

- Sections
- Une base de données unique, des accès différenciés
 - Un recours obligatoire à des algorithmes sophistiqués
 - Protéger les données personnelles
 - Une veille en temps réel grâce aux flux bruts

Aller plus loin Commentaires



Faire parler les données de santé

24.03.2016, par Guillaume Garvanèse

© A. DUCLOS/SIPA

Une récente loi prévoit d'ouvrir l'accès aux données de santé des assurés sociaux. Une occasion pour certains d'étudier les maladies, leur évolution et l'efficacité des traitements, et pour d'autres de s'inquiéter de la protection nécessaire de la vie privée et du secret médical.

Avec près de 1,2 milliard de feuilles de soins par an, le Système national d'information inter-régimes de l'assurance maladie (Sniiram) est l'une des plus grosses bases de données de santé au monde. Ce foisonnement d'informations fait le bonheur de nombre de chercheurs. Mais son accès est très encadré, trop pour certains. L'article 193 de la récente loi de modernisation de notre système de santé prévoit d'ouvrir en partie l'accès à ce fichier géré par la Caisse nationale d'assurance maladie des travailleurs salariés (Cnamts). Tandis que certains voient dans cette ouverture un formidable moyen, grâce aux algorithmes de big data, de mener des études de pharmacovigilance ou d'évaluer les soins et les politiques de santé, d'autres pensent qu'elle risque de porter atteinte à la vie privée, au secret médical, ou d'être exploitées par les assureurs pour recalculer les primes payées par leurs clients.

Une base de données unique, des accès différenciés

Âge, sexe, lieu de résidence, détail de toutes les prestations remboursées, données relatives aux séjours... Le recensement des divers soins fournis à près de 65 millions d'assurés sociaux intéresse de nombreux acteurs de la santé et de la recherche. Un arrêté de juillet 2013 fixe la liste des 25 organismes ayant accès aux bases de données du Sniiram.

Pour des raisons historiques évidentes, le principal exploitant du Sniiram est l'Assurance maladie elle-même. Et ses champs d'action sont très étendus. « Dans le cadre de la gestion du risque, nos observations s'étendent du suivi d'événements indésirables, comme l'étude réalisée en 2013 en partenariat avec l'Agence nationale de sécurité du médicament (ANSM) sur le risque thromboembolique des pilules de 3^e génération, à l'observation des parcours de soins en passant par la promotion de référentiels de bon usage des soins auprès des professionnels de santé », précise Claude Gissot, directeur de la stratégie, des études et des statistiques à la Cnamts. Du côté des patients, la collecte des tarifs a permis à l'Assurance maladie de publier les tarifs des professionnels et de proposer des programmes d'accompagnement spécifiques aux malades. En outre, la Cnamts scrute les bases de données à la recherche d'anomalies révélatrices d'actes frauduleux. « D'un point de vue macroéconomique, nous surveillons et analysons de près l'évolution des dépenses, car la question de la soutenabilité financière du système de santé à court et moyen terme est au cœur de nos missions », ajoute Claude Gissot.

“Le recensement des soins fournis à près de 65 millions d'assurés sociaux intéresse de nombreux acteurs de la santé et de la recherche.”

Bien entendu, tous les organismes ne disposent pas des mêmes accès. Les données individuelles ne sont accessibles qu'à l'Assurance maladie, aux agences régionales de santé dans la limite de leur périmètre géographique, à la Haute Autorité de santé (HAS), à l'Agence nationale de sécurité du médicament (ANSM) et à l'Institut national de veille sanitaire (INVS). D'autres institutions, telles que l'Institut national du cancer, les fédérations hospitalières, l'Institut national de la santé et de la recherche médicale (Inserm), l'Institut de recherche et documentation en économie de la santé (Iredes) ou encore le CNRS, ne bénéficient quant à elles que d'un accès aux bases de données agrégées comme le suivi des dépenses, le suivi de l'activité et l'échantillon général des bénéficiaires (EGB), qui permet d'analyser les parcours individuels.

Pour les autres organismes, après avoir obtenu l'autorisation de la Commission nationale de l'informatique et des libertés (Cnil), l'Institut des données de santé (IDS) autorise ou non l'accès aux données après examen de l'objet de la recherche. Entre janvier 2014 et juin 2015, les CHU ont représenté à eux seuls 38 % des demandes à l'IDS. Les recherches portaient majoritairement sur le suivi de pathologies comme Alzheimer, l'asthme ou le cancer.

En plus des données à accès réglementé du Sniiram, l'Assurance maladie publie en open data certains jeux de données strictement anonymes. À ce jour, 17 jeux de données sont publiés sur la plateforme du gouvernement, parmi lesquels la base des dépenses d'Assurance Maladie ou la base des personnes souffrant d'affections de longue durée (ALD).

Un recours obligatoire à des algorithmes sophistiqués

L'exploitation du Sniiram, qui est composé de bases de données thématiques (appelées datamarts) créées selon des besoins spécifiques, constitue un gigantesque défi technique. D'abord en raison de sa taille – près de 500 Téraoctets –, qui en fait l'une des plus grandes bases de données de santé au monde. Ensuite à cause de sa structure hétérogène combinant 15 datamarts, un EGB et une base des données individuelles des bénéficiaires (DCIR). Ces caractéristiques rendent le Sniiram inexploitable sans traitement préalable.

La Cnamts s'est donc tournée fin 2014 vers le Centre de mathématiques appliquées (CMAP)^[1] afin d'explorer les possibilités offertes par les technologies du big data au domaine de la santé. L'objet de la recherche menée par le CMAP consiste à faire parcourir cette masse de données par des algorithmes non supervisés à la recherche de signaux faibles. Des algorithmes similaires à ceux utilisés par les grands moteurs de recherche du Web.



© HUMBERT/ESP

Imaginons une courbe des individus par leur sexe et leur âge. Un algorithme classique ne déduirait pas de lui-même que le poids évolue en fonction de l'âge : ses concepteurs doivent lui indiquer ce lien afin que l'algorithme détecte par la suite certaines anomalies, comme une augmentation ou une perte de poids anormale sur une population qui prend un certain type de médicament. Cette approche est limitée à la validation d'hypothèses préétablies.

En revanche, « la propriété d'un algorithme non supervisé sera de découvrir par lui-même, sans apprentissage préalable, que le poids est corrélé à l'âge », explique Emmanuel Bacry, chercheur au CMAP et responsable du partenariat pour l'école polytechnique. Cette différence est fondamentale car, en explorant la base Sniiram, ces algorithmes autonomes seront en mesure d'effectuer des rapprochements auxquels personne n'avait pensé et, espèrent les chercheurs, de détecter des anomalies si infimes soient-elles. Dans notre précédent exemple, l'algorithme découvrirait de lui-même comment le poids des individus évolue avec l'âge, plus détecterait tout seul que certaines personnes prenant un certain médicament révoquent pas de la même manière. « De tels algorithmes mettront peut-être un jour en évidence des différences de résultats de santé entre des parcours de soins de patients liés à des questions d'orientations de pratiques professionnelles hétérogènes », espère Claude Gissot. Ces algorithmes savent utiles pour repérer les signaux faibles, découvrir des facteurs jusqu'ici cachés et formuler de nouvelles hypothèses ; hypothèses qui doivent toutefois être ensuite validées de manière classique.

Protéger les données personnelles

L'existence de la base Sniiram pose inévitablement l'épineuse question de la sécurité des données personnelles. L'ouverture aux organismes privés permise par l'article 193 de la future loi de modernisation de santé a ainsi fait bondir les associations d'usagers qui y voient un danger pour la protection des données personnelles. La loi impose néanmoins des garde-fous afin d'éviter que les assureurs privés ne se servent pour profiler les individus et adapter leurs tarifs en fonction. Elle prévoit notamment que les opérateurs privés et publics pourront bénéficier d'un accès aux données pseudonymisées à condition d'une part de justifier de l'intérêt public de leur usage et d'autre part qu'elles ne soient pas exploitées dans le but de promouvoir ou de modifier les contrats d'assurance.

“La loi impose des garde-fous afin d'éviter que les assureurs utilisent la base du Sniiram pour profiler les individus.”

La pseudonymisation automatique des données est assurée par l'application de la procédure Foin (Fonction d'occultation des informations nominatives), élaborée en 1996 par le Cnamts. Cette dernière repose sur une fonction dite de hachage qui consiste à transformer de manière irréversible le numéro de sécurité sociale, le sexe et la date de naissance du patient. Une première application de Foin intervient au moment où les informations partent des caisses locales, puis un second traitement est effectué à leur arrivée dans le Sniiram, au Centre national de traitement de l'informatique de la Cnamts à Evreux.

Le processus d'ouverture d'un jeu de données doit répondre aux grandes règles de l'anonymisation. Ces règles permettent d'empêcher la réidentification des données en les dégradant tout en veillant à en préserver l'information utile aux études de santé. Malgré tout, le risque de réidentification reste techniquement présent. « On ne sait jamais de quel type d'informations auxiliaires l'entité qui analyse les données dispose », explique Kévin Huguenin.

« Des chercheurs de l'université du Texas à Austin sont déjà parvenus à identifier des utilisateurs de Netflix en comparant le contenu de sa base anonymisée avec les notes attribuées sur le site Internet IMDb pourvus ». On pourrait donc essayer de croiser les données ouvertes par l'Assurance maladie avec les informations disponibles sur les assurés (localisation, âge, symptômes, etc.) en particulier les informations laissées sur les réseaux sociaux et les forums tels que PatientsLikeMe ou Doctissimo. »



© O. MARK/FOTOLIA.COM

L'Assurance maladie a sollicité le Centre de mathématiques appliquées pour qu'il explore les possibilités offertes par les technologies du big data dans le domaine de la santé.

Une veille en temps réel grâce aux flux bruts

Enfin, en plus de l'accès aux bases d'archives de l'Assurance maladie, certains acteurs aimeraient pouvoir exploiter les flux bruts, ce flot continu d'informations qui partent des caisses locales à destination du Sniiram. Analyser ces flux permettrait par exemple de mettre en place des veilles en temps réel. Toutefois, personne n'y a accès, celui-ci étant contrôlé par l'Assurance maladie. Certains regrettent cette situation, tel Jean-Yves Robin, directeur général d'OpenHealth Company. Cette société a créé un réseau de plus de 6 000 pharmacies qui lui transmettent les données de vente de médicaments. À partir de ce panel, la société propose une série de veilles sur la vaccination grippale, l'étenue de la gastro-entérite ou encore les manifestations allergiques. Ce procédé n'est pas propre à cet opérateur privé. L'INVS organise sa veille en temps réel à travers des réseaux comme Oscour (Organisation de la surveillance coordonnée des urgences) ou Sursaud (Surveillance sanitaire des urgences et des décès).

Mais s'il est très probable que le suivi des flux bruts permette d'étayer des stratégies commerciales et marketing en matière de médicament, d'autres doutent que ce seul suivi permette de prévenir des problèmes sanitaires. En effet, le suivi de la consommation des soins rattachée au patient est nécessaire et il se fait en rapprochant les données de suivi des médicaments avec les données issues du Programme de médicalisation des systèmes d'information (PMSI) via l'accès au Sniiram. Des dispositions d'accès qui figurent justement dans la nouvelle loi santé.

Notes

1. Unité CNRS/École polytechnique.

ALLER PLUS LOIN 0 commentaire

Mots-clés: Sniiram, Cnamts, Assurance Maladie, Soins, Santé, Base De Données, Vie Privée, Secret Médical, Anonymisation, Foin, Cnil, Flux Bruts

Partager cet article: Facebook (635), Twitter, Email, Google+ (4)

Auteur: Guillaume Garvanèse

Formé à l'Institut de journalisme de Bordeaux et à l'école de photographie Efest, Guillaume Garvanèse est journaliste et photographe, spécialisé dans les domaines de la santé et du social. Il a notamment travaillé pour le groupe Le Monde.

En savoir plus sur l'auteur

Prométée, la plateforme de extrême Comment nos cellules ont-elles appris à respirer?

Voir aussi: Vivant

Commentaires: 0 commentaire Flux RSS

Pour laisser votre avis sur cet article Connectez-vous, rejoignez la communauté du Journal CNRS

Se connecter Créer un compte