



Enjeux pour la santé publique du traitement des données massives diverses

Margo Bernelin
Sonia Desmoulin

Chargées
de recherche CNRS,
UMR 6297 CNRS,
université de Nantes

« **L**a médecine de demain ne sera pas la même que celle d'aujourd'hui : prédictive, personnalisée, numérique, elle devra sans cesse s'adapter aux nouveaux enjeux et aux nouvelles technologies¹. » C'est par cette phrase que la Stratégie nationale de santé 2018-2022 indique vouloir faire de l'introduction du numérique en santé, et notamment des *données massives*², l'avenir des politiques publiques en santé et positionner la France dans un mouvement global. Ainsi, tant la médecine que la santé publique doivent bénéficier des avancées du numérique. Dans ce cadre, le développement d'outils algorithmiques exploitant des masses de données diverses au service de la santé des populations semble constituer, avec les nouvelles capacités d'archivage, une innovation phare de ce début de XXI^e siècle. Cela se combine avec la possibilité d'exploiter l'ensemble des bases de données publiques et privées constituées notamment par la collecte de données issues d'objets connectés. Les données exploitées peuvent être très variées : signaux biologiques, dépenses de santé, observance des traitements, facteurs environnementaux, habitudes de vie, etc. Aussi, le volume des données traitées et leur diversité sans précédent ouvrent la voie à de possibles transformations majeures en santé publique.

De nombreuses perspectives

Les pouvoirs publics attendent des études menées sur les données collectées l'élaboration de nouveaux plans de prévention en santé, des analyses inédites en matière de risques épidémiques, des traitements mieux ciblés pour les patients et la découverte de voies de recherche insoupçonnées en santé publique, par exemple, de nouveaux facteurs de risques associés à des pathologies pour développer des mesures préventives. Le contexte pandémique né de la propagation du virus SARS-Cov-2 (Covid 19) a montré l'intérêt porté par les pouvoirs publics à l'analyse de données, afin d'anticiper l'évolution de la situation. En France, Epidemap, outil de modélisation issu de trois modèles traitant respectivement des données sur la répartition des habitants (bâtiment par bâtiment), des simulations de déplacements et de

données d'épidémiologie a été conçu pour établir des scénarios de propagation du virus en fonction d'options de confinement à domicile ou de déconfinement. Il a participé à légitimer les mesures de gestion de crise³. Pour lutter contre la transmission virale, les autorités publiques ont également soutenu le développement d'applications mobiles ou de plateformes telles que StopCovid ou ContactCovid. Ces outils visent à collecter des données pour informer les personnes concernées d'un contact possible, durant les jours précédents, avec une personne dépistée ultérieurement, mais aussi pour informer les autorités publiques et les institutions sanitaires sur les parcours géographiques des personnes (dépistées ou non) et sur la détermination des cas contacts. Le relatif échec de l'application StopCovid (beaucoup moins utilisée en France qu'en Allemagne, par exemple) n'a pas remis en question la volonté du gouvernement français, puisqu'une nouvelle version de l'application mobile est prévue. Rebaptisée TousAntiCovid, elle inclut une dimension informative supplémentaire sur les risques sanitaires (symptômes de la maladie, statistiques...) et sur les lieux de dépistages, en plus de l'alerte sur les « cas contacts ».

Toutefois, ces évolutions s'accompagnent de nouvelles interrogations. En sus des enjeux de protection des droits et libertés individuels, en lien avec la protection de la vie privée et des données personnelles, des questions relatives aux frontières et aux acteurs de la santé publique sont ainsi posées. La délimitation des champs de la santé publique était déjà sujette à discussion, en ce qu'ils pouvaient être restreints à l'approche populationnelle et aux dimensions institutionnelle et organisationnelle, ou inclure plutôt, de manière large, tous les domaines couverts par le Code de la santé publique (y compris la médecine). L'approche par le croisement de données diverses en grand volume semble élargir encore les perspectives et donner une importance particulière au concept d'exposome récemment introduit dans le Code de la santé publique et voulant que « l'intégration sur la vie entière de l'ensemble des expositions qui peuvent influencer la santé humaine⁴ » constitue un objectif primordial des politiques de santé publique. Si toutes sortes de données, constamment collectées, peuvent être compilées et croisées pour concevoir de nouvelles politiques et de nouvelles mesures de prévention, la détermination de ce qui relève spécifiquement de la santé publique devient ardue. Tout relève alors possiblement de la santé publique.

1. Stratégie nationale de santé 2018-2022. Ministère des Solidarités et de la Santé, Paris : 2017, p. 63.

2. Il n'existe pas de terme légal ou consensuel en France pour la traduction de l'expression big data. Deux expressions sont couramment employées : mégadonnées et données massives. Toutes deux ont l'inconvénient de ne pas suffisamment insister sur la caractéristique de diversité des données traitées qui, en sus de la rapidité de traitement grâce aux nouveaux outils informatiques et techniques algorithmiques, constitue la spécificité de l'approche par le big data. Sur le concept, voir le chapitre initial de ce dossier.

3. <https://www.data.gouv.fr/fr/organizations/epidemap/>

4. Art. L. 1411-1 CSP

Une intervention grandissante du secteur privé

Quant à la dimension organisationnelle, elle est également bousculée par l'idée qu'il ne serait peut-être plus nécessaire de passer par des institutions et des réseaux créés par l'État ou par les autorités publiques pour traiter les alertes et transmettre les messages de prévention. La place du secteur privé dans les activités de collecte, de stockage et de traitement des données de santé (applications en ligne, plateformes nationales) interroge. Alors que par le passé la collecte et l'analyse des données à des fins de santé publique étaient essentiellement opérées par des services de l'État à des fins d'intérêt général, elles s'appuient aujourd'hui plus souvent sur le concours d'entreprises privées non spécialisées en santé. Les plateformes développées par des réseaux sociaux pour suivre la pandémie de Covid-19 en sont un exemple⁵. Désormais, les entreprises du numérique ou les assureurs s'impliquent au-delà de la simple collecte de données, par la transmission de messages à caractère sanitaire auprès de leurs abonnés-clients-assurés. Un tel constat n'est pas anodin. D'une part, une collecte si volumineuse de données personnelles par des sociétés privées enrichit encore un peu plus leur catalogue de données, faisant peser un risque accru sur la vie privée des citoyens dont les vies sont documentées. D'autre part, la qualité et la fiabilité des informations sanitaires offertes par ces entreprises peuvent être questionnées. En effet, comment s'assurer de la fiabilité du contenu d'une application qui partage des informations sur les moyens de lutte contre certaines pathologies ou virus ? Le passage de la donnée/mesure à l'information et à la donnée probante est ici crucial. Par exemple, les informations offertes par l'agence Santé publique France s'appuient sur le concours avéré de chercheurs dont les moyens d'investigation peuvent être retracés.

La place grandissante du secteur privé se fait également sentir dans les activités de conservation des données. Des entreprises privées telles que Google, Amazon ou Microsoft proposent leurs services. C'est dans ce contexte que le Health Data Hub, plateforme nationale des données de santé, appelée à jouer un rôle central pour la recherche en santé publique, a annoncé que la société étatsunienne Microsoft serait en charge de l'hébergement des données de santé, bien qu'elles soient souvent considérées comme un « trésor national ». Ce choix a été principalement justifié par les capacités importantes d'hébergement de cette société. Sous réserve que les conditions d'archivage respectent le droit européen en matière de protection des données personnelles, le droit français ne s'oppose pas à cette imbrication « public/privé », mais elle a pour conséquence de rendre l'État dépendant du secteur privé en matière de santé et notamment de santé publique. De nombreuses critiques ont été formulées à cet égard, pointant l'inopportunité de choisir une

entreprise étrangère qui dispose déjà d'une puissance financière et commerciale considérable.

Au final, l'introduction des données massives diverses en santé publique paraît renouveler profondément cette discipline centenaire. En effet, les acteurs de la santé publique se diversifient notamment au profit du secteur privé et d'entreprises dont les activités étaient jusqu'alors fort éloignées du champ de la santé. La discipline semble également se renouveler dans ses pratiques, les contributions à ce dossier montrant que les recherches en santé publique évoluent sous l'attraction de la disponibilité nouvelle de données nombreuses et variées. Les objectifs de prévention des risques et de promotion de la santé de la population semblent néanmoins perdurer. Il reste toutefois à vérifier si, dans l'avenir, l'évolution vers toujours plus de stratification (en sous-groupes de populations en fonction de caractéristiques sociales, environnementales, physiologiques ou génétiques notamment) et vers une médecine « de précision » (adressant des messages préventifs et des propositions thérapeutiques individualisés) ne remettra pas en cause l'idée même d'une approche populationnelle globale et de choix collectifs. ●

5. Covid-19 Interactive Map & Dashboard, plateforme proposée aux abonnés de Facebook. https://covid-survey.dataforgood.fb.com/survey_and_map_data.html