



# Données massives et santé publique

**Margo Bernelin**  
**Sonia Desmoulin**

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

**D**ans son avis n° 130 sur les données massives publié en mai 2019, le Comité consultatif national d'éthique pour les sciences de la vie et de la santé (CCNE) indiquait : « *La révolution numérique s'étend à tous les domaines et elle partage, avec l'économie et l'environnement, la caractéristique d'être un champ transversal qui innerve tout le champ de la santé*<sup>1</sup>. » Les « données massives » semblent constituer un élément clé de cette « révolution » annoncée (et sans doute déjà entamée). Traduction française (parfois contestée) de la formule *big data*, l'expression « données massives » renvoie à une « *composante essentielle des sciences et technologies du numérique et en particulier de l'apprentissage machine, de la robotisation et des nouveaux moyens de communication*<sup>2</sup> ». Il s'agit non pas uniquement de la réunion de mesures en volumes très importants, mais aussi de l'accumulation et du croisement de données issues de sources variées, traitées par des algorithmes informatiques toujours plus performants. Grâce à ces traitements, les données massives sont susceptibles de livrer rapidement des informations autrefois difficiles d'accès, d'opérer des rapprochements inattendus et d'offrir des modèles prédictifs inédits.

Dans le champ de la santé, on peut ainsi envisager de croiser des données sur les remboursements des soins, des données de dossiers médicaux ou d'objets connectés, des données génétiques et pharmacologiques, des données de cohortes expérimentales ou de collections scientifiques, mais aussi des données d'établissements pour personnes âgées ou handicapées, des données environnementales ou d'urbanisme, etc. Les perspectives ouvertes par une telle manne semblent prometteuses. On ne compte plus les articles de revues médicales et scientifiques sur l'avènement d'une médecine de précision, adaptée aux caractéristiques spécifiques de chaque patient. De curative, la médecine devient aussi prédictive, pour anticiper le développement de pathologies en fonction du profil des patients. Mais qu'en est-il en santé publique, entendue comme approche populationnelle et collective de la santé ? Les « données massives » constituent-elles ici un nouvel eldorado, comme elles sont présentées en médecine, ou forment-elles plus simplement un nouveau substrat pour des méthodes déjà connues s'appuyant depuis longtemps sur des données en (plus ou moins) grand nombre ?

Après avoir posé dans un chapitre introductif les jalons nécessaires à la compréhension des termes et des enjeux du débat, en définissant les concepts, en pointant les enjeux pour les personnes, d'une part, et pour la santé publique, d'autre part, le dossier entend soulever clairement la question du caractère révolutionnaire ou non de l'usage du *big data* en santé publique. En effet, le recours à ces données opère-t-il une véritable révolution ? Selon quels critères et avec quels moyens ? Les contributions au chapitre 2 reviennent sur ces questionnements et soulignent que si la santé publique s'est construite sur le traitement de données, des éléments changent profondément la situation :

la diversité des sources traitées, la réutilisation de données collectées et le brouillage des frontières entre les approches « pour comprendre » et les approches « pour prévoir », la volonté apparente de faire advenir une « santé publique de précision » ou l'intervention d'acteurs privés intéressés à l'exploitation des bases de données. Ce mouvement transforme la santé publique en conduisant ce champ à adopter, par exemple, de nouvelles méthodes de recherche, à questionner le niveau véritablement populationnel de la santé publique, mais aussi son encadrement juridique actuel. Les chapitres suivants prolongent et reprennent ces réflexions dans des domaines particuliers.

Le chapitre 3 aborde les enjeux des données génétiques. La génétique, apparemment moins centrale pour la santé publique, puisqu'elle vise à analyser l'ADN et l'ARN des patients à des fins de diagnostic et de pronostic de l'évolution pathologique, constitue le fer de lance de la « médecine des données » et pourrait apporter d'importants éléments en matière de santé publique. Le plan France médecine génomique évoque notamment l'intérêt de l'étude du génome par groupes de population, voire en population entière. Les contributions du chapitre 3 montrent toutefois que les questions demeurent nombreuses, s'agissant notamment de l'information à transmettre au patient (lors des analyses ou en raison de nouvelles découvertes) ou des ressources du droit en vigueur.

Le chapitre 4 expose les enjeux de la création d'un système national des données de santé avant de questionner sur l'usage des données massives dans des domaines classiques de la santé publique : hygiène, épidémiologie, vigilance sanitaire. Dans ces champs, l'usage de données massives diverses permet d'entrevoir de nouvelles pistes de recherche et de nouveaux moyens d'action afin que la puissance publique apprécie mieux les décisions à prendre. Qu'il s'agisse de contrôler les infections associées aux soins de santé, de prévenir ou de limiter le développement des épidémies, de lutter contre l'obésité ou de rendre plus efficace les systèmes de vigilance sanitaire, l'approche par la collecte systématique de données et le croisement de bases de données paraît prometteuse. Recours au *big data* ou organisation systématique de la collecte et du partage des données y apparaissent comme des outils potentiellement intéressants, mais dont il ne faut pas négliger les limites et les conséquences.

Enfin, les tribunes de ce dossier offrent des points de vue sur les recherches et les actions en la matière. Les recherches en cours sont illustrées avec la présentation d'une étude sur les soins bucco-dentaires menée avec l'appui des données massives par l'observatoire régional de la santé des Pays de la Loire, d'une part, et avec l'utilisation de données massives à des fins de prévention d'effets secondaires d'un médicament, d'autre part. Une troisième et dernière tribune expose plus spécifiquement les aspects historiques et organisationnels autour de la promotion et de l'utilisation des données de santé par les institutions publiques françaises. Ces tribunes viennent ainsi éclairer les enjeux concrets du recours aux données massives en santé publique, montrant leur actualité, leur complexité, les bénéfices attendus mais aussi leurs limites. ■

1. CCNE, « Données massives et santé : état des lieux, prospective et nouvelles questions éthiques », avis n° 130, 29 mai 2019, p. 11.

2. *Ibid.*