

Les Données de Vie Réelle (Real-World Data) jouent un rôle croissant en santé. L'appariement de plusieurs bases de données complémentaires pourrait faciliter la connaissance médicale et la prise de décision.

Les procédures d'appariement et d'intégration de bases différentes doivent néanmoins franchir certains obstacles techniques, éthiques, et réglementaires.

L'appariement entre les données cliniques (dossiers médicaux ou registres) et les données médico-administratives du système national de données de santé (SNDS) mérite une attention toute particulière.

Quel est l'intérêt des données du SNDS pour la recherche?

En 1999, le Système National d'Information InterRégimes de l'Assurance Maladie (SNIIRAM) a été lancé, afin de connaître et d'évaluer la consommation de soins médicaux et leurs coûts en France.

Les principaux intérêts du SNIIRAM (remplacé maintenant par le SNDS) comprennent la surveillance sanitaire (y compris l'épidémiologie des maladies), l'analyse des pratiques médicales, le suivi de la mise en œuvre des politiques de santé, l'économie de la santé, et la pharmacoépidémiologie. Une limite du SNDS est l'absence de données cliniques (pas de diagnostic, sauf en cas d'hospitalisation ou d'ALD -Affection de Longue Durée), ou de résultats pour les examens paracliniques.

Des algorithmes ont néanmoins été développés pour identifier des maladies précises, en combinant codes diagnostiques, procédures médicales et traitements. Par ailleurs, actuellement, les données socio-démographiques sont limitées, sans information sur la profession ni les revenus.

L'utilisation du SNDS est cependant rendue complexe par sa taille et sa structure, déterminés par des priorités comptables. Les analyses nécessitent donc des compétences analytiques pointues et une bonne compréhension des consommations de soins, et de leur facturation. Il est également important que toute investigation initiale soit conduite en bonne interaction entre « data scientists » (analystes et statisticiens) et experts médicaux du domaine.

Pourquoi faut-il considérer que données cliniques et les données du SNDS sont complémentaires ?

Dans une étude clinique, les objectifs prioritaires du projet et les contraintes budgétaires déterminent souvent les données collectées. Par exemple, une étude clinique conçue pour évaluer le taux de réponse global sur quelques mois est parfois peu informative pour le taux de survie. Par ailleurs, des informations peuvent être obtenues dans le SNDS qui ne sont pas recueillies au cours de l'étude clinique, comme les procédures de soins primaires ou secondaires effectuées avant ou après la période d'étude.

Les données de consommations de soins peuvent également être utilisées comme source exhaustive et valide (sans biais de mémorisation) de variables explicatives dans des modèles prédictifs de réponse aux traitements. Cela permettrait par exemple d'identifier des sous-groupes où les traitements ont une efficacité et une sécurité optimales, ou de prédire les résultats à long terme. En outre, les données d'utilisation des soins consommés pendant la période d'étude permettent des vérifications a posteriori de la qualité des données rassemblées dans les essais.

Les données du SNDS peuvent également être chaînées avec des registres nationaux, par exemple afin de vérifier la représentativité des populations incluses.

Comment chaîner les données cliniques au SNDS ?

D'un point de vue technique, le chaînage peut être probabiliste, en utilisant un nombre minimal de variables communes (« balises ») pour assurer une correspondance individuelle entre les deux bases, avec une faible probabilité d'erreur. Le chaînage peut également être déterministe lorsque l'identifiant du patient SNDS est collecté parmi l'ensemble des données cliniques.

Avec la méthode probabiliste, le succès d'un appariement dépendra de la possibilité d'identifier des variables communes aux deux ensembles de données, qui soient suffisamment spécifiques. Dans le cas d'une collecte primaire de données cliniques (étude de terrain) chaînée ensuite aux données du SNDS, les objectifs de l'étude doivent être d'emblée précisés et le protocole de l'étude doit préciser la nécessité d'un chaînage au SNDS.

Quelle que soit la méthode de chaînage, les règles de confidentialité des patients doivent être strictement appliquées et doivent être conformes au RGPD et aux procédures de la CNIL, ainsi qu'aux bonnes pratiques pharmacoépidémiologiques.

A propos de PELyon

PELyon travaille depuis plusieurs années sur les chainages entre sources de données, en ayant par exemple réalisé un chaînage entre des dossiers médicaux électroniques, des informations déclarées par les patients, et le SNDS dans le cadre du projet européen ASTROLAB, ou en réalisant le chaînage entre le registre national de la mucoviscidose et le SNDS. Les résultats de ces chainages donnent toujours lieu à publication dans des revues scientifiques.

SOURCE : HEALTHCARE DATA INSITUTE :

<https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUK Ewjxr8yHiYnpAhW95OAKHVo CwwQFjAAegQIBxAC&url=https%3A%2F%2Fhealthcaredatainstitute.com%2F&usg=AOvVaw0I8ZRmgzHajqUg-kpfRcpC>