

SOMMAIRE

- **Croiser les bases médico-administratives** avec les bases de données cliniques
Farid Benabdallah, Directeur des Affaires médicales d'Ethicon – J&J
- **Le circuit** est clairement identifié
Mylène Girard, Responsable de la mission d'accès aux données de santé à la DREES
- **Le guichet unique de l'INDS**
Dominique Polton, présidente de l'Institut national des données de santé (INDS)
- **Un pouvoir de simplification renforcé pour la CNIL**
Hélène Guimiot-Bréaud (Cnil)
- **Pour faciliter l'utilisation des données de santé par les chercheurs**
Frédérique Lesaulnier (Inserm)
- **Bien connaître l'historique d'enregistrement des données**
Alain Weill (Cnamts)
- **L'avenir s'oriente vers une approche mixte**
Stéphane Bouée (AFCROS)
- **L'apport des e-cohortes**
Guy Fagherazzi (Inserm – cohorte E4N)



Les sources de données de santé – un panorama toujours plus riche

Comme pour faire suite à la formidable ouverture sur le monde de notre société, l'accès aux données de santé publiques issues des bases de données médico-administratives a une large place au sein de la Loi de modernisation du système de Santé du 26 janvier 2016.

Sous le haut pilotage de la DRESS, l'Etat pose désormais et depuis Avril 2017, l'INDS comme guichet unique de réception des dossiers des chercheurs et le CERES comme bureau d'experts pour juger, sous 1 mois, la recevabilité des demandes d'accès aux bases déjà constituées. L'autorisation in fine reste du domaine de la CNIL, qui voit la fusion des chapitres IX et X et l'élargissement de ses compétences de sanction - tous se préparant au règlement européen sur la protection des données personnelles applicable en mai 2018.

Les réponses aux questions de recherche passent essentiellement par la mise en place d'études de cohorte prospectives, elles-mêmes nourries désormais de multiples moyens de collecte de données, adaptés aux champs et aux populations d'étude – eCRF, objets connectés, croisement de données cliniques et observationnelles, données de bases privées ou associatives – permettant un meilleur ciblage des patients et facilitant leur entrée dans des études interventionnelles.

Le montage de telles études est fastidieux et couteux mais reste essentiel. Une structure de services est créée pour les chercheurs de tous horizons à l'initiative de l'INSERM. Les CROs se posent comme experts de la faisabilité et de la gestion de ces projets complexes.

Si ces bases de données sont un enjeu majeur pour toute la communauté des chercheurs issue du domaine public et privé, c'est un pan entier de la recherche en santé qui se transforme, élargissant le champ des possibles dans un univers qui reste très complexe et pas encore rodé.



Adeline Meilhoc,
DPsy, AFCROS
www.afcros.com

Croiser les bases médico-administratives avec les bases de données cliniques



Farid Benabdallah
Directeur des Affaires
médicales d'Ethicon - J&J

En tant qu'industriel, nos attentes sont liées aux exigences que nous avons à remplir vis-à-vis des autorités de santé, avec notamment les études nécessaires pour l'accès au remboursement. Mais nous avons aussi besoin, et c'est une exigence nouvelle, de pouvoir anticiper les différentes situations liées à la commercialisation et au suivi de nos produits sur le long terme. Il nous faut, entre autres, intégrer les problématiques de commercialisation dans les stratégies de développement pré-marquage CE et de péri-marquage CE. Nous participons aussi à des projets de santé publique dans le domaine de l'obésité ou de la prévention des cancers colorectaux, en partenariat avec les autorités de santé ou avec les praticiens experts dans le domaine concerné. L'accès à des bases de données médico-administratives serait très intéressant dans ce cadre. Il faudrait aussi pouvoir croiser ces bases avec les bases de données cliniques et cette question mériterait d'être débattue. Nous espérons aussi avoir plus d'éclaircissements sur les possibilités légales et réglementaires en termes d'accès et d'utilisation de ces bases de données médico-administratives via des organismes de recherche ou d'autres partenaires.

Le circuit est clairement identifié



Mylène Girard
Responsable de la mission
d'accès aux données de
santé à la DREES

La Direction de la Recherche, des Etudes, de l'Evaluation et des Statistiques (DREES) pilote les orientations stratégiques du système national des données de santé (SNDS) au sein du ministère des Solidarités et de la Santé. Le système est opérationnel depuis début avril et a nécessité l'élaboration de nombreux textes réglementaires destinés à permettre l'utilisation des données de santé par des acteurs publics et privés dans le respect du cadre mis en place par l'article 193 de la loi de modernisation du système de santé du 26 janvier 2016. Le circuit est maintenant clairement identifié avec une procédure classique d'accès et une procédure spécifique pour les organismes ayant une mission de service public. L'architecture juridique repose sur deux décrets d'application en date du 26 décembre 2016. Le premier décret réforme complètement l'accès aux données de santé et met en place les procédures nécessaires tandis que le deuxième organise, en contrepartie d'un certain nombre de dispositions, des accès permanents pour les organismes ayant une mission de service public¹. Plusieurs autres arrêtés sont encore en cours de préparation pour que l'architecture soit complète.

Les deux briques majeures du dispositif ont aussi été mises en place avec les publications de l'arrêté transformant l'Institut des données de santé en Institut national des données de santé (INDS) et de l'arrêté nommant les membres du Comité d'expertise pour les recherches, études et évaluations en santé (CEREES). L'INDS a un rôle de guichet unique tandis que le CEREES a un rôle d'expertise sur la qualité scientifique des protocoles déposés par les demandeurs. L'accès aux données se fait sur la base du critère d'intérêt public. Il existe aussi des finalités interdites. Les industriels n'ont pas le droit de se servir des données de santé pour la promotion commerciale et les assureurs ne peuvent pas les utiliser pour moduler des contrats individuels ou collectifs.

La procédure classique d'accès et le dispositif global sont en train de se mettre en place. Les accès permanents se déploient depuis avril et le CEREES vient d'être installé. Dans le cadre du comité stratégique de filière Industries et technologies de santé, la DREES copilote depuis plusieurs mois, avec le Leem et le Snitem, les travaux visant à faciliter l'accès aux données de santé à des fins de santé publique, de recherche et de développement industriel. Ces travaux ont donné lieu à la publication d'un rapport en mars dernier pour accompagner l'utilisation du SNDS. Ce rapport est composé de fiches explicatives détaillant le cadre légal et réglementaire du SNDS et de préconisations pour l'accès aux données de santé. Le rôle des bureaux d'études et des laboratoires de recherche, qui sont instaurés comme intermédiaires, y est également abordé.

(1) La liste des organismes publics pouvant avoir un accès permanent au SNDS a été publiée dans le cadre du décret du 26 décembre 2016, publié au Journal Officiel du 28 décembre 2016.

Le guichet unique de l'INDS



Dominique Polton,
présidente de l'Institut
national des données de
santé (INDS)

Tous les textes permettant à l'INDS de fonctionner et de remplir sa mission principale sont maintenant sortis. Nous sommes en train de construire à un rythme soutenu les dispositifs et les outils nécessaires pour ouvrir l'accès aux données de santé à davantage d'acteurs et accélérer les délais dans la fluidité et la transparence. Concernant les accès au système national des données de santé (SNDS), on distingue les données anonymes qui ne présentent aucun risque de réidentification et les données qui comportent un risque de réidentification. Les premières seront en libre accès et aucune procédure d'autorisation ne sera nécessaire. Mais il faut encore construire les outils permettant de les mettre à disposition. Ceci implique un important travail technique qui nécessitera peut-être d'établir des priorités.

Deux types d'accès sont possibles pour les données comportant un risque de réidentification. Certains organismes assurant des missions de service public pourront avoir des accès permanents, selon un périmètre défini, avec un historique et un cahier géographique d'accès, conformément aux dispositions fixées par le décret n°2016-1871. Pour les autres, l'accès se fera sur demande et les circuits d'autorisation prévoient trois types de procédures :

- Si la demande fait intervenir un recueil de données auprès des personnes, elle sera examinée par un comité de protection des personnes.

- S'il s'agit de réutilisation de données déjà constituées, et notamment du SNDS, dans la procédure standard classique, le porteur de projet demande un accès aux données du SNDS à l'INDS. Celui-ci a ensuite sept jours pour envoyer le dossier au Comité d'expertise pour les recherches, études et évaluations en santé (CEREES) qui dispose d'un mois pour donner un avis. Si le CEREES n'a pas émis d'avis à l'issue de ce délai, la réponse est réputée positive. L'autorisation finale est ensuite délivrée par la CNIL, comme c'était déjà le cas auparavant.

- La loi prévoit aussi trois procédures simplifiées, plus rapides que le circuit normal. Des méthodologies de référence pourront être élaborées. Les porteurs de projet pourront également demander une décision unique en cas de besoins très récurrents et formatés qui se répètent. Cette solution leur permettra de répéter les mêmes types d'exploitation sans avoir à redemander d'autorisation à chaque fois. Enfin nous réfléchissons, et ceci est à construire avec la CNIL, à la possibilité de procédures allégées dans le cas où les données sont appauvries et présentent moins de risques de ré-identification ou dans le cas d'échantillons particuliers. Ici, l'INDS pourrait ouvrir directement des accès sans passer par la CNIL et ce, dans des conditions homologuées et cadrées par la CNIL. Le but ultime est de fluidifier le système et d'alléger la charge de la CNIL.

Outre son rôle de guichet unique par lequel transitent toutes les demandes d'autorisation d'accès aux données de santé, l'INDS est aussi l'instance chargée d'évaluer le caractère d'intérêt public du projet soumis. Nous nous y préparons avec un comité de réflexion. L'INDS est un groupement d'intérêt public au service de tous les utilisateurs et sa mission est avant tout de les accompagner pour démultiplier l'accès aux bases et leurs utilisations, sachant que les besoins des utilisateurs (secteur public, industriels, bureaux d'études, chercheurs, enseignants, journalistes, professionnels de santé, usagers...) sont assez différents et nécessiteront la construction d'outils variés.

Un pouvoir de simplification renforcé pour la CNIL



Héléne Guimiot-Bréaud
(Cnil)

Nous sommes à une période charnière. La production de normes a été importante au cours des dernières années avec, au niveau européen, les règlements sur les essais cliniques et sur la protection des données à caractère personnel, et au niveau français, la loi Jardé et la loi de modernisation du système de santé. Cette dernière a créé le système national des données de santé (SNDS), et a transformé le CNDP en INDS. Elle a aussi un impact important sur la loi Informatique et Libertés puisque les deux chapitres (chapitres 9 et 10) qui s'appliquaient en matière de recherche en santé vont fusionner et créer un nouveau cadre légal pour la réalisation de recherches à partir de données de santé.

Le nouveau chapitre 9 de la loi Informatique et Libertés fusionne des dispositions qui s'appliquaient à des types de recherches très différents. Jusqu'à présent, ce chapitre s'appliquait aux études réalisées dans le cadre des dispositions du code de santé publique et le chapitre 10 s'adressait à des études très spécifiques, réalisées à partir de bases telles que le PMSI ou le Sniiram, dont on étudie les données figées à un instant t. La fusion des chapitres 9 et 10 doit simplifier le fonctionnement du système d'autorisation pour les utilisateurs. Les études réalisées dans le cadre des dispositions du code de la santé publique (recherches interventionnelles, recherches interventionnelles à risques et contraintes minimales, recherches non interventionnelles) doivent d'abord être soumises à un CPP puis à la CNIL. Les études correspondant à des recherches n'impliquant pas la personne humaine sont soumises au CEREES, par l'intermédiaire du guichet unique de l'INDS puis transmises à la Cnil pour autorisation. Le fonctionnement du système est donc simplifié pour les utilisateurs dans la mesure où, une fois le dossier transmis à l'INDS, les transferts se font directement entre institutions.

La loi de modernisation du système de santé a aussi renforcé le pouvoir de simplification de la Cnil. Elle maintient non seulement sa possibilité d'homologuer des méthodologies de référence pour les études relevant du chapitre 9 mais y ajoute maintenant la capacité d'homologuer des conditions d'accès à des jeux de données agrégées ou à des échantillons. Cette homologation est réalisée en collaboration avec le CEREES, auquel ces conditions d'accès sont au préalable soumises pour avis. Ces autorisations uniques représentent un vrai gain de temps pour les utilisateurs. En revanche, il faut être très attentif sur leurs critères d'attribution. Le responsable du traitement des données doit toujours être le même pour une même étude réalisée chaque année avec une même question et les mêmes catégories de données. Si l'une d'elle change, avec par exemple la demande d'une base supplémentaire, une demande de modification sera nécessaire. Autre point important, les catégories de destinataires doivent aussi rester les mêmes.

Pour faciliter l'utilisation des données de santé par les chercheurs



Frédérique Lesaulnier
(Inserm)

La France est un des rares pays au monde à disposer de bases médico-administratives concernant de façon exhaustive la totalité de sa population. Ces bases détenues par des organismes publics sont alimentées de façon permanente et continue, contiennent des données codées de manière standardisée et concernent des domaines stratégiques pour la recherche en santé. Leur utilisation revêt des enjeux de santé publique majeurs, qu'on les utilise seules ou appariées avec d'autres bases de données, telles que celles des cohortes ou des registres de morbidité. Dans ce contexte, le décret du 26 décembre 2016 relatif au traitement de données à caractère personnel SNDS a reconnu un rôle particulier à l'Inserm en tant que « coordinateur national d'infrastructures de recherche utilisant des données de santé » : celui de pouvoir assurer la réalisation des extractions et la mise à disposition des données du SNDS pour des traitements mis en œuvre à des fins de recherche, d'étude et d'évaluation dans le domaine de la santé, selon des modalités prévues dans le cadre d'une convention conclue avec la CNAM. Dans le prolongement de ce rôle spécifique, l'Inserm pilote une infrastructure nationale au bénéfice de l'ensemble de la communauté des chercheurs qui a vocation à servir d'interface entre le SNDS et cette communauté afin d'optimiser l'utilisation de ces données à des fins de recherche ; Les chercheurs qui souhaitent utiliser ces données complexes et massives qui ont été collectées au départ dans une logique administrative et non à des fins de recherche sont confrontés à des difficultés multiples face auxquelles les services d'assistance sont insuffisants et le besoin de mutualisation des services et des compétences patents. L'objectif est de proposer, en partenariat avec d'autres institutions et réseaux d'utilisateurs experts (INRIA, Redsiam...), aux chercheurs qui le souhaitent une palette de services à la carte sur différents aspects : scientifiques (préparation des requêtes, développement d'outils de requêtage, d'algorithmes d'identification des pathologies, documentation), techniques (proposition d'espaces sécurisés...) et réglementaires et éthiques (aide au montage du dossier, élaboration de normes simplifiées adaptées aux besoins, mise en conformité avec le RGPD...) pour faciliter l'utilisation de ces données majeures pour la santé publique. Elle permettra également la remontée coordonnée des besoins en assurant des interactions entre les organismes de recherche et l'INDS sur les usages et besoins de la recherche afin de faciliter l'utilisation des données de santé en vie réelle.



Bien connaître l'historique d'enregistrement des données



Alain Weill (Cnamts)

En tant qu'utilisateur de données de santé en vie réelle au long cours, je peux témoigner des progrès enregistrés. Peu de gens croyaient à leur utilité il y a une vingtaine d'années, maintenant il paraît tous les semaines une publication internationale ou un rapport réalisé à partir des données de santé en vie réelle. De nombreuses études en France se sont appropriées ce type de données et on observe une montée en charge importante au cours des deux ou trois dernières années. Mais, l'utilisation du SNDS, du Sniiram ou du PMSI nécessite une bonne connaissance de ces données enregistrées en routine pour d'autres objectifs. Il est en particulier important de connaître l'historique d'enregistrement des variables selon les bénéficiaires et les régimes d'assurance maladie. Les utilisateurs se servent parfois de variables qui n'ont peut être pas toute la fiabilité souhaitée, un certain nombre de données ayant été recueillies par tous les régimes mais n'étant saisies dans les bases que depuis un nombre d'années restreint. De plus, de nouvelles variables apparaissent chaque année au niveau du PMSI, tandis que d'autres disparaissent. On peut citer l'exemple d'une variable devant fournir une information sur les maladies rares. Celle-ci devait être enregistrée en routine par les hôpitaux, mais le bilan réalisé au bout de deux ans a montré que seule une dizaine d'entre eux l'enregistraient de façon fiable et systématique, ce qui a rendu impossible l'utilisation de cette variable qui aurait pourtant été précieuse pour l'étude d'un certain nombre de maladies rares. A l'inverse, jusqu'en 2011, il n'était pas possible de chaîner les informations d'une mère accouchant dans un établissement avec celles de l'enfant à sa naissance. Ces opérations sont maintenant réalisables et permettent de recueillir des informations de santé très importantes sur la mère (maladies avant, pendant et après la grossesse, médicaments pris pendant la grossesse...) et sur l'enfant à la naissance (présence de malformations congénitales...). Néanmoins, cette variable n'existait que dans les établissements publics en 2011 et les données ne sont complètement exhaustives qu'à partir de 2013. Ces éléments techniques doivent absolument être connus avant de lancer un tel type d'études et les exemples cités ne sont que quelques uns parmi des centaines d'autres.



L'avenir s'oriente vers une approche mixte



Stéphane Bouée (AFCROs)

Les études les plus utilisées sont des études ad hoc dont les méthodologies extrêmement variables peuvent aller d'une enquête sur la population générale pour des maladies fréquentes jusqu'à des études réalisées auprès des investigateurs (médecins, pharmaciens, autres professionnels) ou des études rétrospectives sur dossiers dans le cas des maladies rares. Ces études présentent l'avantage d'être construites spécifiquement pour le projet à réaliser. Leur inconvénient réside dans leur coût qui peut être réduit avec l'utilisation d'autres sources de données déjà disponibles telles que le SNDS, les objets connectés, les bases publiques spécifiques d'une pathologie, les grandes cohortes, les réseaux sociaux de patients et les bases privées issues de logiciels implémentés à partir des médecins ou des pharmaciens. Mais l'utilisation de ces sources se heurte souvent à des problèmes de complexité et de délai pour la contractualisation et l'accès aux données.

On oppose souvent les avantages et les inconvénients de la réalisation d'une étude ad hoc à ceux de l'utilisation des données du SNDS. L'avenir s'oriente plutôt vers une approche mixte, en utilisant ces 2 sources de façon conjointe. A minima, il peut s'agir de vérifier la représentativité d'une étude ad hoc avec les données du SNDS. Pour aller plus loin, on peut réaliser deux études lorsque l'étude post-inscription nécessite le suivi du sujet de façon prospective : 1) Une étude ad hoc rétrospective pour renseigner l'indication ou le bon usage d'un produit au moment de l'inclusion du patient et 2) Le suivi de patients sera renseigné en utilisant une cohorte historique à partir du SNDS : celle-ci sera exhaustive et ne comportera pas de perdus de vue. On peut aussi aller plus loin en chaînant, au moyen du NIR (numéro d'inscription au répertoire des personnes physiques ou numéro de sécurité sociale), le SNDS avec des données collectées sur des patients inclus dans une étude ad hoc. Les dossiers déposés à l'INDS devraient d'ailleurs rapidement inclure de telles demandes de chaînage. Cette disposition pourrait aussi s'appliquer aux essais cliniques. La possibilité d'utiliser le SNDS dans ce cadre constituerait une chance pour l'attractivité de la France pour la recherche clinique.



L'apport des e-cohortes



Guy Fagherazzi
(Inserm – cohorte E4N)

Notre équipe de recherche gère actuellement deux grandes cohortes, E3N et E4N. La première est une cohorte de 100 000 femmes suivies depuis 27 ans, à partir de laquelle on peut réaliser de nombreuses études en pharmaco-épidémiologie. La deuxième, en cours de construction, est le prolongement de cette première cohorte. Concrètement, l'objectif est de suivre les familles des 100 000 femmes de la cohorte E3N sur trois générations. Outre ces 100 000 femmes recrutées depuis 1990, nous avons recruté en 2014 20 000 conjoints, pères biologiques des enfants des femmes E3N. La CNIL a autorisé cette année le recrutement de la deuxième génération pour laquelle nous espérons recruter 50 000 personnes d'ici un an. Le recrutement de la troisième génération débutera en 2018. Le suivi de la première génération est très simple avec l'envoi classique de questionnaires. Les 2ème et 3ème générations seront suivies uniquement en ligne avec une e-cohorte, ce qui ouvre le potentiel pour des études avec des objets et des dispositifs médicaux connectés. Cette étude étant une étude familiale, notre équipe de recherche dispose de biobanques avec de la salive pour faire des extractions d'ADN. De plus, notre plateforme de collecte de données a été conçue comme un data hub, c'est-à-dire que les questionnaires et les formulaires en ligne seront envoyés par SMS au membres de la cohorte qui pourront y répondre sur leur portable et leurs réponses directement intégrées dans les bases de données.

Les grandes cohortes comme E4N permettent de combiner des sources d'informations très variées et de coupler ainsi le SNDS avec des données biologiques, des biomarqueurs et des informations issues d'applications et d'objets connectés. En termes de recherche, notre laboratoire va développer des travaux sur l'hérédité, les aspects transgénérationnels de la santé, le mode de vie, les comportements, les aspects psycho-sociaux, l'environnement socio-économique et leurs liens avec la santé. Notre autre axe majeur est celui de l'intégration des nouvelles technologies pour la collecte des données et pour l'évaluation en recherche médicale. L'objectif est de développer un suivi en ligne grâce aux e-cohortes et de dépasser un cadre purement observationnel pour passer à un mode de plus en plus interventionnel. La cohorte E4N est l'équivalent d'un « living lab » : les participants sont interrogés et pourront entrer dans des études randomisées destinées à évaluer l'impact et l'efficacité d'un objet connecté ou d'un dispositif médical. Des projets sont en cours de développement avec l'entreprise Visiomed dans le cadre d'un partenariat public-privé.

