

LES VRAIS CHIFFRES DU CORONAVIRUS



Antoine Jardin

14/05/2020

De nombreux indicateurs ont été utilisés pour suivre l'évolution de la pandémie due au coronavirus qui se propage à travers le monde. La tentation est grande de considérer que l'on ne connaît pas « les vrais » chiffres de la pandémie (nombre de cas, nombre de décès, nombre de masques...). Antoine Jardin, chercheur au Cesdip et enseignant à Sciences Po et à l'Université de Versailles Saint-Quentin, revient sur ce que sont les indicateurs de l'épidémie, comment ils sont mesurés et compilés, mais aussi comment ils sont interprétés et utilisés dans la définition des politiques publiques.

Compter les cas

Un des premiers enjeux du suivi et de la réponse à l'épidémie implique de dénombrer les personnes infectées par le virus. Dans un premier temps, ce comptage s'est effectué à partir de la démarche des malades qui se sont présentés aux services sanitaires pour recevoir des soins. Les premiers comptages consistaient à compiler le nombre de cas reçus et, dans un second temps, à dépister dans leur entourage les personnes susceptibles d'être également contaminées. Les caractéristiques de la propagation du Covid-19 rendent difficile cette opération apparemment anodine. Les chercheurs se sont rapidement aperçus que de nombreuses personnes infectées pouvaient transmettre le virus avant de déclarer des symptômes de la maladie, mais aussi que les symptômes étaient parfois si faibles que les personnes infectées ne sollicitaient pas d'aide médicale auprès d'un médecin ou d'un service d'urgence et, enfin, qu'une part indéterminée des personnes atteintes ne présentait parfois aucun symptôme.

Au-delà de ces difficultés inhérentes, qui amènent à une sous-estimation probable du nombre de personnes malades sur un territoire donné à différents moments, les stratégies de dépistage varient d'un pays à l'autre, provoquant des mesures difficilement comparables, parfois même d'une institution à l'autre (en France, par exemple, certains centres sont en mesure d'effectuer des tests virologiques chez une plus grande part des cas suspects que d'autres, les dépistages n'ont pas pu dans un premier temps être disponibles dans les laboratoires de biologie

médicale ou dans les Ehpad).

La difficulté à réaliser systématiquement des tests s'est posée dès l'émergence du virus à Wuhan et dans la province de Hubei. Dans le courant du mois de janvier dernier, les médecins chinois ont changé leur méthode de comptage des cas, en incluant des malades non testés mais dont les symptômes cliniques, détectés à partir de scanner thoraciques, étaient conformes au Covid-19. Selon la disponibilité en kits de tests dans le temps et dans l'espace, la systématique des tests a pu évoluer. En conséquence, un même indicateur peut connaître des transformations telles qu'il ne mesure pas la même chose à différentes périodes et dans différents lieux.

À cela s'ajoute le fait que les tests de détection de la présence du virus dans l'organisme ne sont, comme toutes les procédures, pas des tests « parfaits ». En effet, ils peuvent ne pas déceler une personne contaminée lors des premiers stades de la phase d'incubation. Par ailleurs, ils ne sont pas en mesure d'indiquer si un individu a été contaminé par le passé. Des incertitudes demeurent par ailleurs sur l'immunité acquise par les personnes infectées mais guéries, dont le corps émet toujours des fragments de virus et qui paraissent, pour cette raison, comme « infectées » lors de ces tests. Une étude récente du centre de prévention des maladies de Corée du Sud portant sur 250 personnes a révélé que ces individus n'étaient plus contaminants même s'ils présentaient toujours des fragments de virus dans leur corps.

Les statistiques des tests : un outil à manier avec rigueur

Face à cette situation, les tests sérologiques paraissent, à première vue, comme un outil idéal, qui permet de décompter la part totale de la population ayant déjà été infectée par le virus à partir de la présence d'anticorps dans le sang. Cependant, ces tests sont produits par de très nombreuses sociétés, avec des caractéristiques différentes. Face à l'urgence et à la forte demande, les procédures de mise sur le marché de ces tests ont été rendues moins contraignantes, notamment aux États-Unis. Une liste régulièrement mise à jour de ces tests est publiée par l'université Johns-Hopkins. On distingue notamment deux caractéristiques importantes : d'une part la sensibilité, d'autre part la spécificité du test. La sensibilité renvoie au fait d'être capable de détecter une personne infectée comme l'étant effectivement (autrement dit, une faible sensibilité est associée au risque de faux négatif, c'est-à-dire de personnes testées comme négatives alors qu'elles sont en réalité contaminées). La spécificité, elle, renvoie au fait d'être capable de bien garantir qu'une personne testée positive est véritablement infectée. La spécificité renvoie donc au risque de faux positifs, c'est-à-dire le fait de considérer une personne comme ayant été infectée alors qu'elle ne l'a pas été. Les valeurs de sensibilité et de spécificité peuvent avoir un impact considérable lorsque le taux de personnes infectées dans la population globale (que l'on nomme « taux de prévalence »)

est faible. Ainsi, l'une des premières études sérologiques, réalisée dans le comté de Santa Clara en Californie, a d'abord fait grand bruit en présentant le nombre de personnes déjà contaminées comme très important, entre cinquante et quatre-vingts fois le nombre de cas mesurés à partir des tests PCR. Cependant, les travaux du statisticien de l'université de Columbia, Andrew Gelman, ont montré par la suite que ces premiers résultats n'étaient pas fiables, en raison de la marge d'erreur sur la sensibilité et la spécificité du test utilisé.

Chaque étude locale est donc particulière, il est difficile de les comparer terme à terme. Les travaux des différentes équipes apparaissent donc comme les pièces mélangées de différents puzzles et ne peuvent s'assembler de façon précise ni évidente. Par la comparaison des différents articles publiés, les chercheurs espèrent parvenir à conjecturer, par tâtonnement, des fourchettes de valeurs raisonnables pour des paramètres de l'épidémie que l'on ne parvient pas à connaître parfaitement. Si de nombreuses visualisations, tableaux de bord et comparaisons ont été effectués à partir du chiffre des personnes contaminées chaque jour dans les différents pays du monde touchés par la pandémie, la rigueur de ces informations est donc toute relative, appelant à une grande prudence d'interprétation et rendant par là même difficile la définition des stratégies et la mise en œuvre des plans de lutte contre la diffusion du virus.

Compter les décès

Une statistique *a priori* plus facile à constater que celle des infections est la mesure des décès provoqués par le virus. D'une part, si l'infection est souvent susceptible d'être pas ou très peu symptomatique, les décès sont, par définition, visibles et recensés. L'existence d'un état civil dans la plupart des pays du monde amène à compter en continu le nombre et l'identité des individus qui perdent la vie. Le point hebdomadaire réalisé par Santé publique France dans notre pays compile les statistiques des décès en milieu hospitalier, dans les Ehpad et dans les établissements médicaux sociaux (EMS). En revanche, les décès au domicile ne sont pas inclus dans ces données. Pour produire une estimation complémentaire de l'impact de la propagation du virus sur la mortalité, l'Insee a diffusé le nombre de décès recensés quotidiennement par l'état civil. Par comparaison avec la mortalité des années précédentes, on peut mettre en évidence une surmortalité associée au Covid-19. Cependant, cette mesure n'est pas une mesure totalement précise des décès attribuables au Covid-19 non plus dans la mesure où d'autres causes de décès ont décliné sous l'effet du confinement. Ainsi, la mortalité routière a chuté au mois de mars 2020 par rapport au mois de mars 2019 (passant de 255 à 154 décès). Mais d'autres causes de décès ont pu connaître une évolution différente, si l'on pense par exemple aux personnes ayant connu des accidents vasculaires cérébraux (AVC) ou des accidents cardiaques sans solliciter une prise en charge médicale. Par ailleurs, mesurer l'impact du coronavirus pourrait impliquer le décompte des

décès directement causés par cette maladie, mais aussi des décès indirects liés à l'encombrement des services hospitaliers ou au renoncement aux soins. Enfin, si nous comptons aujourd'hui les décès survenus à court terme suite à l'infection, plusieurs travaux médicaux indiquent que le Covid-19 est susceptible de laisser des séquelles, tout comme les longs séjours des malades dans les services de réanimation. Ainsi, la mortalité à long terme pourrait être plus importante que celle mesurée aujourd'hui.

On perçoit rapidement qu'aucun indicateur n'est parfait et que la diversité des sources et des méthodes rendent particulièrement difficiles les comparaisons dans l'espace et parfois même dans le temps. L'impossibilité de définir exactement les nombres de cas et de décès n'est pas un problème spécifique au Covid-19, c'est une caractéristique globale des épidémies. Mais le problème se pose de façon plus flagrante dans le cas d'une pandémie se propageant rapidement, désorganisant profondément et durablement les sociétés, pour laquelle il n'existe ni traitement ni vaccin et dont l'émergence toute récente implique une mauvaise connaissance du virus.

Modéliser dans l'incertitude

Ne pouvant pas mesurer précisément les paramètres essentiels de la propagation du virus, les modélisateurs qui tentent de calculer les caractéristiques et l'évolution de la maladie font face à de nombreuses incertitudes et paramètres inconnus. C'est finalement à tâtons, en calant les modèles théoriques sur des données éparses que sont réalisées les projections sur la trajectoire de l'épidémie. De nombreux travaux s'inspirent notamment des taux de mortalité mesurés à bord du bateau de croisière Diamond Princess dans lequel les passagers, tous dépistés, ont été confinés alors que le coronavirus circulait à bord. L'expérience récente du porte-avions américain Theodore Roosevelt et du groupe aéronaval du Charles-de-Gaulle fournissent des données de même nature. Cependant, dans chacun de ces cas, les passagers ne présentent pas des caractéristiques démographiques et médicales représentatives de la population générale. L'effet de l'âge semble particulièrement important, aussi il est difficile de comparer les scénarios de diffusion du coronavirus dans un pays comme l'Italie, où près de 22,6% de la population est âgée de soixante-cinq ans et plus et un pays comme l'Afrique du Sud où ce groupe ne représente que 5% des habitants.

Un adage célèbre attribué au statisticien britannique George Box indique que tous les modèles sont faux mais que certains sont utiles. Les positions dogmatiques qui visent à voir dans ces techniques soit des vérités révélées soit des mirages mathématiques sont donc également fausses et dangereuses. Certaines caractéristiques des pratiques sociales donnent aujourd'hui particulièrement du fil à retordre aux statisticiens, notamment la mesure de la densité des

relations sociales (qui évoluent selon le comportement des gens), tout comme les pratiques d'hygiène.

Tous ces critères doivent amener à une grande rigueur et à une grande prudence dans la présentation des résultats. Ainsi, chaque modèle est assorti d'un seuil de confiance et d'une marge d'erreur quand on estime la part de la population déjà infectée ou le nombre de contaminations quotidiennes attendues chaque jour à partir du 11 mai 2020. Ces marges parfois importantes sont souvent négligées dans la diffusion journalistique des travaux scientifiques, ce qui contribue à la fois à la confusion et à la méfiance du grand public tout en laissant un espace aux théories complotistes. Ainsi, la principale publication estimant le nombre de personnes infectées en France, réalisée et publiée par l'Institut Pasteur, produit une estimation de 3,7 millions de personnes infectées, avec une marge d'erreur au seuil de confiance de 95% allant de 2,3 à 6,7 millions d'individus, un écart du simple au triple.

Du besoin d'une science véritablement ouverte

On l'a vu, l'émergence du SARS-CoV-2 a donné suite à une effervescence scientifique considérable, les travaux de virologie et d'épidémiologie du monde entier ont été convoqués pour tenter d'apporter leur pierre à l'édifice. Le mouvement de diffusion des publications « preprint », c'est-à-dire mises en ligne sans avoir été vérifiées par d'autres chercheurs ni publiées dans les revues spécialisées, s'est rapidement développé pour faire face à l'urgence de la situation. Dans un contexte de crise, chacun tente de développer des travaux rapidement, d'obtenir des résultats dans les plus brefs délais et de les rendre disponibles au public le plus large. C'est en soi une bonne chose, et un des aspects positifs de la crise du coronavirus est d'avoir stimulé un travail scientifique très réactif, sans précédent à l'échelle mondiale.

Mais publier les résultats et présenter les conclusions d'une étude ne sont pas des pratiques suffisantes.

Encore faut-il que les autres chercheurs puissent continuer à vérifier, juger et évaluer les travaux de leurs pairs et collègues de par le monde. Nous souffrons aujourd'hui d'une grande opacité sur les données et les protocoles de recherche. Ainsi, il paraît indispensable de demander que toutes les enquêtes financées par de l'argent public rendent disponibles les données brutes à partir desquelles sont effectués les calculs et obtenues les conclusions. Ce mouvement de science transparente est devenu indispensable. Si les controverses scientifiques et le développement de théories, voire d'interprétations, rivales sont inévitables et même nécessaires au progrès de la connaissance, l'opacité n'aide en rien à ce que l'on puisse sélectionner au cours du temps les

résultats les plus robustes et les études les plus complètes et les plus fiables. S'il existe bien une logique de compétition, qu'il s'agisse des études, de la capacité de compréhension, d'anticipation, d'identification de traitement ou de développement d'un vaccin, c'est bien, au bout du compte, par une démarche cumulative et empirique que progresse notre connaissance scientifique.

La crise du coronarivus montre que les indicateurs apparemment évidents ne le sont pas toujours, que la mesure de la situation sociale exige toujours des opérations et des choix, parfois instables dans le temps et dans l'espace. Si la recherche comparée est extrêmement riche et utile, elle est également difficile à développer. De plus, la crise du Covid-19 va nécessairement impacter de très nombreux indicateurs (taux de chômage, travail de la police et de la gendarmerie, réponse pénale, participation électorale), dont l'évolution va mettre en question les procédures d'évaluation des politiques publiques, car leurs valeurs pour les années à venir seront fortement perturbées par les événements d'aujourd'hui. Répondre à l'urgence sanitaire et sociale révèle ainsi la nécessité de travailler, avec les indicateurs les plus stables et les plus maîtrisés possible, sur le temps long. Elle pointe également la nécessité de rapprocher les citoyens et les élus de la démarche scientifique, si l'on souhaite que celle-ci puisse être comprise et acceptée dans les sociétés démocratiques.