



Epidémiologie et COVID-19

28 avril 2020

Les questions scientifiques à résoudre concernant l'épidémiologie du COVID-19 sont simplement celles que tout citoyen se pose à propos de cette maladie. Simples à formuler, elles sont difficiles à résoudre :

- Combien y a-t-il de cas ? de morts ?
- Quelle est l'histoire « naturelle » de la maladie ?
- Comment doit-on traiter la maladie ?
- Quelle prévision faire sur la suite de l'épidémie de COVID-19 ?

1- Combien de cas ? Combien de morts ?

Les cas : Pendant des semaines, en France comme dans beaucoup de pays, on a caractérisé l'extension de l'épidémie en publiant les nombres de cas « identifiés », c'est-à-dire les nombres de sujets chez lesquels un test PCR avait été effectué et trouvé positif. Ceci n'avait aucune valeur statistique, faute de définition de la population testée, ne reflétant qu'une pratique (non formalisée) de tests qui – notamment- ne pouvaient prendre en compte les éventuels sujets asymptomatiques infectés.

Le bulletin hebdomadaire de Santé Publique-France¹ précise désormais bien les populations concernées pour lesquelles on dispose d'une information. Pour l'instant, les données sont recueillies à partir du réseau Oscour pour les hospitalisations avec « suspicion d'infection par COVID-19 » et, en ce qui concerne les patients en réanimation, à partir du SI-VIC (système d'Information pour le suivi des victimes) et 194 services sentinelles de réanimation. Signalons enfin que des données de « surveillance syndromique » sur l'infection telle qu'observée en médecine de ville sont en cours d'acquisition grâce au réseau Sentinelles et à SOS-Médecins.

En pratique, un document complet établi par l'INSEE, le ministère des solidarités et de la santé, et Santé Publique France² énumère l'ensemble des outils permettant le suivi de l'épidémie de COVID-19 en France en nombres de cas et de décès.

Santé Publique France fournit les estimations du nombre de sujets à ayant été hospitalisés pour COVID 19, et de décès liés à l'épidémie : à la date du 21 Avril il y a eu 82 916 sujets *hospitalisés pour COVID-19, et 20 796 décès*. Ces chiffres sont actualisés hebdomadairement sur le site de Santé Publique France. En revanche, le nombre total des personnes contaminées est inconnu, puisque –par exemple- l'immense majorité des (nombreux) asymptomatiques n'a pas été testée. Comment connaître ce nombre total ? évidemment, ce serait en disposant d'un test fiable qu'on appliquerait à de grands échantillons représentatifs de la population (voir la fiche de la Cellule de crise Corona de l'Académie des sciences relative aux tests). Faute d'une telle information directe, des estimations de ce nombre sont obtenues par modélisation ; elles indiquent que la fraction de la population française qui a été infectée serait très faible : le groupe de modélisation de l'Imperial College³ a indiqué le 30 Mars 2020

¹ COVID-19 : Point épidémiologique hebdomadaire, Santé publique France, 23 Avril 2020

² Le suivi de l'épidémie de COVID-19, INSEE, Mars 2020

³ Flaxman et col., <https://doi.org/10.25561/77731>, Mars 2020



que ceci concernerait entre 1.1% et 7.4% de la population française. Deux travaux français très récents non encore publiés, mais publics, ont analysé l'ensemble des données françaises : l'un⁴ estime (au 5 Avril) que la proportion nationale d'infectés serait comprise entre 1 et 6% ; l'autre⁵ fournit des estimations par région et indique que la fraction de personnes infectées varie d'environ 12% (avec un intervalle possible entre 7% et 21%) en Ile-de-France et dans le Grand Est, à moins de 2% en Ile-de-France et Bretagne. Cette fraction est plus importante chez les hommes que chez les femmes.

Les décès : Les données publiées relatives aux décès par COVID-19 furent, au début de l'épidémie, incomplètes, omettant par exemple la part importante des décès se produisant dans les EHPAD. La raison en était le mode de recueil de l'information qui n'était en place que dans les hôpitaux. Les sources d'information sont diverses, y compris en qualité. Une source directe, et spécifique du COVID-19, est celle des services de réanimation (cf. plus haut). Une application permettant d'avoir l'information dans les EHPAD et établissements médico-sociaux vient d'être mise à leur disposition. Par ailleurs, l'évolution quotidienne de la mortalité « toutes causes » est également étudiée, s'appuyant sur les systèmes généraux de déclaration des décès : celui de l'échantillon de 3 000 communes suivie par l'INSEE, celui des certificats de décès déclarés informatiquement à l'Inserm par les médecins (mais ceci concernerait moins de 20% des décès...). Notons que la simple analyse des évolutions de la mortalité toutes causes, et sa comparaison avec les années précédentes nécessitent des méthodes statistiques avancées : par exemple, la surmortalité due au COVID-19 en 2020 ne peut être évaluée qu'en estimant celle due à la grippe, qui connaît de grandes variations d'une année à l'autre.

Pour l'instant, il n'est pas possible de donner des chiffres précis pour les taux de mortalité par COVID-19. On connaît à peu près les nombres de morts (seulement à peu près, car par exemple dans les EHPAD, le diagnostic n'est pas toujours posé, ni certifié). Mais on ne connaît pas le dénominateur de ces taux, c'est à dire la taille de la population à risque à laquelle ces morts doivent être rattachés. Ceci est la conséquence de l'existence –en proportion inconnue- de sujets porteurs asymptomatiques du COVID-19. L'information sur la mortalité du COVID publiée par le Centre chinois de surveillance des maladies transmissibles s'est appuyée sur les 72 314 cas observés avant le 11/2/2020 : elle mène à une estimation du taux de mortalité général de 2,3% (très inférieur, donc, à celui du SARS en 2003-2004) ; aucune mort en dessous de l'âge de 9 ans ne fut trouvée, 8% chez les 70-79 ans, et 14,8% chez les plus de 80 ans. En France, les résultats récents du groupe de modélisation de l'Institut Pasteur, trouvent une mortalité générale de 0,53% des infectés (plus forte chez les hommes que chez les femmes, et variant de 0,001% chez les moins de 20 ans, à 8,3% chez les plus de 80 ans. Remarquons que la forte mortalité par COVID-19 chez les sujets âgés peut être mise en parallèle avec les chiffres de mortalité générale, en dehors du COVID-19 : ainsi, les données 2018 de l'INSEE indiquent que le taux de mortalité d'un français de 45 ans était de 0,4%, et celui d'un français de 85 ans était de 8,3%, soit plus de 20 fois supérieur.

2- L'histoire naturelle de la maladie

En épidémiologie des maladies transmissibles, on partitionne classiquement la population en sujets « susceptibles » (S), « exposés » (E), infectieux (I) et guéris (R, pour « recovered »). Les sujets S peuvent être contaminés (d'où le terme « susceptibles »); les sujets E sont contaminés, mais pas encore infectieux. Les sujets R sont supposés immunisés. On désigne par « période de latence » le temps passé dans le stade E (exposé), par période infectieuse celui passé dans le stade I (infectieux). La « durée d'incubation » est le temps qui s'écoule entre l'infection et l'apparition des premiers symptômes.

⁴ Di Domenico et col., in www.epicx-lab.com/covid-19.html, rapport 9 du 12 Avril 2020

⁵ Salge et col., <https://hal-pasteur.archives-ouvertes.fr/pasteur-02548181>, archivé le 20 Avril 2020



Évidemment, il s'agit là de simplifications qui doivent être discutées : quand déclare-t-on qu'un patient est « guéri », par exemple ? De plus il y a une énorme variabilité interindividuelle. La durée d'incubation du COVID-19 ne peut être résumée à sa valeur moyenne. Il faut connaître sa distribution statistique, notamment la forme de cette distribution (est-elle unimodale ?) et identifier les cofacteurs qui influencent cette durée. Pour cela, il faut des méthodes statistiques, et avant tout un recueil soigneux de données représentatives. Les données nécessaires seraient celles de « cohortes » de sujets infectés dont on connaîtrait le plus précisément possible le moment de contamination. Ceci est évidemment difficile et coûteux. Des situations très particulières fournissent des sources précieuses d'information. Ainsi, le cas du *Diamond Princess*, bateau de croisière mis en quarantaine au Japon en février, avec 3 711 passagers est souvent cité. Il a permis en particulier de montrer que (sur ce bateau, et dans cette population de croisiéristes), il y a eu 18% d'asymptomatiques parmi les personnes infectées (qui furent 712, environ $\frac{1}{4}$ des passagers). Plus récemment ⁶, la proportion d'asymptomatiques parmi des infectés jeunes a été trouvée encore plus forte, supérieure à 50% dans l'équipage du porte-avions Charles De Gaulle (2 100 marins testés sur 2 300, dont 1 081 ont été trouvés positifs).

L'analyse la plus récente⁷ de la durée d'incubation indique que sa valeur moyenne est d'environ 5 jours (comme c'est le cas dans le SARS), et que 95% des sujets étudiés (181 personnes, symptomatiques, de la région d'Hubei) avaient eu une durée d'incubation comprise entre 4,5 et 11,6 jours (avec des incertitudes). Il s'agit de résultats précieux acquis avec les bonnes méthodes biostatistiques par une équipe confirmée. Cependant, on ne peut savoir comment peuvent varier ces résultats en fonction des caractéristiques individuelles des patients (telles : l'âge, la sévérité de l'infection, leurs facteurs de risque...). L'absence de connaissance précise est bien sûr encore plus grande sur la période de latence (temps passé dans le stade E), ou la durée d'infectiosité. On a –en revanche- des informations sur l'ordre de grandeur de la durée s'écoulant entre infection et hospitalisation, et sur la durée passée en réanimation (de l'ordre de 15 à 20 jours, ce qui fait comprendre qu'il faut attendre longtemps avant d'espérer observer l'effet du confinement). Mais là non plus, on ne dispose pas d'information épidémiologique permettant de prédire ces durées en fonction de caractéristiques individuelles des patients.

3- Comment doit-on traiter la maladie ?

L'épidémiologie est en principe la clé pour (tenter de) répondre à cette question : c'est elle qui a développé l'approche de médecine fondée sur les preuves (« evidence based medicine ») qui rationalise les décisions thérapeutiques. Des essais thérapeutiques européens et français concernant le COVID-19 sont actuellement en cours, et leurs résultats sont attendus bientôt. Mais l'épidémie de COVID-19 est particulière de deux points de vue :

- d'abord, positivement, il semble que les délais et procédures nécessaires à décision, et à la mise en œuvre de ces essais thérapeutiques aient été considérablement améliorés par rapport à la pratique habituelle dont on regrette souvent les lenteurs.
- ensuite, au moins en France, on a observé chez certains une remise en cause de l'« essai randomisé », au motif qu'il induit –ou induirait- des retards par rapport à des évaluations empiriques en essais (et erreurs !). Constituer un groupe non traité serait de plus inacceptable au vu du risque immédiat. Cette façon de voir a reçu des soutiens forts auprès d'une partie de l'opinion publique en dehors du monde scientifique, beaucoup moins souvent en son intérieur.

⁶ Chiffres de la Ministre des armées, le 17 Avril 2020

⁷ Lauer et coll, *Annals of Internal Medicine*, 20 Mars 2020



Les essais multicentriques comparatifs devraient apporter des éclaircissements sur les molécules testées et sur le moment de la maladie où elles auront été testées, précoce ou tardif. Les mécanismes combattus ne sont en effet pas les mêmes selon qu'on essaie de limiter l'installation du virus dans l'organisme ou le développement dramatique, une fois installé, de ses effets inflammatoires suraigus.

4- Quelle prévision faire sur la suite de l'épidémie de COVID-19 ?

La modélisation est le seul outil de prédiction de l'avenir d'une épidémie. Les modèles de prévision se font grâce aux données disponibles dans le présent. La situation actuelle du COVID-19 n'est pas meilleure du tout que celles des autres pathologies étudiées dans le passé, on vient de le voir : données éparses, sources insuffisantes, manque de connaissances sur l'histoire naturelle de la maladie, etc., limitent dès le départ la qualité des modélisations. Au mieux, le défenseur de ces méthodes dira que c'est la meilleure façon d'utiliser l'information disponible, aussi faible soit-elle. Par ailleurs, la culture mathématique des biologistes et médecins, beaucoup plus grande que dans le passé, leur interaction avec les collègues mathématiciens et physiciens, et –enfin– l'accès généralisé à des outils informatiques puissants expliquent la grande fréquence de modèles publiés sur cette épidémie. Une culture de la modélisation des épidémies s'est répandue, y compris en dehors du monde des épidémiologistes. Une preuve en est par exemple le R_0 , un paramètre de la diffusion des épidémies, qui est désormais souvent évoqué même dans la presse de grande audience et est perçu comme la signature de la gravité d'une épidémie. R_0 est le « nombre de reproduction » d'une épidémie. R_0 est le « nombre de cas secondaires » qu'une personne infectieuse produit lorsqu'elle est introduite dans une population totalement susceptible. Une épidémie ne peut démarrer que si $R_0 > 1$. Une épidémie de grippe est souvent caractérisée par un R_0 compris entre 1,5 et 2,5. Les R_0 publiés pour le SARS étaient un peu supérieurs à 1,5. La rougeole est connue pour avoir un R_0 élevé (10 à 20). Le nombre de reproduction R (ou R_t , pour signaler qu'on est au temps t) a la même définition que R_0 , mais l'infectieux est introduit dans la population telle qu'elle est à l'instant t . Du coup, maîtriser une épidémie (par exemple, via du confinement), c'est faire tomber R_t au-dessous de 1. On montre que dans une population théorique où l'agent infectieux a une valeur R_0 , lorsque la fraction de personnes ayant été infectées atteint $1 - 1/R_0$, le nombre de reproduction devient < 1 , et l'épidémie s'éteint.

Les méthodes d'estimation de R_0 sont multiples, selon la nature des données disponibles.

Au début de l'épidémie de COVID-19, le R_0 de l'épidémie fut trouvé aux alentours de 2 : l'OMS, le 23 Janvier 2020 annonça entre 1,4 et 2,5. Li⁸ publia (à partir de l'analyse de 350 villes chinoises) une valeur de 2,38... Mais au fur et à mesure que l'épidémie avance, de nouvelles valeurs beaucoup plus élevées sont trouvées, ou supposées. Ainsi, l'équipe de l'Imperial College analysant les données de 11 pays européens a récemment trouvé une valeur R_0 de 3,87 [3,0 à 4,7]. Les deux équipes françaises dont les travaux sont sous presse trouvent des valeurs comparables : le premier trouve $R_0 = 3,0$ (intervalle de précision : entre 2,8 et 3,2). Celui du groupe de modélisation de Pasteur trouve 3,3. Si ces valeurs sont exactes, alors le % d'immunisés par l'infection à rechercher pour obtenir l'« immunité de groupe » et contenir ainsi l'épidémie serait de l'ordre de 70 à 75%, un chiffre impossible à atteindre naturellement compte tenu des chiffres actuels d'infection estimés (cf. plus haut). Notons en revanche qu'un résultat de l'étude récente du groupe de modélisation de l'Institut Pasteur est encourageant : selon lui, le taux de reproduction R prédit à la date du 11 Mai serait de 0,5, donc nettement en dessous de 1.

Deux remarques peuvent cependant être faites à propos du R_0 dans le contexte de l'épidémie de COVID-19 : d'abord, si on prend l'exemple du SARS, on a remarqué qu'un déterminant majeur de l'épidémie a été l'existence de « super-contaminateurs » (« super spreading events »), lesquels sont

8 Li et col., Science, 16 March 2020



négligés dans le modèle du R_0 . Y-a-t-il des super-contaminateurs dans le COVID-19 ? il semble que oui, même si les données précises manquent. Ensuite, et surtout, R_0 n'est pas le seul paramètre à prendre en compte pour prévoir la possibilité de contrôle d'une épidémie : un paramètre clé⁹ est θ qui est le % du temps d'incubation de la maladie où le sujet est contagieux. Dans le cas du SARS, θ était petit (de l'ordre de 0,11), d'où le succès des méthodes de contrôle qui furent utilisées. Dans le cas de la grippe (qui a un R_0 comparable à celui du SARS), θ est grand et on ne peut maîtriser l'épidémie par isolation, quarantaine des cas, etc. Dans le cas du COVID-19, on ne dispose pas de l'information précise sur la valeur de θ , faute d'histoire naturelle documentée de la maladie, mais tout indique -hélas- que les sujets sont contagieux avant leurs premiers signes, rendant leur contrôle difficile. De plus l'existence probable d'un grand nombre d'asymptomatiques, néanmoins contagieux, complique beaucoup ce modèle.

En conclusion

Nous en sommes à la naissance de l'épidémie de COVID-19. Force est de constater que l'on dispose encore de bien peu de données épidémiologiques sur cette maladie : les données critiques particulièrement recherchées sont celles de l'histoire naturelle de la maladie (et en particulier en ce qui concerne la phase infectée-asymptomatique). On n'a pas non plus de registres de malades permettant de calculer les pronostics individuels en fonction de facteurs de risque génétiques, ou environnementaux (par exemple, un rôle protecteur de la nicotine est répétitivement indiqué). Cette absence de registres et de cohortes est sans doute une situation habituelle pour une nouvelle maladie. Cependant, il y a une course à mener si l'on veut répondre assez vite aux questions simples que cette fiche abordait. C'est ce qui se produit, et plusieurs cohortes, avec comité d'accès aux données, sont en cours de développement dans le monde, et en France (notamment la cohorte French COVI19 mise en place par REACTing).

Cette fiche a été conçue et rédigée par la cellule de crise Coronavirus de l'Académie des sciences. Créée à l'initiative de Pascale Cossart, Secrétaire perpétuel de l'Académie, celle-ci réunit des académiciens experts du domaine : Jean-François Bach, Pierre Corvol, Dominique Costagliola, Pascale Cossart (coordinatrice), Patrick Couvreur, Olivier Faugeras, Daniel Louvard, Félix Rey, Philippe Sansonetti, Alain-Jacques Valleron.

Les informations qui figurent sur cette fiche ont été produites collectivement et sont susceptibles d'évoluer. Elles seront éventuellement réactualisées en fonction des avancées des connaissances scientifiques.

⁹ Fraser et col., PNAS, 101, 6141-6146