

NOTEBOOK PYTHON ICI

Objectif

Je m'intéresse dans cet article aux différentes métriques utilisées pour le suivi de la propagation du virus SARS-COV-2 et de la maladie COVID.

Les politiques de gestion de la situation liée à ce virus cherchent *in fine* à éviter deux phénomènes, mesurés par deux métriques :

- Un dépassement des ressources hospitalières critiques
 - Mesuré par le taux d'occupation des lits de réanimation
- Une surmortalité en population générale
 - Mesurée par le nombre de décès hospitaliers liés au COVID

D'autres métriques sont annonciatrices d'une évolution favorable ou défavorable de la situation, et ce de part la succession temporelle des événements pour un individu :

Transmission du virus => (éventuelle) Maladie => (éventuelle) Maladie grave => (éventuelle) Admission en hospitalisation d'urgence => (éventuel) passage en réanimation => (éventuel) décès.

Je m'intéresse aux métriques annonciatrices suivantes :

- Nombre de tests positifs réalisés par les laboratoires et pharmacies en ville
- Taux d'incidence des Infections Respiratoires Aigües mesuré par le réseau Sentinelles
- Le nombre de passages aux urgences pour suspicion de COVID
- Le nombre d'actes médicaux SOS Médecins pour suspicion de COVID

Ces différentes métriques mesurent la situations à différents points de la succession temporelle définie ci-dessus. Le point de mesure n'est d'ailleurs pas toujours clairement défini : par exemple, un test positif peut survenir pour la première fois lors de la transmission ou lors de l'apparition de la maladie ou même lors de l'hospitalisation.

Je définis également une métrique théorique calculée rétrospectivement à partir du nombre de décès liés à la COVID par tranche d'âge. Le point de mesure de cette métrique est exclusivement à la transmission du virus. Je l'appelle "Nombre théorique d'infections".

Cette métrique ne peut pas être utilisée pour un suivi puisque par définition, elle ne peut être calculée que dans le futur, une fois que tous les décès imputables aux infections d'un jour donné soient survenus...

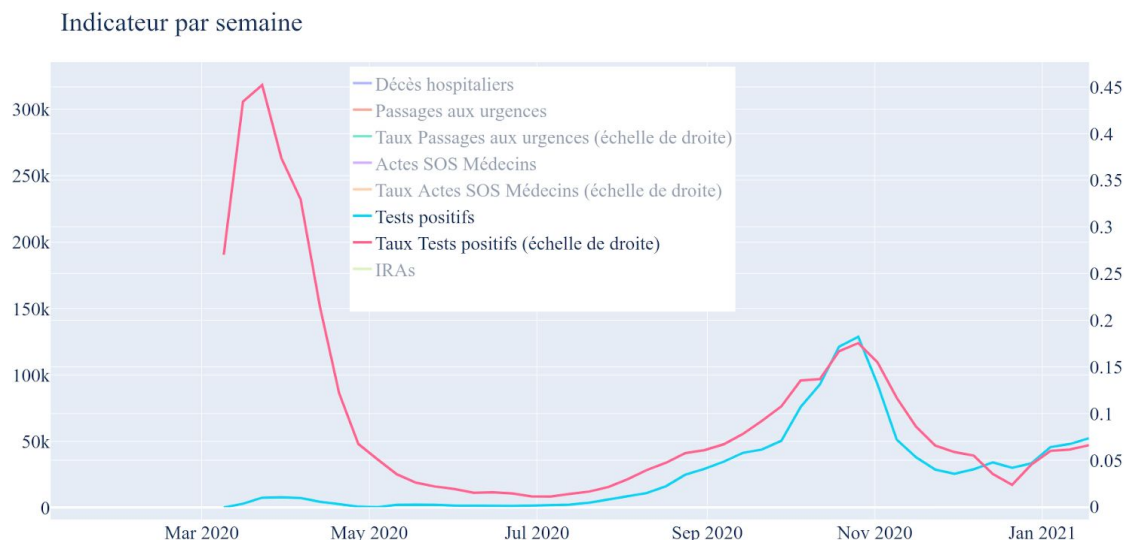
Cependant, je compare ce nombre théorique d'infections aux métriques annonciatrices afin de déterminer laquelle d'entre elles permet le meilleur suivi.

Sources de données

Métrique	Source	Fichier	Description
Tests positifs	SI-DEP via Santé Publique France data.gouv.fr/fr/datasets/donnees-relatives-aux-resultats-des-tests-virologiques-covid-19/	https://www.data.gouv.fr/fr/datasets/r/b4ea7b4b-b7d1-4885-a099-71852291ff20 Et à partir du 29 Mai 2020 https://www.data.gouv.fr/fr/datasets/r/001aca18-df6a-45c8-89e6-f82d689e6c01	Format CSV Nombre de tests par région, sexe et résultat (positif ou négatif)
Taux d'incidence IRA	Réseau Sentinelles https://www.sentiweb.fr/france/fr/?page=maladies&mal=25	https://www.sentiweb.fr/datasets/incidence-REG-25.json	Format JSON Par semaine et région : estimation (haute, basse, moyenne) absolue et par 100.000 habitants
Passages aux urgences	SOS Médecins et OSCOUR via Santé Publique France https://www.data.gouv.fr/fr/datasets/donnees-des-urgences-hospitalieres-et-de-sos-medecins-relatives-a-lepidemie-de-covid-19/	https://www.data.gouv.fr/fr/datasets/r/d2af5160-a21d-47b7-8f30-3c20dade63b1	Format CSV Par jour, région et classe d'âge et type (COVID ou pas) : nombre d'actes SOS Médecins et nombre de passages en urgence
Actes SOS Médecins			
Décès hospitaliers	Santé Publique France https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19	https://www.data.gouv.fr/fr/datasets/r/08c18e08-6780-452d-9b8c-ae244ad529b3	Format CSV Par jour, région et tranche d'âge de 10 ans: nombre de décès hospitaliers liés à la COVID

Afin d'éviter tout biais lié à l'hétérogénéité des régions françaises, j'illustrerai mes propos et mes calculs sur le **périmètre Ile de France**.

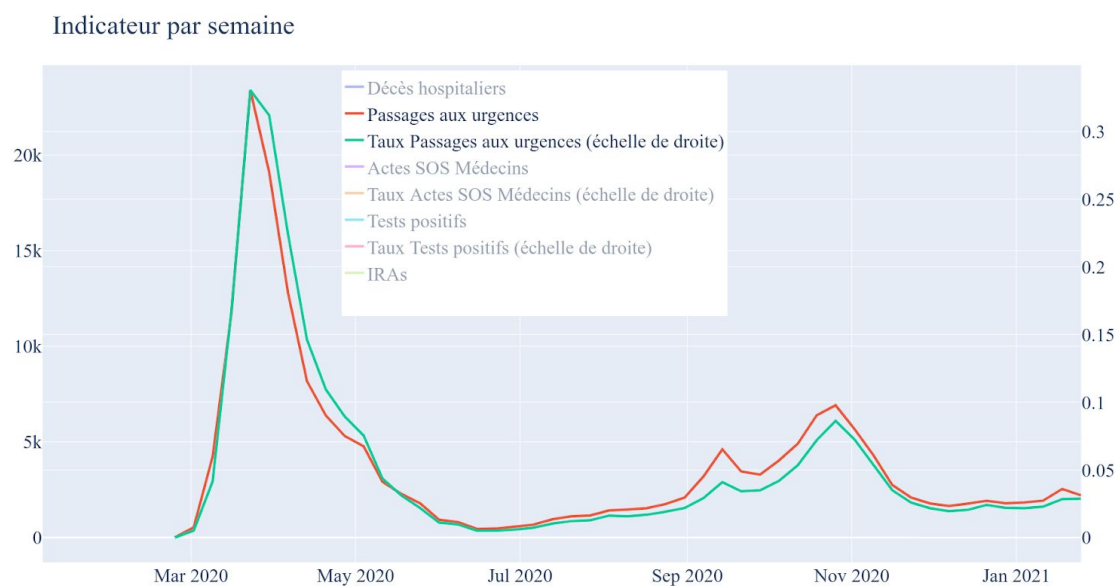
Où en est-on à fin janvier 2021 ?



Le nombre de tests positifs est en faible croissance depuis début janvier 2021

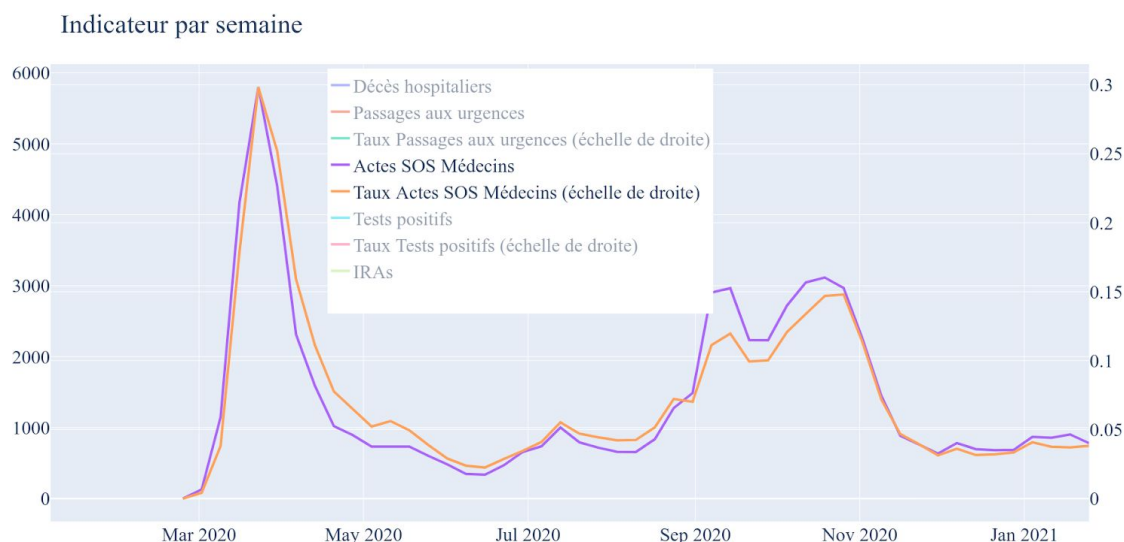
Le taux de tests positifs suit à peu près la même dynamique, mais seulement à partir de juillet 2020. En effet, avant cette date, les tests en ville n'étaient pas généralisés, et relativement peu de tests étaient pratiqués.

Deux pics sont visibles, le premier la semaine du 23 mars 2020, le deuxième la semaine du 26 octobre 2020.



Le nombre de passages en urgences est quasiment stable depuis début décembre 2020. Le taux de passages aux urgences (passages COVID rapportés aux passages toute cause) suit parfaitement la dynamique de la métrique.

Les mêmes deux pics sont visibles : la semaine du 23 mars 2020, le deuxième la semaine du 26 octobre 2020, en plus d'un pic moins marqué la semaine du 14 septembre 2020



Mêmes constatations sur la métrique des actes SOS médecins : une stabilité depuis début décembre 2020 et un pic la semaine du 23 mars 2020 en plus de deux autres pics. Le pic d'octobre 2020 est avancé d'une semaine et le troisième pic est plus marqué.



Les infections respiratoires aiguës ont une dynamique légèrement différente. Je retrouve le pic de la semaine du 23 mars 2020, mais il y a plutôt un plateau entre le 14 septembre et le 26 octobre, et une hausse notable depuis début janvier 2021

Je constate donc que les 4 métriques annonciatrices donnent des signaux concordants en mars 2020 mais moins concordants en automne, voire contradictoires les dernières semaines.

Calcul du nombre d'infections théorique

Je vais essayer comme annoncé plus haut de calculer une métrique théorique, non observable en temps réel car dépendante des décès du futur.

Je pars d'une simplification qui consiste à dire qu'un décès COVID survient après un délai fixe de l'infection par le virus, et survient à hauteur du taux de létalité de la maladie :

$$Infection_{(date\ décès - \text{délai décès})} = \frac{Décès_{(date\ décès)}}{Taux\ de\ létalité}$$

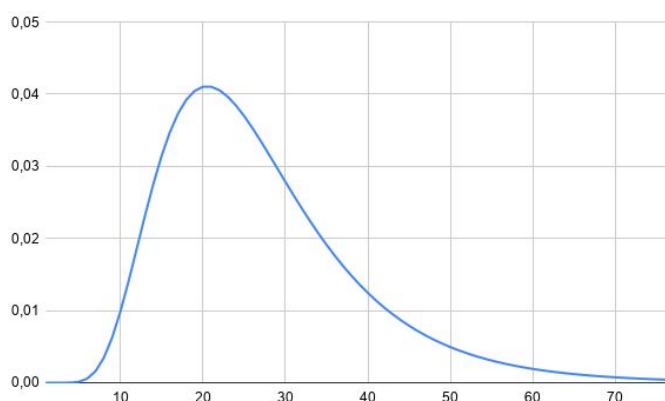
Or, bien entendu, le délai entre l'infection et la date de décès éventuel n'est ni prévisible, ni fixe.

L'article suivant mentionne une loi de probabilité de la distribution de ce délai :

<https://www.cebm.net/covid-19/estimating-the-infection-fatality-ratio-in-england/>

Ce papier de trois chercheurs des universités de Leeds et d'Oxford utilise une distribution log-normale ayant une moyenne de 26.8 jours et un écart type de 12.4 jours :

Distribution du délai en jours entre l'infection et le décès COVID



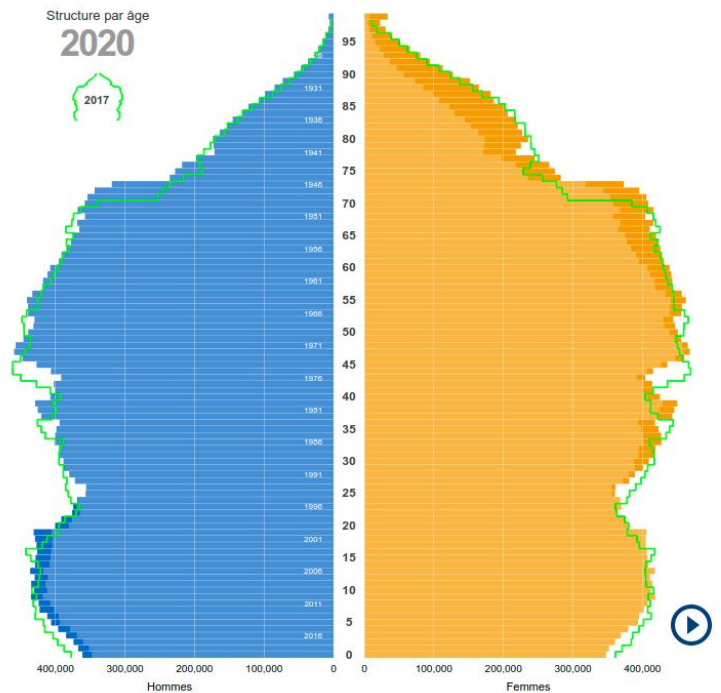
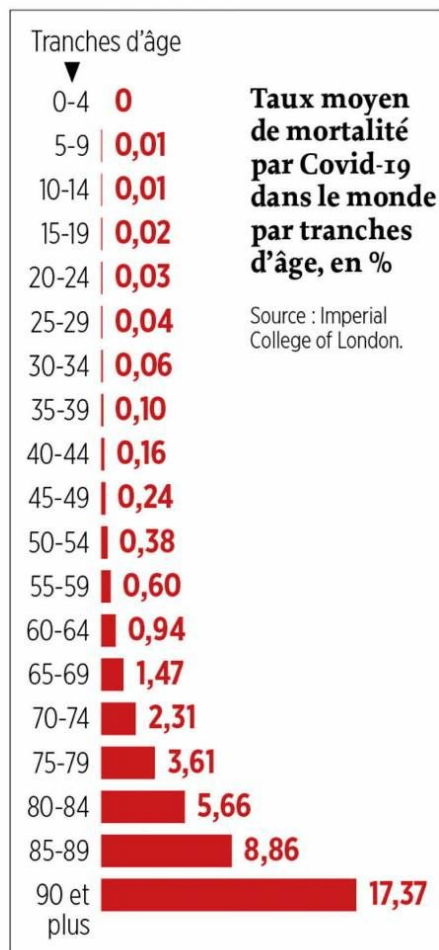
Je retiens que le décès survient dans 98.5% des cas en 9 semaines ou moins de l'infection, et 99.5% en 11 semaines ou moins.

Aussi, le taux de létalité (IFR en anglais pour Infection Fatality Ratio) est largement dépendant de l'âge de la personne décédée.

Heureusement, j'ai pu accéder à un résumé dans [Le Point](#) de cette étude de l'Imperial College London :

<https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-34-ifr/>

Cette étude établit la dépendance suivante entre l'âge et le taux de létalité (et non taux de mortalité comme indiqué à tort dans le graphe) :



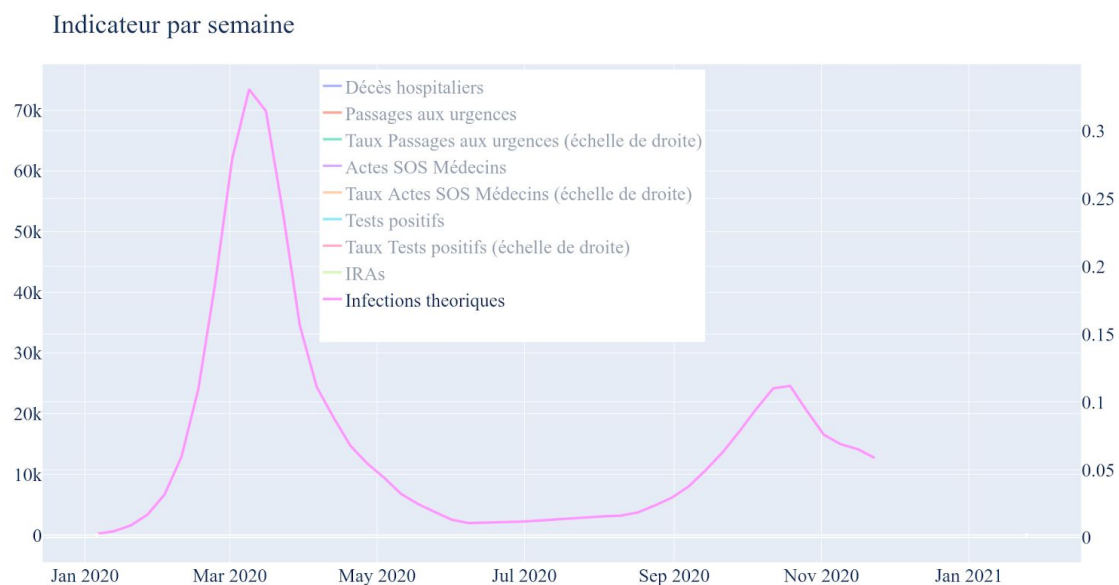
Les taux ci-dessus sont par tranches d'âge de 5 ans. Afin de les avoir par tranche d'âge de 10 ans comme dans les données de décès de Santé Publique France, j'agrège ces taux en pondérant par la [pyramide d'âge 2020](#) (aperçu ci-dessus). Ceci donne les taux de létalité suivants :

0 à 19 ans	0.0150 %
20 à 29 ans	0.0350 %
30 à 39 ans	0.0804 %
40 à 49 ans	0.2021 %
50 à 59 ans	0.4892 %
60 à 69 ans	1.1984 %
70 à 79 ans	2.8161 %
80 à 89 ans	7.0167 %
90 ans et plus	17.370 %

Une meilleure définition du nombre d'infections théorique pour une semaine S donnée est donc :

$$Infections_S = \sum_{tranches\ d'âge} \frac{\sum_{semaine\ 1\ à\ 11} Proba_{semaine} \times Décès_{S + semaine}}{Taux\ létalité_{tranche\ d'âge}}$$

L'implémentation me donne la courbe suivante, contenant un pic la semaine du 9 Mars et 16 Mars, ainsi qu'un autre pic les semaines du 12 octobre et 19 octobre 2020.



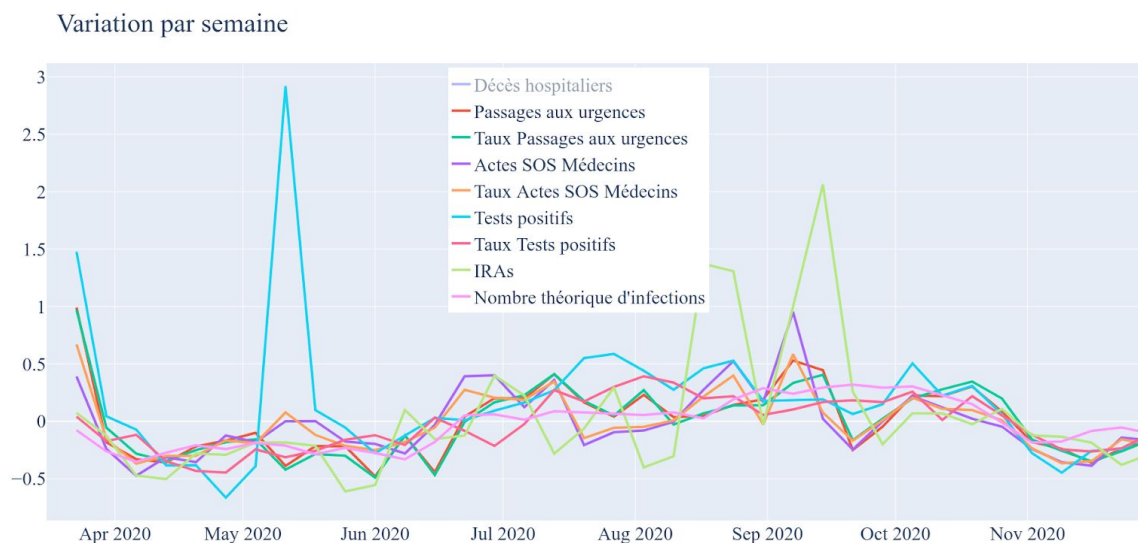
Comparaison des métriques

Afin de comparer les métriques annonciatrices au nombre théoriques d'infections, je commence par constater qu'elles sont décalées d'une ou deux semaines par rapport à cette dernière.

A part le décalage, les différents ordres de grandeur rendent difficile la comparaison des métriques.

Je procède donc à la transformation de chaque métrique d'une série d'indicateurs bruts en série de variations "de semaine en semaine".

Voici l'aperçu des séries de variations (38 semaines, de celle du 16 mars 2020 à celle du 30 novembre 2020) :



Je compare alors chaque série de variations avec la série de variation issue du nombre d'infections théoriques, en prenant soin de séparer les 19 semaines avant le 1er août 2020 des suivantes, afin de tenir compte de la non significativité de la métrique "tests positifs" avant cette date.

J'effectue la comparaison avec la métrique [RMSE](#), après avoir fait le décalage d'une semaine.

Ceci donne :

Avant le 1er août 2020

Métrique	Erreur RMSE
Taux Passages aux urgences	0.152195
Passages aux urgences	0.155777
Taux Tests positifs	0.167478
Taux Actes SOS Médecins	0.186467
IRAs	0.206977
Actes SOS Médecins	0.231752
Décès hospitaliers	0.304702

Après le 1er août 2020

Métrique	Erreur RMSE
Taux Tests positifs	0.147515
Taux Passages aux urgences	0.167091
Passages aux urgences	0.197988
Tests positifs	0.214832
Taux Actes SOS Médecins	0.224122
Actes SOS Médecins	0.290431
IRAs	0.68271

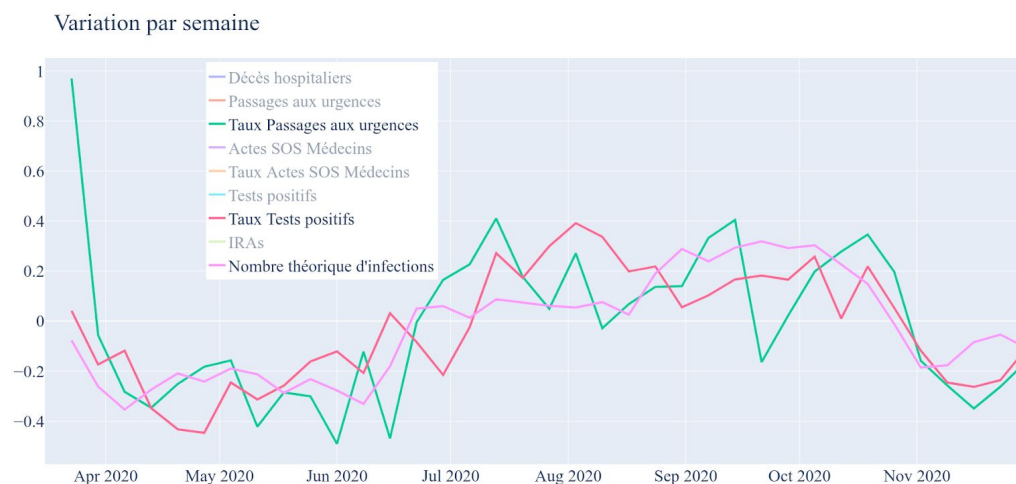
Et sur toute la période

Métrique	Erreur RMSE
Taux Passages aux urgences	0.161206
Taux Tests positifs	0.165328
Passages aux urgences	0.177289
Taux Actes SOS Médecins	0.203575
Actes SOS Médecins	0.259319
IRAs	0.497456

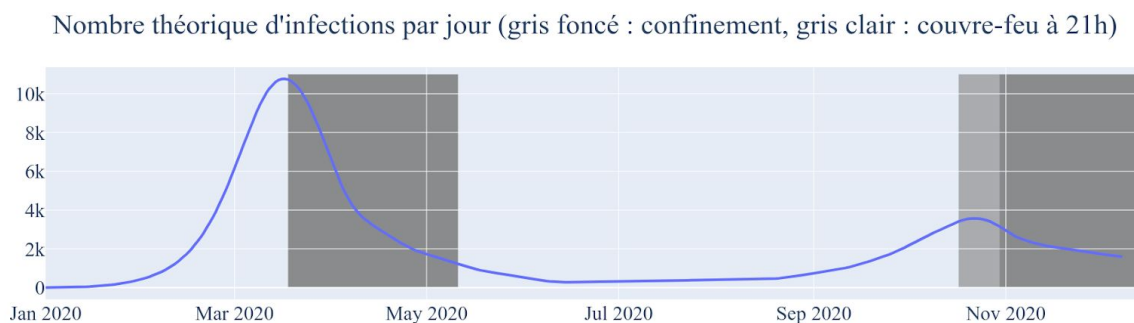
J'en conclue que les deux meilleures métriques annonciatrices pour suivre la situation sont :

- Le taux de passages aux urgences
- Le taux de tests positifs, **et non pas le nombre de tests positifs**

Pour s'en convaincre visuellement :



Remarques sur la métrique “Nombre théorique d’infections”



La première constatation que je fais est qu'en mars 2020, le nombre théorique d'infections ralentit sa croissance (point d'inflexion) le 4 mars, soit 2 semaines avant le confinement, et atteint son pic le 17 mars, le jour même du confinement qui débute à midi.

En automne, l'inflexion se produit le 14 octobre, soit 3 jours avant le couvre-feu à 21h en Ile de France. Le pic se produit le 22 octobre soit 8 jours avant le deuxième confinement.

Ces constatations sont sensibles au calcul du nombre théorique d'infections, et donc à la loi de probabilité du délai entre le jour d'infection et le jour de décès. Pour rappel j'ai utilisé une loi log-normale de moyenne 28.6 jours et d'écart type 12.4 jours.

D'après cette loi, 90% des décès surviennent dans les 6 semaines suivants l'infection.

J'ai pu calculer le nombre théorique d'infections jusqu'à la semaine du 30 novembre, donc jusqu'au 6 décembre 2020 inclus, avec un total de 723.000 infections en Ile de France.

6 semaines après cette date, c'est-à-dire le 18 janvier 2021, l'Ile de France déplorait 12800 décès hospitaliers dus à la COVID.

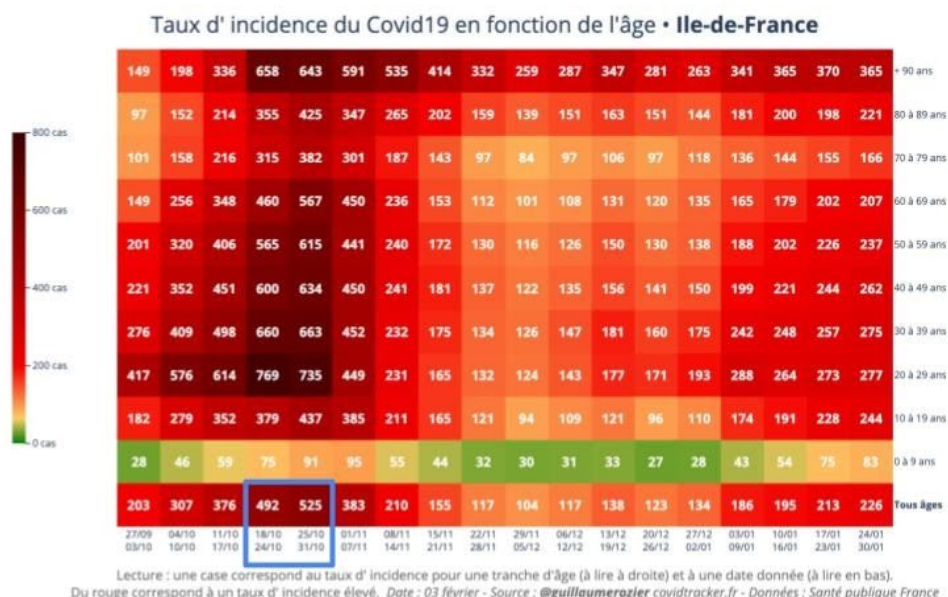
Ceci me donne un taux de létalité en Ile de France de $12800 / 723.000 = 1.77\%$

L'ordre de grandeur de ce chiffre semble cohérent avec les enseignements de [l'article du Point](#).

Ce nombre théorique d'infection, rapporté à la population de l'Ile de France (12.200.000), me permet d'avoir un taux d'incidence par 100.000 habitants qui a connu 2 pics :

- **615 en mars 2020 (~73.800 infections la semaine du 16 mars 2020)**
- **204 en octobre 2020 (~24.400 infections la semaine du 19 octobre 2020)**

A ce stade je ne peux pas m'empêcher de constater que ce taux d'incidence que j'obtiens avec le nombre théorique d'infections est inférieur en octobre au taux d'incidence calculé à partir du nombre de tests positifs : **~200 versus ~500 soit 2.5 fois plus faible**



Ma conclusion à ce stade est que le taux d'incidence des tests positifs surestime de 150% le taux d'incidence de la véritable maladie COVID...

Cette conclusion est assez perturbante puisque le nombre de tests positifs ("nombre de cas") est l'indicateur le plus médiatisé.

Pour avoir un troisième point de contrôle, je reviens au nombre d'infections respiratoires aiguës estimés par le réseau sentinelles.

Pour l'Ile de France, elles sont autour de 20.000 par semaine en octobre 2020.

Cependant, il faut avoir à l'esprit que ces IRAs ne sont pas nécessairement diagnostiquées chez des personnes malades COVID.

Le bulletin de Santé Publique France de fin 2020 donne la proportion de 19% sur la base de tests pour un échantillon de malades :

- En semaine 52 (du 21 au 27 décembre 2020), le **taux d'incidence de consultations pour une IRA a été estimé à 67/100 000 habitants** (intervalle de confiance à 95% : [52-82]) en France métropolitaine, stable par rapport à celui de la semaine 51 : 56/100 000 habitants (IC95% : [48-64]) (Figure 1).
- En semaine 52, sur les 13 prélèvements effectués chez des patients présentant une IRA, 2 se sont avérés positifs pour le SARS-CoV-2, aucun n'était positif pour un rhinovirus ou pour un virus grippal.
- Depuis la semaine 37, sur les 520 prélèvements effectués chez les patients qui ont été testés pour les différents virus respiratoires surveillés sur prélèvement nasopharyngé, 211 (42%) étaient positifs pour un rhinovirus, 98 (19%) pour le SARS-CoV-2, 4 pour un métapneumovirus et 1 pour un virus respiratoire syncytial (VRS). Un prélèvement s'est avéré positif pour un virus grippal de type B/Victoria en semaine 49.

Ceci me donne donc $20.000 \times 19\% = 3.800$ malades COVID en Ile de France ayant connu une infection respiratoire aiguë.

Or il est admis que 80 à 85% des malades du COVID font des formes bénignes, donc les COVID + IRAs représentent dans un scénario haut 15% des malades COVID.

Je trouve donc $3.800 / 15\% = \sim 25300$ malades COVID en Ile France par semaine en octobre 2020.

Ceci me donne un **taux d'incidence de 207 par 100.000 habitants**, presque identique au taux d'incidence à la même période déduit du nombre théorique d'infections.

Je suis donc conforté dans ma conclusion : le taux d'incidence des tests positifs exagère largement la prévalence de la maladie COVID.

Pourquoi le nombre de tests positifs surestime-t-il la prévalence de la COVID ?

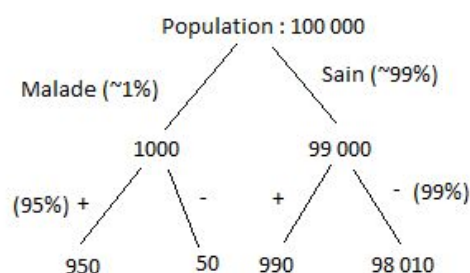
Je n'ai pas de qualification pour parler des tests d'un point de vue médical.

Je peux cependant noter que, pour une métrique de suivi d'un phénomène, les tests de dépistage COVID accumulent beaucoup de mauvais points depuis le début de la situation :

- évolution de la disponibilité (d'abord uniquement en hôpital, jusqu'à la généralisation en ville)
- évolution des conditions de prescription (selon le fait d'être prioritaire, symptomatique, cas contact, avec ou sans ordonnance du médecin, etc...)
- la variété de technologies (d'abord uniquement RT-PCR puis des tests antigéniques, retrait de certaines marques peu fiables, etc...)
- l'évolution de l'appétence selon le remboursement ou la médiatisation

Ces éléments me laissent penser que cet "outil de mesure" a été trop instable dans le temps pour pouvoir suivre une situation donnée (de surcroît nouvelle). Ceci étant dit, la comparaison faite des métriques montre que le taux de tests positifs corrige en partie ces biais.

Par ailleurs, comme l'explique [cet article pédagogique sur le théorème de Bayes](#), même à des niveaux de [sensibilité et spécificité](#) élevés (95% et 99% respectivement), un test produit autant de faux positifs que de vrais positifs lorsque le taux de prévalence du phénomène que l'on cherche à mesurer est faible. L'hypothèse de prévalence est de 1% dans l'article, sachant qu'en Ile de France je trouve une prévalence de 0.2% en octobre et 0.6% en mars.



Je ne peux pas transposer cet élément tel quel dans mon analyse puisque les tests réalisés en France ne sont pas aléatoires. En examinant les bulletins de Santé Publique France d'octobre 2020, je trouve que ~30% des personnes testées déclarent des symptômes et ~35% (pas forcément les mêmes) sont des cas contacts de personnes testées positives. Ceci étant dit, je retiens de ce raisonnement le risque de faux positifs dans les tests en population générale.

Je termine sur les tests de dépistage en soulignant une particularité du test RT-PCR que l'on comprend à la lecture des [articles de vulgarisation](#) : le résultat du tests dépend de la valeur CT (Cycle Threshold, ou le nombre maximum de cycle d'amplification du matériel génétique avant de détecter la présence de la séquence recherchée).

Autrement dit, deux tests RT-PCR ne sont pas nécessairement équivalents, puisqu'ils peuvent avoir été positifs avec un nombre de cycles différents.

Pour moi, cette non constance du test RT-PCR en tant qu'outil de mesure s'ajoute à sa liste de mauvais points.

Une notice d'information du 13 janvier 2021 publiée par l'Organisation Mondiale de la Santé a alerté sur les faux positifs et la sensibilité au nombre de cycles d'amplification pour les tests RT-PCR :

WHO guidance [Diagnostic testing for SARS-CoV-2](#) states that careful interpretation of weak positive results is needed (1). The cycle threshold (Ct) needed to detect virus is inversely proportional to the patient's viral load. Where test results do not correspond with the clinical presentation, a new specimen should be taken and retested using the same or different NAT technology.

WHO reminds IVD users that disease prevalence alters the predictive value of test results; as disease prevalence decreases, the risk of false positive increases (2). This means that the probability that a person who has a positive result (SARS-CoV-2 detected) is truly infected with SARS-CoV-2 decreases as prevalence decreases, irrespective of the claimed specificity.

La notice a été retirée du site internet de l'OMS, mais reste consultable à travers [archive.org](https://web.archive.org/web/20210121170101/https://www.who.int/news/item/14-12-2020-who-information-notice-for-ivd-users) <https://web.archive.org/web/20210121170101/https://www.who.int/news/item/14-12-2020-who-information-notice-for-ivd-users>

Synthèse

A partir des décès hospitaliers dus à la COVID en Ile de France, et en utilisant des hypothèses de l'Imperial College et des des universités de Leeds et d'Oxford, je calcule un nombre théorique d'infections par semaine de Février à Novembre 2020.

Ce nombre me permet d'estimer le taux d'incidence de la COVID en Ile de France à 200 pour 100.000 habitants en octobre 2020, cohérent avec un taux d'incidence estimé à partir de l'indicateur "Infections respiratoires aiguës" du réseau sentinelles, mais bien en dessous du taux d'incidence des tests positifs qui surestime d'un facteur 2.5

La progression de ce nombre théorique d'infections peut être le mieux prédite par la progression du taux d'admission en urgence pour suspicion de COVID, mais aussi par la progression du taux de tests positifs - en non pas le nombre absolu de tests positifs.

Enfin, je constate dans la progression de ce nombre théorique d'infection que les ralentissement de la croissance surviennent avant les confinements.