

Big data : premier succès dans l'alerte sanitaire

Une méthode originale promet de débusquer les effets secondaires des médicaments dans les archives de l'Assurance-maladie.

LE MONDE SCIENCE ET TECHNO | 30.01.2018 à 14h00 • Mis à jour le 30.01.2018 à 17h59 | Par David Larousserie

(journaliste/david-larousserie/)

« *C'est le plus beau sujet de recherche de ma vie !* », s'enthousiasme Emmanuel Bacry, directeur de recherche au CNRS, attaché à l'université Paris-Dauphine. Ce mathématicien, également responsable de l'initiative data science de l'École polytechnique, a été autorisé, depuis 2014, à toucher l'un des diamants que rêve d'approcher tout spécialiste des mégadonnées, ou big data : le Sniiram.

Sous ce nom, qui signifie Système national d'information interrégimes de l'Assurance-maladie, sont rassemblées plus de un milliard de feuilles de soins annuelles des quelque 65 millions d'assurés français. Ses plus de 100 téraoctets (soit 100 000 gigaoctets) contiennent les remboursements, l'âge, le sexe, l'adresse, les prescriptions, les consultations... de tous les assurés.

L'ensemble est une vraie mine pour les économistes de la santé, les sociologues ou les statisticiens, qui, comme Emmanuel Bacry, traquent des effets secondaires rares des médicaments. L'Assurance-maladie elle-même ne s'en prive pas. Ses équipes ont confirmé le lien entre l'antidiabétique Mediator et des atteintes des valves cardiaques. Ou bien les liens entre pilules de troisième et quatrième génération et des embolies pulmonaires.

Un antidiabétique repéré

Emmanuel Bacry, et son collègue Stéphane Gaïffas, en collaboration avec des spécialistes de l'Assurance-maladie, viennent, eux, de repérer un autre problème, décrit dans [un article soumis \(https://arxiv.org/abs/1712.08243\)](https://arxiv.org/abs/1712.08243) à une revue de biostatistique. Un antidiabétique, la pioglitazone, augmente le risque de cancer de la vessie. Les spécialistes s'étonneront, à raison, de l'intérêt d'un tel résultat. La pioglitazone a été retirée de la vente en 2011, justement parce qu'une équipe de l'Assurance-maladie avait repéré cet effet indésirable dans les données du Sniiram.

L'intérêt est donc d'abord dans la méthode – ainsi validée – plus que dans le résultat.

« *Actuellement, chaque médicament doit être testé individuellement. Ce processus, qui nécessite de réunir des spécialistes pour définir précisément la bonne cohorte, est long. Tout comme l'extraction de cette cohorte, en interrogeant la base de données,* résume Emmanuel Bacry. *Avec notre méthode, on peut aller plus vite et tester plusieurs médicaments en même temps, avec un travail sur la cohorte largement simplifié et non spécifique à chaque médicament. Il est aussi possible de mettre en évidence d'éventuelles interactions néfastes.* » La technique s'apparente à du dépistage automatique, sans a priori sur les molécules concernées. De quoi alerter, afin de procéder à des évaluations classiques.

Mais, pour y arriver et gagner ainsi du temps, il a fallu commencer par en « perdre ». Pendant plus de deux ans, jusqu'à six ingénieurs se sont échinés à transformer le monstrueux Sniiram en un géant plus simple à dompter. « *Le Sniiram est adapté à ce pour quoi il a été conçu, les remboursements de santé, mais il est moins pratique pour des études épidémiologiques* », explique le chercheur. Cette vaste base de données est une sorte de tableau (en fait la juxtaposition de plusieurs centaines de tables contenant diverses informations), constitué de lignes de feuilles de soin, quand les statisticiens voudraient plutôt des lignes avec des patients et tout leur parcours. Ce changement d'infrastructure est si lourd que l'équipe ne l'a fait que sur un extrait du Sniiram, de quelques téraoctets, pour prouver leur concept.

Un travail unique au monde

La preuve finale est donc ensuite arrivée sur un cas déjà connu, la pioglitazone, qui, comparé à cinq autres traitements antidiabétiques, est le seul à augmenter les risques. Pour cette dernière phase, les chercheurs ont quitté le domaine du big data. Des 2 millions de patients initiaux, il ne restait plus qu'une cohorte de 1 800 personnes ayant eu un cancer de la vessie. Une seule machine de calcul a alors suffi pour sortir le résultat, quand une vingtaine a été nécessaire pour la transformation du mini-Sniiram.

« *Ce que nous avons fait est unique au monde, car cette base est unique au monde* », affirme Emmanuel Bacry. Les Etats-Unis, au système d'assurance éclaté entre plusieurs acteurs, ne possède pas de telles quantités. Idem pour les pays du nord de l'Europe, aux fichiers très exhaustifs, mais à la population moins grande.

« *Même s'il est encore tôt pour conclure sur l'apport des big data, elles représentent clairement un potentiel dont la mise en œuvre opérationnelle est en développement* », estime Claude Gissot, directeur de la stratégie, des études et des statistiques de l'Assurance-maladie.

La suite, concrétisée dans un nouveau partenariat entre l'Ecole Polytechnique et l'Assurance-maladie, cherchera à découvrir des faits jusque-là inconnus. L'équipe étudie ainsi, à partir d'une cohorte dix fois plus grande, les liens entre la prise de médicaments et les chutes de personnes âgées.