

ETIENNE DECROLY  
JACQUES VAN HELDEN  
MARTINE QUINIO

# COVID-19 : APPROCHES SCIENTIFIQUES & ENJEUX SOCIÉTAUX



# Covid-19 : approches scientifiques et enjeux sociétaux

par Etienne Decroly, Jacques van Helden & Martine Quinio

Campus Saint-Charles, Bâtiment 2, grand amphithéâtre

Inscription : [contact@universite-populaire-de-marseille.net](mailto:contact@universite-populaire-de-marseille.net)

Info et documents : [http://universite-populaire-de-marseille.net/covid-19\\_2020/](http://universite-populaire-de-marseille.net/covid-19_2020/)

Vidéos : <https://www.youtube.com/c/UniversitePopulairedeMarseille/>

Lundi 28/09	<b>Etienne Decroly</b> Virologue, Directeur de Recherches au CNRS	Approches de virologie moléculaire
Jeudi 01/10	<b>Jacques van Helden</b> Professeur de bioinformatique à Aix-Marseille Université	Retracer dans les génomes l'origine et la propagation du virus
<b>Jeudi 08/10</b>	<b>Martine Quinio</b> <b>Professeure de mathématiques à Aix-Marseille Université</b>	<b>Raisonnement statistique et modélisation</b>
Mercredi 14/10	Annick Stevens, Jacques van Helden, Martine Quinio et Etienne Decroly	Covid-19: et maintenant ? a. Table-ronde-débat b. Compléments d'information

*“Savoir pour prévoir afin de pourvoir” (Auguste Comte)*

# *Que peuvent les mathématiques, la statistique?*

- Transmettre afin de partager des connaissances: rôle de passeur, médiateur entre expertise scientifique et population.
- Objectifs de la science statistique:
  - Donner un sens aux informations chiffrées, les transformer en connaissances partagées
  - Éclairer les décideurs et tous les citoyens dans la compréhension des risques et incertitudes

## ■ Santé Publique France

- ❑ Bulletin hebdomadaire à télécharger; [les chiffres clés de la Covid](#)
- ❑ Points épidémiologiques régionaux
- ❑ [Géodes](#) pour cartes, synthèses, graphiques Départements, France, Monde

## ■ Université Johns Hopkins

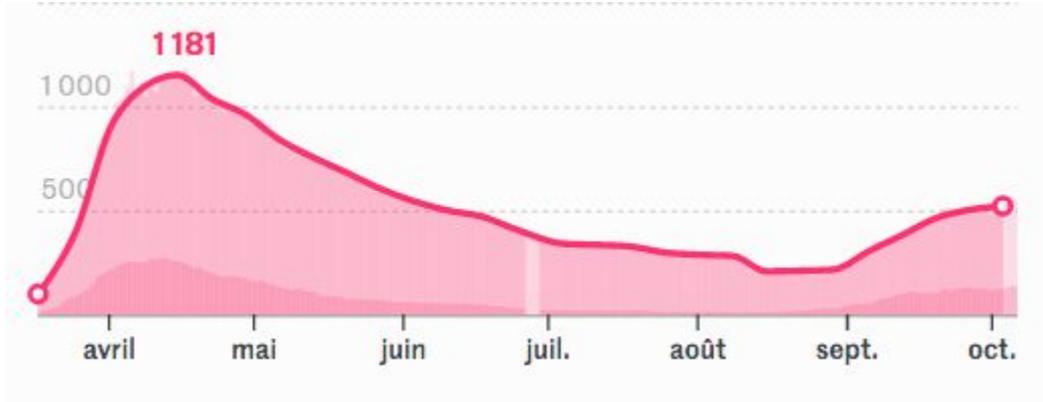
- ❑ <https://coronavirus.jhu.edu/map.html>
- ❑ Cartes interactives, actualisées chaque jour, dans le monde pour visualiser l'épidémie.

## ■ Bulletins hebdomadaires et vidéos de l'IHU (APHM)

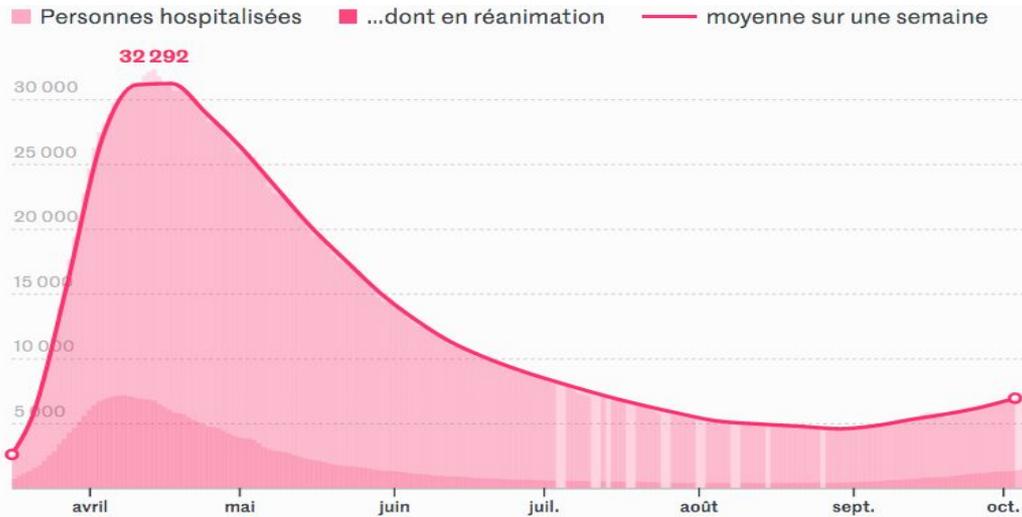
## ■ Dernier bulletin information décryptage de l'APHM

## ■ .....

# Exemple: Hospitalisation et Réanimation au jour le jour (Univ J.Hopkins)



Bouches du Rhône



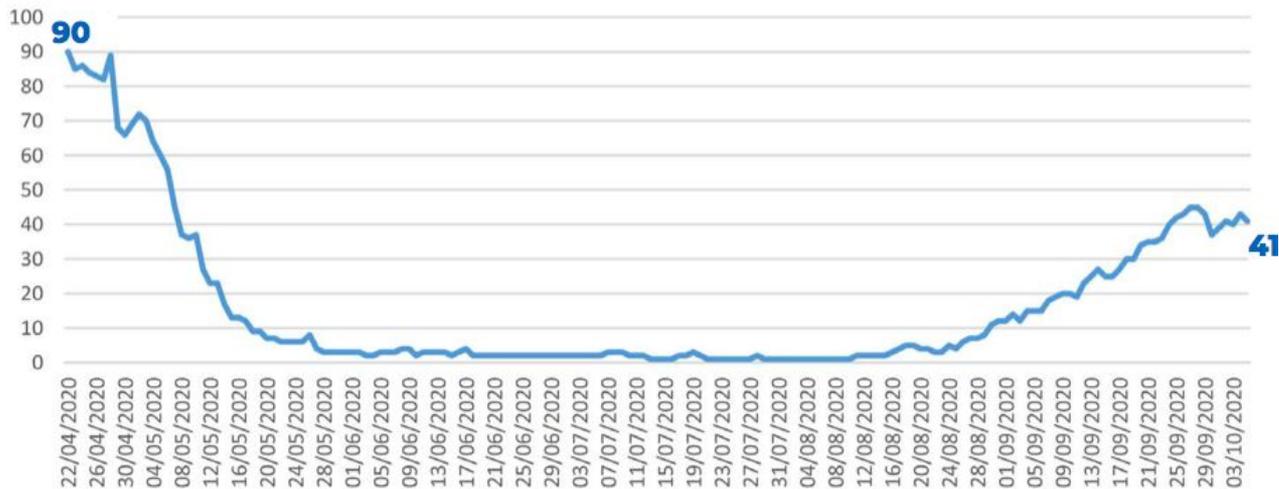
France

## Historique portant sur les indicateurs COVID

Suivi du 22 avril 2020 au lundi 5 octobre 2020

### Patients COVID hospitalisés (dénombrement à minuit)

#### Réanimation



# Hospitalisés Covid APHM (au 5/10)

## Unités dédiées COVID hors Réanimation



- Forte demande de pédagogie: demande des médecins, demande de tous.
- Responsabilité individuelle et collective
- Que peut le raisonnement statistique pour éclairer décideurs, médecins (point de vue collectif), citoyens (point de vue individuel) ?

## En quoi consiste la science statistique?

- Organiser les données numériques afin de les “lire” et leur donner un sens.
- Pratique ancienne (dénombrer pour prévoir)
  
- Raisonnements basés sur les probabilités (jeune branche des mathématiques)
- Statistique de prévision: Science des “erreurs contrôlées”
- A partir d’observations, formuler des hypothèses
- Allers-retours: “observation, expérience (réalité) - modélisation (lois) - retour à l’expérience”

- Définir certains mots et concepts jetés dans les médias
- A partir des derniers documents source disponibles,
  - Lire et interpréter des graphiques, tendances.
  - Prendre un peu de recul par rapport aux informations quotidiennes
  - Aborder les raisonnements statistiques dans leur diversité
- Exemples: comment “savoir” (avec un certain degré de vraisemblance) si:
  - Un traitement A est-il plus efficace qu’un traitement B ?
  - Il y a une corrélation entre des “événements” (Bar, resto et COVID) ?

# Plan du cours

1. Introduction : principe de la science statistique
2. Nécessité de vérifier les sources d'information et d'abord, de les choisir.
3. Dictionnaire de mots utiles, définitions usuelles.
4. Vérité apparente, incertitude, risques: incohérences apparentes?
5. Efficacité des outils mathématiques pour comprendre des informations (Bulletin SpF)
6. Les raisonnements statistiques : qu'est-ce qu'une hypothèse scientifique ?
7. Les raisonnements bayésiens, basés sur une « croyance » en une hypothèse.
8. Généralisation, Modèle de probabilité ;
9. Les raisonnements de type « fréquentiste » consistant à comparer deux groupes (cas testés/cas témoins) (exemple: Resto/Covid)
10. Les tests de dépistage ; valeur prédictive d'un test
11. Un exemple de collaboration mathématiques-médecine avec le projet de tests groupés *-pooling-*
12. Accepter la complexité; mettre de la complexité dans la simplicité apparente.

- Etude d'une certaine population Pop
  - $N_{pop}$  taille de la population
- "Cas" : personne infectée
  - qu'on le sache ou pas
  - qu'elle présente ou pas des symptômes (symptomatique/asymptomatique)
- Nombre de cas
  - $N_{cas}$  nombre de cas dans la population à un instant  $t$
- Prévalence
  - $T_{prev}$  Taux (ou proportion) du nombre de personnes infectées rapporté à la population étudiée.

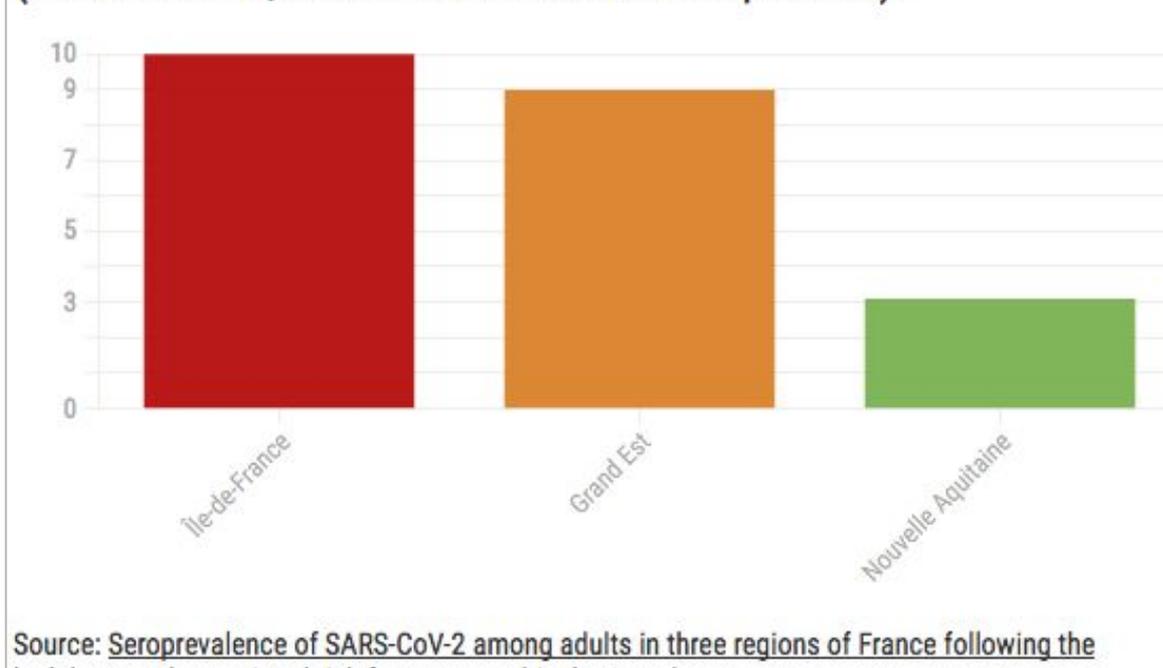
$$T_{prev} = \frac{N_{cas}}{N_{pop}}$$

# Séroprévalence Covid-19 par région

## Etude Institut Pasteur

- Publication: 18 septembre 2020
- Période d'observation: mi Mai - 20 Juin
- Sur 15 000 personnes
- Axe vertical : taux de séroprévalence "au long la vie" (infection passée)

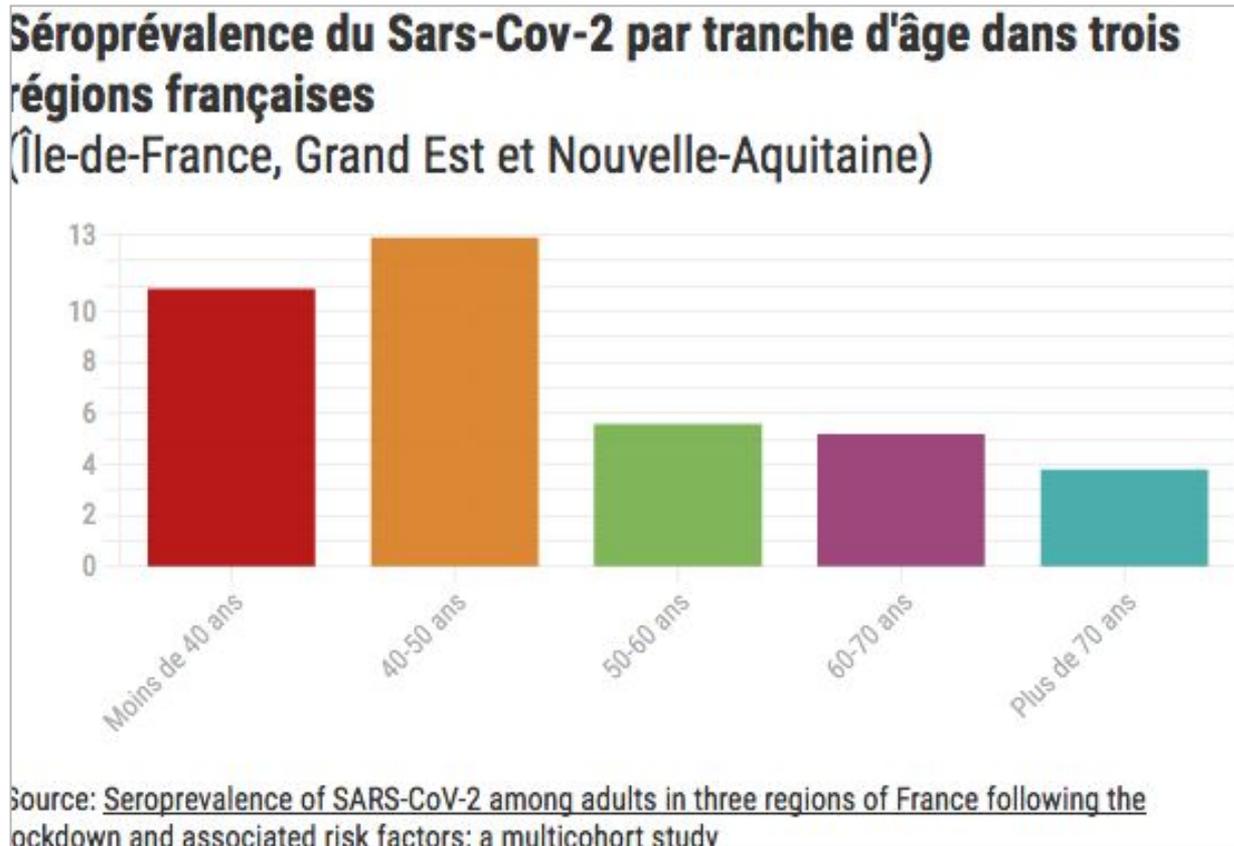
## Séroprévalence du Sars-Cov-2 dans trois régions françaises (Île-de-France, Grand Est et Nouvelle-Aquitaine)



Carrat, F., de Lamballerie, X., Rahib, D., Blanche, H., Lapidus, N., Artaud, F., *et al.* 2020. *Seroprevalence of SARS-CoV-2 among adults in three regions of France following the lockdown and associated risk factors: a multicohort study.* Infectious Diseases (except HIV/AIDS).

doi: <https://doi.org/10.1101/2020.09.16.20195693>

# Séroprévalence Covid-19 par tranche d'âge



Carrat, F., de Lamballerie, X., Rahib, D., Blanche, H., Lapidus, N., Artaud, F., *et al.* 2020. Seroprevalence of SARS-CoV-2 among adults in three regions of France following the lockdown and associated risk factors: a multicohort study. *Infectious Diseases (except HIV/AIDS)*.

doi: <https://doi.org/10.1101/2020.09.16.20195693>

# Qu'est-ce que l'incidence ?

- Incidence
  - $N_{inc}$  nombre de nouveaux cas dans une population *sur une période donnée*
- Taux d'incidence
  - $T_{inc}$  rapport entre l'incidence et la taille de la population

$$T_{inc} = \frac{N_{inc}}{N_{pop}}$$

## ■ Nombres de:

- $N_T$  nombre de personnes testées dans une population pendant une période  $t$  (semaine)
- $N_+$  nombre de personnes testées positives
- $N_-$  nombre de personnes testées négatives

## ■ Taux de positivité

- $T_+$  proportion de personnes positives parmi les testées

$$T_+ = \frac{N_+}{N_T}$$

$$T_- = \frac{N_-}{N_T}$$

QUESTION : POSITIF, SUR QUEL CRITÈRE ?

## *Mesures et estimations*

# Hospitalisation Covid à Marseille au 5 octobre (APHM)



**170 patients  
COVID**  
(sur 200 lits  
disponibles)

<b>Réanimation</b>	<b>41 / 47</b>
<b>HÔPITAL NORD</b> (Pôle MUSCA)	<b>17 / 20</b>
Réa. ancienne USC DRIS	10 / 10
Réa. MIR B	4 / 6
Réa. MIR B 2	3 / 4
<b>TIMONE ADULTES</b>	<b>20 / 23</b>
RUSC (Pr Gainier)	5 / 6
Réa T2 (Pr Bruder)	14 / 15
Réa RPPF (Pr Albanese)	1 / 2
<b>CONCEPTION</b>	<b>4 / 4</b>
Réa brûlés	4 / 4
<b>Autres Unités dédiées</b>	<b>128 / 143</b>
<b>TIMONE ADULTES</b>	<b>91 / 98</b>
Pôle MIT	69 / 75
Unité UMAP	22 / 23
<b>HÔPITAL NORD</b>	<b>15 / 16</b>
Unité COVID Maternité Nord	1 / 2
HC.PNEUMOLOGIE-HN	14 / 14
<b>TIMONE ENFANTS</b>	<b>1 / 4</b>
Unité 14e	1 / 4
<b>CONCEPTION</b>	<b>21 / 25</b>
HC OBSTETRIQUE 1ER-CO	1 / 1
UNITE COVID CONCEPTION	20 / 24

# Que mesure-t-on ?

- Note: Santé Publique France mesure un “taux d’incidence” SUR les personnes testées.

Sous estimation du réel taux d’incidence puisque  $N+ < N_{inc}$

Taux d’incidence estimé par SPF

$$T_{inc}^{SPF} = \frac{N+}{N_{pop}}$$

Taux d’incidence

$$T_{inc} = \frac{N_{inc}}{N_{pop}}$$

Taux de positivité

$$T_{+} = \frac{N+}{N_T}$$

## QUE PEUT-ON MESURER? QUE PEUT-ON ESTIMER?

Quelles informations nous donnent ces taux?

Quelles sont les données fiables?

Quelles sont les données manquantes?

# Questions sur la pertinence des indicateurs

- Question : le nombre de nouveaux cas confirmés stable ou diminue, mais le taux de positivité augmente : comment interpréter ?

R: On a diminué le nombre de tests

- Question : l'augmentation du taux de positivité est-elle indicative ?
  - R: Seulement si la diminution du nombre de tests est "neutre" (non-biaisée)
- Y aurait-il des *biais de détection* ?
  - R: Oui:
    - présentation spontanée au dépistage/ tirage au sort
    - Diversités locales de politique de test

- Voici un extrait du [bulletin hebdomadaire de Santé publique France](#)
- Université Johns Hopkins publie chaque jour

## Période de la semaine pertinente

### Chiffres clés en France

S39 S38\* Evolution

#### Indicateurs hebdomadaires semaine 39 (du 21 au 27 septembre 2020)

Nombre de nouveaux cas confirmés de COVID-19 (SI-DEP)	70 458	(73 965)	- 5%
Taux de positivité (%) pour SARS-CoV-2 chez les patients symptomatiques (SI-DEP)	13,4	(7,8)	+6,1 point
Nombre d'actes SOS Médecins pour suspicion de COVID-19	5 070	(6 967)	-27%
Nombre de passages aux urgences pour suspicion de COVID-19 (réseau OSCOUR®)	5 101	(6 669)	-23%
Nombre de nouvelles hospitalisations de patients COVID-19 (SI-VIC)	4 204	(3 657)	+15%
Nombre de nouvelles admissions en réanimation de patients COVID-19 (SI-VIC)	786	(599)	+31%
Nombre de décès liés au COVID-19 (incluant les décès en hospitalisation et décès en EHPA et autres EMS)	436	(332)	+31%

#### Indicateurs cumulés du 1<sup>er</sup> mars au 29 septembre 2020

Nombre de cas de COVID-19 ayant été hospitalisés (SI-VIC)	124 377
Nombre de décès liés au COVID-19 (incluant les décès en hospitalisation et décès en EHPA et autres EMS)	31 893

## Des indicateurs plus fiables?

Catherine Hill (épidémiologiste, biostatisticienne)

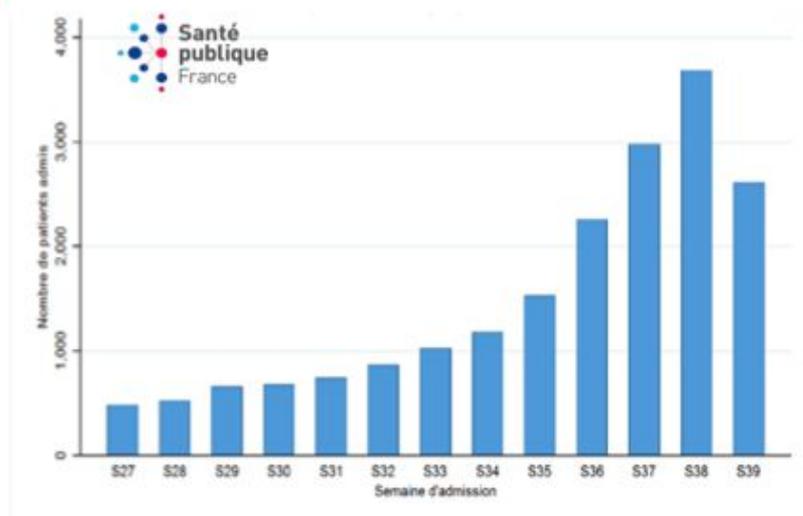
*"Il faut faire attention avec le taux d'incidence, car plus vous allez dépister de gens, plus le nombre de cas positifs va être grand et plus ce taux va augmenter.*

*Le taux d'incidence, défini comme le nombre de personnes testées positives rapporté au nombre d'habitants, dépend énormément du périmètre de la population testée. Si on ne teste que les personnes symptomatiques qui le sollicitent on va sous-estimer considérablement la vraie incidence qui est la fréquence des contaminations dans la population."*

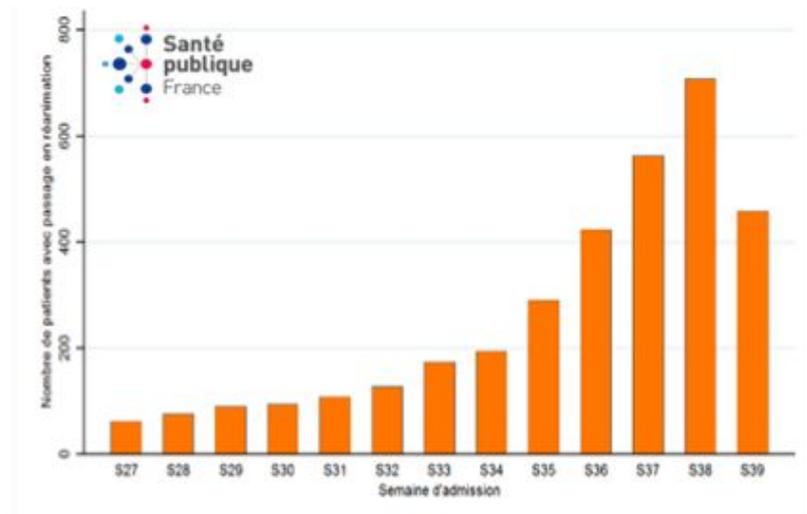
(Catherine Hill, épidémiologiste, sur France Info le 21/7/2020)

**Figure 31. Nombre hebdomadaire de nouvelles admissions A) à l'hôpital et B) en service de réanimation de patients COVID-19, selon la date d'admission à l'hôpital, depuis le 29 juin 2020, France, données au 29 septembre 2020 (source : SI-VIC)**

**A**

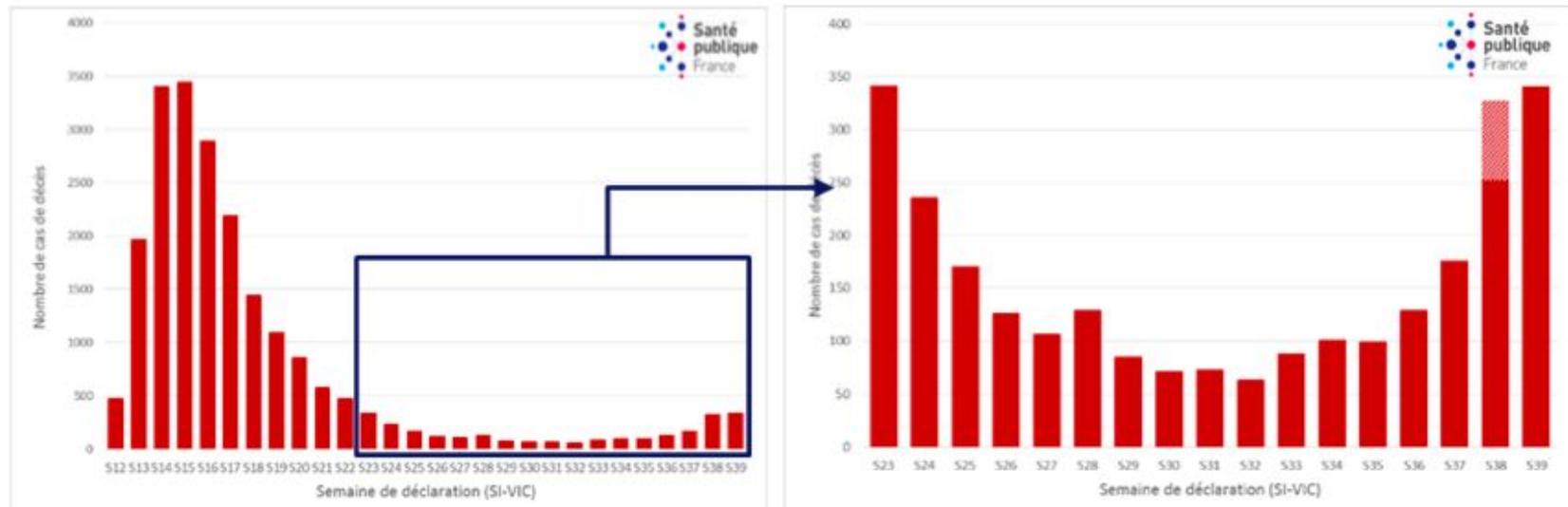


**B**



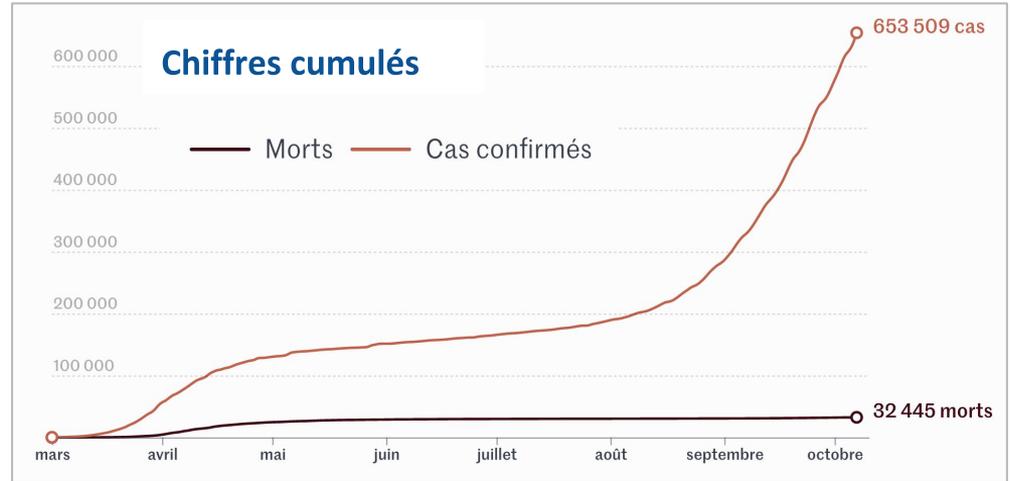
*S39 : données non consolidées*

Figure 29. Nombre hebdomadaire de nouveaux décès survenus au cours d'une hospitalisation pour COVID-19 selon la date de déclaration, depuis le 19 mars et depuis le 1<sup>er</sup> juin 2020, France, données au 29 septembre 2020 (source : SI-VIC)



# Nombre de cas et de décès en France (Johns Hopkins via Le Monde)

Chiffres de l'Université Johns Hopkins  
Graphiques sur le site Web du Monde



[https://www.lemonde.fr/les-decodeurs/article/2020/05/05/corona-virus-age-mortalite-departements-pays-suivez-l-evolution-de-l-epidemie-en-cartes-et-graphiques\\_6038751\\_4355770.html](https://www.lemonde.fr/les-decodeurs/article/2020/05/05/corona-virus-age-mortalite-departements-pays-suivez-l-evolution-de-l-epidemie-en-cartes-et-graphiques_6038751_4355770.html)

- On ne mesure pas:
  - ❑ Le taux d'incidence réel (dans toute la population, taux de nouveaux cas sur semaine)
  - ❑ La séroprévalence réelle (taux de personnes ayant été infectées)
- Rôle de la statistique: mesures, estimation
  - ❑ On pourrait estimer les indicateurs inconnus si on disposait d'échantillons de personnes testées représentatifs de la population
  - ❑ On donnerait alors les estimations avec une marge d'erreur contrôlée. (fourchette)

## Taux mesurés semaine 39 (27/9)

Le taux de testés en France dans la semaine du 27/9:

1,3% de la population

Le “taux d’incidence” (rapporté aux testés):

105 pour 100000 habitants:

0,105%

Le taux de positivité (rapporté aux testés):

$P(T+) = 7,6\%$

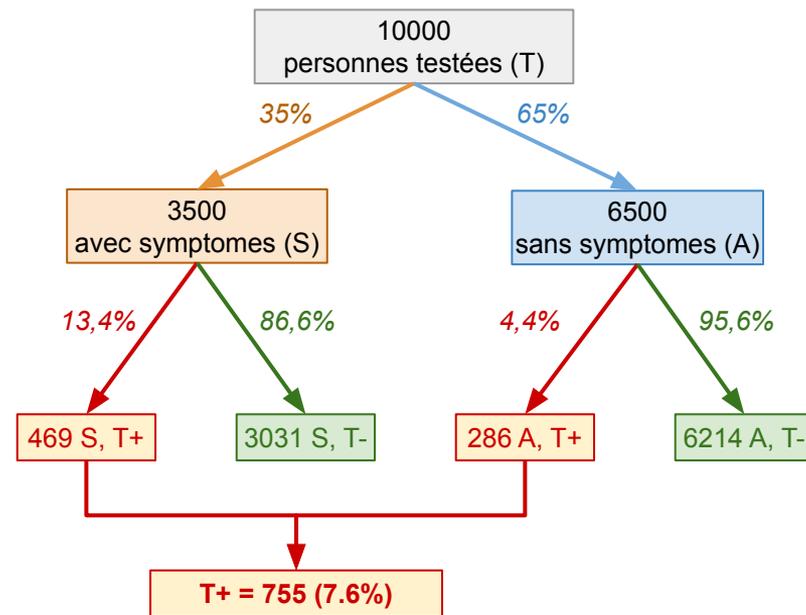
# Arbre de probabilités des taux de positivité

Données Santé publique France (1/10/2020):

*“ Parmi les testés : 65% se déclarent sans symptôme et représentent 38% des positifs. Le taux de positivité chez les symptomatiques est 13,4%. Le taux de positivité chez les sans symptômes est 4,4% ...*

*Le taux de positivité est **7,6%** “*

Complicé !



# Efficacité des probabilités

Parmi les testés:

Taux de positivité = 7,55%

Taux de personnes avec symptômes S

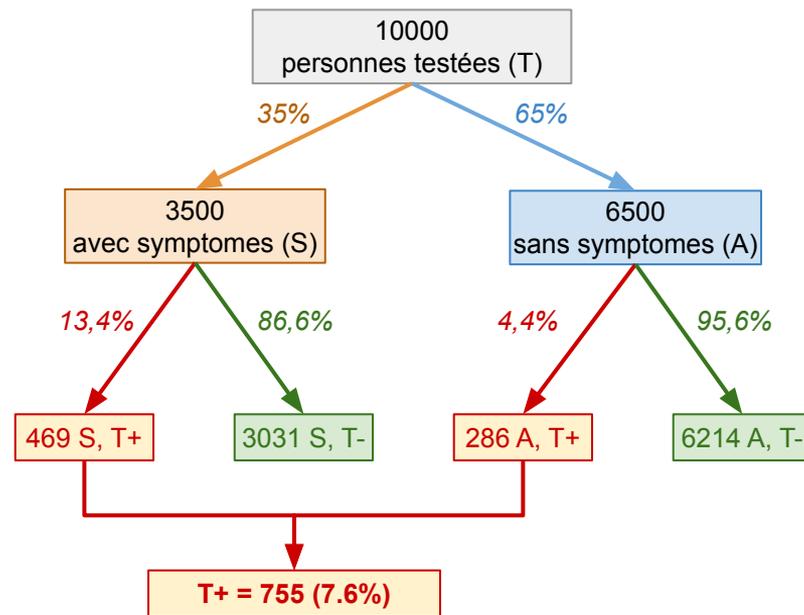
Notations, traduction en probabilité

Noté:  $P(T+) = 7,55\%$

Noté:  $P(S) = 35\%$

Difficulté: S peut désigner des réalités différentes: ici, contraire de Non S

**SpF donne** "Parmi les testés : ceux qui se déclarent sans symptôme représentent 38% des positifs: noté?"



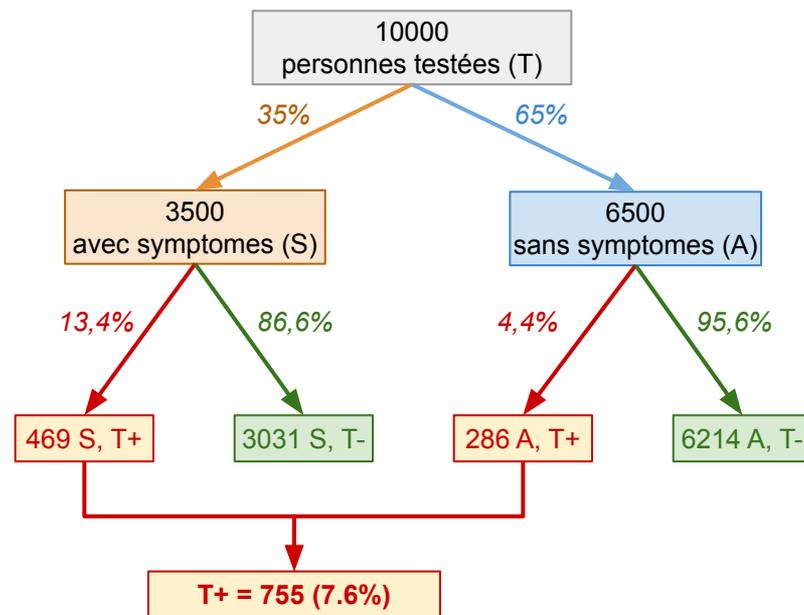
# Efficacité des probabilités, (suite)

Calculons avec l'arbre la

- **Proportion de personnes avec symptôme parmi les positifs:**
  - $469/755=62\%$
- **Donc la proportion de personnes avec symptôme parmi les positifs est 62%**
  - Noté  $P(S | T+) = 62\%$
  - Le dénominateur: nombre de personnes positives

Inversion :

- Proportion de personnes positives parmi les personnes avec symptôme:
  - Noté  $P(T+ | S) = 13,4\%$
  - Le dénominateur: nombre de personnes avec symptôme



$$P(S | T+)$$

Le numérateur est le nombre de personnes à la fois positives et avec symptôme:  $N(S \text{ et } T+)$

Le dénominateur: nombre de personnes positives,  $N(T+)$

$$P(T+ | S)$$

Le numérateur est le nombre de personnes à la fois positives et avec symptôme:  $N(S \text{ et } T+)$

Le dénominateur: nombre de personnes avec symptôme,  $N(S)$

Puisque les numérateurs sont identiques, c'est que

$$N(S \text{ et } T+) = P(S|T+) \cdot N(T+) = P(T+ | S) \cdot N(S)$$

**Et en divisant par le nombre de testés:  $N(T)$ , on a**

## Formule de Bayes:

$$P(T+ | S) \cdot P(S) = P(S | T+) \cdot P(T+)$$

Il y a 4 termes dans cette formule: 3 suffisent pour calculer le 4 ième

# Formule de Bayes

$P(S|T+)$  Probabilité d'avoir des symptômes si l'on a été testé positif

$$P(S|T+) = \frac{N(S \text{ et } T+)}{N(T+)}$$

En divisant par le nombre de testés, on a les probabilités

$$P(S|T+) = \frac{P(S \text{ et } T+)}{P(T+)}$$

Ou:

Ou:

$$P(S \text{ et } T+) = P(S|T+) \cdot P(T+)$$

$$P(T+ \text{ et } S) = P(T+ | S) \cdot P(S)$$

Probabilité (*a posteriori*) d'avoir une observation de symptômes étant donnée une infection M notée  
 $P(S|M)$

Probabilité d'avoir une infection M si observation de symptômes S (probabilité *a priori*) notée:

$P(M|S)$   
(ce qui intéresse le patient!)

Par exemple:

S est l'observation de symptôme(s)

- Maux de tête, fièvre, raideur de la nuque
- Perte d'odorat, toux..

M: infection

- Méningite?
- Covid?

$$P(M|S) \cdot P(S) = P(S|M) \cdot P(M)$$

## Exemples:

- **Observation S: Toux+fièvre**
- **M: infection Covid?**
  
- P(S|M) de l'ordre de 60%: degré de vraisemblance de symptômes S si infection M, ou infection
  
- P(M|S) de l'ordre de 13% : degré de vraisemblance de l'infection M si symptômes S
  
- Remarque: Ces probabilités dépendent du contenu de S

■ **Ne pas confondre!**

**La formule de Bayes  $P(H|O) \cdot P(O) = P(O|H) \cdot P(H)$**

**est une expression qui exprime le lien entre:**

- **le degré de vraisemblance d'une hypothèse H suite aux observations O.**
- **le degré de vraisemblance d'une observation O sous cette hypothèse H.**

# Comment valider une hypothèse par comparaison?

Une hypothèse, suite à une observation, est un pari de généralisation, de reproductibilité:

*« Expérimenter, c'est poser une question, et on ne conçoit jamais une question sans l'idée qui sollicite la réponse. Je considère donc, en principe absolu, que l'expérience doit être instituée en vue d'une idée préconçue ( .... ) C'est qu'il ne suffit pas d'observer, il faut se servir des observations et pour cela, il faut généraliser...qu'est ce donc qu'une bonne observation ?*

*C'est celle qui nous fait connaître autre chose qu'un fait isolé ; c'est celle qui nous permet de prévoir, c'est à dire celle qui nous permet de généraliser »*

*(Henri Poincaré, Sciences et Hypothèses)*

La statistique par tests d'hypothèses permet, avec un risque de se tromper contrôlé, de valider ou pas un modèle. (tentative d'expliquer un phénomène)

Pour tester une hypothèse, constitution de 2 groupes tirés au sort:

- (1) Le groupe de “cas” (ou groupe “traité”)
- (2) Le groupe de “témoin” (ou groupe “contrôle”)

Par exemple:

- (1): Traité avec médicament A (tiré au sort dans cette population)
- (2): Traité avec médicament B (tiré au sort dans cette population)

## *Hypothèse nulle/hypothèse alternative*

- Question: Le traitement B est-il plus efficace que le traitement A?
  - Hypothèse nulle: Le traitement B est aussi efficace que le traitement A
- Question: Y-a-t-il indépendance statistique entre “être gros fumeur” et “avoir un cancer des bronches”
  - Hypothèse nulle: “être fumeur” n’a pas d’influence sur “cancer bronches”

Travaux de Nathan et Hill années 70 ... à l’origine des lois anti-tabac 30 ans plus tard.

Importance de la précision dans la formulation des hypothèses

# Une étude US: lien “Covid positifs” et “Fréquentation bars, restaurants”

Ce qui a été rapporté dans les médias: 26 SEPT

*« une étude montre que les gens malades du Covid ont fréquenté les restaurants 2,8 fois plus que les autres et les bars 4 fois plus que les autres ».*

**QU' EST-CE QUE CELA VEUT DIRE?**

# Différentes hypothèses

(1) Parmi les gens malades, il y a 4 fois plus de personnes ayant fréquenté les bars que de personnes ne les ayant pas fréquentés ?

ou :

(2) Parmi les gens ayant fréquenté les bars, il y a 4 fois plus de personnes malades que de personnes non malades du Covid ?

(3) Les gens malades et ayant fréquenté les bars sont 4 fois plus nombreuses que les personnes n'ayant pas fréquenté les bars et non malades du Covid ?

# Fréquentation des restaurants et résultats des tests Covid-19

## Etude de référence

- Etude américaine fin Juillet sur 314 personnes venues spontanément se faire tester:
- Parmi elles, 154 testées positives, symptomatiques. (“groupe de cas”) et 160 personnes testées négatives (“groupe témoin”).
- Les 314 personnes ont été ou pas exposées à un risque potentiel : être allé au restaurant dans les 15 jours précédents.

## Effectifs observés

Resto/Cas		Positifs	Négatifs	Marges
	<i>Symboles</i>	<i>T+</i>	<i>T-</i>	<i>T = T+ + T-</i>
<b>Restaurant</b>	<i>R</i>	63	44	107
<b>Pas de restaurant</b>	<i>nonR</i>	91	116	207
<b>Marges</b>	<i>= R + nonR</i>	154	160	314

Tableau des résultats de tests Covid-19 (colonnes) en fonction de la fréquentations de restaurant (lignes)

Fisher KA, Tenforde MW, Feldstein LR, et al. Community and Close Contact Exposures Associated with COVID-19 Among Symptomatic Adults  $\geq 18$  Years in 11 Outpatient Health Care Facilities – United States, July 2020. MMWR Morb Mortal Wkly Rep 2020;69:1258–1264.

DOI: <http://dx.doi.org/10.15585/mmwr.mm6936a5external>

# Lien entre “ Fréquentation des restaurants ” et “ Résultat test Covid ” ?

Pour évaluer s’il y a un lien, on pose l’hypothèse appelée *Hypothèse nulle* :  $H_0$

$H_0$  : “Pas de lien entre R et T+”

Choix de  $H_0$  important: ici

$H_0$  : « Les événements R et T+ sont indépendants »,

au sens où la réalisation de l’un (R) n’a pas d’influence sur la réalisation de l’autre (T+).

Sous cette hypothèse, on peut alors construire un tableau, appelé tableau des effectifs théoriques (sous l’hypothèse  $H_0$ )

## Effectifs observés

Resto/Cas		Positifs	Négatifs	Marges
	<i>Symboles</i>	$T+$	$T-$	$T = T+ + T-$
<b>Restaurant</b>	$R$	63	44	107
<b>Pas de restaurant</b>	$nonR$	91	116	207
<b>Marges</b>	$= R + nonR$	154	160	314

## Effectifs théoriques sous $H_0$

Resto/Cas		Positifs	Négatifs	Marges
	<i>Symboles</i>	$T+$	$T-$	$T = T+ + T-$
<b>Restaurant</b>	$R$	52.5	54.5	107
<b>Pas de restaurant</b>	$nonR$	101.5	105.5	207
<b>Marges</b>	$= R + nonR$	154	160	314

# Si $R$ n'a aucune influence sur $T+$

- Seul le terme en rouge est nécessaire, grâce aux contraintes des marges qui ne changent pas : on dit qu'il y a un seul **degré de liberté** :
- **Comment calculer le terme en rouge ?**
- **C'est le nombre de personnes à la fois  $T+$  et Resto. (au *pro rata*):**

## Par une règle de proportionnalité

- Si Resto ou Non Resto n'a aucune influence sur l'effectif des  $T+$ , alors le nombre de personnes à la fois  $T+$  et Resto est :
  - **Proportion de  $T+$  (soit, ) rapportée au nombre total (107) de Resto :**

$$(154/314) \cdot 107 \text{ soit, } 52,5$$

## Effectifs observés

Resto/Cas		Positifs	Négatifs	Marges
	<i>Symboles</i>	$T+$	$T-$	$T = T+ + T-$
Restaurant	$R$	63	44	107
Pas de restaurant	$nonR$	91	116	207
Marges	$= R + nonR$	154	160	314

## Effectifs théoriques sous $H_0$

Resto/Cas		Positifs	Négatifs	Marges
	<i>Symboles</i>	$T+$	$T-$	$T = T+ + T-$
Restaurant	$R$	52.5	54.5	107
Pas de restaurant	$nonR$	101.5	105.5	207
Marges	$= R + nonR$	154	160	314

# Interprétation des écarts: significatifs ou pas?

Il s'agit de savoir si les différences entre ces tableaux sont significatives (ou juste dues à la variabilité, au hasard) :

- On mesure un écart total (appelé Khi deux) entre les données observées et les théoriques ; des tables de lecture basées sur des lois de probabilité permettent alors de calculer la probabilité que l'écart dépasse une valeur seuil : (la p.value)
- Les calculs donnent un indice d'écart calculé valant 6,28.
- Or d'après les tables, la probabilité de dépasser cette valeur serait d'environ 1,5% (p.value) si nous étions sous hypothèse nulle (autrement dit, si le taux d'infection était indépendant de la fréquentation des bars et restaurants)

*Mathématiques:*

*Pour mesurer des écarts entre les tableaux, terme à terme:*

*L'indice de Khi deux est la somme des carrés des écarts entre les termes des tableaux rapportés aux effectifs théoriques*

- On peut rejeter l'hypothèse d'indépendance des deux événements:
- Sur la base de *cette étude*, “aller au restaurant” et “être testé positif 15 jours plus tard” sont significativement liés.

## Etude contestable ?

- Pas de tirages au sort: présentation spontanée de personnes symptomatiques
- Les conditions de restauration (terrasse, intérieur) ne sont pas prises en compte.
- Le nombre de passages au bar et restaurant n'est pas pris en compte.
- Pour les bars, l'étude de référence a été faite sur 21 personnes: effectifs faibles
- Les différences d'âges dans le groupe de cas et le groupe témoin très importantes

Cependant:

- D'autres études (SpF?) vont dans le même sens : absence de masques quand on mange ou boit, diffusion de gouttelettes

# Effectifs selon âge des personnes de l'étude

## Age group, yrs

18–29

T+: 44

T- : 39

30–44

T+: 46

T- : 62

45–59

T+: 46

T- : 35

≥60

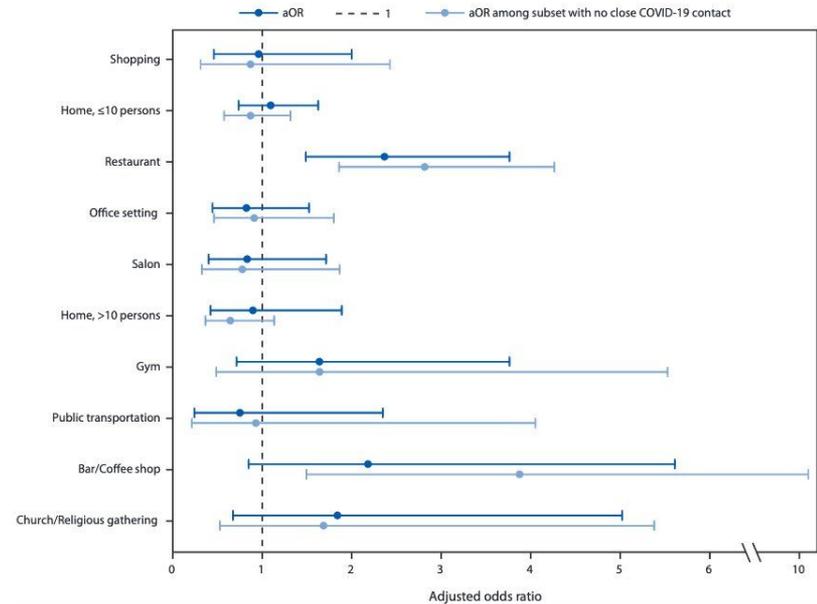
T+:18

T- :24

# Estimation d'une mesure inconnue

- On utilise les méthodes d'estimation statistique pour estimer une valeur inconnue:
- Exemple: le taux de retour des médecins généralistes concernant le nombre d'actes "Covid"
- Choix d'un échantillon représentatif
- Les résultats sont toujours présentés avec une double incertitude :
  - L'incertitude autour d'une mesure estimée : la mesure est donnée dans une fourchette.
  - Cette fourchette ne contient la valeur estimée qu'avec une certaine confiance.

- Une hypothèse est validée avec contrôle du risque de se tromper
- Cette incertitude est-elle compatible avec les mathématiques ?
  - Oui, même si les résultats sont donnés sous forme de probabilité;
  - L'incertitude fait partie du raisonnement statistique, mais elle est incompatible avec une communication médiatique simpliste ;
  - **Acceptons la complexité**



- Les mathématiques, la science statistique sont un outil de
  - rigueur et raisonnement,
  - modélisation,
  - aide aux décisions
- Mais :
  - On ne peut pas tout quantifier;
  - Il faut accepter ou refuser les bénéfices/risques en connaissance;
  - Il faut accepter la complexité : les incohérences comme la vérité sont (souvent) apparentes.

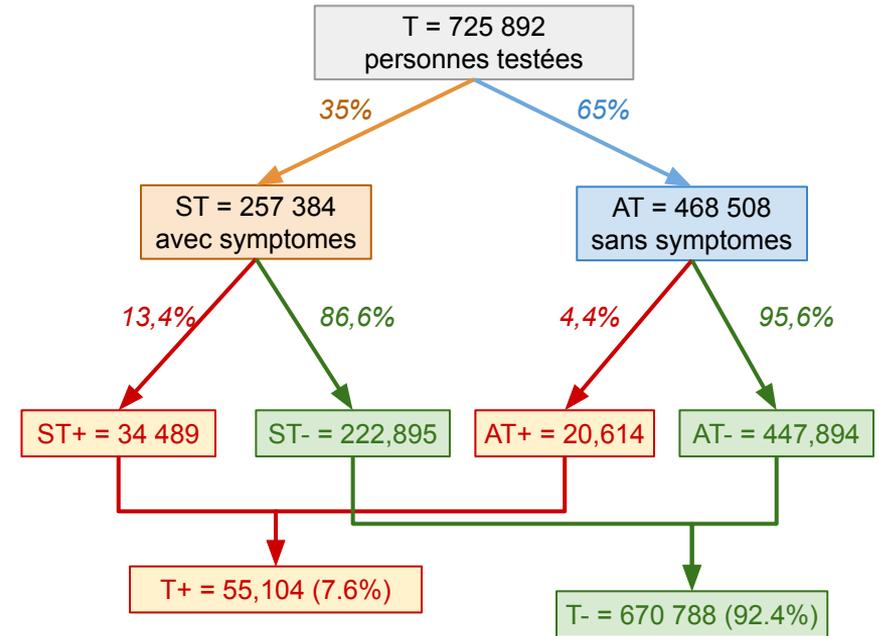
*«Il y a une fissure dans toutes les choses et c'est par là que la lumière entre» »  
(Eri De Luca)*

## *Matériel supplémentaire*

# Arbre de probabilités des taux de positivités (réel au 1/10)

Données Santé publique France (1/10/2020):

- “Parmi les testés : 65% se déclarent sans symptôme et représentent 38% des positifs.
- Le taux de positivité chez les symptomatiques est 13,4%. Le taux de positivité chez les sans symptômes est 4,4%.
- Parmi les positifs, 38% sont sans symptômes”



# Taux de reproduction $R$

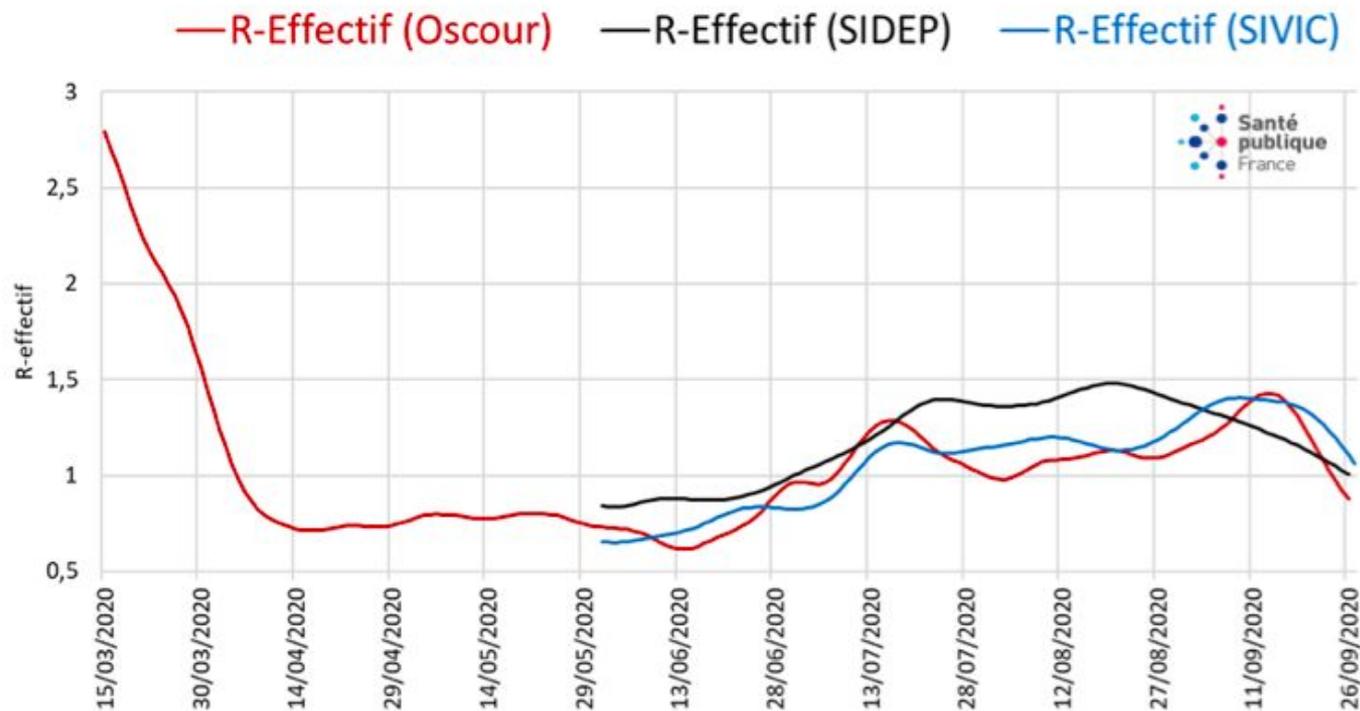
- C'est le nombre MOYEN de personnes qu'un individu infecté est susceptible de contaminer.
- Croissance exponentielle si  $R$  supérieur à 1
- Décroissance exponentielle si  $R$  inférieur à 1

Ce n'est qu'une moyenne ...

Si  $R=2$ : plus de 19000 contaminées en 10 itérations....théoriquement (décalage)

Hypothèse de SUPER CONTAMINATOR

Figure 24. Trajectoire du nombre de reproduction effectif (R effectif) à partir des tests PCR positifs au SARS-COV-2, des passages aux urgences avec suspicion de COVID-19 et des hospitalisations pour COVID-19 en France métropolitaine du 15 mars au 26 septembre 2020 (Sources : SI-DEP, OSCOUR® et SI-VIC)



# Létalités estimées



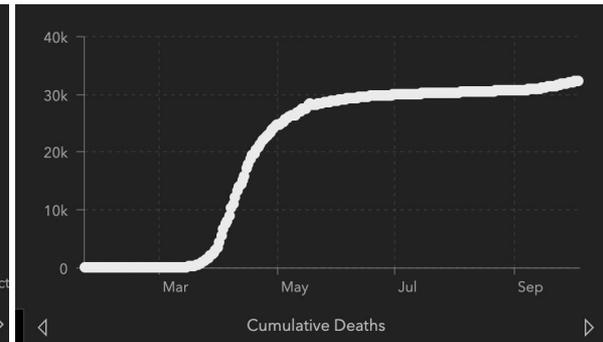
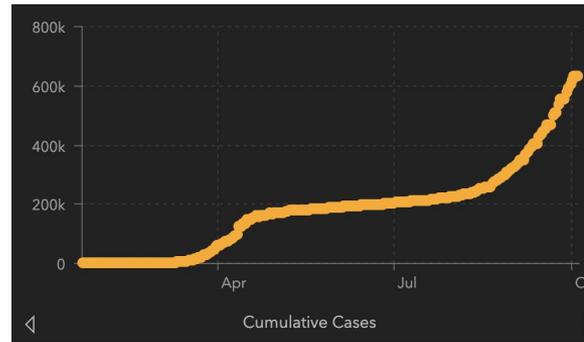
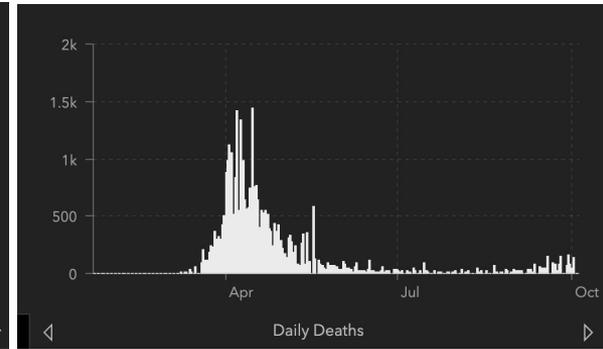
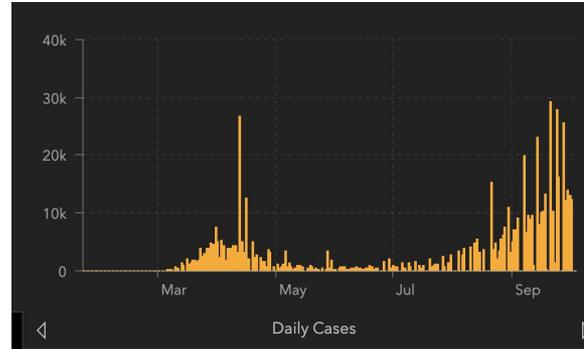
\* Estimation basée sur des données cliniques

\*\* Estimation basée sur une modélisation

\*\*\* Estimation basée sur une enquête sérologique

Le taux de mortalité (IFR) est la proportion de personnes infectées par le virus du Covid-19 qui vont mourir de la maladie. Les estimations portent sur des régions limitées, et peuvent varier en fonction de la démographie, du système de santé et de la méthodologie de l'étude.

# Nombre de cas et de décès en France (Johns Hopkins)

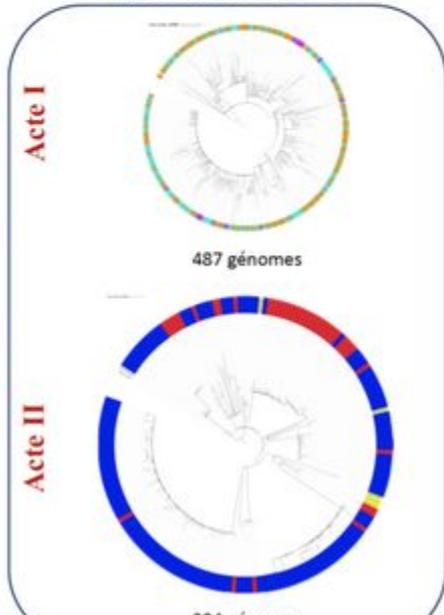


## 06/10/2020

**Coronavirus Monde**  
 Nombre de décès (pop. 7,8 milliards)

**1 042 954**

### Diversité des mutants / génome COVID-19



### Assistance Publique – Hôpitaux de Marseille (AP-HM) / IHU

**Acte I**  
 (du 27/01/2020 au 14/06/2020)

Nombre de tests réalisés

**141 240**

Nombre de patients AP-HM / IHU testés

**57 522**

Nombre de patients AP-HM / IHU positifs

**6 901 (9.0%)**

Nombre de décès

**164 (2.4%)** dont 40 en réanimation

Hospitalisation

**1 633 (23.7%)**

Réanimation

**288 (4.2%)**

### IHU Méditerranée Infection

Nombre de patients traités (Hydroxychloroquine ± azithromycine)

**3 070 (HDJ)**

Nombre de décès avec > 3 jours de traitement (Hydroxychloroquine ± azithromycine)

**18 (0.6%)**

### Assistance Publique – Hôpitaux de Marseille (AP-HM) / IHU

(du 15/06/2020 au 31/08/2020) **Acte II** (du 01/09/2020 au 05/10/2020)

Nombre de tests réalisés

**66 595**      **72 569**

Nombre de patients AP-HM / IHU testés

**36 807**      **36 871**

Nombre de patients AP-HM / IHU positifs

**2 485 (6.8%)**      **3 959 (10.7%)**

Hospitalisation Réanimation      Hospitalisation Réanimation

**171 (6.9%)**      **39 (1.6%)**      **252 (6.4%)**      **91 (2.3%)**

Nombre de décès

**21 (0.8%)**      **36 (1.0%)**

dont 1 en réa      dont 9 en réa

### IHU Méditerranée Infection

Nombre de patients traités (Hydroxychloroquine ± azithromycine)

**620 (HDJ)**      **1 767 (HDJ)**

Nombre de décès avec > 3 jours de traitement (Hydroxychloroquine ± azithromycine)

**1 (0.1%)**      **11 (0.6%)**

## 1) Sensibilité Se, spécificité Sp:

**La sensibilité Se : C'est le taux de personnes testées positives parmi les personnes infectées :**

**Se=**

(idéal : Se=100%)

**2) La spécificité Sp : C'est le taux de personnes testées négatives parmi les personnes saines :**

**Sp=**

Notations probabilités

- $Se = P(T+ | Inf)$
- $Sp = P(T- | Non Inf)$

**A partir de cet arbre de probabilité on a, par de simples multiplications :**

**Le pourcentage de vrais positifs : 3,5% (=70%.5%) ;**

**Le pourcentage de faux positifs : 1,9%**

**Le total de positifs : 5,4%**

## *Valeur prédictive positive VPP (et VPN)*

**La VPP est 65% donc il y a 35% sains parmi les positifs.**

**Un test positif se trompe une fois sur 3 en moyenne.**

**Si le test est positif, on teste de nouveau, c'est pourquoi le nombre de tests n'est pas égal au nombre de personnes.**

**La VPN=98,4%, très peu de faux négatifs**

Principe, si les tests sont assez sensibles:

Regrouper les écouvillons de 5 (ou 10 ou 50?) personnes testées et ne faire qu'un seul test avec un kit de dépistage.

Garder les prélèvements individuels:

- Si ce test unique est négatif, tous les tests sont négatifs on a économisé 4 (ou 9 ou 49) kits
- Si le test est positif, on teste alors les 5 et on a “gaspillé” un kit
- Bilan: on fait un test ou 6 tests
- Espérance du nombre de tests à faire selon la prévalence...