



Données massives et prévention du risque épidémique

Paul Véron
Maître de conférences en droit privé à l'université de Nantes, Laboratoire droit et changement social (UMR 6297)

La découverte de l'origine microbienne ou virale des maladies infectieuses a constitué une avancée majeure pour la santé des populations au tournant du xx^e siècle. Le développement de la vaccination et des mesures d'hygiène ont ainsi permis un bond d'espérance de vie d'environ vingt ans entre 1900 et 1950. Grâce à ces progrès, certaines maladies infectieuses ont presque disparu en Europe. C'est le cas de la variole. En France, l'obligation vaccinale contre cette maladie, imposée à partir de 1902, a été supprimée par la loi du 2 juillet 1979. Toutefois, de nouvelles épidémies se sont répandues en Europe et dans le monde. Outre le sida apparu dans les années 1980, on peut citer le SARS, l'encéphalopathie spongiforme bovine (la vache folle), le virus Ebola, le virus Zika, différentes formes de gripes, ou très récemment le SARS-Cov2, nouveau Coronavirus dont la maladie est dénommée Covid-19.

Quels outils juridiques de gestion du risque épidémique ?

Le droit français compte plusieurs dispositifs visant à lutter contre les épidémies. Outre les recommandations ou obligations vaccinales, on citera le système des déclarations obligatoires des maladies contagieuses, les mesures de désinfection, les contrôles sanitaires aux frontières, les dispositifs d'isolement des malades. Plus récemment, l'état d'urgence sanitaire déclaré à la suite de la crise de la Covid-19 a été l'occasion d'un florilège de mesures, parfois extrêmement contraignantes pour les libertés, et visant à lutter contre la diffusion du virus : fermetures d'établissement et cessations d'activités, port du masque obligatoire et restrictions du droit de circulation, allant jusqu'au confinement à domicile.

L'usage de données en grand nombre : un phénomène nouveau ?

Champ traditionnel de l'intervention étatique, la santé publique représente selon l'OMS « l'ensemble des efforts des institutions publiques pour améliorer, promouvoir, protéger et restaurer la santé de la population grâce à une action collective ». Ces actions collectives passent traditionnellement par le recueil de données et l'établissement de statistiques, pour comprendre et mesurer l'état de santé des populations (par exemple identifier le taux de couverture vaccinale) ou pour surveiller la diffusion de maladies. Lors de la grande peste du xiv^e siècle existaient déjà des registres paroissiaux visant à recenser le nombre de malades sur les différents territoires du Royaume. La démarche n'est donc pas nouvelle.

Le recueil de données s'illustre notamment à travers le système des maladies à déclaration obligatoire. Le

Code de la santé publique prévoit que les médecins et responsables des services et des laboratoires d'analyses de biologie médicale sont tenus de transmettre à l'autorité sanitaire les données individuelles dont ils disposent concernant deux grandes catégories de maladies : celles qui nécessitent une intervention urgente locale, nationale ou internationale, et celles dont la surveillance est nécessaire à la conduite et à l'évaluation de la politique de santé publique. En tout, une trentaine de maladies figurent sur la liste élaborée par les pouvoirs publics. La grande majorité d'entre elles sont des maladies infectieuses transmissibles. Sont concernées, entre autres, le chikungunya, le choléra, la dengue, la diphtérie, les fièvres hémorragiques africaines (dont Ebola), la fièvre jaune, le paludisme, la peste, la poliomyélite, la rage, la rougeole, la rubéole, la Creutzfeldt-Jakob, le typhus, le zika. Étonnamment, la Covid-19 n'a pas été ajoutée. Par ailleurs, depuis 1984, le réseau Sentinelles de recherche et de veille sanitaire suit plusieurs maladies infectieuses et alerte sur les épidémies grâce à la contribution de centaines de médecins généralistes et pédiatres répartis sur tout le territoire. Ces professionnels rapportent au moins une fois par semaine le nombre de cas observés pour plusieurs maladies transmissibles courantes, dont les syndromes grippaux, la varicelle et le zona. Les données sont transmises, via un réseau sécurisé, à l'Institut d'épidémiologie et de santé publique Pierre Louis (commun à l'Inserm et à la Sorbonne) et à l'agence nationale de santé publique Santé publique France, qui a absorbé en 2016 l'ancien Institut de veille sanitaire (InVS).

Des évolutions à la fois quantitatives et qualitatives dans l'utilisation des données peuvent néanmoins être relevées. D'une part, la lutte contre le risque épidémique associe aujourd'hui des acteurs privés. Des collaborations entre les autorités sanitaires et les entreprises privées ou multinationales se font jour, à l'image du programme Google Flu, fruit d'un partenariat entre l'Institut fédéral de veille sanitaire américain formé par les Centers of Diseases Control and Prevention (CDC) et la société Google. D'autre part, on observe un double changement d'échelle et de nature des données utilisées, très diverses, y compris des données qui n'ont initialement pas été collectées pour une finalité d'ordre médical ou sanitaire. Avec Google Flu, il s'agissait de tirer profit de millions de traces issues des requêtes des utilisateurs du moteur de recherche sans que lesdites traces aient été à l'origine conservées pour cette utilité. Une autre illustration concerne l'usage des données détenues par les aéroports de différents pays sur les déplacements des usagers à une échelle

locale ou mondiale. Ainsi du simulateur GLEAM (Global Epidemic and Mobility Model), mis au point par une équipe de chercheurs et destiné à prédire la dissémination de certaines épidémies au niveau mondial, en exploitant entre autres les données de transport aérien, la densité de population et de flux de voyageurs entre différentes zones géographiques. Cet outil a notamment été utilisé, en coordination avec l'OMS, pendant la crise du virus Ebola afin d'évaluer en temps réel le risque d'importation de cas dans différents pays. D'autres outils pourraient être cités, dont le logiciel HealthMap, développé par des épidémiologistes et des informaticiens américains en 2006, avec pour objectif d'identifier la naissance de foyers d'épidémies dans le monde, en se fondant là encore sur des sources très diverses (notes de départements sanitaires et d'organismes publics, rapports officiels, données issues d'Internet et de réseaux sociaux, bulletins d'informations locaux, etc.). Au Yémen, un système informatique a été conçu pour permettre aux acteurs humanitaires d'anticiper les épidémies de choléra. Dans plusieurs endroits du pays, des pluies importantes submergent le réseau d'égout, endommagé par la guerre, ce qui favorise la propagation de la bactérie. Pour prédire le risque sur chaque zone du pays, le logiciel croise des données pluviométriques avec d'autres, relatives à la densité de population et à l'accès à l'eau potable.

L'évolution concerne également les modalités de collecte des données, avec notamment l'utilisation croissante d'objets connectés. On peut cette fois prendre des exemples plus locaux qui concernent un volume de données moins important, à l'image des expériences visant à comprendre le phénomène de diffusion d'une maladie infectieuse à l'échelle d'une ville, d'un hôpital, ou encore d'une cour d'école. Au CHU de Lyon, cinq cents patients et professionnels ont été équipés de capteurs électroniques (puces RFID) afin d'enregistrer l'ensemble de leurs contacts – plusieurs millions – sur une période de six mois, dans le but d'améliorer la compréhension de la dissémination des staphylocoques et éventuellement proposer de nouvelles stratégies d'hygiène. La récente crise sanitaire a également vu naître différentes applications de traçage numérique de « cas contacts » de personnes contaminées, dont StopCovid, ayant fait l'objet d'un encadrement réglementaire spécifique. Les personnes alertées – celles qui ont volontairement installé l'application – sont encouragées à se faire tester et le cas échéant à rester confinées, afin de réduire la propagation du virus. Le système repose sur une collecte automatisée de données de localisation ou relatives aux interactions sociales individuelles.

Quels bénéfices pour quels risques ?

L'exemple de Google Flu permet d'illustrer à la fois l'apport potentiel de l'utilisation des *big data* et de leur traitement algorithmique, mais surtout les biais, limites et risques qui peuvent y être associés. Ce programme a été créé par Google en 2008 dans le but de mieux

anticiper et prévenir l'apparition de foyers de grippe sur le territoire américain. La grippe a en effet la particularité d'être un virus sujet à des mutations fréquentes, ce qui explique que les vaccins doivent constamment s'adapter. Sa diffusion rapide impose en outre une réactivité des autorités sanitaires sur deux plans : en amont, quant à la détection de la nouvelle souche du virus (c'est ce qui nous intéresse ici) ; en aval, sur le temps nécessaire à la fabrication du vaccin. Compte tenu de ces contraintes, il peut arriver qu'au moment où le nouveau vaccin est disponible, le virus en circulation soit différent ou encore que la vaccination de masse ne soit possible qu'après que le pic épidémique soit passé.

Aux États-Unis, les CDC (déjà évoqués) reposent sur un réseau de 150 laboratoires biologiques, qui transmettent les types, éventuellement les séquençages, des virus détectés, et un réseau d'environ 2 500 médecins couvrant le territoire des États-Unis et chargés de signaler à l'autorité sanitaire les maladies « semblables à la grippe ». Un échantillon de patients ainsi détectés est ensuite examiné de manière plus approfondie, afin de déterminer quelle proportion souffre réellement de la grippe. Si ce système est efficace, il suppose néanmoins que les patients prennent rendez-vous chez un médecin, le voient en consultation et que ce dernier rapporte avoir constaté un cas de grippe, ce qui peut prendre plusieurs jours ou semaines. C'est dans ce contexte que les CDC et Google se sont entendus pour développer un système d'alerte avancé avec pour but de donner, presque en temps réel, une indication sur l'avancée de la grippe. L'idée était la suivante : en se penchant sur la période 2003-2007, Google et les CDC ont observé une correspondance entre d'une part l'augmentation des requêtes en lien avec la grippe sur le moteur de recherche Google, et d'autre part, l'augmentation des déclarations des cas de grippe faites auprès des CDC. La confrontation des deux ensembles de données a fait ressortir une correspondance pour quarante-cinq termes de recherche ayant un rapport logique avec la grippe (complications, remèdes, symptômes, toux, etc.). En d'autres termes, grâce à cet algorithme, en observant la fréquence d'apparition de certains termes dans les requêtes, il serait possible de détecter le début de la phase de propagation de la grippe en temps quasi réel, et ainsi de gagner plusieurs semaines dans l'identification d'une nouvelle souche du virus.

Plusieurs limites de l'outil ont toutefois été relevées. Premièrement, il n'est pas prévu pour remplacer les déclarations faites par les médecins, pour au moins deux raisons. D'une part, le modèle a été élaboré à partir des données des CDC puisque l'algorithme repose sur la comparaison entre les requêtes générales des utilisateurs de Google et ces déclarations. D'autre part, le programme permet d'établir une corrélation entre le nombre de requêtes effectuées sur Google et le nombre de déclarations faites à l'organisme de veille sanitaire, mais pas un rapport de causalité. Il permet seulement de presumer une augmentation des cas de malades.



Deuxièmement, il peut exister un risque important de surestimation du risque. En effet, les recherches ne sont pas forcément effectuées par les personnes ressentant les symptômes de la grippe. Des paramètres internes et externes peuvent conduire à une augmentation du nombre des requêtes, tels que la forte médiatisation du sujet ou des déclarations faites par les pouvoirs publics. C'est précisément cette surestimation du risque qui a posé problème dans le cas de la pandémie grippale particulièrement virulente de l'hiver 2012-2013. En janvier 2013, à New York, le modèle prédit presque le double d'infections par rapport à ce que les médecins rapportent finalement. Ces prédictions faussées peuvent avoir des conséquences sur les comptes publics, par exemple des commandes massives de vaccins sur la base de ces estimations inexactes. Des changements dans le fonctionnement des moteurs de recherche intervenus postérieurement à la création du modèle – tels qu'un système de saisie semi-automatique – peuvent de même influencer sur son utilisation.

Troisièmement, des tentatives de manipulation du modèle ne sont pas exclues : un fabricant de médicament

antigrippal pourrait par exemple chercher à augmenter artificiellement le nombre de recherches d'un terme faisant partie du modèle.

À partir du début de l'année 2013, le dispositif Google d'alerte avancée sur la grippe a été adopté par vingt-neuf pays et étendu à une autre maladie, la dengue. Toutefois, en raison de ses résultats décevants, il n'est plus utilisé depuis 2015. En France, l'application StopCovid – reposant sur le volontariat – n'a guère rencontré davantage de succès. Moins de 3 % des Français l'avaient téléchargée au milieu de l'été 2020, nombre très insuffisant pour espérer impacter les chaînes de transmission du virus. Outre les questions relatives à la confidentialité et à la fiabilité des données recueillies, plusieurs limites de l'outil ont été soulignées, en particulier au regard du sous-équipement en téléphone mobile des personnes âgées, particulièrement exposées, et des enfants, le plus souvent porteurs asymptomatiques. De quoi relativiser les discours prometteurs des pouvoirs publics et de l'industrie du numérique pour ces dispositifs technologiques de prévention – et demain de surveillance ? – sanitaire. ■

Hygiène et éducation alimentaire à l'heure des *big data* : aide à la décision et gains pour la santé ?

Marine Friant-Perrot

Maître de conférences-HDR à la faculté de droit et de sciences politiques, université de Nantes

L'étude des effets de l'alimentation sur le corps remonte à l'Antiquité, mais l'éducation alimentaire et les principes de diététiques ont connu un regain d'intérêt aux XVIII^e et XIX^e siècles sous l'influence des hygiénistes. L'idée que l'environnement de vie et les habitudes individuelles peuvent largement influencer sur le développement de maladies a ainsi favorisé la diffusion de principes d'hygiène publique inspirés du courant néohippocratique. Si la construction moderne du droit de la santé publique a marqué une rupture avec ce discours hygiéniste – ce n'est plus l'individu qui est comptable de sa santé envers la collectivité, mais bien cette dernière qui lui doit la protection de sa santé –, on a vu resurgir depuis les années 2000 divers « repères », « guides », « recommandations nutritionnelles » dans un contexte de forte augmentation du surpoids et de l'obésité, qui renouent avec ce modèle fondé sur la responsabilisation individuelle. Dans notre pays, près de la moitié des adultes et 17 % des enfants sont en surpoids (respectivement 17 % et 4 % sont obèses) et, au-delà des conséquences sanitaires (diabète, maladies cardiovasculaires...), le coût financier en résultant est évalué par le Trésor public à 20 mil-

liards d'euros. Parmi les mesures recommandées par l'OMS pour améliorer l'alimentation des populations figurent des leviers juridiques de nature à modifier les comportements alimentaires dans un sens plus vertueux (information et éducation nutritionnelles, taxation nutritionnelle, réglementation du marketing alimentaire, reformulation des produits...). Au sein de cette panoplie de mesures, la France comme l'Union européenne ont privilégié les mécanismes informationnels et éducatifs en se fondant sur le principe qu'une personne informée sur la composition nutritionnelle des aliments saura faire des choix conformes à sa santé. En témoigne le programme national relatif à la nutrition et à la santé (PNNS), dont la France s'est dotée dès le début des années 2000. Au plan européen, le règlement n° 1169/2011 dit Inco (information des consommateurs sur les denrées alimentaires) a suivi cette voie en prévoyant que toutes les denrées alimentaires préemballées doivent obligatoirement comporter une déclaration nutritionnelle qui doit aider le consommateur à opter pour une alimentation équilibrée. Cette déclaration peut être complétée par une forme d'étiquetage simplifiée. Il s'agit en France du Nutri-Score, logo