
Utiliser le Système National des Données de Santé pour des analyses pharmaco-épidémiologiques : quelques défis méthodologiques à relever!

Emmanuel Oger*¹

¹Centre régional de pharmacovigilance, pharmaco-épidémiologie et information sur le médicament, Upres-ea 7449 reperes, C HRU de Rennes, Rennes – Je ne comprends pas – France

Résumé

Le Système National des Données de Santé contient des données de remboursement de prestations de soins, notamment les médicaments délivrés en pharmacie et des données médicales sur les séjours hospitaliers avec un codage diagnostique et des informations sur les actes réalisés, diagnostiques ou thérapeutiques. Ces informations existant pour la quasi-totalité de la population française et sur une profondeur de champ de plus de dix ans maintenant, on comprend l'intérêt suscité : il est possible de reconstituer des cohortes d'individus quasi exhaustives et sans biais d'attrition.

Le Système National des Données de Santé ne contient pas de données cliniques pourtant fondamentales : motif de prescription d'un produit de santé (son indication médicale), habitudes de vie ou résultat d'examen biologique ou d'imagerie pour ne citer que quelques exemples.

Outre les difficultés techniques liées à la volumétrie des données et à leur structure (il s'agit d'une base comptable ayant évolué par sédimentation de strates), sans oublier les quelques obstacles réglementaires, la reconstitution de données longitudinales de qualité pour une analyse épidémiologique soulève quelques défis. Pour ne citer que trois questions, retenons celles-ci : (1) peut-on qualifier une exposition à un produit de santé à partir de traces temporelles (les délivrances)? Comment gérer les changements d'exposition à un produit de santé au cours du temps, changements certainement pas dus au hasard mais dont les raisons ne sont pas disponibles dans les données? (2) Peut-on pallier le manque d'information médicale en couplant les données du Système National des Données de Santé avec d'autres données issues de cohortes, de registres ou d'essais cliniques ? Un couplage sans clé commune fiable (se serait trop facile!) est-il possible et quelle conséquence sur le résultat de l'étude? (3) Comment analyser la survenue d'un événement de santé qui ne serait pas brusque et grave (et conduisant alors à une hospitalisation facile à identifier)?

Les très grandes bases de données ne feraient-elles qu'amplifier les questions de gestion des biais que se doit de discuter tout épidémiologiste sans se décourager et en toute humilité? Et une dernière question, que devient l'inférence lorsque toute la population est là ?

*Intervenant