



tout autant parler de santé publique 4 P, celle-ci ayant dans ses attributions classiques la prévention, la participation (démocratie sanitaire, associations de patients) et la prédiction. Reste la personnalisation : en réalité, en santé publique, on peut chercher comme en marketing à « segmenter » les populations, c'est-à-dire à identifier certains ensembles de personnes qui présentent des caractéristiques similaires par rapport à un problème de santé. Le concept est voisin de celui de stratification du risque : tout le monde n'est pas exposé de la même façon aux mêmes risques de développer une pathologie donnée. À l'inverse, la personnalisation « vraie » – un diagnostic strictement unique, personnel, un traitement unique, personnel – n'existe pas et n'a probablement pas beaucoup de sens en général. De fait, la médecine 4 P est elle aussi basée sur une approche de groupe : le groupe des personnes partageant des caractéristiques similaires, impliquant un même traitement.

Pour l'heure, la santé publique de précision semble

se démarquer selon deux grandes dimensions de la médecine 4 P :

- la précision est une précision en termes d'échelles des mesures, par exemple l'échelle spatiale – en effet, les bons résultats d'un indicateur, comme la mortalité infantile selon les régions, peuvent masquer des hétérogénéités géographiques majeures si l'on y regarde à une plus petite échelle que la région ou le pays, et la précision révèle alors une aggravation de l'indicateur en certains lieux, et une amélioration en d'autres, donc une aggravation des inégalités géographiques de santé ;
- ce qui a toujours distingué l'approche sociale de l'approche individuelle, à savoir qu'il existe des caractéristiques propres au collectif, au social, déterminants de l'état de santé individuel et populationnel, qui ne sauraient se réduire à la somme des caractéristiques individuelles. Il ne s'agit au fond que d'une réactualisation, voire une exacerbation de la tension classique entre préférences et approches individuelles, et préférences et approches collectives. ●

## Données massives en santé publique : quels enjeux pour les personnes ?

**Margo Bernelin**  
**Sonia Desmoulin**  
Chargées de  
recherche CNRS,  
UMR 6297 CNRS,  
université de Nantes

**A**u cœur de la politique de santé publique, « l'observation épidémiologique et la surveillance de l'état de santé des populations » s'appuie désormais « sur les nouveaux outils d'exploitation des données », ainsi que le proclame Santé publique France sur son site Internet. La gestion de la crise sanitaire née de la propagation du virus SARS-Cov-2 (Covid 19) illustre bien l'intérêt de la collecte et du traitement des données massives pour la protection de la santé des populations, à des fins de diagnostics, de suivi de la progression pandémique et de recherche en santé publique. L'exemple montre aussi que sont traitées à la fois des données personnelles, comme le résultat des tests de dépistage PCR, et des données non personnelles, telles que les relevés de présence du virus dans les eaux usées de grandes villes. Pour être efficace, une telle approche suppose de collecter énormément de données, de natures diverses, et issues de plusieurs bases. Caractérisée notamment par les concepts de *volume* et de *variété*, cette démarche a logiquement des répercussions sur les droits des personnes. La protection de la vie privée et des données personnelles est difficile à tenir. La tension, bien connue en santé publique, entre protection des droits et libertés individuels et protection de la population est ainsi réactivée. En effet, s'appuyant en partie sur le traitement de données personnelles, l'usage des données massives en santé publique nécessite le concours des individus au risque de leur vie privée.

### L'utilisation des données personnelles de santé

L'usage des données massives en santé publique s'appuie en partie sur la collecte et le traitement de données personnelles, c'est-à-dire de données qui permettent d'identifier directement ou indirectement des individus. Ainsi, noms, adresses, identifiants et autres numéros spécifiquement attachés à une personne sont autant de données personnelles, qui peuvent être utilisées en santé publique. Au sein de ces données, les données de santé tiennent une place particulière en ce qu'elles informent sur l'état de santé passé, présent ou futur d'un individu, qu'il s'agisse de sa santé mentale ou physique. Par exemple, les résultats sanguins ou de radiographie ou les données de remboursement des soins révèlent l'état de santé d'un individu et sont, par conséquent, des données personnelles de santé. Des données plus anodines peuvent également fournir des informations sur l'état de santé, comme une adresse qui mettrait en évidence l'exposition à un environnement pollué et au risque de développer des pathologies spécifiques. Le traitement de ces données variées en masse pourrait, par exemple, mettre en lumière des liens entre la prise d'un médicament, une zone géographique de domiciliation et l'expression d'effets secondaires.

Une double difficulté naît de cette démarche. La première tient à l'impératif de protection de la vie privée des personnes. Étant susceptibles de révéler des informations sur le comportement, l'état de santé et les conditions de

vie des individus, les données personnelles font l'objet d'un encadrement juridique spécifique (loi Informatique et libertés en France depuis 1978 ; Règlement européen sur la protection des données personnelles (RGPD) adopté en 2018). S'agissant des données de santé, les risques d'usages aboutissant à une discrimination néfaste pour la personne concernée sont importants. En conséquence, ces données sont considérées comme sensibles et ne doivent pas, en principe, être recueillies ni utilisées, sauf pour des motifs dérogatoires tels que les soins vitaux, les services de prestations ou de contrôle par les organismes d'assurance maladie, le traitement par les agences régionales de santé ou les traitements nécessaires à la recherche, aux études et évaluations dans le domaine de la santé. Très souvent recherché, le consentement exprès de la personne permet également le traitement de telles données. Cependant, la protection conférée par le consentement peut être discutée dès lors que la personne est en situation de vulnérabilité (un patient en attente de prise en charge ou de soins par exemple) et/ou que le consentement ne correspond en réalité qu'à un acte irréflecté (comme un clic mille fois répété sur un ordinateur ou un téléphone portable). Ainsi, en matière sanitaire, un équilibre délicat doit être recherché entre protection des droits des personnes et objectif de protection de la vie et des populations.

La seconde difficulté vient de la possibilité d'explorer différentes bases de données et, en croisant les éléments ainsi collectés, de remonter de manière indirecte à des informations potentiellement sensibles pour les personnes. Dès lors, la distinction entre les données personnelles et les autres données est mise à l'épreuve. Les recherches en informatique montrent que les techniques d'anonymisation ne sont pas toujours efficaces (par exemple pour les données génétiques) ou qu'elles trouvent rapidement leurs limites lorsque plusieurs bases de données peuvent être croisées.

### La tension entre protection des droits individuels et protection de la population

Si l'impératif de protection de la population semble ici prendre le pas sur le souci de préserver l'intimité des individus et les usages discriminatoires, cette préoccupation ne disparaît pas. Elle se traduit dans la mise en place de statuts spéciaux (hébergeurs de données de santé) et d'outils spécifiques (système centralisé sous la responsabilité d'autorités publiques). Cette tension entre protection individuelle et protection collective n'est pas nouvelle. En effet, dès le xx<sup>e</sup> siècle des règlements nationaux imposaient aux médecins de déclarer les cas de rougeole ou de scarlatine. Les règlements sanitaires municipaux pouvaient également imposer aux hôteliers et aubergistes les mêmes obligations de déclaration. Ainsi, des données personnelles étaient déjà collectées afin d'endiguer une épidémie par l'organisation de quarantaines.

Aujourd'hui, la protection des données personnelles de santé devient d'autant plus difficile à assurer que les

données des patients ne sont plus seulement traitées à un échelon local (celui du CHU par exemple), mais à un niveau national. Une grande partie des données est concentrée au sein de la plateforme nationale des données de santé Health Data Hub (données sur les soins pris en charge par l'assurance maladie, données de remboursement des soins...). À ces données pourraient s'ajouter d'autres, plus hétérogènes : adresses, géolocalisation, contacts, photos, etc. Certains réseaux sociaux proposent, par exemple, de collecter des données de géolocalisation afin de suivre l'évolution de la pandémie de Covid-19. Ainsi, il ne s'agit plus simplement de déclarer aux institutions nationales une pathologie avérée, mais de collecter par différents biais un nombre important de données très diverses auprès de patients mais aussi, et surtout, de personnes en bonne santé.

Bien que les objectifs portés par la santé publique soient louables, le flot des données collectées, notamment sur les déplacements des malades ou sur les contacts avec des personnes à risque, interroge sur notre capacité à maintenir un cadre protecteur pour la vie privée et à éviter que les données collectées ne soient ensuite réutilisées à des fins discriminatoires ou commerciales. De graves failles de sécurité sont régulièrement signalées, comme dans l'exemple donné par l'association Greenbone Networks sur les données d'IRM et de scanner<sup>1</sup>. Or, les données peuvent intéresser de nombreux acteurs privés, comme les assureurs afin d'analyser le risque « santé » de chaque assuré, mais aussi les banques ou certains employeurs.

Par conséquent, le concours des individus à la santé publique à l'ère des données massives pose des questions particulières que le seul impératif de solidarité dans le partage des données ne saurait balayer. L'emploi des données massives a toutefois d'indéniables effets positifs sur les droits individuels, notamment en matière de droit à l'information, ainsi que le montre la création du « Tableau de bord de suivi de l'épidémie de Coronavirus en France », centralisant les données de Santé publique France, de la Direction générale de la santé et les données sur les transferts de patient. Un même outil peut servir des finalités collectives et de protection des personnes.

### Conclusion

L'usage des données massives pour la santé publique remplace plus que jamais l'individu au cœur de la promotion et de la protection de la santé de la population. Cependant, cette mobilisation n'est pas sans risques, notamment concernant le dévoilement d'informations sur le comportement, l'état de santé et la vie privée des personnes. Le thème de ce dossier invite donc à réinterroger le cadre et les objectifs de la santé publique, ainsi qu'à tester la robustesse des équilibres établis à la lumière du recours croissant aux données massives. ●

1. <https://www.lemondeinformatique.fr/actualites/lire-plus-d-un-milliard-d-images-medicales-exposees-77688.html>