} \sqrt{\frac{\frac{1}{\lambda_{\tilde{t}}} \nabla f(\hat{q}_{\mathcal{T},a,s})^T \hat{\Sigma}_{\mathcal{T},a,s} \nabla f(\hat{q}_{\mathcal{T},a,s})}{n_{\tilde{t}}}} \right] \right) \xrightarrow[n_t \to \infty]{p.s} 1 - \alpha$$

We can then multiply the bounds of the confidence interval by the associated population size to determine a level α confidence interval for the number of observed deaths during year $\tilde{t}$ at given age and sex.

Similarly, following the same method as before, we have:

$$\mathbb{P}\left(q_{\tilde{t}}^P \in \left[\tilde{q}_{\tilde{t}} \pm z_{1-\alpha/2} \sum_{a,s} \frac{N_{\tilde{t},a,s}}{n_{\tilde{t}}} \sqrt{\frac{\frac{1}{\lambda_{\tilde{t}}} \nabla f(\hat{q}_{\mathcal{T},a,s})^T \hat{\Sigma}_{\mathcal{T},a,s} \nabla f(\hat{q}_{\mathcal{T},a,s})}{n_{\tilde{t}}}}\right]\right) \xrightarrow[n_t \to \infty]{p.s} 1-\alpha$$

We can then multiply the confidence interval bounds by the associated population size to determine a level α confidence interval for the number of observed deaths during year $\tilde{t}$.

3.5 Excess Mortality Tests

3.5.1 Excess Mortality Test for Given Age and Sex

In this subsection, we aim to determine if there has been a significant difference in mortality in $\tilde{t}$ for a given age and sex.

Therefore, we need to test:

$$\text{H0} : q_{\tilde{t},a,s} = q_{\tilde{t},a,s}^P \text{ against H1} : q_{\tilde{t},a,s} \neq q_{\tilde{t},a,s}^P$$

To do this, we will establish a level α confidence interval for the quantity $q_{\tilde{t},a,s} - q_{\tilde{t},a,s}^P$ to determine its significance and quantify it.

We have a consistent estimator for this purpose:

$$\hat{q}_{\tilde{t},a,s} - \tilde{q}_{\tilde{t},a,s} \xrightarrow[n_t \to \infty]{p.s} q_{\tilde{t},a,s} - q_{\tilde{t},a,s}^P$$

Furthermore, $\hat{q}_{\tilde{t},a,s}$ and $\tilde{q}_{\tilde{t},a,s}$ are asymptotically independent as we do not use the same data to determine them, and these data are independent.
In fact, $\hat{q}_{\tilde{t},a,s}$ uses data from $\tilde{t}$, while $\tilde{q}_{\tilde{t},a,s}$ uses data from previous years, and these data are assumed to be independent.

We then have:

$$\mathbb{P}\left(q_{\tilde{t},a,s} - q_{\tilde{t},a,s}^P \in \left[\hat{q}_{\tilde{t},a,s} - \tilde{q}_{\tilde{t},a,s} \pm z_{1-\alpha/2} \sqrt{\frac{\frac{1}{\lambda_{\tilde{t}}} \nabla f(\hat{q}_{\mathcal{T},a,s})^T \hat{\Sigma}_{\mathcal{T},a,s} \nabla f(\hat{q}_{\mathcal{T},a,s}) + \hat{\Sigma}_{\tilde{T},a,s}}{n_{\tilde{t}}}}\right]\right) \xrightarrow[n_t \to \infty]{p.s} 1-\alpha$$

We can then determine if the difference between $q_{\tilde{t},a,s}$ and $q^P_{\tilde{t},a,s}$ is significant at level α. To do so, we need to check if 0 belongs to the confidence interval determined earlier. If it does not, then the difference between the two is significant.

To quantify this difference, we can multiply the confidence interval bounds by the associated age and sex population size.

3.5.2 Global Excess Mortality Test

In this subsection, we aim to determine if there has been a significant difference in mortality during year $\tilde{t}$ across the entire population.

Therefore, we need to test:

$$\text{H0} : q_{\tilde{t}} = q^P_{\tilde{t}} \text{ against H1} : q_{\tilde{t}} \neq q^P_{\tilde{t}}$$

To do this, we will establish a level α confidence interval for the quantity $q_{\tilde{t}} - q^P_{\tilde{t}}$ to determine its significance and quantify it.

By reasoning in the same manner as before, we obtain:

$$\hat{q}_{\tilde{t}} - \tilde{q}_{\tilde{t}} \xrightarrow[n_t \to \infty]{p.s} q_{\tilde{t}} - q^P_{\tilde{t}}$$

and

$$\mathbb{P}\left(q_{\tilde{t}} - q^P_{\tilde{t}} \in \left[\hat{q}_{\tilde{t}} - \tilde{q}_{\tilde{t}} \pm z_{1-\alpha/2} \sum_{a,s} \frac{N_{\tilde{t},a,s}}{n_{\tilde{t}}} \sqrt{\frac{\frac{1}{\lambda_{\tilde{t}}} \nabla f(\hat{q}_{\mathcal{T},a,s})^T \hat{\Sigma}_{\mathcal{T},a,s} \nabla f(\hat{q}_{\mathcal{T},a,s}) + \hat{\Sigma}_{\tilde{T},a,s}}{n_{\tilde{t}}}} \right] \right) \xrightarrow[n_t \to \infty]{p.s} 1-\alpha$$

We can then determine if the difference between $q_{\tilde{t}}$ and $q^P_{\tilde{t}}$ is significant at level α. To do so, we need to check if 0 belongs to the confidence interval determined earlier. If it does not, then the difference between the two is significant.

To quantify this difference, we can then multiply the confidence interval bounds by the associated population size.

4 Application of the Model: Results

Now that we have established a solid theoretical framework to model our experiment, we can proceed with the practical implementation of the defined model.

4.1 Practical Implementation of the Model

All of our code files are available here. In the interest of scientific rigor, we have made each step of our implementation openly accessible and meticulously documented.

To make our predictions, we use two databases available on the INSEE website:
• the T2 dataset, which provides the mortality quotient associated with each exact age for 100,000 individuals of given sex and birth year.
• the T69QMORT dataset, which provides the mortality quotient associated with each age and sex for 100,000 individuals.

After some modifications, we obtain several matrices (represented by DataFrames) containing our $\hat{q}_{t,a,s}$.

We can then define our different functions f that we use to obtain our $\tilde{q}_{t,a,s}$:
• f_1 defines the quotients for the year 2020 as those of the year 2019. This basic approach facilitates both the application of our model and provides an initial estimate of the expected number of deaths in France in 2020. Given that mortality quotients are relatively stable from one year to another, it is common to use those from a previous year to estimate the number of deaths in the following year.
• f_2 defines the quotients for the year 2020 as the average of the quotients for 2017, 2018, and 2019. This method is adopted by the authors of our article. This choice is justified by the relatively stable mortality quotients and a recent trend of slowing growth rates. Mortality quotients for a given age and sex have increased more slowly in the last 5 years than in the early 2000s. It is therefore reasonable to consider only the past 3 years and exclude the earlier ones.

For each database and each function, we create multiple matrices containing our $\tilde{q}_{t,a,s}$. We also create associated matrices containing the bounds of confidence intervals for $\tilde{q}_{t,a,s}$.

We can then determine the predicted number of deaths and the associated confidence intervals, which we summarize in a recap table similar to the one produced by the authors of the article. As a reminder, here is the authors' final table:

Tableau 1a : Evaluation de la surmortalité en 2020 (3 années de référence : 2017-19)

Classe d'âge	Population en million d'habitants	Décès attendus	Décès observés	Décès en excès sur classe d'âge	Décès en excès par âge
0-64 ans	51,68 (79,36 %)	93 211 [91 612 ; 94 810]	92 478	(0 % [0 % ; 0 %])	- 705 (-0,76 % [-0,74 % ; -0,77 %])
65-99 ans	13,44 (20,64 %)	536 279 [528 940 ; 543 618]	561 538	17 920 (3,34 % [3,3 % ; 3,39 %])	17 013 (3,17 % [3,13 % ; 3,22 %])
Somme	65,12 (100 %)	629 490 [628 011 ; 630 970]	654 016	23 046 (3,66 % [3,65 % ; 3,67 %])	16 309 (2,59 % [2,58 % ; 2,6 %])

Figure 3: Results from the article

And here are the results we obtain:

	Population	Décès attendus	Décès observés	Différence de décès	Pourcentage de différence
(0, 64]	51821513	[93716, 100149]	92478	[-7671, -1238]	[-7.66 %, -1.32 %]
(64, 100]	13443795	[484223, 496919]	561538	[64619, 77315]	[13.0 %, 15.97 %]
Total	65265308	[577939, 597068]	654016	[56948, 76077]	[9.54 %, 13.16 %]

Figure 4: Results for the first database and the function f_2

4.2 Results and Interpretation

We can now discuss our results and compare them with those obtained by the authors of the article. All our results are available in the appendix.

In all four of our applications, we predict excess mortality in the year 2020. Specifically, we predict an additional 50,000 to 100,000 deaths compared to what was expected in 2020. We observe that the second database seems to underestimate mortality quotients, as our predictions appear low (between 529,575 and 561,188 expected deaths for function f_1, compared to over 600,000 deaths in 2019). Therefore, we will consider the results from the first database, which seem more realistic.

Indeed, the prediction made using f_2 gives us a 95% confidence interval for the expected number of deaths equal to [577,939, 597,068], which means **there was an excess mortality between 9.54% and 13.16%**, approaching the excess mortality announced by INSEE, which was 9%. Our estimation thus opposes the authors' results even more clearly than those of INSEE.

Nevertheless, it is interesting to note a strong dichotomy between those under 65 and those over 65. Indeed, according to prediction f_1, there was no excess mortality among those under 65 (as 0 is within the 95% confidence interval for the difference between expected and observed deaths). However, prediction f_2 goes even further, suggesting there was even under-mortality among those under 65, with between 1,238 and 7,671 fewer deaths than expected. Several factors could explain these results: various lockdowns, reinforcement of protective measures, improved health practices, etc. However, for those over 65, the excess

mortality is much higher **(over 10%)**, leading to an overall increase in excess mortality. Thus, there seems to be a consensus on this point: whether it's the authors, INSEE, or us, we all observe this dichotomy based on age. This doesn't just mean that elderly people are more severely affected by infectious diseases like Covid-19 (which is obvious), it means that Covid-19 did not lead to excess mortality in young people.

In summary, when we compare our results obtained with the first database and the function f_2 (to follow the method used in the article) with the results of the article authors, we observe clear differences. While our results are rather similar for those under 65: there was no excess mortality among those under 65 in 2020, our results differ significantly for those over 65. Like the authors of the article, we predict excess mortality among those over 65, but our excess mortality is much higher **(over 13% for us compared to 3.17% in the article, nearly 5 times more)**. This leads to a difference in overall excess mortality, again much higher according to us **(over 9% according to us compared to 2.59% according to the article).**

Thus, we are compelled to refute the conclusions put forward by the authors of the article. No, Covid-19 was not merely a seasonal flu episode, as it caused significant excess mortality, especially among individuals over 65. In this regard, we align and support the perspective of INSEE.

5 Limitations of the Model and its Implementation

In this section, we consider potential limitations and areas for improvement that we could identify concerning the statistical model and its implementation.

5.1 Limitations of the Statistical Model

First and foremost, we assumed that individuals were all independent and that the death of one individual had no impact on the probability of death for another individual. Furthermore, we also assumed the independence of different mortality quotients across years. However, this seems hardly conceivable in a real-world context. The "harvesting effects" provide a good example. When a group of people experiences a significant number of deaths in one year, they typically experience fewer deaths in the following year (note that the authors mention the harvesting effect in their introduction but do not indicate implementing it in their model). This is especially true for the elderly and flu epidemics, which tend to be more severe every two to three years. Addressing these assumptions would lead to a significant complexity in the model, particularly for the variance/covariance matrices

(Σ), which would no longer be diagonal matrices, as covariances would no longer be zero.

We also assumed that population counts grow at the same rate and converge to $\lambda_t \in \mathbb{R}$ as we approach infinity, which is not very restrictive.

5.2 Application Limitations

Firstly, the databases we were able to use are not always ideal. Indeed, the second database significantly underestimated mortality quotients from previous years, which prevented us from obtaining accurate results. Additionally, we would have liked to obtain geographical data to study excess mortality based on geographical area. However, we did not find a database with this kind of information.

We would have also liked to have other functions f to predict mortality quotients for 2020. It may be possible to implement linear regression on mortality quotients for the last 10 years to estimate our mortality quotients for 2020. This is a new trend in mortality quotient estimation, and we would have liked to have the time to implement this kind of method.

To replicate what was done in the article, we restricted ourselves to Metropolitan France and individuals aged from 0 to 99 (unlike INSEE, which considers the entire French population). Thus, any comparison of our results with those of INSEE carries a slight bias. Nevertheless, we can reasonably think that these differences would marginally alter the orders of magnitude of our results.

6 Annexes

6.1 Our Results

Tableau 1a : Evaluation de la surmortalité en 2020 (3 années de référence : 2017-19)

Classe d'âge	Population en million d'habitants	Décès attendus	Décès observés	Décès en excès sur classe d'âge	Décès en excès par âge
0-64 ans	51,68 (79,36 %)	93 211 [91 612 ; 94 810]	92 478	(0 % [0 % ; 0 %])	- 705 (-0,76 % [-0,74 % ; -0,77 %])
65-99 ans	13,44 (20,64 %)	536 279 [528 940 ; 543 618]	561 538	17 920 (3,34 % [3,3 % ; 3,39 %])	17 013 (3,17 % [3,13 % ; 3,22 %])
Somme	65,12 (100 %)	629 490 [628 011 ; 630 970]	654 016	23 046 (3,66 % [3,65 % ; 3,67 %])	16 309 (2,59 % [2,58 % ; 2,6 %])

Figure 1: Article Results

	Population	Décès attendus	Décès observés	Différence de décès	Pourcentage de différence
(0, 64]	51821513	[93716, 100149]	92478	[-7671, -1238]	[-7.66 %, -1.32 %]
(64, 100]	13443795	[484223, 496919]	561538	[64619, 77315]	[13.0 %, 15.97 %]
Total	65265308	[577939, 597068]	654016	[56948, 76077]	[9.54 %, 13.16 %]

Figure 2: Results for the first database and the function f_2

	Population	Décès attendus	Décès observés	Différence de décès	Pourcentage de différence
(0, 64]	51821513	[90616, 101739]	92478	[-9261, 1862]	[-9.1 %, 2.05 %]
(64, 100]	13443795	[488035, 509574]	561538	[51964, 73503]	[10.2 %, 15.06 %]
Total	65265308	[578651, 611313]	654016	[42703, 75365]	[6.99 %, 13.02 %]

Figure 3: Results for the first database and the function f_1

	Population	Décès attendus	Décès observés	Différence de décès	Pourcentage de différence
(0, 64]	51821513	[90312, 96641]	92478	[-4163, 2166]	[-4.31 %, 2.4 %]
(64, 100]	13443795	[454539, 466901]	561538	[94637, 106999]	[20.27 %, 23.54 %]
Total	65265308	[544851, 563542]	654016	[90474, 109165]	[16.05 %, 20.04 %]

Figure 4: Results for the second database and the function f_2

	Population	Décès attendus	Décès observés	Différence de décès	Pourcentage de différence
(0, 64]	51821513	[86210, 97156]	92478	[-4678, 6268]	[-4.81 %, 7.27 %]
(64, 100]	13443795	[443365, 464032]	561538	[97506, 118173]	[21.01 %, 26.65 %]
Total	65265308	[529575, 561188]	654016	[92828, 124441]	[16.54 %, 23.5 %]

Figure 5: Results for the second database and the function f_1

6.2 Definitions

Here are some important definitions to better understand the article by our authors and the topics addressed in this project.

Excess mortality: the difference between the observed number of deaths during a significant health event and the expected number of deaths for the same period.

Mortality rate: for a given group and period, it's the ratio of the number of individuals in the group who died during the period to the total population of that group during the period.

Expected number of deaths: estimated by the number of living individuals at the start of the year multiplied by the expected mortality rate. It's calculated using a statistical model capable of predicting this expected number.

Comorbidity: the coexistence of two mental or physical diseases without established causality, which in our case, prevents determining the true cause of a patient's death.

Harvesting effect: during a significant health event, older and vulnerable people die in a targeted manner over a relatively short period. This leads to a temporary depletion of this population category, resulting in under-mortality in the following years (alternation between highly deadly and less deadly flu seasons). Therefore, these "respite" years should not be used as a reference, which is what INSEE does.

Confidence interval: any observed value within the confidence interval of the predicted value is considered expected, regardless of observed differences. Conversely, the excess or deficiency of the studied variable is the sum of values beyond the bounds defined by the confidence interval around the predicted value.

6.3 Key Figures

Here are all the key figures to keep in mind to better understand the magnitudes and current demographic dynamics:

Increase in the number of deaths: Since 2004, the number of deaths has been increasing on average by 2% per year.

Comorbidity and Covid-19: 90% of people admitted to intensive care and 65% of people who died from Covid-19 had at least one comorbidity.

Median age at Covid-19 death: The median age at hospital death for Covid-19 patients is 85 years, and nearly 92.5% of people were aged 65 and over.

Population and deaths in 2020: In 2020, those aged over 65 accounted for 19.45% of the population but 84.3% of the deaths. Conversely, those under 65 represented 80.5% of the population but only 15.7% of the deaths.

Average annual mortality rate: 1.004% in 2020; 0.922% in 2019; 0.920% in 2018; 0.918% in 2017. In 2020, this translated to 0.17% for those under 65, 4.18% for those over 65, and 35% for those over 95.

Metropolitan population in 2020: 65,123,843 individuals.

Deaths in 2020: There were 654,016 deaths, including 64,632 deaths attributed to Covid-19.

Excess mortality due to Covid-19: Covid-19 is estimated to have caused an excess mortality ranging from 2.59% to 3.66% in 2020.

6.4 Literature Review

Link 1: How many deaths would there have been in France without the Covid-19 epidemic?

Natalie Blanpain, INSEE:

The number of deaths in France generally increases each year due to population growth and aging, but the risks of dying at an older age are decreasing. INSEE has published an assessment of the increase in deaths in 2020 caused by the Covid-19 epidemic (+9% compared to 2019). The impact of this epidemic on mortality can be evaluated by comparing the expected deaths if the epidemic had not occurred. This article proposes a method to calculate this expected number of deaths in 2020.

People aged 65 and over accounted for 84% of deaths in 2019, while their share in the population is 4 times lower. However, mortality rates have generally decreased over the years due to progress in various areas, which means people are living longer. The Covid-19 epidemic in 2020 interrupted this downward trend in mortality rates.

Population aging tends to increase the number of deaths, while declining mortality rates tend to reduce them. Between 2010 and 2019, the first trend dominated the second, leading to an increase in deaths. Methods to predict deaths in 2020 using only previous mortality rates alone may underestimate the expected deaths due to the downward trend in mortality rates and increased life expectancy.

The goal of the text is to estimate the number of deaths that would have occurred in 2020 without the Covid-19 epidemic. To do this, demographers assume that mortality rates would continue to evolve at the same pace as in recent years. Based on the 2010-2019 period (+0.8 months increase in life expectancy per year), we would have expected 621,900 deaths, 8,700 more than in 2019. This increase comes from population aging (+13,800 deaths) and the extra day from the leap year 2020 (+1,900 deaths), tempered by declining mortality rates (-7,000 deaths). However, in 2020, 669,000 deaths were actually recorded, 47,100 more than expected.

However, the number of deaths attributed to COVID-19 is higher than the impact of the health crisis on total deaths, as it includes deaths of vulnerable people who would have died in 2020 even without the epidemic, but were anticipated due to the epidemic. The impact of the health crisis, which measures the gap between expected and observed deaths, does not take into account these anticipated deaths, but it is reduced by avoided deaths. Official definitive statistics on medical causes of death are needed to clarify the comparison between the number of deaths attributed to COVID-19 and the impact of the health crisis on total deaths.

Link 2: 2020: An Unprecedented Increase in Deaths in 70 Years

Sylvie Le Minez, Valérie Roux, INSEE:

This article reports that the two successive waves of Covid-19 in 2020 did not have the same duration or intensity in different regions in mainland France and overseas. The

first wave occurred in spring 2020, with a 27% increase in deaths (+27,300 deaths) in the March/April period compared to 2019. The peak of deaths during this period was reached on April 1, 2020, with 2,810 deaths in one day, and it decreased in April. Between May and August 2020, the number of deaths returned to the average of previous years. The second wave, which lasted from September to December 2020, was less intense but lasted longer (+34,300 deaths), an increase of 17% compared to the same period in 2019. The peak of deaths was reached on November 7 with 2,340 deaths in one day.

Furthermore, the two successive waves had a different impact in different regions. The first wave was more deadly in Île-de-France and the Grand Est than anywhere else in France. The first wave began in Hauts-de-France in March 2020 with a peak of +61% in Oise. In the Grand Est, between March and April, deaths were 55% higher than in 2019. In Bourgogne Franche-Comté, the increase in mortality was half that of the Grand Est in spring and equivalent to that of Hauts-de-France. Finally, after a slow start, from March 16, Île-de-France recorded a much larger increase in mortality than elsewhere, with a 91% excess mortality between mid-March and April.

During the autumn of 2020, the second wave of Covid-19 was stronger in regions that had been less affected in the spring, especially Auvergne-Rhône-Alpes (+38%), where the excess was less than that of Île-de-France but lasted longer. Other central and southern regions of France, such as Bourgogne-Franche-Comté and Provence-Alpes-Côte d'Azur, were also particularly affected by this second wave.

In the overseas territories, excess mortality varies greatly between departments. For example, in Mayotte, the excess deaths were 25% throughout 2020 compared to 2019, while they were zero for Martinique and Réunion, and even negative for French Guiana.

In conclusion, the article states that in 2020, the number of deaths increased in most European countries. France is in a median position with a 9% increase in deaths. France's neighboring countries saw a larger increase in deaths, except for Germany and Luxembourg. Seven European countries recorded an increase in deaths equal to or greater than 14%: Spain, Poland, Belgium, Slovenia, Bulgaria, the Czech Republic, and Italy. In spring, during the first wave of the Covid-19 pandemic, mortality increased in only seven countries in Europe, but with the second wave in the autumn, the health crisis spread more widely.

References

[1] Laurent Toubiana, Laurent Mucchielli, Pierre Chaillot, Jacques Bouaud (2022). The Covid-19 epidemic had a relatively small impact on mortality in France: https://docplayer.fr/216724313-L-epidemie-de-covid-19-a-eu-un-impact-relativement-faible-sur-la-mortalite-e html?fbclid=IwAR2L7bhpnr2USVWq7QXPuHRi1iiisuweZcoUM5o7sEoeeQNRw99ls73dNog

[2] Nathalie Blanpain, INSEE (2022). How many deaths would there have been in France without the Covid-19 epidemic? https://blog.insee.fr/combien-y-aurait-il-eu-de-deces-en-france-sans-lepidemie-de-covid-19/

[3] Sylvie Le Minez, Valérie Roux, INSEE (2022). 2020: An unprecedented increase in deaths in 70 years: https://www.insee.fr/fr/statistiques/5347349?sommaire=5413896

[4] WHO (2021). COVID-19 – WHO Timeline: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19

[5] AFP (2012). Flu saturates emergency services: https://www.lexpress.fr/actualite/societe/sante/la-grippe-sature-les-urgences_1086604.html