

# Big data et action publique

L'usage des données massives dans les domaines tels que l'hygiène, l'épidémiologie, la vigilance sanitaire permet d'entrevoir de nouvelles pistes de recherche et de nouveaux moyens d'action pour éclairer les décisions de la puissance publique.

## Du Système national des données de santé au Health Data Hub : mise en œuvre et évolution

Ouvrir l'accès aux données de santé collectées par des organismes publics afin de tirer profit des potentialités qu'elles offrent est un enjeu sanitaire majeur. Cette ouverture vise à accroître les connaissances relatives à l'offre de soins et à la prise en charge médico-sociale à destination aussi bien des professionnels de santé que des usagers et des citoyens, à contribuer à la recherche et à l'innovation en santé et enfin à favoriser la veille et la sécurité sanitaires, des objectifs dont l'épidémie de SARS-CoV-2 a révélé le caractère essentiel.

La création du Système national des données de santé (SNDS), par la loi de modernisation de notre système de santé<sup>1</sup>, constitue la réponse du législateur, sous l'impulsion de la ministre des Affaires sociales et de la Santé. La France s'est ainsi dotée d'un dispositif de pointe par rapport à de nombreux États de l'Union

européenne (UE), s'appuyant sur un cadre institutionnel, des acteurs et des procédures bien définis.

Le Système national des données de santé a été profondément réformé en 2019 pour tenir compte des besoins des acteurs concernés en termes de données, mais aussi d'un contexte politique favorable au développement du numérique en santé.

### L'accès aux données médico-administratives du SNDS : une réglementation qui repose sur un équilibre entre ouverture des données et respect de la vie privée

Le Système national des données de santé historique, tel qu'il a été mis en place en 2016, a deux missions principales : il rassemble et met à disposition trois bases exhaustives au niveau national de données de santé publique préexistantes et indépendantes les unes des autres. Il est alimenté par :

- la base de données Sniiram (Système national d'information inter-régimes de l'Assurance maladie) contenant les données relatives à toutes les dépenses de l'Assurance maladie ;

### Ève Jullien

Conseillère juridique,  
Sous-direction  
de l'observation  
de la santé  
et de l'assurance  
maladie, Drees

1. Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé. Elle institue le SNDS à l'article 193, dont les dispositions ont été ensuite intégrées aux articles L. 1461-1 à L. 1461-2 du Code de la santé publique.



- la base de données PMSI (Programme de médicalisation des systèmes d'information) contenant les données d'analyse de l'activité des établissements de santé ;
- la base de données du CépiDc (Centre d'épidémiologie sur les causes médicales de décès) contenant les données relatives aux causes de décès recueillies par les collectivités territoriales.

Des données médico-sociales liées au handicap, fournies par les maisons départementales des personnes handicapées viendront bientôt compléter cette base exhaustive<sup>2</sup>.

Le SNDS, à l'image d'un entrepôt, regroupe donc des données de santé au sens strict, des données sur l'état de santé d'un patient (une affectation de longue durée par exemple ou un diagnostic d'hospitalisation) mais également des données portant sur le patient ou son professionnel de santé et enfin des informations relatives aux soins qu'il a consommés.

La Caisse nationale de l'assurance maladie (Cnam) est chargée du rassemblement et de la mise à disposition des données du Système national des données de santé historique. L'Institut national des données de santé (INDS) représentait le guichet unique pour le suivi des demandes de mise à disposition en vue de réaliser une étude, une recherche ou une évaluation dans le domaine de la santé. L'INDS opérait ainsi le suivi des formalités d'accès aux données du Système national des données de santé et ce conformément au cadre réglementaire.

La plupart des acteurs, publics et privés, devaient en effet déposer un dossier de demande d'accès présentant leur projet de recherche et accomplir des formalités obligatoires auprès d'un comité scientifique d'abord, puis de la Commission nationale de l'informatique et des libertés (Cnil). Un comité de protection des personnes (CPP), si le projet porte sur une recherche impliquant la personne humaine, ou le Comité d'expertise pour les recherches, les études et les évaluations dans le domaine de la santé (Cerees), pour celles n'impliquant pas la personne humaine, émet un avis sur les aspects méthodologiques et scientifiques du dossier avant que la Cnil ne se prononce sur le respect de la protection des données. Le Comité d'expertise sur l'intérêt public (CEIP) de l'INDS produit un avis sur ce point. Cette procédure permet de faciliter l'examen par la Cnil des projets de recherche et de réduire les délais d'instruction des demandes.

En revanche, certains organismes publics ou services chargés d'une mission de service public, dont la liste est prévue par voie réglementaire, bénéficient d'un accès dit permanent au Système national des données de santé et sont dispensés de formalités.

2. Un cinquième flux contenant des données de remboursement des complémentaires santé était prévu mais il n'a pas encore été mis en œuvre.

Le Système national des données de santé s'est construit au croisement de deux logiques, une logique de valorisation de cette masse considérable de données de santé ainsi collectée par des procédures homogénéisées et transparentes d'une part, et une logique de protection des patients par l'instauration de garde-fous juridiques et techniques d'autre part.

La réglementation encadre les motifs pour lesquels les données du Système national des données de santé peuvent être utilisées et interdit qu'elles servent à la promotion commerciale de produits de santé et à une éventuelle modification de cotisations ou primes d'assurance d'une personne présentant un risque. La loi formule également l'obligation de « pseudonymiser » les données du Système national des données de santé afin de rendre impossible l'identification directe un patient.

Des règles en matière de sécurité et confidentialité qui s'appliquent aux organismes ayant accès aux données du Système national des données de santé sont également prévues. Un référentiel de sécurité, propre au Système national des données de santé, a en outre été élaboré par le ministère chargé de la santé. Il précise les règles de sécurité que tout système utilisant des données du Système national des données de santé doit mettre en place en termes de procédure d'accès, d'habilitation des utilisateurs, de formations obligatoires et de traçage des traitements effectués, les données étant mises à disposition dans des bulles sécurisées.

Enfin, l'ensemble des principes applicables en matière de protection des données sont en vigueur, non seulement les droits que les personnes dont les données sont utilisées peuvent exercer mais également les obligations encadrant la collecte des données<sup>3</sup>.

### La loi d'adaptation du RGPD a maintenu les grands principes prévalant en France sur l'accès aux données de santé

Le règlement général sur la protection des données (RGPD), dont les dispositions sont en vigueur depuis le 25 mai 2018, approfondit et étend les droits et obligations à la charge des entités traitant des données à caractère personnel. Il accroît notamment la responsabilisation des acteurs en supprimant la plupart des formalités préalables et en faisant porter sur les

3. Les personnes ont le droit d'obtenir les informations relatives au traitement opéré (catégories de données collectées, identité du responsable de la collecte, objectifs du traitement, durée de conservation, destinataires, éventuels transferts ou réutilisations). Elles disposent également d'un droit d'accès (droit de connaître les données qui ont été traitées), de rectification (droit de corriger les données collectées) et d'opposition sous certaines conditions (droit de refuser que les données soient collectées).

S'agissant des obligations prévues par le RGPD, les données doivent être collectées et traitées de manière licite et loyale, pour des raisons déterminées et légitimes ; les données doivent être pertinentes, exactes et la durée de conservation doit être proportionnée à l'usage qui en est fait.

entités la charge de s'assurer, par elles-mêmes, que leurs traitements sont conformes aux dispositions en vigueur. Toutefois, s'agissant des données sensibles, dont les données de santé, le RGPD permet à chaque État de prévoir des procédures particulières.

La France, par la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles, a fait le choix de maintenir l'obligation d'obtenir une autorisation préalable de la Cnil pour traiter des données de santé dans tous les cas où aucun autre cadre ne permet d'y déroger (exceptions législatives, conformité à un référentiel...). Le rôle des instances consultatives (CPP, Cerees et CEIP) a été conservé. Cette volonté de conserver un haut niveau de protection des données de santé a conduit à l'instauration, par cette même loi, du comité d'audit du Système national des données de santé. Ce comité, présidé par le haut fonctionnaire de défense et de sécurité du ministère chargé de la santé, diligente des audits visant à s'assurer du bon respect des règles d'accès et d'usage du Système national des données de santé.

#### **La réforme de 2019 élargit le périmètre du SNDS et modifie la gouvernance des données de santé afin que la France devienne pionnière dans la structuration des données de santé**

Les évolutions du SNDS, introduites par la loi relative à l'organisation et à la transformation du système de santé<sup>4</sup>, s'inscrivent dans la stratégie de développement du numérique et de l'intelligence artificielle, annoncée par le président de la République en mars 2018.

La loi du 24 juillet 2019 fonde l'extension du périmètre du Système national des données de santé à l'ensemble des données de santé dont le recueil est directement ou indirectement financé par des fonds publics, en particulier les données cliniques recueillies dans le cadre du parcours de soins, ce qui inclut les activités de prévention, de diagnostic, de soins et de suivi social et médico-social. De nouvelles sources de données sont ainsi ajoutées aux quatre bases historiques, l'appariement entre ces données devant faciliter leur structuration et démultiplier leurs usages. Ces nouvelles données permettent essentiellement d'apporter de l'information sur l'état de santé précis du patient (résultats d'examen, diagnostics des médecins) et sur ses déterminants de santé (situation sociale, indice de masse corporelle, consommation de tabac...), dont l'absence limitait les usages du Système national des données de santé.

Ces modifications s'accompagnent d'une refonte de la gouvernance des données de santé. L'Institut national des données de santé (INDS) est remplacé par la Plateforme des données de santé (PDS), également désignée par son appellation anglaise Health Data Hub, dont les missions sont élargies. Si elle conserve le

rôle de guichet unique pour les demandes d'accès, la Plateforme peut en outre procéder pour le compte de tiers à des traitements de données et se voit confier un rôle de facilitateur, tant pour les usagers dans l'exercice de leurs droits que pour les porteurs de projets dans leur réalisation.

Une organisation bicéphale est instituée afin de mettre en œuvre le Système national des données de santé élargi, la Cnam et la PDS en deviennent les deux responsables. Le Cerees devient le Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (Cesrees), qui rend désormais des avis plus complets. Les demandes pour les recherches n'impliquant pas la personne humaine sont examinées par un seul comité qui se prononce sur les différents aspects (méthodologiques, éthiques et intérêt public) qui sont généralement imbriqués. Le système des accès permanent a, en revanche, été conservé.

Il est encore tôt pour évaluer l'impact des changements prévus par la loi du 24 juillet 2019 dès lors que tous les textes d'application ne sont pas encore entrés en vigueur et que la Plateforme des données de santé n'est pas complètement opérationnelle. Il importe toutefois de souligner que l'ambition de cette réforme est de faire de la France l'un des pays pionniers dans le domaine de la promotion et de la valorisation des données de santé. Au sein de l'Union européenne, seule la Finlande a mis en œuvre un système relativement comparable. Sans conteste, la refonte du Système national des données de santé confère à la France un atout alors que la Commission européenne a lancé en septembre 2019 une initiative destinée à créer un « espace européen de la donnée de santé ».

#### **Exemple d'utilisation de données du Système national des données de santé**

De nombreuses études mobilisent les données du SNDS. Cette base de données a déjà permis d'améliorer les connaissances sur le système de santé et ses acteurs mais également d'améliorer significativement la pharmacovigilance (depuis les affaires du Médiateur ou de la Dépakine par exemple) ou encore de mieux suivre les parcours de soins et leurs conséquences. On peut par exemple citer l'étude de M. Coldéfy et C. Gandé [15] qui, à partir de la mobilisation des trois sources de données actuellement présentes dans le Système national des données de santé (Sniiram, PMSI et CépiDc), montre que la réduction de l'espérance de vie des individus suivis pour des troubles psychiques sévères atteint en moyenne 16 ans chez les hommes et 13 ans chez les femmes. Si ce phénomène était connu, seule la richesse des données du SNDS permet d'en objectiver l'ampleur, de le décliner par cause de mortalité ou troubles psychiatriques ou encore de suivre son évolution temporelle et géographique. ●

*Les références entre crochets renvoient à la Bibliographie générale p. 57.*

4. Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé (art. 41).



# La prévention par la transparence dans le contrôle des infections liées aux soins

**Lara Khoury**

Ad. E.,  
professeur agrégée,  
faculté de droit,  
université McGill  
(Montréal, Canada)

La transparence et la communication ouverte participent-elles à la prévention en matière de sécurité des soins de santé offerts à la population ? Plusieurs le croient et aspirent à réduire les taux d'accidents évitables associés aux soins de santé par le biais de techniques de partage public des données liées à la sécurité sanitaire. C'est le cas notamment de la province canadienne du Québec, dont le gouvernement situe la transparence et la communication au cœur de la prévention des infections associées aux soins de santé (IAS) dans son plan d'action ministériel 2015-2020 sur la prévention et le contrôle des infections nosocomiales. Le Québec est à l'avant-garde de la divulgation des événements indésirables associés aux soins, à la fois aux patients et aux comités de gestion des risques œuvrant au sein de chaque établissement de soins. Toutefois, nous nous attarderons ici plutôt à la divulgation au public des taux d'infections associées aux soins de santé et à son effet possible sur l'amélioration de la qualité des soins offerts à la population.

## La surveillance des infections associées aux soins de santé au Québec

La divulgation publique des taux d'IAS au Québec a lieu dans le cadre du Programme provincial de prévention et de contrôle des infections (PCI). Bien que le ministre de la Santé établisse les modalités de surveillance, la méthodologie et les outils, la surveillance est décentralisée et confiée aux établissements de santé. Toutefois, des structures locales et régionales font aussi partie du système québécois de surveillance des IAS (CS-PCI ; TC-PCI ; TRPIN).

La collecte de données locales est obligatoire pour sept IAS : les bactériémies nosocomiales panhospitalières (BACTOT), les bactériémies nosocomiales associées aux cathéters centraux aux soins intensifs (BACC-USI), les bactériémies associées aux accès veineux en hémodialyse (BAC-HD), les bactériémies à *Staphylococcus aureus* résistant à la méthicilline (BAC-SA), les diarrhées à *Clostridium difficile* (DACD), les infections à bacilles à Gram négatif producteurs de carbapénémases (BGNPC) et les infections à entérocoque résistant à la vancomycine (ERV). L'obligation de transmission des données est parfois restreinte à certains types d'établissements : par exemple, pour les ERV, les centres hospitaliers de soins généraux et spécialisés enregistrant plus de mille admissions par année ; pour les DACD, les centres hospitaliers de soins de courte durée.

Les établissements de santé communiquent ensuite leurs données locales à l'Institut national de santé publique du Québec (INSPQ) par le biais du Système

d'information pour la surveillance des infections nosocomiales (SI-SPIN), qui permet de connaître à l'échelle provinciale l'incidence des infections associées aux soins de santé. Le SI-SPIN a comme objectif, notamment, de permettre aux établissements de comparer leur taux d'infection, de réduire le taux d'incidence des bactériémies au minimum et d'identifier des situations d'éclosion dans chaque établissement.

Le public n'a pas accès aux données du SI-SPIN. En revanche, les résultats nationaux de chaque programme de surveillance sont analysés et diffusés à la population dans les rapports du Comité sur les infections nosocomiales du Québec (CINQ), un comité faisant partie de l'Institut national de santé publique du Québec (INSPQ). Le CINQ produit annuellement un rapport général, ainsi que des rapports spécifiques pour chaque IAS surveillée. Alors que le rapport général (« Faits saillants ») adopte une forme et un langage accessibles au public, les rapports spécifiques (« Résultats de surveillance ») sont techniques et s'adressent de toute évidence plutôt aux acteurs du système de santé. Le rapport général du CINQ effectue des comparaisons dans le temps, nationales et internationales, basées sur des analyses statistiques. Il émet aussi des recommandations liées à la prévention et au contrôle des infections associées aux soins de santé à l'échelle provinciale, directement fondées sur les données recueillies. Il inclut en outre des données statistiques tel le taux d'incidence, la description des cas et la microbiologie. Des données spécifiques à chaque établissement de santé sont aussi fournies pour chaque IAS surveillée.

Outre le Québec, plusieurs provinces canadiennes partagent leurs taux d'infections associées aux soins de santé avec la population. Les taux et les tendances des IAS dans les établissements canadiens sont aussi accessibles publiquement par le biais du Programme canadien de surveillance des infections nosocomiales (PCSIN). La divulgation publique des indicateurs de sécurité des soins des établissements de santé au Canada — tels les taux d'IAS — s'insère dans un mouvement nord-américain favorable à la divulgation publique obligatoire des taux d'IAS. Ce mouvement a pris naissance aux États-Unis en 2003 à la suite d'une campagne majeure lancée par la Consumer Union, un lobby proconsommateur associé à la revue *Consumer Reports*. Cette campagne, probablement influencée par la publication du rapport de l'Institute of Medicine, *To Err is Human*, provoqua l'adoption dans de nombreux États américains de lois forçant les établissements de santé à faire rapport de leurs taux d'infection, publiquement pour la majorité.

### Informer sur la qualité des données

Est-ce que la divulgation publique des infections associées aux soins de santé influe sur la sécurité sanitaire? Difficile de le savoir. Un premier enjeu à cet égard concerne la qualité des données recueillies et partagées. L'effet du partage des données recueillies est évidemment tributaire de leur fiabilité. Pourtant, le Cinq n'émet presque aucun commentaire dans son rapport annuel général au sujet du bon fonctionnement du système de surveillance et de la qualité des données transmises à la population. Nous n'avons repéré par ailleurs aucune évaluation à cet égard provenant d'autres entités.

Une discussion de la qualité des données visant les taux d'infections associées aux soins de santé, accessible à la population et lui permettant de mieux comprendre les données transmises et leurs limites, serait opportune. Cela est d'autant plus vrai que la littérature américaine et canadienne documente les problèmes qui peuvent affecter la qualité des données sur la sécurité des soins. Cette littérature s'inquiète, par exemple, de la possibilité que les infections soient sous-rapportées ou mal rapportées ou qu'il y ait une absence d'uniformité dans les définitions de ces infections [52, 57]. Elle soulève le besoin de standardiser/ajuster les données pour tenir compte des différences dans les populations, de la portée et du type de procédures médicales concernées, des risques d'infection inhérents au patient, et de la complexité des soins offerts à chaque hôpital [23, 30]. Enfin, on note des enjeux de manque d'uniformité dans les techniques de surveillance [23, 30] ou leur utilisation inégale, ainsi que la nécessité de comprendre les limites de ces techniques. Ce ne sont ici que quelques exemples des mises en garde et critiques exprimées. Une évaluation des facteurs portant atteinte à la qualité des données transmises à la population nous apparaît requise pour leur assurer un véritable impact, et les résultats de cette analyse devraient être inclus dans le partage public des données. Le Cinq et l'INSPQ devraient aussi faire rapport à la population des ajustements périodiques du système de collecte de données, à la lumière de ces facteurs.

### Divulgation des données et amélioration des soins

Un des moteurs du système de divulgation publique est la conviction que les établissements de santé amélioreront le contrôle des infections s'ils dévoilent publiquement leurs taux d'infections. Toutefois, certains auteurs remarquent qu'il n'existe que peu de preuves que la divulgation publique des IAS a une incidence sur les taux d'infections, ou même qu'elle améliore la qualité des soins apportés au patient, ou leur sécurité [30]. Certains effets sont constatés, cependant. Par exemple, Hausteine *et al.* [30] notèrent en 2011 une baisse remarquable et inattendue des infections au SARM au Royaume-Uni après que la divulgation des taux de ces infections soit devenue obligatoire. Au Canada, Daneman *et al.* [20] observèrent en 2012

que la divulgation publique des infections à *C. difficile*, qui débuta en 2008 dans la province de l'Ontario, fut associée à une chute de 26 % de ces infections dans les hôpitaux, sauvant ainsi plus de cent vies par année. Avant la mise en place du système de divulgation obligatoire, soit entre 2002 et 2008, les infections causées par la bactérie *C. difficile* avaient été constamment en hausse. Enfin, aux États-Unis, Cardo *et al.* [8] constatèrent en 2005 que, depuis le début de la divulgation des IAS — à l'époque, celle-ci était volontaire et confidentielle —, les taux de bactériémies associées aux cathéters intravasculaires centraux, de pneumonies associées à la ventilation mécanique et les infections urinaires associées aux cathéters avaient aussi connu des baisses importantes.

Hausteine *et al.* [30] posent l'hypothèse que la divulgation des taux d'infections associées aux soins de santé attire l'attention sur la situation et qu'elle agit comme un outil de renforcement externe, associé à un changement de culture organisationnelle et à une augmentation des activités préventives. Bien que ne pouvant pas déterminer exactement le mécanisme ayant entraîné un déclin des taux d'infections à la suite de la divulgation publique des infections à *C. difficile* en Ontario, les auteurs canadiens Daneman *et al.* [20] émettent l'hypothèse que cette divulgation aurait élevé les infections à *C. difficile* au rang de priorité et motivé les hôpitaux à mieux adhérer aux standards de prévention et de contrôle de ces infections. Aux États-Unis, on note que la demande de données publiques sur les taux d'infections associées aux soins de santé a mené à une amélioration par les hôpitaux de leurs pratiques en prévention des infections, afin de répondre aux attentes des législateurs et des patients, mais qu'il n'y aurait aucun lien clair entre l'imposition d'une obligation de divulgation publique et des procédures améliorées ou des réductions dans les taux d'infections subséquents [42].

### Conclusion

La transparence et la communication ouverte et franche à l'égard des accidents médicaux sont des conditions nécessaires à une culture de prévention des accidents et de promotion de la sécurité des patients dans la délivrance des soins de santé. La divulgation des indicateurs de qualité a comme objectifs avoués non seulement de renforcer l'autonomie des patients en leur permettant de faire un choix éclairé des services de santé qui leur sont offerts, mais également d'améliorer la qualité des soins. Des recherches au Canada, aux États-Unis et au Royaume-Uni démontrent en effet — sans pouvoir établir un lien de causalité — que la divulgation publique des IAS semble contribuer à un changement de culture et d'attitude dans la prévention des risques au sein même des institutions de soins. Les entités participant à la surveillance et à la divulgation des données pertinentes doivent toutefois faire preuve de vigilance afin d'assurer la qualité des données et communiquer ouvertement au public les limites de celles-ci. 📌

Les références entre crochets renvoient à la Bibliographie générale p. 57.



## 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<sup>e</sup> 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<sup>e</sup> 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 présumer 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<sup>e</sup> et XIX<sup>e</sup> 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



ou signal coloriel recommandé aux exploitants du secteur agroalimentaire par l'arrêté du 31 octobre 2017, et qui constitue une mesure phare du nouveau PNNS publié le 20 septembre 2019. Le mangeur informé de ses choix, voire incité à s'alimenter plus sainement, est ainsi perçu comme l'artisan de sa propre santé. Dans cette lutte contre les maladies chroniques d'origine nutritionnelle, la collecte massive de données sur la composition des aliments et sur les comportements alimentaires offre de nouveaux outils informationnels pour prévenir l'obésité et le surpoids [17]. Mais quelle sera sa véritable incidence sur les choix alimentaires et la santé nutritionnelle [31] ?

### Des données massives nouvelles sur l'alimentation dans un contexte de renouvellement des champs classiques

La collecte et l'exploitation de données sur la composition des aliments et sur les comportements alimentaires connaissent une croissance exponentielle depuis les années 2000. À la suite de la crise sanitaire de la vache folle, il est apparu nécessaire de restaurer la confiance des consommateurs en améliorant la transparence sur les ingrédients composant les aliments et en instaurant une traçabilité obligatoire permettant d'identifier l'ensemble des maillons de la chaîne alimentaire. La place du numérique s'est alors accrue dans les États membres de l'Union européenne sous l'effet du règlement 178/2002 relatif à la législation alimentaire. Grâce au code-barres (EAN), toutes les informations relatives au produit, au fabricant et au pays d'origine sont recueillies et permettent de suivre l'itinéraire de la denrée alimentaire jusqu'au consommateur final. Au-delà de la traçabilité des opérateurs imposée par l'article 18 du règlement 178/2002 et de l'objectif initial lié à la préservation de la sécurité sanitaire, ce sont de véritables cartes d'identité numériques des aliments qui sont constituées sur les qualités nutritionnelles ou environnementales du produit.

Dans cette course à la transparence, on recense des bases de données collaboratives, comme Open Food Fact, qui a conclu un partenariat avec Santé publique France pour nourrir la base de calcul du Nutri-Score (plus de 700 000 aliments recensés), des start-up qui développent des applications smartphone comme Yuka ou des initiatives des opérateurs économiques eux-mêmes comme NumAlim. Les métabases de données répertorient les informations publiques qui figurent obligatoirement sur l'étiquetage du produit conformément au règlement Inco de 2011 (ingrédients dont allergènes et additifs, composition nutritionnelle...), mais aussi la présence de signes de qualité (bio...) ou des données privées collectées par les maillons de la filière agroalimentaire (présence de résidus de pesticides...).

Les consommateurs sont en attente de ces informations, notamment pour manger plus sainement car près de 35 % d'entre eux utilisent ces applications. Cette collecte de données sur la composition des

aliments est complétée par de multiples informations sur les comportements alimentaires, qu'il s'agisse des bases de données de la recherche publique (cohorte Nutrinet Santé...), des données recueillies lors des soins (consultations de suivi de l'obésité...), des données répertoriées par les assureurs dans le cadre de contrat de complémentaires santé comportementales (programme Vitality...) ou des données collectées par les personnes elles-mêmes *via* les applications mobiles de coaching et d'éducation nutritionnelle (compteurs de calories, même à partir d'une photographie d'un plat, conseils nutritionnels et culinaires...) et *via* les objets connectés (pèse-personne, frigo intelligent...).

L'arrivée de ce flot de données coïncide avec une redéfinition des frontières entre le champ alimentaire et le champ médical. On assiste, d'une part, à une « médicalisation » du marché alimentaire avec l'augmentation de la consommation des compléments alimentaires et des allégations nutritionnelles et de santé qui vantent les effets positifs de certains aliments sur la santé. D'autre part, les frontières matérielles de la santé publique s'élargissent. Elles intègrent les déterminants de santé liés au mode de vie, notamment à la nutrition et à l'activité physique. Dans les actions de prévention des maladies nutritionnelles, les acteurs privés sont très présents et on assiste à une forme de « marchandisation » de la prévention en santé (marché des applications et des objets connectés du bien-être et de la santé, comme les pèse-personnes connectés, complémentaires santé comportementales). Ce double mouvement s'accompagne d'une porosité croissante entre le droit de la santé et le droit du marché. Selon une rhétorique commune, le « patient-consommateur » est responsabilisé dans la promotion et l'amélioration de sa santé par l'alimentation, il est « acteur du marché » et « acteur de sa santé », informé et éduqué, il est à même de choisir le régime alimentaire qui lui convient.

### Incidence des big data sur les choix alimentaires

Dans ce contexte de « sur-responsabilisation » des personnes dans la gestion de leur alimentation et de leur mode de vie, la collecte de données massives a nécessairement une incidence sur les choix alimentaires. Les dispositifs juridiques mis en œuvre pour améliorer la qualité nutritionnelle de l'alimentation sont essentiellement centrés sur l'individu (informations par la déclaration nutritionnelle, Nutri-Score...) et ne modifient que timidement l'environnement alimentaire pour ne pas heurter les libertés économiques (par exemple la taxe soda en France...). Ils procèdent de l'idée, confortée par l'évolution des connaissances scientifiques (épigénétique, micronutrition...) et du marché des aliments santé, que les individus peuvent maîtriser leur alimentation et adopter un régime alimentaire individualisé vecteur d'une santé parfaite. Dans ce cadre, la transparence sur la composition nutritionnelle des aliments *via* le Nutri-Score

*Les références entre crochets renvoient à la Bibliographie générale p. 57.*



ou les applications comme Yuka, ainsi que la possibilité de mesurer, de comparer ses données nutritionnelles par les applications de *quantified self* constituent des outils normatifs qui incitent les individus à consommer les aliments les plus favorables à leur santé.

Ces possibilités liées aux données massives vont être déployées de manière encore plus importante avec la mise en place du « Store santé » au sein de l'espace numérique de santé (ENS). En vertu de l'article L. 1111-13-1 du Code de la santé publique, chaque usager pourra accéder à ses constantes de santé « éventuellement produites par des applications ou objets connectés ». Une place nouvelle est ainsi octroyée aux données dites de « bien-être » : formant un *continuum* avec les données recueillies lors des soins, elles sont recensées par le patient, qui alimente lui-même son espace numérique de santé et renforce ainsi sa capacité à améliorer son alimentation au bénéfice de sa santé.

Cet effet vertueux de la collecte de données sur ce que nous mangeons nécessite toutefois que chacun puisse se référer à une norme alimentaire définie de manière transparente, objective et fondée scientifiquement. Or, on assiste à une privatisation, voire une captation par le marché, du discours nutritionnel et de la prévention de l'obésité et du surpoids. Les forces créatrices de la norme alimentaire ne sont pas uniquement les autorités sanitaires, mais aussi les géants de l'agroalimentaire, qui tentent de développer leurs propres profils nutritionnels pour échapper au classement défavorable qu'établit par exemple le Nutri-Score pour leurs produits (par exemple tentative de mise en place de l'Evolved Nutrition Label, abandonné en 2018). Les messages sanitaires, les applications et les objets connectés conçus par les opérateurs privés ont aussi tendance à se focaliser sur la promotion de l'activité physique en minorant l'incidence de la consommation alimentaire comme cause de l'obésité et du surpoids.

### **Incidence des *big data* sur la santé nutritionnelle**

Si les *big data* ont une incidence indéniable sur les choix alimentaires, il est plus complexe d'établir leur effet bénéfique sur la santé nutritionnelle.

Les risques sanitaires peuvent d'ores et déjà être mis en lumière. En premier lieu, la pertinence scientifique du processus connaît différentes failles. Il existe d'abord un décalage entre les déclarations et la réalité des consommations alimentaires qui ne peut pas toujours être décelé par les différents outils qui recueillent les données. Les applications sont parfois non conformes aux nouveaux repères nutritionnels, actualisés en 2017 (par exemple concernant la limitation de la consommation des jus de fruit ou des produits laitiers). Elles sont généralement opaques (par exemple Yuka qui ne communique pas le contenu détaillé de son algorithme) et simplificatrices (par exemple absence d'identification des légumes secs ou des produits céréaliers complets dans certaines applications alors même que leur consommation est recommandée selon les nouveaux repères).

Certaines applications promeuvent des régimes amaigrissants alors même qu'ils sont de nature à accroître les risques nutritionnels et que leur interdiction a été préconisée par le HCSP en 2017. En second lieu, on peut craindre l'existence de pratiques commerciales trompeuses contraires à l'article L. 121-2 du Code de la consommation (art. 5213-3 et suivants du Code de la santé publique pour les dispositifs médicaux). À l'exception du régime juridique applicable aux allégations nutritionnelles et de santé, le contenu des messages relatifs à l'hygiène alimentaire ne fait l'objet que d'un contrôle *a posteriori*. Comme beaucoup d'acteurs ne sont pas issus du secteur de la santé (la moitié des entreprises commercialisant des applications en santé sont issues du monde de l'informatique [17]), le risque d'induire le consommateur en erreur, voire de l'amener à adopter des comportements alimentaires pathologiques comme l'anorexie est à craindre. En présence de liens capitalistiques avec des partenaires commerciaux, comme dans le cas des complémentaires santé comportementales qui récompensent en bons d'achat les comportements sains, ces dérives préjudiciables à la santé sont renforcées. Même si le gain escompté peut améliorer l'alimentation des consommateurs, des considérations commerciales interfèrent nécessairement avec la volonté d'améliorer la nutrition. En dernier lieu, une attention particulière doit être portée aux populations spécifiques (enfants et adolescents, femmes enceintes, personnes âgées, végétariens...). Les repères nutritionnels destinés à la population générale ne sont pas toujours adaptés à ces catégories particulières de personnes et, malgré l'existence de recommandations nutritionnelles validées scientifiquement, les applications et objets connectés ne tiennent généralement pas compte de ces besoins spécifiques. On peut aussi s'inquiéter du fait que les adolescents pourront dès 16 ans gérer leurs données de santé sur leur espace numérique de santé et accéder à leur magasin numérique d'applications santé sans considération de leur vulnérabilité, notamment face au diktat de minceur.

Pour s'assurer des gains pour la santé des données collectées sur l'alimentation, il serait nécessaire d'évaluer et de contrôler non seulement la capacité de ces données à orienter les choix alimentaires vers les aliments plus sains, mais aussi à réduire l'incidence des maladies nutritionnelles. Si le bénéfice sanitaire de la collecte et de l'exploitation des données est parfois contrôlé par les pouvoirs publics (Nutri-Score...), les exigences concernant les applications et dispositifs développés par les acteurs privés semblent faibles. La loi du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé ne prévoit que le respect de référentiels de sécurité, d'interopérabilité et d'engagement éthique (art. L. 1111-13-1 du Code de la santé publique), sans exiger la preuve d'un bénéfice sanitaire. Cela obère très largement les espoirs suscités par le développement des *big data* dans le domaine de la nutrition. ●

# Big data en santé : quel intérêt pour les vigilances sanitaires ?

La dynamique des Trente Glorieuses a permis la mise sur le marché de produits de santé de plus en plus performants, mais aussi de plus en plus complexes. Les vigilances sanitaires ont été mises en place pour répondre aux besoins identifiés par des crises successives provoquées par les effets délétères de ces produits. C'est ainsi que les professionnels de santé ont inventé des systèmes pour quantifier les risques tandis que le législateur leur a défini un cadre juridique. Actuellement, le système repose sur l'extraction de résultat à partir de bases de données. Aujourd'hui, l'enjeu consiste à savoir si le traitement automatisé des données massives est un atout à la fois pour la santé publique et pour l'aide à la décision politique.

## Les vigilances sanitaires

### Quel est le contexte qui favorise la prise de conscience des risques associés aux soins ?

Après la Seconde Guerre mondiale, l'augmentation massive des citoyens qui accèdent aux thérapeutiques a fait émerger la notion de risque inhérent aux médicaments et a mis en lumière la nécessité de la gestion de crise. Par exemple, la crise autour des prescriptions de la Thalidomide a mis en lumière les problématiques d'effets indésirables liés aux médicaments (iatrogénie). L'usage massif de stupéfiants et de psychotropes a permis de documenter la plasticité des circuits des neuromédiateurs et d'objectiver les problématiques de dépendance à l'arrêt des traitements. Successivement, trois conférences de l'Organisation mondiale de la santé ont formulé ces problématiques à l'échelon mondial : en 1961 sur la dépendance aux stupéfiants, en 1963 autour de la gestion de la crise de la Thalidomide, en 1972 sur la dépendance aux psychotropes. La réponse scientifique et étatique à ces crises a été de mettre en place des systèmes de détection et de quantification des risques : les vigilances sanitaires.

### Qu'est-ce qu'une vigilance sanitaire ?

Les vigilances sanitaires sont des dispositifs de surveillance épidémiologique qui se déroulent lors de la phase commerciale des produits de santé. Parmi les principaux produits de santé, nous retenons : les médicaments, les dispositifs médicaux, les dispositifs de diagnostic *in vitro*, les produits sanguins labiles, les tissus, les cellules, les organes, les produits cosmétiques, les produits de tatouage.

En dépit de l'absence d'une définition juridique, voici une définition globale des vigilances qui découle de la définition de l'OMS de la pharmacovigilance : les vigi-

lances sont « *la résultante de la science et des activités relatives à la détection, l'évaluation, la compréhension et la prévention des effets indésirables et de tout autre problème lié à l'utilisation des produits de santé et des pratiques de soins* ».

Méthodologiquement, elles sont basées sur des outils épidémiologiques de veille sanitaire autour d'un dispositif d'alerte qui caractérise et valide le signal. Ainsi, la détection, la surveillance et la caractérisation des risques s'effectuent *a posteriori*, de façon prospective et spécifique. Cette évaluation des risques des produits de santé a pour but de permettre une prescription ou une utilisation la plus pertinente possible au regard de l'état actuel des connaissances scientifiques. Les missions des vigilances sanitaires sont complémentaires des autres dispositifs d'évaluation des produits de santé. En France, le législateur a choisi d'individualiser chaque vigilance et de lui attribuer une agence sanitaire référente. La codification législative des vigilances est un processus long de quatorze années qui a débuté en 1994.

L'objectif d'une vigilance est *in fine* de collecter le maximum de données vérifiables et inattendues sur les produits de santé une fois ceux-ci commercialisés. Cela vient en complément des études préalables à la commercialisation, qui visent à écarter des produits de santé présentant trop d'effets indésirables pour la santé humaine. La compilation de ces données est nécessaire pour établir les recommandations les plus adéquates au regard des connaissances scientifiques. Par retour d'information, les professionnels de santé peuvent améliorer la pertinence de leurs actes et des schémas de soins.

### Que détecte une vigilance sanitaire ?

Les vigilances sanitaires détectent un effet indésirable ou un incident imputable à un produit de santé. Dans un premier temps, la déclaration ascendante du professionnel de santé ou du patient vers la structure experte a pour but de caractériser le signalement. À ce stade, le signal est validé et un degré d'imputabilité lui est attribué par des experts de la vigilance. Au niveau national, les agences sanitaires dédiées à chaque vigilance analysent l'ensemble des signalements. Lors de la mise en évidence de signaux convergents, l'agence émet alors une alerte descendante à destination des professionnels qui détaille la conduite à tenir adéquate en fonction du problème identifié.

### Quelles sont les principales limites des vigilances sanitaires ?

La principale faiblesse du système réside dans son manque d'exhaustivité structurelle. En effet, il existe bien

**Caroline Huchet-Kervella**  
Pharmacienne  
hospitalière



une obligation légale (loi n° 2002-303 du 4 mars 2002) de déclaration des effets indésirables qui s'applique à l'ensemble des professionnels de santé. Contrairement au secteur libéral, l'effectivité de la mise en place des vigilances au sein des établissements de santé est un élément recherché depuis la première version de la certification. Par ailleurs, les vigilances figurent dans les contrats d'amélioration de la qualité et de l'efficacité des soins (Caques), qui lient les établissements de santé et les agences régionales de santé depuis 2016. Ainsi, pour le secteur hospitalier, les vigilances sont abordées comme outil managérial au service de l'amélioration de la qualité des soins. Toutefois, il existe une vision janusienne des vigilances. D'un côté les experts, qui analysent les situations cliniques et rendent compte aux agences. De l'autre les cellules qualités des établissements, qui analysent le *feed-back* négatif de la déclaration pour améliorer la qualité des soins. Par ailleurs, le fait de favoriser les échanges entre ces deux types d'expertise pourrait mettre en lumière des leviers managériaux motivationnels qui font sens.

Actuellement, l'acte de déclarer un effet indésirable ou un incident n'est pas assez valorisé dans le système de soins pour ce qu'il est. La capacité à détecter un effet indésirable est avant tout la capacité des professionnels de santé à questionner la situation dans laquelle se retrouve le patient par rapport à celle escomptée. C'est utiliser le doute comme moteur pour agir, comme le soulignent les travaux de la sociologue Anne-Chantal Hardy. Les résultats issus de l'enquête nationale sur les événements indésirables liés aux soins (Eneis) de 2004 et 2009 [41] mettent en évidence une sous-déclaration massive des événements indésirables comparable à ce qui est décrit dans les autres pays européens et nord-américains. Pour le manager de santé, ces travaux mettent en avant un manque de motivation pour agir, ce qui insinue une perte de sens probable pour l'action de déclarer.

### Quels sont les principaux biais à la déclaration ?

Le mécanisme du doute, qui conduit à la remise en cause d'une étape du processus de soins, doit passer au travers de différents biais qui conduiront le professionnel à déclarer.

Le plus facilement appréhendable est le biais d'identification. En effet, pour mettre en lumière un effet indésirable, il faut l'identifier et établir un diagnostic différentiel qui fait soupçonner le lien causal entre le produit de santé et l'effet. En particulier, s'il est facile de mettre en évidence des effets indésirables fréquents décrits dans la littérature, il devient moins évident d'identifier la connexion de symptômes apparemment non reliables entre eux. En plus du biais d'identification, Anne-Chantal Hardy [29] identifie deux biais principaux à l'origine de la sous-déclaration.

Tout d'abord, le biais de connaissance qui est à l'origine d'un sentiment de trahison car il entraîne une remise en cause du savoir académique et de la transmission

par les pairs. Ce sentiment de trahison est d'autant plus inconfortable que le professionnel a également le sentiment qu'il trahit la confiance que le patient lui a accordée.

Pour évacuer cet inconfort, un autre biais entre en jeu : celui de la minimisation, par lequel le professionnel dévalorise la plainte du patient. C'est comme si le fait de ne pas accorder de la valeur à la parole diminuait l'intensité de l'effet indésirable.

Enfin, il existe un biais structurel de sous-déclaration du fait de l'absence de valorisation et de reconnaissance de la mise en évidence d'effets indésirables au niveau des structures de soins. En effet, en dépit de la mise en place de la déclaration en ligne des effets indésirables, la déclaration exhaustive pour permettre une analyse pertinente prend du temps pour lequel il n'est pas prévu de contrepartie financière.

### Bases de données et vigilances sanitaires

#### Quel est l'intérêt d'avoir recours aux bases de données massives dans le cadre des vigilances sanitaires ?

L'intérêt principal d'un recours aux données massives serait d'avoir accès à des données complémentaires pour renforcer la prise de décision. Aussi, le but serait d'identifier tous les signaux, y compris ceux que les biais précédemment évoqués ne permettent pas de déceler.

Au niveau national, une telle approche permettrait d'accroître la force des recommandations de la Haute Autorité de santé (HAS), mais aussi celles des sociétés savantes. Au niveau local, elle permettrait de mettre à disposition des professionnels de santé des informations plus complètes et transparentes au sujet des produits de santé.

Voici deux exemples qui illustrent les types de requêtes envisageables pour compléter les données manquantes. Dans le champ de la médecine libérale, un raccourcissement du délai entre deux consultations ou l'arrêt d'une thérapeutique généralement prescrite au long cours sont deux marqueurs de l'inattendu dans le processus de soins. Traiter de telles données pourrait faire remonter des signaux pertinents. Lors d'une prise en charge hospitalière, la présence d'un effet indésirable lié aux soins peut se détecter lorsque l'allongement de la durée de séjour est statistiquement significatif par rapport à celle attendue et définie par la tarification à l'activité dans les données agrégées des groupes homogènes de malades. Ainsi, au niveau de l'établissement de santé, il peut aussi être envisagé de définir à l'avance des mots-clés qui peuvent être le signe de la présence d'un effet indésirable. Une fois ce thésaurus défini, il serait à la base des requêtes dans les comptes rendus d'hospitalisation issus des résumés d'unités médicales et générés à chaque fin d'hospitalisation. Dans ces deux exemples, l'automatisation du traitement ne va toutefois pas de soi. La requête statistique générerait des séjours qu'il faudrait analyser manuellement pour confirmer le caractère iatrogène. Au vu des biais cognitifs évoqués

Les références entre crochets renvoient à la Bibliographie générale p. 57.

plus haut, il semble ardu de bâtir une intelligence artificielle capable de prendre en compte l'implicite dans les comptes rendus médicaux. L'augmentation significative des dossiers à analyser générerait d'importants besoins en ressources humaines pour pouvoir soit analyser, valider et imputer la déclaration, soit la réfuter.

En revanche, au niveau national, des projets de systèmes algorithmiques mobilisant les données massives au service de la pharmacovigilance sont déjà en cours de développement. En effet, cette vigilance est idéale pour mettre en place ce type de programme. Cela s'explique par une diffusion à large échelle des médicaments, par la connaissance des effets indésirables lors des tests qui précèdent la commercialisation. Enfin, les médicaments font partis des produits de santé qui présentent la période d'exploitation la plus longue. Depuis 2013, il est possible pour les chercheurs d'exploiter les données contenues dans le Système national interrégimes de l'Assurance maladie (Sniiram) et désormais dans le Système national des données de santé (SNDS). L'objectif est de pouvoir répondre à des problématiques non solvables lors des études précliniques : études des signaux non signalés par les professionnels de santé, études chez des populations particulières (personnes âgées, grossesses) ou pour réaliser des études de pharmaco-épidémiologie. Par ailleurs l'avantage secondaire de ce type de recherche est de permettre de répondre aux problématiques de disparités territoriales de prescription et de documenter à l'échelle macro les différents leaders d'opinion de prescription. À partir de l'exploitation de ces données, des études pluridisciplinaires, associant des sociologues, pourraient mieux documenter les relations soignants-soignés autour de certaines thérapeutiques.

#### Quels sont les principaux obstacles méthodologiques ?

Dans le cadre de l'utilisation des bases de données, la définition de la faisabilité de la problématique est une étape fondamentale. En effet, l'architecture de la base de données permet de savoir quelles problématiques peuvent être résolues, en les distinguant de celles pour lesquelles il est impossible d'obtenir une réponse. Après validation de la problématique, il est nécessaire de retraiter les données pour permettre un

usage statistique. Dans une démarche qualité (type ISO 9001), cette opération doit être décrite au préalable. Par ailleurs, la non-exhaustivité de certaines données est l'un des problèmes lors de la mise en place de base de données massives. Toujours dans une démarche qualité et en vue d'explicitier les possibles biais d'analyse, il est utile de préciser les modalités qui conduisent à ne pas retenir certaines données. L'idéal est donc de composer des équipes pluridisciplinaires (médecin, pharmacien, ingénieur, statisticien) autour d'une problématique commune.

#### Quels sont les principaux obstacles managériaux ?

L'arrivée des données massives dans les pratiques courantes va impliquer d'importants changements dans les pratiques des professionnels de santé. En effet, pour beaucoup de nos concitoyens, la méconnaissance de l'intelligence artificielle et des bases de données conduit à une réticence réactionnelle aux changements. Dans un contexte de perte de sens dans le travail, ces bouleversements technologiques induisent plus ou moins consciemment la peur de voir le travail de l'homme remplacé par celui de la machine. Bien que l'augmentation des données à traiter conduit à une probable augmentation des besoins en professionnels experts, les leviers motivationnels et la conduite du changement passent également par la prise en compte de la perte de sens au travail, qui s'étend au secteur de la santé et qui s'est accéléré avec la mise en place de la tarification à l'activité à partir de 2008. Les travaux de Nathalie Angelé-Halgand et Thierry Garrot [1] analysent les mécanismes de contrôle des ressources humaines pour satisfaire aux objectifs d'équilibre budgétaire. Ils font le parallèle entre l'introduction du New Public Management dans le monde de la santé et les mécanismes de régulation du panoptique de Bentham décrit par Michel Foucault [28]. Ils introduisent la notion de perte d'initiative des équipes soignantes, et par extension de sens, dans la prise en charge des patients. Les injonctions de devoir rendre compte budgétairement en temps quasi réel induit un biais d'attention des équipes soignantes. Cela se traduit jusque dans la façon de concevoir les relations entre soignants tout comme la relation soignant-soigné. ●