

16

Chiffres à l'appui: objectivité dans l'appréciation du COVID et incohérence de l'étude ComCor

Vincent PAVAN

02/03/2021

Plan de la présentation

- 2020: comment organiser la comparaison de la mortalité avec les années précédentes ?
 - méthode
 - exploitation reproductible des données de l'INSEE
 - présentation des résultats
- L'étude ComCor par Fontanet & al: incohérence et suspicion méthodologique quant aux prétendus facteurs de risques liés à la fréquentation des bars et des restaurants.
 - présentation de la méthode ComCor
 - notion de facteur de risques (OR et CI)
 - incohérence des résultats
 - suspicion méthodologique

Mortalité en 2020: comment la comparer aux années précédentes.

Question

Comment serait-on mort en 2020 si on avait appliqué les taux de mortalité des années précédentes ?

- Pour chaque année $Y = k, k \in [1962, 2020]$ et pour chaque tranche d'âge $A = \theta, \theta \in [0, 100+]$ on définit le taux de mortalité par classe d'âge

$$\psi(\theta, k) = \frac{M(\theta, k)}{P(\theta, k)}$$

avec $M(\theta, k)$ les personnes de l'année k qui sont mortes à l'âge θ et P l'ensemble des personnes qui ont l'âge θ l'année k

- Ensuite, on recalcule la mortalité de 2020 avec le taux de mortalité des années précédentes

$$M_a(\theta, k) = \psi(\theta, k) * P(\theta, a), \quad a = 2020$$

fichiers utilisés

Pour faire ces calculs, on utilise uniquement les fichiers de l'INSEE (pyramide des âges et mortalité), en particuliers les derniers fichiers mis à jours rapportant la mortalité de l'année 2020.

- Pour les nombres $M(\theta, k)$, les fichiers utilisés sont les suivants
 - Fichier INSEE: Séries longues sur les décès - Tableaux de séries longues sur les décès - T74 (de 1962 à 2019): lien = (<https://www.insee.fr/fr/statistiques/4771989?sommaire=4772633>)
 - Fichier INSEE : Décès quotidiens en 2020 par âge, sexe et lieu : lien = (https://www.insee.fr/fr/statistiques/fichier/4487988/2021-02-05_deces_sexe_age_lieu.zip).
- Pour les nombres $P(\theta, k)$ les sont les suivants:
 - de 1962 à 2017 inclus : Fichier INSEE lien = (<https://www.insee.fr/fr/statistiques/4503164?sommaire=4503178>)
 - de 2018 à 2021 : Fichier INSEE: lien = (<https://www.insee.fr/fr/statistiques/5007688>)

Programmes et exploitation des fichiers

Reproductibilité

Les programmes sont codés en PYTHON et sont accessibles et reproductibles via une fichier .Rmd : tout est transparent et exécutable par n'importe qui, qui peut également vérifier l'ensemble des lignes et code et la manière avec laquelle sont tracés les graphiques.

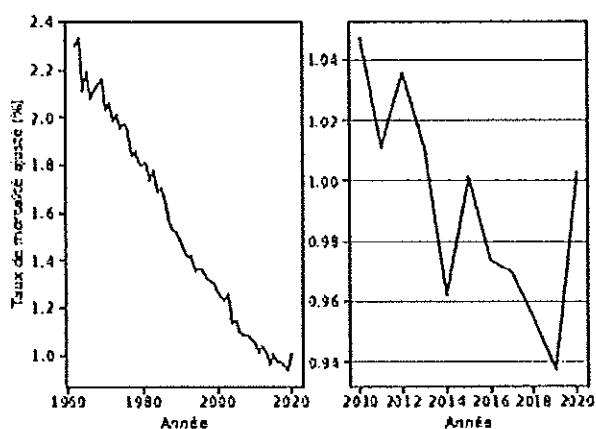
```
def extraire_age_data_structure(fil_name):
    # lecture de la feuille v/s contenu de fichier
    df = pandas.read_excel(fil_name, 'PA - 0', engine='openpyxl')
    for leg in df.legs():
        df[leg] = pandas.to_numeric(df[leg], errors='coerce')
    data = df.values
    age_data_structure = []
    # Parcours des années
    for i in range(0, len(data)):
        head = data[i][0]
        vector_PA = []
        if not pandas.isna(head):
            for j in range(2, len(data[i])):
                value = data[i][j]
                vector_PA.append(value)
            age_data_structure.append(data[i])
    return age_data_structure

# fonction principale d'extraction des données des âges des décès par
# génération
def create_age_data_structure():
    fil_name = path_file_data + '1992_2018'
    age_data_structure = extraire_age_data_structure(fil_name)
    create_data_structure('Age des décès par génération
de 1992 à 2018', 'Age par génération', 'total', 'total', 'total', 'total',
1), age_data_structure)
print('Info : Extraction des âges des décès par génération de 1992 à 2018
extraits et stockés dans le fichier final')
return age_data_structure

age_data_structure = create_age_data_structure()

# INFO : Typhoides des âges des décès par génération de 1992 à 2018
extraits et stockés dans le fichier final
```

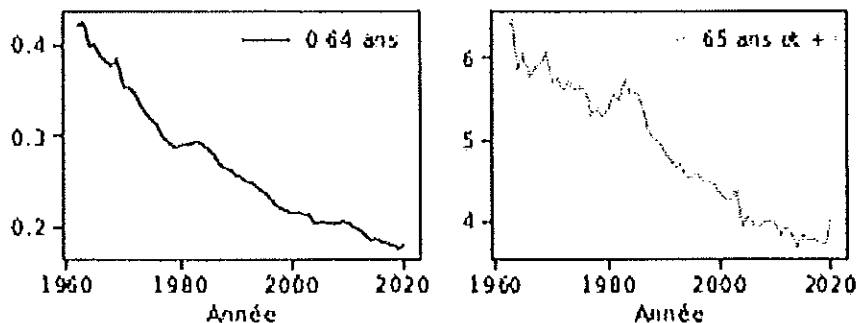
Les résultats (1)



Comme on peut le voir:

- sur les temps longs on meurt de moins en moins (effet structurel)
- sur les temps courts, on voit qu'il y a légèrement moins de mortalité qu'en 2020, mais on voit que l'année 2015 est totalement identique à l'année 2020 en terme de mortalité.
- eu égard aux faibles mortalités des années précédant 2020 (en particulier 2019), on a manifestement également un effet de moisson.

Les résultats (2)



Lorsque l'on compare avec la méthode, la mortalité par groupe d'âge (baby boom > 65 et non baby boom < 65) on constate que

- pour les moins de 65 ans, 2020 est une année de mortalité historiquement basse.
- pour les plus de 65 ans, la mortalité est revenue au niveau des années 2000-2010. Malgré l'amélioration structurelle des soins, la population a vieilli et on commence à atteindre la limite de l'espérance de vie.

Conclusions

L'année 2020 n'est donc ni particulièrement anormale, ni particulièrement significative du point de vue de la mortalité générale. Dire qu'il y aurait un danger particulier d'avoir vécu en 2020 à cause du COVID est donc sans fondement statistique.

L'étude ComCor: l'interview de Jean Castex

Extrait de l'interview de Jean Castex par Sonia Mabrouk le 15 décembre 2020.

- **Sonia Mabrouk:** Sur la crise sanitaire, on en vient, Monsieur le Premier Ministre, aux fondements mêmes de vos arbitrages face à l'épidémie. Pourquoi est-ce qu'il n'y a pas une grande étude scientifique, rigoureuse française sur les principaux lieux de contamination dans notre pays ?
- **Jean Castex:** D'abord on dispose d'études scientifiques internationales
- **Sonia Mabrouk:** Internationales
- **Jean Castex:** Oui mais enfin le virus est à peu près le même partout. Ces études sont engagées, l'une va être publiée je crois cette semaine sous l'autorité du Pr. Fontanet, qui confirme d'ailleurs les éléments issus des études internationales. Donc on va disposer d'une étude nationale.
- **Sonia Mabrouk:** Mais Monsieur le Premier Ministre, sur la grande étude, pour faire accepter des mesures, il faut qu'elles soient comprises, si vous aviez publié cette grande étude, ça fait un an que le virus circule, pourquoi il n'y a pas cette étude ?

L'interview de Jean Castex (suite)

- **Jean Castex:** Madame, encore une fois, nous avons les éléments scientifiques: le virus n'est pas différents en Allemagne aux Etats-Unis (...)
- **Sonia Mabrouk:** il y a des modes de vie différents
- **Jean Castex:** les lieux de contaminations sont connus. Et vous verrez d'ailleurs... Je dispose des pré-éléments de l'étude du Pr. Fontanet, elle confirme, elle confirme la littérature scientifique.
- **Sonia Mabrouk:** Mais est-ce qu'elle confirme vos arbitrages ? Est-ce que cette étude confirme vos arbitrages et vos choix pour la fermeture ?
- **Jean Castex:** Oui. Oui Madame. Oui Bien sûr.
- **Sonia Mabrouk:** Donc elle indique que les principaux lieux de contamination...
- **Jean Castex:** sont des lieux de contamination privés (...) les restaurants et les cafés - évidemment - c'est-à-dire en fait, pardonner moi il peut arriver que la science rejoigne le bon sens, c'est-à-dire les lieux où vous ne pouvez pas garder votre masque, les lieux qui sont plutôt clos couverts, les lieux où vous restez un temps assez long ensemble, des lieux où les interactions sont plus fortes qu'ailleurs, donc vous avez les salles de sports, vous avez les discothèques (...)
- :

L'interview de Jean Castex (suite)

- **Jean Castex:** bon alors s'agissant des restaurateurs, là aussi je vous ferai la même réponse que pour le secteur de la culture, c'est un crève-coeur pour moi madame de fermer les restaurants. C'est un crève coeur mais c'est nécessaire et l'étude du professeur Fontanet le confirmera.

ComCor

C'est bien "l'étude" de Fontanet et al. qui permettra au premier ministre de justifier la fermeture des restaurants et des bars. Cette étude pose problème

En fait ce n'est pas une étude mais un sondage qui compare: une cohorte de "Malades" et une cohorte de "Témoins" et leurs habitudes. Les "Malades et les Témoins doivent être semblables sur des critères (âges, sexe, géographie) et différents sur d'autres critères (restaurants, transports en communs etc.). Les différences entres malades et témoins sont interprétées en terme de risque, de protection ou de neutralité.

Le principe de l'étude ComCor

Le principe de l'étude est un principe classique dans le domaine des risques qui consiste à tester ce qu'il est convenu un facteur de risque (Odd Ratio en anglais) et à calculer ensuite un intervalle de confiance (IC)

- On choisit un échantillon de personnes ayant le cancer du poumon: la cohorte des cas.
- on choisit un échantillon de personnes n'ayant pas le cancer: c'est la cohorte des témoins.
- On choisit un critère d'exposition (par exemple fumer la cigarette). On sépare alors, pour chaque cohorte ceux qui sont exposés (ils fument) de ceux qui ne sont pas exposés (ils ne fument pas)

	<i>Cas(cancer)</i>	<i>Témoins(sains)</i>
<i>Exposés (fumeurs)</i>	$a = 10$	$b = 3$
<i>Non exposés (non fumeurs)</i>	$c = 2$	$d = 9$

Calcul de l'OR et du CI

On définit alors le risque d'exposition (OR) par la formule

$$OR = \frac{ad}{bc} = \frac{10 * 9}{2 * 3} = 15$$

Auquel on ajoute un intervalle de confiance (CI)

	x
estimate	15.000000
lower	2.023857
upper	111.173880

On a deux possibilités:

- Si 1 est dans l'intervalle de confiance, on exclut le fait que l'exposition soit un facteur de risque
- Si 1 n'est pas dans l'intervalle de confiance, on conclut que l'exposition est un facteur de risque (si $OR > 1$) ou de protection (si $OR < 1$)

Le principe

Les chiffres qui posent problème sont ceux fournis à la page 25 du document de Fontanet. Nous les reproduisons ici:
COMCOR - Comparaison cas témoins – Réunions, lieux visités et loisirs

Caractéristiques	Cas (n=3426)	Témoins (n=1713)	OR univarié	OR multivarié
Avez-vous fréquenté des restaurants ou bars ?				
Non	2516 (73.4)	1444 (84.3)	1 (réf)	1 (réf)
Oui	910 (26.6)	269 (15.7)	2.00 (1.85-2.18)	1.91 (1.72-2.11)
Avez-vous fréquenté des bars ?				
Non	2975 (86.8)	1644 (96.0)	1 (réf)	
Oui	219 (6.4)	69 (4.0)	1.75 (1.52-2.03)	
Données manquantes	232 (6.8)	0 (0.0)	empty	
Avez-vous fréquenté des restaurants ?				
Non	2452 (71.6)	1460 (85.2)	1 (réf)	
Oui	742 (21.7)	253 (14.8)	1.90 (1.74-2.08)	
Données manquantes	232 (6.8)	0 (0.0)	empty	
Avez-vous fréquenté des discothèques, des soirées, etc. ?				
Non	3344 (97.6)	1688 (98.5)	1 (réf)	1 (réf)
Oui	82 (2.4)	25 (1.5)	1.47 (1.16-1.82)	1.12 (0.82-1.46)
Avez-vous fréquenté des commerces (achats alimentaires, prêt-à-porter...) ?				
Non	658 (19.2)	308 (18.0)	1 (réf)	1 (réf)
Oui	2768 (80.8)	1405 (82.0)	0.92 (0.84-1.01)	0.84 (0.76-0.94)

Un premier problème (1)

Il existe cependant un problème de taille: nous allons voir en fait que si la fréquentation **des bars OU des restaurants** (i.e. l'un ou l'autre ou les deux), **des bars seuls, des restaurants seuls** constitue selon les chiffres un facteur de risque dans la transmission du virus, **en revanche la fréquentation des bars ET des restaurants (i.e. les deux obligatoirement) NE REPRESENTE QUANT A ELLE AUCUN FACTEUR DE RISQUE PARTICULIER DANS LA TRANSMISSION DU VIRUS.**

Principe mathématiques pour la reconstruction du "ET"

$$\begin{aligned} \text{Card}(A \cup B) &= \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cap B) \\ \Leftrightarrow \text{Card}(A \cap B) &= \text{Card}(A) + \text{Card}(B) - \text{Card}(A \cup B) \end{aligned}$$

Avec A les personnes qui fréquentent les bars et B celles qui fréquentent les restaurants

Reconstruction des personnes allant dans les bars et dans les restaurants

Avez-vous fréquenté des bars OU des restaurants (Q1)	Témoins	Cas
NON	1444	2516
OUI	269	910

Avez-vous fréquenté des bars (Q2)	Témoins	Cas
NON	1644	2975
OUI	69	219
Données manquantes	0	232

Avez-vous fréquenté des restaurants (Q3)	Témoins	Cas
NON	1460	2452
OUI	253	742
Données manquantes	0	232

REconstruction de la fréquentation de bars et des restaurants

```
Q4 <- matrix(nrow = 2, ncol = 2)
Q4[2,1] = Q2m[2,1] + Q3m[2,1] - Q1[2,1]
Q4[2,2] = Q2m[2,2] + Q3m[2,2] - Q1[2,2]
pop1 = sum(Q2m[1:2,1])
pop2 = sum(Q2m[1:2,2])
Q4[1,1] = pop1 - Q4[2,1]
Q4[1,2] = pop2 - Q4[2,2]
Q4Tbl = data.frame(c("NON", "OUI"),Q4)
colnames(Q4Tbl) <- c("Avez-vous fréquenté les bars ET les restaurant", "Tém
knitr::kable(Q4Tbl)
```

Avez-vous fréquenté les bars ET les restaurant	Témoins	Cas
NON	1660	3305
OUI	53	121

CONCLUSION

on obtient la mesure du OR avec son intervalle de confiance à 95%

	x
estimate	1.1466903
lower	0.8258817
upper	1.5921149

CONCLUSION

Comme on peut le voir, l'intervalle de confiance à 95% de ce facteur de risque contient 1: la conclusion s'impose: si le sondage permet de prévoir un facteur de risque associé à la fréquentation des bars OU des restaurants, en réalité il permet aussi de dire qu'il n'y a aucun risque à fréquenter les bars ET les restaurants. Il s'agit évidemment d'une incohérence MAJEURE qui permet, à elle seule de montrer que le sondage ne permet absolument pas de conclure sur l'augmentation supposé d'un facteur de risque dans la fréquentation des bars et des restaurants.

Bizzarerie

- des biais inquiétants !

COMCOR - Comparaison cas témoins - Caractéristiques socio-démographiques

Caractéristiques	Cas (n=3426)	Témoins (n=1713)	OR univarié	OR multivarié
Age				
18-28 ans	547 (16.0)	269 (15.7)	1 (réf)	1 (réf)
29-38 ans	768 (22.4)	194 (11.3)	1.94 (1.80 - 2.09)	1.61 (1.41-1.79)
39-48 ans	896 (26.1)	499 (29.1)	0.87 (0.83 - 0.92)	0.71 (0.64-0.79)
49-58 ans	735 (21.4)	428 (25.0)	0.83 (0.78 - 0.89)	0.86 (0.78-0.95)
59-68 ans	352 (10.3)	128 (7.5)	1.31 (1.21 - 1.42)	2.23 (1.91-2.60)
69 ans et plus	129 (3.8)	195 (11.4)	0.31 (0.26 - 0.36)	0.70 (0.56-0.90)

- des données manquantes de façon incompréhensible !

COMCOR - Comparaison cas témoins - Réunions, lieux visités et loisirs

Caractéristiques	Cas (n=3426)	Témoins (n=1713)	OR univarié	OR multivarié
Avez-vous fréquenté des restaurants ou bars ?				
Non	2516 (73.4)	1444 (84.3)	1 (réf)	1 (réf)
Oui	910 (26.6)	269 (15.7)	2.00 (1.85-2.18)	1.91 (1.72-2.11)
Avez-vous fréquenté des bars ?				
Non	2975 (86.8)	1644 (96.0)	1 (réf)	
Oui	219 (6.4)	69 (4.0)	1.75 (1.52-2.03)	
Données manquantes	232 (6.8)	0 (0.0)	empty	
Avez-vous fréquenté des restaurants ?				
Non	2452 (71.6)	1460 (85.2)	1 (réf)	
Oui	742 (21.7)	253 (14.8)	1.90 (1.74-2.08)	
Données manquantes	232 (6.8)	0 (0.0)	empty	
Avez-vous fréquenté des discothèques, des soirées, etc. ?				
Non	3344 (97.6)	1688 (98.5)	1 (réf)	1 (réf)
Oui	82 (2.4)	25 (1.5)	1.47 (1.16-1.82)	1.12 (0.82-1.45)
Avez-vous fréquenté des commerces (achats alimentaires, prêt-à-porter...)?				
Non	658 (19.2)	309 (18.0)	1 (réf)	1 (réf)
Oui	2768 (80.8)	1405 (82.0)	0.92 (0.84-1.03)	0.84 (0.76-0.94)