

Covid-19 : « Il nous manque des données essentielles sur les personnes testées positives »

Tribune

Thomas Breda

Economiste

L'économiste Thomas Breda avance, dans une tribune au « Monde », que le seul moyen d'arrêter d'imposer aux Français « des interdits drastiques sur la base de présomptions » est de connaître l'environnement et la situation personnelle des personnes contaminées.

Publié le 30 novembre 2020 à 02h35 - Mis à jour le 30 novembre 2020 à 06h59 Temps de Lecture 5 min.

Tribune. Les gens se contaminent-ils dans les restaurants ? Dans les bars ? Au supermarché ? A l'école ? Dans les taxis ? Au travail ? A l'heure où le déconfinement se dessine, ces questions sont déterminantes pour décider ce qu'on peut de nouveau autoriser et ce qu'il faut encore interdire ; et nombreux sont ceux qui se demandent pourquoi, à l'heure du big data et malgré tous les efforts qui semblent déployés, on n'a toujours pas de réponses claires. En tant que chercheur dont le métier consiste principalement à analyser des données statistiques, je tente ici de proposer quelques explications.

Il y a essentiellement deux approches pour comprendre la propagation du virus. La première est médicale et biologique. Il s'agit d'étudier le virus directement, ses propriétés, par exemple sa survie sur différents matériaux, dans l'air, dans les lieux clos... Ou encore la contagiosité des personnes avec et sans symptômes, des enfants, etc. A partir des connaissances qui progressent doucement grâce à la recherche médicale, on commence à se faire une idée des risques principaux de contamination.

Examiner qui a été infecté

La seconde approche est épidémiologique à partir de données statistiques. Il s'agit d'examiner qui a contracté le virus (et quand) afin d'en tirer des conclusions. Le principe est simple : si les enfants propagent fortement le virus parmi les adultes, alors on devrait voir que les parents ont été davantage infectés, et plus tôt dans les vagues épidémiques. Idem pour les restaurants : s'ils sont des lieux de propagation, alors les gens qui ont eu le virus devraient être plus nombreux que la moyenne à être allés au restaurant la semaine précédant leur infection. Et pour savoir si les masques les protègent effectivement, on peut examiner si les serveurs sont surreprésentés parmi les personnes ayant contracté le virus récemment.

Il y a bien sûr quelques précautions à prendre et des pièges méthodologiques à éviter, mais ces intuitions peuvent assez vite se transformer en certitudes quantifiées si l'on calibre les modèles adéquats sur des données suffisamment riches.

Cette seconde approche statistique est essentielle pour confirmer les résultats de la recherche médicale. En effet, certaines situations peuvent être en théorie très propices à la propagation du virus, mais en pratique s'avérer peu dangereuses. [C'est le cas des écoles, par exemple](#) : même si les enfants sont asymptomatiques, des études ont montré qu'ils contractent aussi le virus et peuvent en sécréter une quantité importante. Pour autant, il ne semble pas que la propagation se fasse massivement par les enfants, pour des raisons que la recherche médicale semble encore mal comprendre.

Interroger un large échantillon de population

Le seul moyen de transformer des intuitions en certitudes, et donc d'arrêter d'imposer à 67 millions de Français des interdits drastiques uniquement sur la base de présomptions, c'est d'avoir des données individuelles sur les personnes ayant contracté le virus, permettant de savoir ce qu'elles font en général (métier, situation familiale, etc.) et, si possible, ce qu'elles ont fait juste avant d'avoir été diagnostiquées positives. Or, nous n'avons toujours pas de telles données.

Pourtant, ce n'est pas si compliqué. Il y a essentiellement deux moyens de récolter ces données. La première, et aussi la plus coûteuse, consiste à constituer un large échantillon représentatif de la population et à lui poser tout un tas de questions à plusieurs moments dans le temps, notamment concernant la santé et les symptômes du Covid-19. Cela s'appelle une « cohorte épidémiologique ».

De ce côté-là, la France fait figure de bonne élève : l'Institut national de la santé et de la recherche médicale (Inserm) et le ministère de la santé ont lancé sur ce modèle, dès le mois d'avril, l'enquête EpiCov auprès de 200 000 personnes, dont 100 000 auront des tests sérologiques permettant ainsi de détecter aussi les asymptomatiques.

Cela devrait permettre d'apprendre beaucoup de choses sur la propagation réelle du virus. Cependant, du fait de logiques de compétition entre institutions publiques, un délai de neuf mois a été retenu avant que l'ensemble de la communauté des chercheurs puisse accéder aux données, même dans un cadre très contrôlé (ce qui est toujours le cas pour les recherches à partir de données sensibles). A l'heure des confinements à répétition, ne devrait-on pas plutôt tout faire pour aider les chercheurs à collaborer et à travailler tous ensemble le plus tôt possible sur ces questions si urgentes ?

Comprendre comment le virus se propage

Venons-en à la seconde approche. Il s'agit cette fois de récolter des données un peu riches directement sur les personnes positives au SARS-CoV-2 : leurs métier, situation familiale, activités récentes, comment elles pensent avoir contracté la maladie, etc. Cela n'a pas été fait : on se contente de descendre l'arbre des contaminations en cherchant les cas contacts, mais on ne fait pratiquement rien pour le remonter, ce qui permettrait pourtant de bien mieux comprendre comment le virus se propage.

Avec plus de 1 million de personnes détectées positives, on pourrait avoir des données formidables permettant de savoir quels profils de personnes sont surreprésentés parmi les malades, qui a été infecté en premier, etc. Il suffirait en effet de comparer les personnes qui ont contracté le virus avec les données extrêmement riches de la statistique publique sur la prévalence de différents métiers, les activités typiques des individus durant une semaine donnée et le temps qu'ils y passent, etc.

« Les chercheurs doivent faire des hypothèses très fortes pour conclure que les gens se contaminent dans les restaurants, car ils n'ont aucune information sur cela »

Donnons un exemple de l'intérêt de telles données à l'échelle mondiale. Mi-novembre est sortie dans la revue *Nature* (la plus prestigieuse revue scientifique) ce qui est peut-être à ce stade l'étude épidémiologique la plus aboutie, et qui conclut notamment que les restaurants sont des lieux importants de propagation ([« Mobility network models of COVID-19 explain inequities and inform reopening »](#), Serina Chang, Emma Pierson, Pang Wei Koh et al., *Nature*, 10 novembre 2020). L'étude est évidemment parfaitement réalisée à partir de données extrêmement riches (récoltées grâce aux téléphones portables puis anonymisées) permettant de connaître heure par heure l'ensemble des déplacements individuels à l'échelle d'une ville.

Pour autant, concernant l'épidémie elle-même, les chercheurs ne disposent que du nombre total de cas quotidiens dans la ville et n'ont presque aucune information sur les personnes contaminées. Ils doivent donc faire des hypothèses très fortes pour conclure que les gens se contaminent dans les restaurants, tout simplement parce qu'ils n'ont aucune information sur cela, même indirecte, dans leurs données. Tant que cette information n'aura pas été collectée à grande échelle, on ne pourra avoir aucune certitude.

Les données récoltées sur les personnes positives seraient évidemment anonymisées et leur participation rendue facultative. Si on demandait aux personnes positives si elles sont d'accord pour répondre à un court questionnaire anonyme afin d'aider la recherche sur la propagation du Covid-19, la grande majorité accepterait sans hésiter. Il n'est donc peut-être pas trop tard pour bien faire.

Thomas Breda est chercheur au CNRS, à l'École d'économie de Paris et à l'Institut des politiques publiques.