

Comment appairier des données cliniques avec les données du SNDS ?

Lucie de Léotoing et
Patrick Blandin du groupe
AFCROS-DM

Il est désormais possible de rapprocher des données médicales de données médico-administratives, à condition toutefois de respecter un cadre légal strictement défini. Des spécialistes de l'AFCROS nous expliquent les modalités pratiques d'appariement de ces données.



Lucie de Léotoing



Patrick Blandin

Une voie particulièrement intéressante pour l'épidémiologie ou pour l'étude des produits de santé en vie réelle est l'enrichissement de données d'enquêtes ou de données cliniques recueillies directement auprès des personnes, par appariement avec le SNDS (Système National des Données de Santé). Ces projets combinant les données de sources diverses se multiplient grâce aux nouvelles conditions d'appariement de bases de données prévues par la loi. Ils peuvent être utiles aux différentes étapes de développement d'un dispositif médical (DM) et permettent de rapprocher des données médicales de données médico-administratives, afin d'aborder l'aspect relatif au financement et au coût des prestations. Mais comment peut-on en pratique appairier les bases de données entre elles ?

Quelques projets de grande ampleur s'appuient déjà largement sur des appariements de bases de données nationales, comme les grandes cohortes épidémiologiques Constances et Elfe ainsi que les programmes de suivi post-professionnels Spirale et Espri. Des enquêtes en population, telle que l'enquête Handicap-santé de l'INSEE, ont également récemment testé la faisabilité d'un appariement avec la base de données du SNIIRAM.

Un préalable indispensable à tout appariement de bases de données est l'identification des sujets. En effet, les bases composant le SNDS ne comportent aucune donnée directement identifiante. Les données sont rattachées aux personnes par un procédé cryptographique irréversible attribuant de manière unique à chaque personne née en France un Numéro d'Inscription au Répertoire (NIR). Ce numéro est communément appelé « numéro de sécurité sociale » et est formé de 13 chiffres, obtenus par un procédé cryptographique (qui ne permet pas de remonter à l'identité du patient) : le sexe (1 chiffre), l'année de naissance (2 chiffres), le mois de naissance (2 chiffres) et le lieu de naissance (5 chiffres). Les 3 chiffres suivants correspondent à un numéro d'ordre qui permet de distinguer les personnes nées au même lieu à la même période ; une clé de contrôle à 2 chiffres complète le NIR.

Deux méthodes sont donc possibles. La première est celle de l'appariement direct des sujets repérés par le NIR, commun aux deux bases, lorsqu'il est colligé dans la base clinique. La deuxième méthode est celle de l'appariement probabiliste, dans le cas où le NIR n'est pas reporté. Quelle que soit la méthodologie retenue, la base à appairier au SNDS doit faire/avoir fait l'objet d'une autorisation de la CNIL. Un paragraphe dédié à l'explication de la métho-

dologie d'appariement et des modalités de transfert sécurisé des données doit être rédigé dans le protocole qui est soumis au CEREES et à la CNIL.

Appariement direct par l'utilisation du NIR

L'article 193 de la loi de modernisation de notre système de santé comporte un article dérogeant à la nécessité d'un décret en Conseil d'État pris après avis de la CNIL sur les modalités d'utilisation du NIR. De fait, une base de données contenant le NIR des patients peut désormais être appariée directement au SNDS.

Que la base de données à appairier soit une base à vocation de recherche (étude clinique, registre par ex.) ou non (base de soins par ex.), le NIR ne doit en aucun cas figurer directement dans la base de données mais être enregistré et crypté dans une base différente. Le recours à un tiers d'appariement n'est pas obligatoire ; après avis favorable du CEREES et autorisation de la CNIL, l'investigateur transmet directement la table des NIR à la CNAM pour procéder à l'extraction des données du SNDS correspondant aux NIR de la table. Cette table doit répondre à une norme stricte de format et nécessite de renseigner le NIR de l'ouvrant droit, la date de naissance, le sexe et le NIR du bénéficiaire ainsi que le nom du projet afin de constituer l'identifiant SNDS du bénéficiaire utilisé pour les extractions de la base SNDS. Ce fichier suit un procédé de double anonymisation (c'est ce procédé qui est à l'origine de l'assouplissement de la loi sur l'utilisation du NIR) avant que l'algorithme ne permette l'extraction de la totalité des données du bénéficiaire. En pratique, recourir à un tiers de confiance facilite la démarche entre l'investigateur, la CNAM et le responsable de la mise en œuvre du traitement. Ce tiers de confiance est chargé de transmettre à la CNAMTS le NIR et les identifiants des patients dans l'étude, il doit être indépendant du responsable de la mise en œuvre.

Le projet de loi relatif à l'organisation et à la transformation du système de santé voté le 22 mars dernier a permis de mettre en place une plateforme d'exploitation des données de santé appelée Health Data Hub (HDH) qui reprend les missions actuelles de l'Institut national des données de santé (INDS). Les promoteurs peuvent désormais soumettre leurs projets auprès du HDH chargé de "réunir, organiser et mettre à disposition les données du SNDS". Pour ce faire, le HDH devra "procéder, pour le compte d'un tiers, à des opérations nécessaires à la réalisati-

tion d'un traitement de données issues du SNDS pour lequel ce tiers aura obtenu une autorisation".

Appariement probabiliste

Il s'agit d'une méthode d'appariement indirect. Elle doit être mise en place lorsque la base de données (registre, essai clinique, entrepôt de données) à appairier au SNDS ne contient pas le NIR (ce dernier n'a pas fait l'objet d'une demande auprès de la CNIL lors de la constitution de la base). Cette méthode nécessitant la présence de variables discriminantes, la difficulté est d'en prendre suffisamment pour éviter les erreurs d'appariement, mais en nombre limité pour éviter les échecs d'appariement. Il convient donc dans un premier temps d'identifier les variables communes entre la base de données externe et le SNDS. Cette étape nécessite de bien connaître les variables recueillies dans la base clinique. Ensuite, il sera impératif de vérifier si les variables communes sont suffisamment discriminantes pour rapprocher les individus entre les bases. A minima, les variables d'intérêt suivantes peuvent être utilisées : mois et année de naissance ; sexe ; code postal du lieu de naissance ; date de prestations de soins. C'est ainsi la confrontation d'informations sur les dates et natures des contacts ainsi que les informations liées à l'identité du patient qui peuvent permettre de rapprocher les deux bases. Un point d'attention doit cependant être porté sur le code postal du lieu géographique de

résidence. Les patients actualisent peu souvent leur carte vitale. De ce fait, la CNAM a observé une distorsion entre le lieu réel de résidence et le code postal recueilli via la carte vitale. De même, la date d'une prestation de soin hospitalier n'est pas forcément facilement repérable. Il faut prévoir d'embler un intervalle de confiance dû au fait que l'acte est inséré dans un séjour hospitalier. Via le PMSI, on dispose des dates de sortie et par déduction des dates d'entrée.

L'investigateur transmet à la CNAM les variables d'intérêt cryptées. La CNAM décrypte cette table à partir d'une clé préalablement envoyée, réalise l'appariement des variables avec le SNDS et extrait les données du SNDS correspondant à la demande. Ces données sont ensuite cryptées puis mises à disposition de l'investigateur via un flux sécurisé sur le portail de la CNAM. Cette méthode permet d'appairier plus de 80 % des patients d'une base. La CNAM a déjà conduit ce type d'appariements probabilistes.

Ainsi, si les bases constitutives du SNDS sont chaînées entre elles, la loi permet également un appariement ponctuel du SNDS avec d'autres bases extérieures, telles que registres / cohortes, essais cliniques, entrepôts de données... Ces appariements permettent de combiner les données médico-administratives à des données cliniques, quasi-absentes du SNDS. Une amélioration de l'interopérabilité des systèmes d'information permettra l'essor de ces appariements et en facilitera les modalités.

www.afcros.com

DeviceMed INFO

Un tiers d'appariement est un organisme intermédiaire qui permet à au moins un utilisateur d'utiliser des données auxquelles il n'a pas accès directement, notamment parce que certaines d'entre elles lui sont interdites ou qu'elles comportent des données identifiantes.