



## Mortalité – Létalité – Visibilité par tranches d'âge

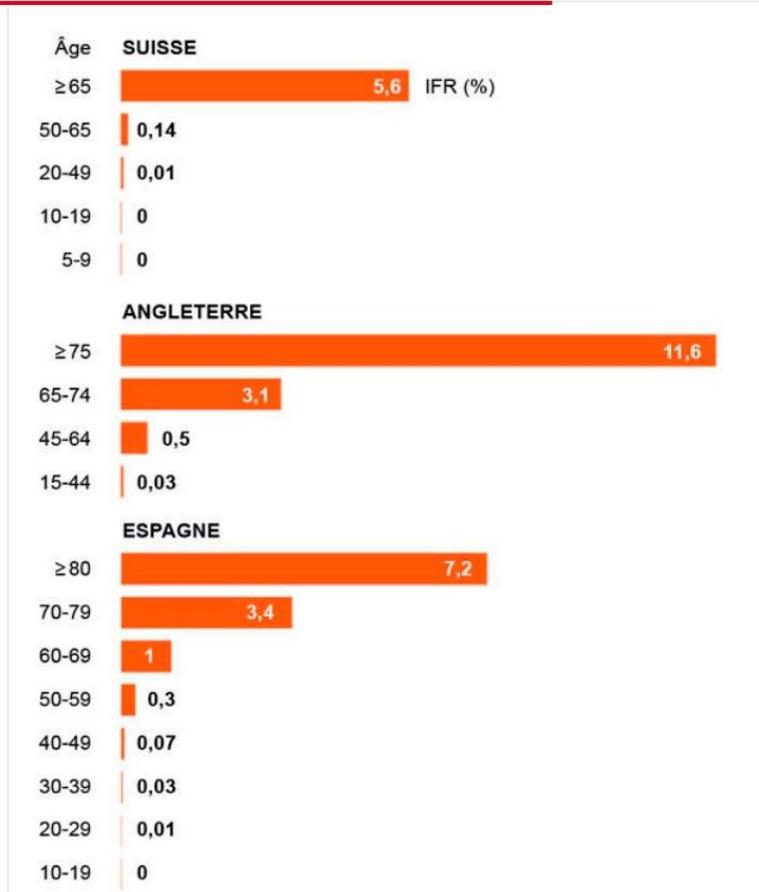
### Identification de la tranche « invisible »

<https://corona-circule.github.io/lettres/>

Nous poursuivons notre approche vers un modèle avec sous-populations (ou « compartiments »). Les données disponibles sont sur le site <https://www.data.gouv.fr/fr/organizations/sante-publique-france/>, sous forme de volumineux fichiers qui présentent leur évolution, jour après jour, selon les différentes tranches d'âge et les régions. Comme précédemment nous suivons en priorité les données de décès (tableau données hospitalières\_ tranches d'âge) et nous commençons ici par le suivi de la mortalité par tranche d'âge.

L'importance des différentes tranches est facilement extraite des statistiques de l'INSEE au 01/01/2020, indiquées sur le tableau ci-dessous. Nous y avons ajouté les valeurs de létalité (probabilité d'issue fatale pour une personne contaminée) dont l'estimation sera détaillée à la page suivante :

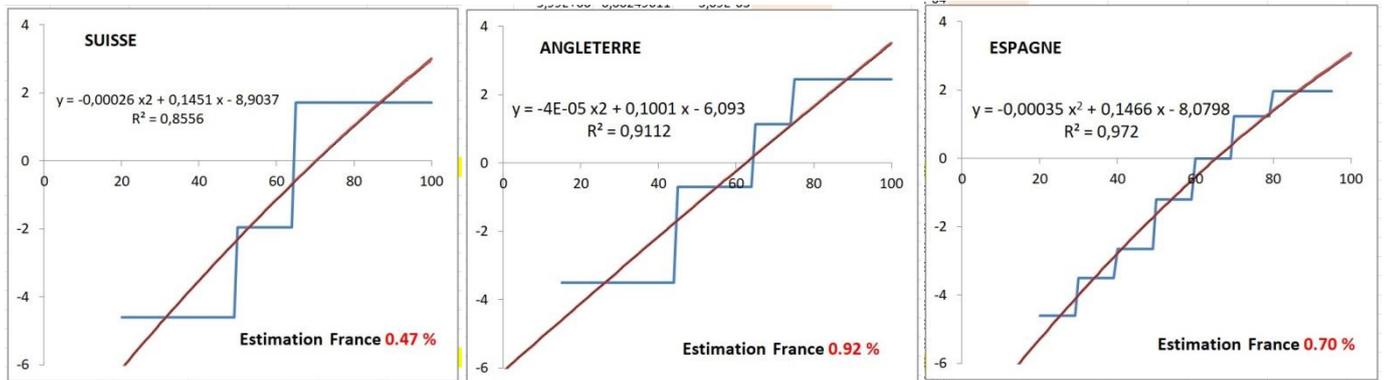
Age (ans)	0 - 9	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 - 79	80 - 89	> 90
Million habitants	7.76	8.33	7.47	8.28	8.58	8.79	8.00	5.69	3.22	0.91
Létalité (%)	3.9 E-4	1.53 E-3	5.80 E-3	2.08 E-2	7.00 E-2	0.223	0.669	1.90	5.10	13.0



**Estimation de la létalité.** Le site de Pour La Science a récemment publié une étude de la létalité dans trois pays européens. Nous la supposons, en première approche, applicable en moyenne à la France.

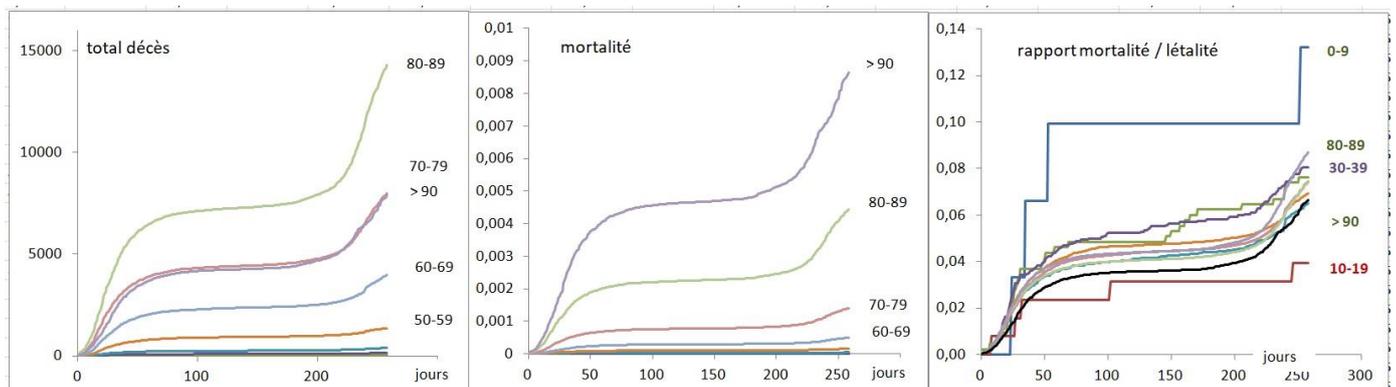
Pour chacun de ces trois pays nous avons tracé les données de la figure ci-contre, en échelle logarithmique et ajusté une loi empirique sur la dépendance du taux de létalité en fonction de l'âge.

Puis nous avons calculé la valeur moyenne de la létalité sur chaque tranche d'âge. Et nous avons finalement exprimé, en tenant compte de la pyramide des âges, la valeur moyenne qui en résulterait pour la France. Cette valeur d'estimation est indiquée en rouge sur chacune des figures.



Les données de l'Espagne sont les plus détaillées. Celles de la Suisse, quoique moins détaillées, fournissent une loi très proche de celle donnée par l'Espagne, et conduisent à une estimation de la moyenne en France proche de la valeur retenue jusqu'ici 0.53 %. La loi résultant des données de l'Angleterre a été écartée. Et nous avons pris choisi une loi intermédiaire entre celles données par la Suisse et l'Espagne en pondérant les coefficients de chaque loi empirique par les facteurs 0.73 et 0.27 respectivement. Cette dernière loi donne exactement, pour la France en moyenne, la valeur « fétiche » 0.53 %.

Cette connaissance des divers taux de létalité permet d'analyser les données de décès attribués à la pandémie dans chaque tranche d'âge. Nous avons d'abord reporté ces données, à diverses échelles, sur la figure ci-dessous :



Le total des décès de Mars à Novembre (figure de gauche) est très faible dans les premières tranches : 4 décès dans la tranche 0-9 ans, 5 dans la tranche 10-19 ans. Il croit régulièrement en fonction de l'âge, sauf pour la dernière tranche (> 90 ans). Cela s'explique par la population réduite de cette dernière tranche. En effet :

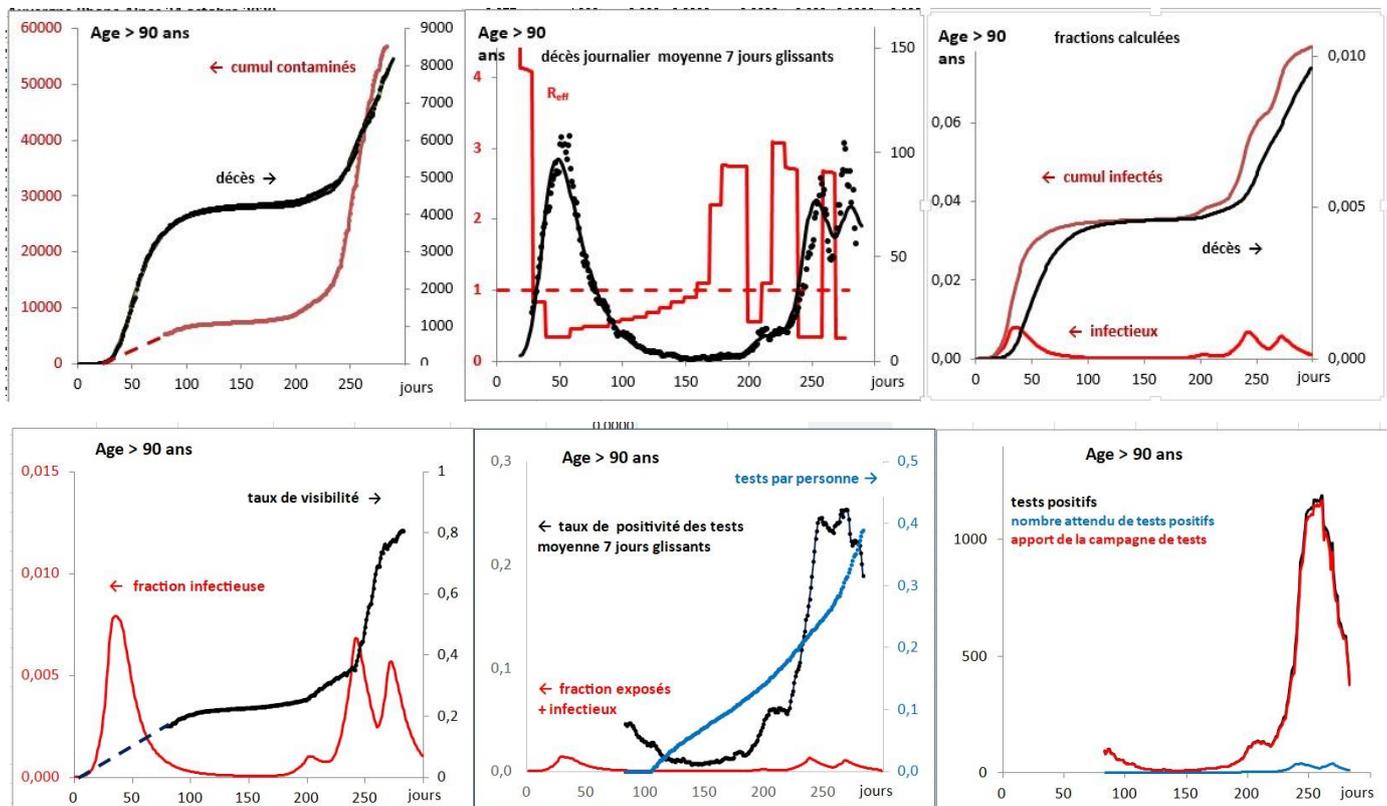
Sur la figure centrale, la présentation en termes de mortalité, c'est-à-dire nombre de décès divisé par la population de chaque tranche, conduit à un échelonnement « normal » des courbes, conforme à l'élévation de la létalité avec l'âge.

La figure de droite présente l'évolution du rapport mortalité / létalité qui donne de manière approximative le taux d'avancement de la contamination (fraction contaminée) de chaque tranche d'âge. Nous avons précisé *approximatif* parce que la date de contamination précède notablement celle du décès ; pour sortir de cette approximation il est nécessaire de tenir compte de la dynamique de la pandémie, comme nous le ferons dans la suite de cette lettre. Il ne faut pas oublier non plus que ce résultat dépend directement des divers du taux de létalité que nous venons de déterminer. L'absence de logique évidente dans l'échelonnement de ces courbes laisse à penser que l'âge n'est probablement pas le facteur prépondérant qui gouverne la dynamique et l'état d'avancement de la pandémie.

Nous avons procédé à l'analyse des données des huit tranches d'âge au-dessus de 20ans. Les deux plus basses ont un nombre (heureusement !) trop faible de décès pour pouvoir être traitées. La procédure d'ajustement reste manuelle et chaque tranche est traitée indépendamment des autres. Les données de tests seront extraites de la base de données déjà citée, dont la mise à jour est faite régulièrement (pourvu que ça dure !) depuis le 13 mai

2020. Comme nous l'avons fait pour les régions, les données de cas seront reconstituées à partir de cette date en intégrant le nombre de tests positifs (recensés en personnes). Le nombre de cas à la veille du 13 mai, inconnu, sera reconstitué « au doigt mouillé » de manière à donner une apparence de logique à la courbe du taux de visibilité (voir lettre 20). Nous allons épargner au lecteur le catalogue complet des résultats des huit tranches, qui sera donné dans une lettre annexe, et nous allons nous concentrer sur trois d'entre elles :

### Tranche > 90 ans :

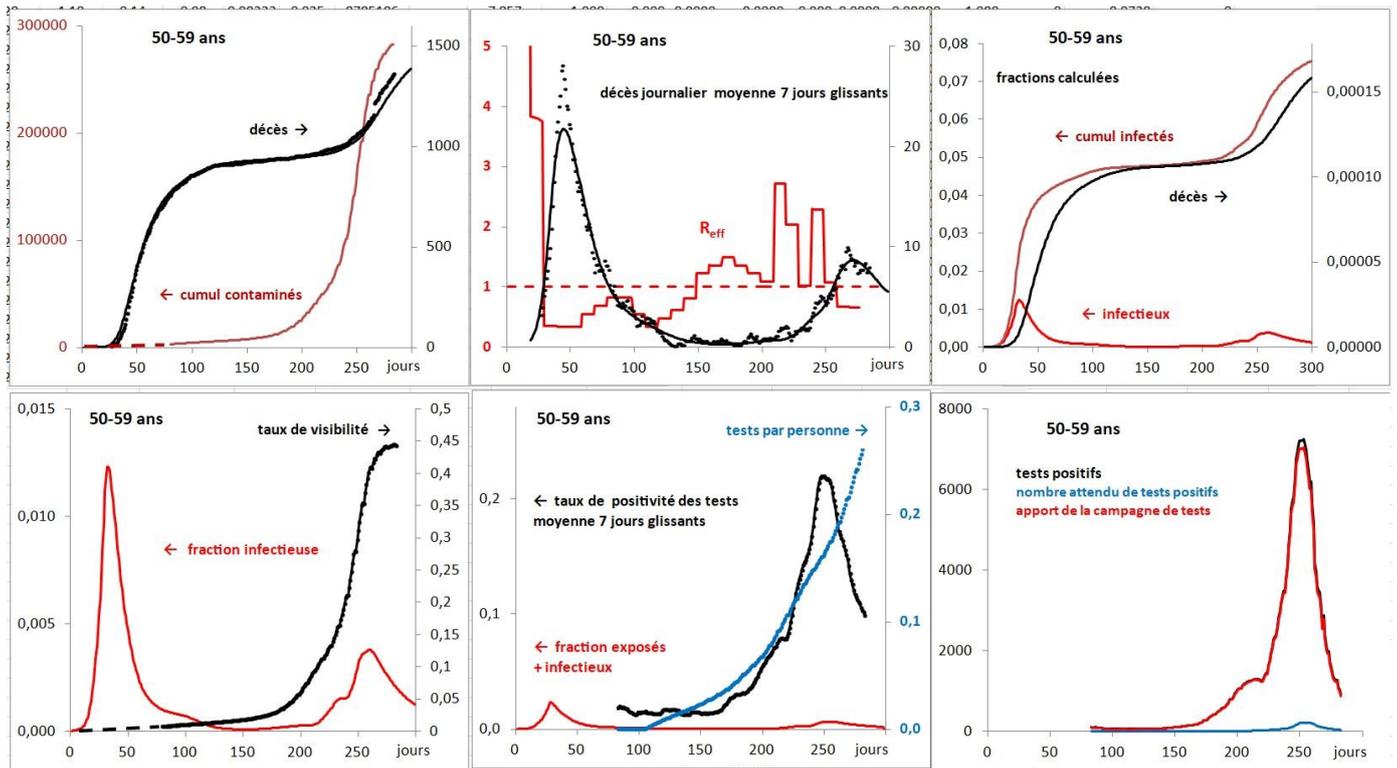


Rappelons, pour la figure en haut à droite, que nous présentons systématiquement la fraction décédés et la fraction contaminés avec un rapport d'échelle égal au taux de létalité ; les courbes se rejoignent quand la pandémie a cessé de flamber.

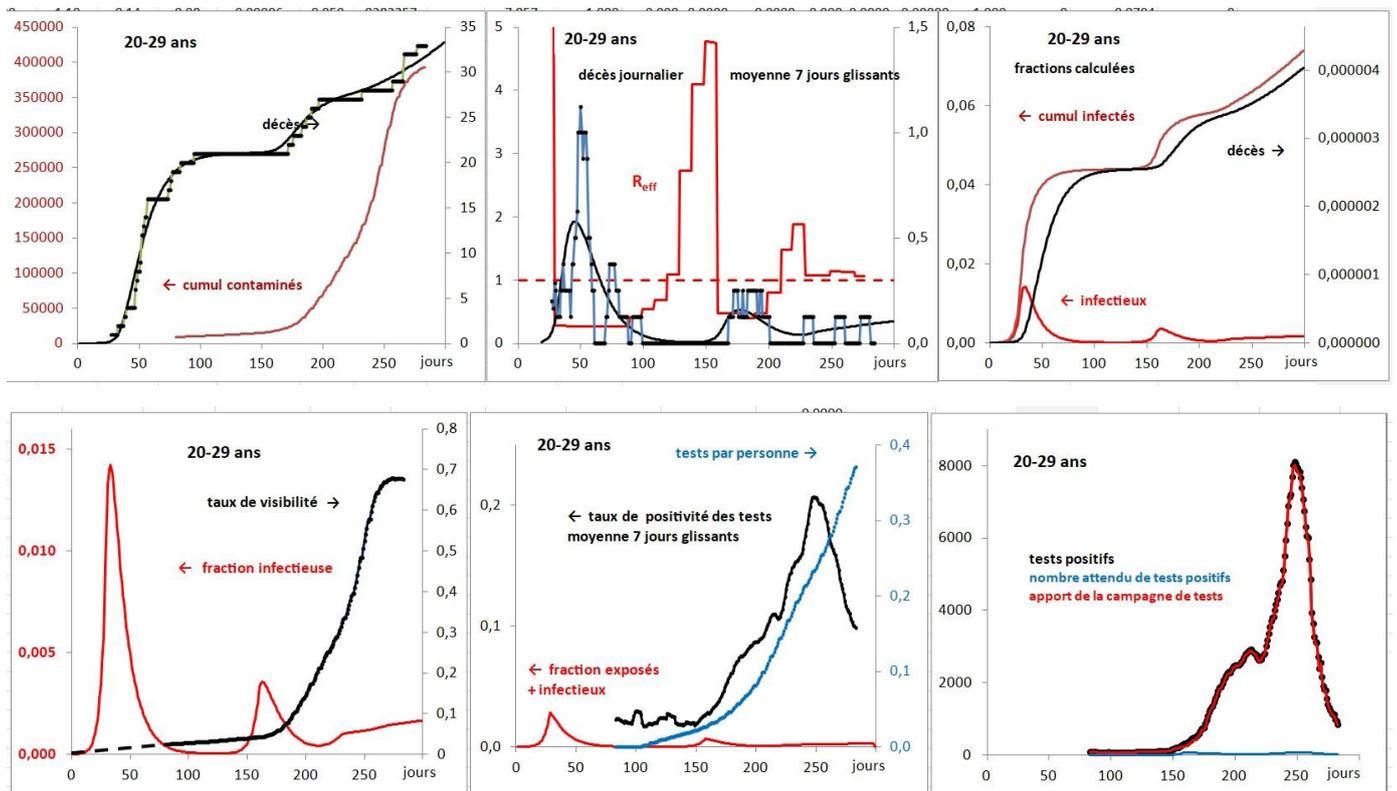
On retiendra la valeur très élevée du taux de visibilité de cette population, la moyenne nationale étant voisine de 0.20. Nous attribuons cette forte visibilité, dès le début de l'épidémie, au suivi médical quasi-systématique (en EHPAD par exemple) de cette tranche âgée de la population. On remarquera aussi le dédoublement de la deuxième vague de contamination (figure centrale en haut), qu'il nous est difficile de reproduire finement en ajustant le niveau des paliers de 10 jours de notre fonction  $\beta(t)$  (dont découle la courbe de  $R_{eff}$ ). Des essais avec des paliers plus étroits sont envisagés.

Remarquons aussi l'importance relative du rattrapage sur la courbe des cas (et de la courbe de visibilité). Nous aimerions évidemment nous affranchir la procédure « au doigt mouillé » dont ce rattrapage est issu. La connaissance du nombre de cas par tranche d'âges de Mars à Mai nous l'aurait permis... Quoi qu'il en soit, pour tester la vraisemblance de cette procédure, nous avons fait la somme de tous ces rattrapages (39000), et trouvé seulement un tiers du total de cas recensés au 13 mai sur le territoire national (140000). Les 2/3 manquants peuvent difficilement être totalement attribués aux deux tranches non analysées. Au final notre procédure « au doigt mouillé » a sensiblement sous-estimé ces rattrapages, et par conséquent surestimé la visibilité dans les différentes tranches. La visibilité de cette tranche la plus âgée reste néanmoins très élevée (> 60% en éliminant tout rattrapage).

## Tranche 50-59 ans



## Tranche 20-29 ans



Malgré la faiblesse des chiffres donnés dans la tranche 20-29 ans, nous avons pu y identifier les deux vagues de l'épidémie. La deuxième vague y semble significativement plus précoce que dans les autres tranches. Cependant elle est basée sur un petit nombre de décès et n'est pas clairement confirmée par les données de tests.

Nous avons rassemblé à la page suivante les résultats de l'analyse qui nous semblent les plus significatifs (valeurs au 30/11/2020) :

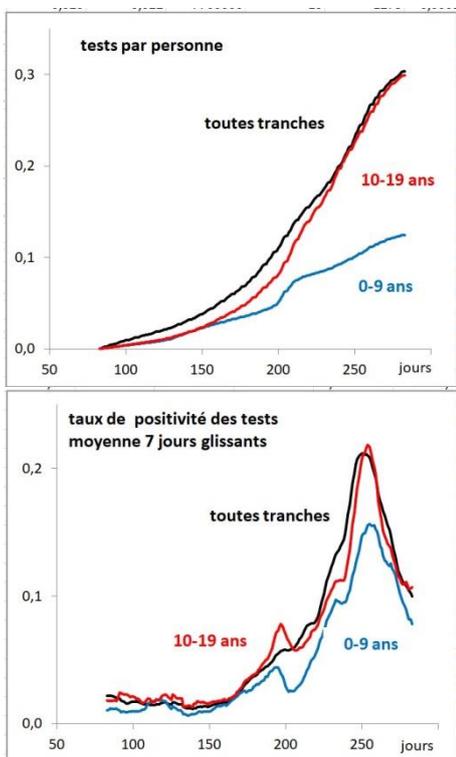
Age (ans)	0-19	20 - 29	30 - 39	40 - 49	50 - 59	60 - 69	70 - 79	80 - 89	> 90	Moyennes		
										20-90+	Nationale	0-19
Million habitants	16.09	7.47	8.28	8.58	8.79	8.00	5.69	3.22	0.91	50.94	67.03	16.09
Fr. infectieuse (%)	-	0.15	0.17	0.34	0.20	0.44	0.27	0.30	0.33	0.27	0.62	1.7
Fr. contaminée (%)	-	7.1	8.7	7.0	7.3	8.4	8.4	9.95	7.7	7.9	17.1	46
Visibilité (%)	-	67	44	47	44	28	26	31	81	43.2	19.6	-67

Pour autant que la présente analyse soit fiable (on sait qu'elle est perfectible comme nous tous), le désaccord entre les moyennes calculées sur les 8 tranches analysables et les moyennes nationales déterminées directement, est frappant : **il manque plus de la moitié des cas, et la visibilité obtenue est le double de la valeur attendue.**

Ces deux constatations forment un ensemble cohérent que l'on est tenté d'attribuer à l'effet des deux tranches qui ont échappé à l'analyse, les 0-9 et 10-19 ans. Nous avons indiqué en rouge dans la colonne de droite du tableau, les valeurs qu'il faudrait leur donner pour aboutir à des moyennes nationales correctes, et force est de constater que ces valeurs sont déraisonnables. Une absence totale de visibilité de cette tranche manquante permettrait de réduire la valeur moyenne de l'ensemble à environ 33 %, encore loin de la valeur attendue, 20 %.

Avons-nous surestimé l'impact des tests sur le nombre de cas ? Nous avons admis que chaque test positif augmentait d'une unité le nombre de cas. Ce « théorème de Trump » n'est évidemment valable que si les statistiques sont données en nombre de personnes testées, ce qui est bien précisé sur le site de Santé Publique France et nous l'avons éprouvé sur les données nationales (lettre n° XX). Au risque de nous répéter encore une fois, seule une campagne de tests sur une base vraiment aléatoire pourrait fournir la réponse.

Ou bien, faudra-t-il corriger les données de létalité ? Quand les données plus précises seront disponibles, ou quand les données de cas seront disponibles par tranches d'âge, nous ferons le nécessaire.



Notre attention ayant été attirée sur la mauvaise connaissance de deux des dix tranches d'âge, nous avons consulté les données de tests, voir figure ci-contre. Les données des tranches y sont comparées à celles du reste de la population, et l'on constate que la tranche 0-9 ans est relativement peu testée par rapport à l'ensemble de la population.

La figure du bas montre que la circulation du virus dans les deux tranches basses est pratiquement aussi active que dans le reste de la population. On remarque aussi une forte croissance pendant tout le mois d'août, que ne présente pas le reste de la population. A analyser par des sociologues... Ces deux tranches actives peuvent considérées comme le **moteur invisible de l'épidémie** : forte dynamique et faible impact sur le nombre total de cas graves et décès. Il est **urgent et nécessaire** que nos pouvoirs publics s'y intéressent plus spécifiquement et y consacrent en **priorité** une **campagne de tests massive et aléatoire**.

Portez-vous bien, en respectant au mieux le confinement, les gestes barrière, en particulier le masque partout où il est requis. Et naturellement n'oubliez pas le nettoyage fréquent de vos mains. Et continuons à réfléchir à une stratégie prudente pour les fêtes de fin d'année.

François VARRET, Physicien, Professeur Emérite à l'Université de Versailles Saint-Quentin

Mathilde VARRET, Chargée de Recherche INSERM (Génétique, Biologie) Hôpital Bichat.