

### Limites de la recherche de précision en santé publique

Bien que cette conception de la santé publique de précision soit séduisante, justifie-t-elle une refondation de la santé publique ? Mieux comprendre la complexité des mécanismes par lesquels les déterminants sociaux influent sur la santé s'inscrit dans la continuité de recherches en santé publique qui intègrent des approches systémiques. Mais surtout, cette conception se démarque de toute la littérature sur la santé publique de précision. Par suite, il semble qu'il y ait plus d'inconvénients que de bénéfices à conserver ce vocabulaire de la précision, associé aux notions d'individualisation, de stratégie ciblée et à la génomique.

Pour finir, il est important d'interroger la valeur et la pertinence d'une priorité donnée à la recherche de plus de précision pour améliorer la santé publique. L'approche de précision véhicule d'une part l'idée d'un privilège donné à la mesure quantitative, elle-même associée à celle de la supériorité des sciences naturelles sur les sciences humaines et, d'autre part, l'illusion que l'on pourrait se rapprocher d'une forme de certitude.

Or l'intérêt de la santé publique ne tient-il pas à ce qu'elle est très pluridisciplinaire et qu'elle complète la biomédecine par des approches qualitatives de sciences humaines ? Surtout cette insistance sur la précision court le risque de laisser de côté les facteurs qui ne peuvent être ainsi mesurés et pour lesquels pourtant une intervention est efficace. Rose souligne que, dans le champ de la santé publique, rien ne peut jamais être certain et que la certitude ne saurait être un prérequis pour l'action. Si plus de précision c'est être avant tout attentif à l'individu, à l'hétérogénéité interindividuelle et à l'exactitude des résultats, et si c'est défendre la priorité de la connaissance sur l'action, les fondements de la santé publique sont remis en cause. Le propos n'est pas ici de faire l'apologie de l'imprécision en santé publique ni de défendre l'idée que la santé publique doit toujours privilégier le collectif sur l'individuel. Toutefois, se centrer sur l'objectif de précision est porteur d'implicites qui peuvent nuire à l'âme même de la santé publique, si on considère que la santé de la population dont elle s'occupe n'est pas réductible à la simple somme des santés des individus. ●

## Données massives et santé publique : entre redéfinitions et ruptures normatives

La stratégie nationale de santé 2018-2022, socle politique des projets de lois en matière de santé pour le quinquennat en cours, énonce que « *le développement des innovations numériques, technologiques et organisationnelles en santé est un enjeu clé pour l'évolution des pratiques professionnelles, l'accélération du virage ambulatoire, la qualité du suivi des patients chroniques ou le partage de l'information par les acteurs du système de santé et du médico-social*<sup>1</sup> ».

Le numérique et le traitement des données ont ainsi pris une place centrale au sein des dispositifs juridiques mis en place depuis 2018. À cet égard, il est certain que la combinaison « numérique/données » permet des avancées importantes en matière de santé publique, qu'il s'agisse du suivi de la progression d'une épidémie, de la détection de facteurs de risques associés à des pathologies ou encore de la mise en lumière d'effets secondaires de médicaments. Le recours aux données massives en santé publique doit révolutionner la matière en favorisant des gains de temps dans la recherche, en faisant émerger de nouveaux champs de recherche et de nouvelles cibles de prévention. De tels bénéfices reposent sur la collecte et la conservation des données

(des vivants mais également des défunts), dans des volumes sans précédents, par des opérateurs privés ou publics. Ils nécessitent le plus souvent une mise en commun des bases de données ainsi que leur exploitation par des algorithmes. Tandis que le champ de la santé publique se trouve, à tout le moins, transformé par ces nouvelles opportunités, qu'en est-il du droit en la matière ?

### Une redéfinition des objectifs du droit de la santé publique

À l'image des transformations décrites, le droit se trouve plus que jamais orienté vers la collecte des données, les lois dernièrement votées en faisant une priorité nouvelle. En effet, jusqu'à présent, le traitement des données était pensé quasi uniquement par le prisme des données personnelles et encadré par la loi Informatique et libertés (LIL, 1978). Or, le Code de la santé publique (CSP) s'ouvre aujourd'hui à des dispositifs juridiques particuliers visant le traitement des données, qu'elles soient personnelles ou non, telles que les données d'activité des hôpitaux et les données scientifiques. Ainsi, alors qu'autrefois la question du traitement des données était rattachée aux nécessités de dénombrement, puis de vigilance (appliqué aux médicaments, aux matériaux, à la traçabilité), la collecte des données sort de ces

**Margo Bernelin**  
Chargée de recherche  
CNRS, UMR 6297  
CNRS, université de  
Nantes

1. Ministère des Solidarités et de la Santé, Stratégie nationale de santé, déc. 2017, p. 63.



cadres spécifiques pour prendre une place plus générale et centrale au sein du Code de la santé publique. Cette transformation est mise en évidence par la création en 2016 du Système national des données de santé (SNDS), dont la composition et le fonctionnement sont régis par le Code de la santé publique. Le SNDS regroupait, en 2016, l'accès à cinq bases de données dont celle de l'Assurance maladie sur le remboursement des soins et la base de données nationale sur les causes de décès. Selon les règles gouvernant l'accès aux données, les données doivent être au service exclusif de la santé.

En 2019, le législateur poursuit ce mouvement enclenché en réformant le SNDS pour accorder une place toujours plus importante à la collecte et à la conservation des données. Pierre angulaire de la réforme : l'élargissement du nombre des données disponibles au sein du SNDS. Ainsi, aux bases de données déjà accessibles viennent s'ajouter, entre autres, toutes les données collectées à l'occasion de soins remboursés par l'Assurance maladie. Par conséquent, mesures de taille, de poids ou des résultats d'examen cliniques sont autant de données pouvant remonter au SNDS car consignées par les professionnels de santé lors de soins remboursés par l'Assurance maladie.

La réforme de 2019 est aussi l'occasion d'entériner un projet pilote lancé quelques mois plutôt : le Health Data Hub, plateforme nationale des données de santé chargée, notamment, de mettre à disposition les données du SNDS élargi, mais aussi de financer des projets

de recherche sur ces mêmes données. Cette réforme n'est pas anodine : la collecte des données en grand nombre n'est plus un simple moyen au service de la santé publique au sein du Code de la santé publique, mais devient une fin en soi. À cet égard, la réforme simplifie la création d'entrepôts de données de santé. Ces derniers, créés par des hôpitaux, visent le regroupement de données diverses, y compris des données de santé. Ici, les nouvelles règles autorisent l'aspiration des données du SNDS à des fins d'enrichissement de tels entrepôts. Par conséquent, la réforme insiste sur la collecte des données en grand nombre, leur duplication au sein de bases existantes pour des recherches futures, et cela sans préciser davantage les types de recherches pouvant justifier un accès aux données.

Cette place centrale accordée à la collecte des données par le droit occulte même les difficultés de terrain, laissant en suspens la question du financement de ces collectes et de la mise en œuvre d'une interopérabilité dans l'accès aux données. De même la question des droits des individus est facilement évacuée au profit d'une anonymisation/pseudonymisation des données, pourtant largement faillibles.

### Un droit perméable aux acteurs privés

Sous l'impulsion des données massives, le droit de la santé publique offre aujourd'hui une place plus importante aux acteurs privés alors même que la santé publique a longtemps été réservée aux services publics. Le légis-

#### Le SNDS en 2016

- Les données d'analyse de l'activité des hôpitaux, du programme de médicalisation, des systèmes d'information.
- Les données du Système national d'information inter-régimes de l'Assurance maladie.
- Les données sur les causes de décès.
- Les données médicosociales des maisons départementales des personnes handicapées.
- Un échantillon représentatif des données de remboursement par bénéficiaire transmises par des organismes d'assurance maladie complémentaire et défini en concertation avec leurs représentants.

#### Le SNDS en 2019

- Les données d'analyse de l'activité des hôpitaux, du programme de médicalisation, des systèmes d'information.
- Les données du Système national d'information inter-régimes de l'Assurance maladie.
- Les données sur les causes de décès.
- Les données médicosociales des maisons départementales des personnes handicapées.
- Un échantillon représentatif des données de remboursement par bénéficiaire transmises par des organismes d'assurance maladie complémentaire et défini en concertation avec leurs représentants.
- Les données destinées aux professionnels et organismes de santé recueillies à l'occasion des activités donnant lieu à la prise en charge des frais de santé en matière de maladie ou de maternité et à la prise en charge des prestations en cas d'accident de travail et de maladie professionnelle.
- Les données relatives à la perte d'autonomie lorsqu'elles sont appariées avec les bases de données précédemment citées.
- Les données à caractère personnel des enquêtes dans le domaine de la santé lorsqu'elles sont appariées avec les bases de données précédemment citées.
- Les données recueillies lors des visites médicales et de dépistage obligatoires effectuées par les médecins et infirmiers de l'Éducation nationale.
- Les données recueillies par les services de protection maternelle et infantile dans le cadre de leurs missions.
- Les données de santé recueillies lors des visites d'information et de prévention auprès des travailleurs.

lateur avait ainsi largement entendu confier les missions de prévention, de surveillance de la population à des agences sanitaires publiques, à des réseaux publics ou encore à des organismes, bien que de droit privé, à but non lucratif tels que les mutuelles. Cependant, l'écosystème autour du traitement des données massives a conduit à faire émerger la possibilité d'associer un plus grand nombre d'acteurs privés à la santé publique. Cette attraction vers le secteur privé a été mise en évidence par les discussions parlementaires de 2019 sur la création de la plateforme nationale des données de santé. La question posée était la suivante : cette plateforme, au rôle stratégique de pilotage de la collecte et de l'accès aux données en santé, devait-elle prendre la forme d'une structure publique ou d'une structure privée ? Certains députés avaient ainsi proposé la création d'une société par actions simplifiées (SAS), laquelle aurait pu être majoritairement détenue par l'État, garant de la finalité de la plateforme et gage de confiance pour le public. Porteuse de souplesse, une SAS devait offrir une attractivité plus importante pour la plateforme, capable de conclure des partenariats avec le public ou le privé plus rapidement qu'une institution publique. Cependant, les discussions parlementaires ont mis en lumière la crainte d'une défiance du public vis-à-vis d'une SAS dans la gestion de données personnelles aussi sensibles que celles relatives à la santé. Partant, c'est bien une structure publique qui fut privilégiée avec la création d'un *groupement d'intérêt public*.

La question de la présence d'acteurs privés au

sein du champ de la santé publique fut relancée en décembre 2019 avec l'hébergement informatique des données du SNDS. Extrêmement volumineuses, les bases associées nécessitent, pour être regroupées et interrogées, d'être conservées par des hébergeurs disposant de capacités de conservation et de traitement des données considérables. Le choix s'est alors porté sur l'entreprise américaine Microsoft pour héberger ces données. Les critiques n'ont alors pas tardé, faisant valoir que les données de santé des Français devaient, pour plus de protection, être hébergées par une entreprise française ou à tout le moins européenne. Ces critiques ont également avancé que le SNDS se trouverait dépendant d'une entreprise possédant une capacité commerciale écrasante. Pourtant un tel choix n'est pas étonnant dès lors que le droit ne s'y oppose pas. En effet, le droit de la santé publique, qui encadre strictement l'hébergement des données de santé, imposant aux hébergeurs d'obtenir une certification, n'interdit pas qu'une entreprise privée, y compris étrangère, propose de tels services. Ainsi, du fait de l'introduction des données massives en santé publique, le code du même nom s'ouvre peu à peu à la présence d'acteurs privés, lesquels apparaissent comme des renforts, plus ou moins bien accueillis, de la puissance publique.

Pour conclure, la rencontre entre « données massives » et « santé publique » marque un véritable tournant pour le droit, dont les objectifs s'avèrent aménagés et réorientés vers la collecte des données et dont les acteurs se trouvent diversifiés. ●

#### Nicolas Savy

Maître de conférences à l'université Toulouse III, Institut de mathématiques de Toulouse, UMR 5219

#### Anne Mayère

Professeure à l'université Toulouse III, laboratoire Certop, UMR 5044, directrice adjointe de l'Iferiss

#### Anja Martin-Schol

Maître de conférences à l'université Toulouse III, laboratoire Certop, UMR 5044

#### François Lambotte

Professeur à l'université catholique de Louvain, Institut langage et communication

## La fabrique des données à l'épreuve des programmes de *big data*

Nous proposons d'interroger ici la fabrique de données dans le contexte du *big data*, en prenant exemple auprès de B. Latour [37], qui a investigué la fabrique du droit en observant la façon dont il se construit. C'est posé que, s'agissant des données, la question des finalités, du « pour quoi », n'est pas dissociable du « comment ». Les discussions autour des « données massives » laissent entendre l'avènement d'un nouveau régime de « fabrique des données » qui viendrait amplifier les précédents du fait d'évolutions techniques. Qu'est-ce qui différencie les données de santé, telles que produites selon les standards de recherche, des « données massives », qui sont au cœur de projets conséquents (Health Data Hub) ? Nous verrons que ces fabriques se différencient quant à la spécification des données, aux logiques de traitement, et à la question des « biais » inhérents à toute fabrique de données, nécessairement partielle et éventuellement partielle.

Avant d'aller plus en avant, évoquons la notion de *big data* et la différence avec le terme de données massives au cœur de ce dossier. Le premier V, volumétrie, de la règle des 5V couramment employée pour parler du *big data*, laisse entendre que les données massives y seraient incluses. Encore faut-il s'accorder sur la notion de « massives ». Selon G. Saporta [51], massive est entendue comme « trop gros pour entrer dans la machine », nécessitant une adaptation (si possible) des techniques et des modélisations statistiques usuelles. Par exemple, le volume d'information est tel qu'il est impossible de mettre en place, en une seule étape, une régression logistique. Cette technique, couramment employée en statistique, vise, pour une variable binaire (hypothèse 1 = malade/hypothèse 0 = non malade), à déterminer l'influence d'un ensemble de facteurs (l'âge, le sexe, le poids...) sur la probabilité d'observer l'hypothèse 1. Comme l'ont étudié les sociologues des sciences et