Valorisation des pratiques

Réutilisation des données médico-administratives dans les armées

G. DESJEUX[(1)], J.-M. LOREAU[(2)], F. DELON[(3)], P. GUITTARD[(4)]

(1,2,3) Centre d'Épidémiologie et de Santé Publique des Armées, Marseille
(4) Direction de la Médecine des Forces, Tours

Auteur correspondant : gdesjeux@bdmxp.fr

Open access DOI
10.17184/eac.9358

Reçu
le 9 décembre 2024

Accepté
le 1er avril 2025

Publié
le 11 juin 2025

Conflit d'intérêt
Les auteurs ne déclarent pas de conflit d'intérêt.

Résumé :
Les bases de données médico-administratives sont des apports précieux pour les études en santé. Leur développement est ancien, mais la mise à disposition au début des années 2000 des bases de remboursements de l'assurance maladie a permis aux chercheurs d'avoir accès à des données standardisées et exhaustives, couvrant le territoire métropolitain et les départements d'outre-mer. Dans les armées, ces bases sont également disponibles.
En outre, il existe des bases médico-administratives spécifiques permettant d'étudier plus spécialement la population des militaires. Le passage d'une donnée médico-administrative à une donnée de santé nécessite une appropriation des règles de constitution de la base et de ses aspects réglementaires.
L'identifiant unique et pseudo-anonymisé permet un chaînage entre différentes bases, constituées pour une finalité autre, multipliant ainsi les axes de recherche. Les progrès dans les outils d'interrogation de base de données de grand volume optimisent les calculs et la vitesse de réponse pouvant faire de ses bases des aides précieuses pour la surveillance épidémiologique.

Mots-clés : Base de données médico-administratives. Médecine d'armée.

Abstract:
USING PRE-EXISITING GENERAL MEDICO-ADMINISTRATIVE DATA FOR ARMED FORCES PURPOSES
Healthcare databases are valuable assets for health studies. Data collection in healthcare started a long time ago, and the availability of health insurance reimbursement databases in the early 2000s has given researchers access to standardized

and comprehensive information covering mainland France and French overseas departments.
In the army, these databases are also available, some of them specifically processing medico-administrative matters, particularly useful for studying the military population. Converting medical-administrative data into health data requires an understanding of the rules governing the database structures and the legal framework.
Using a unique, pseudonymized identifier makes it possible to link different databases that were originally created for other purposes, thereby multiplying research possibilities. Advances in high-volume database query tools facilitate calculations and response times, making these databases valuable tools for epidemiological surveillance.

Keywords: Healthcare database. Military medicine.

Introduction

L'analyse des données « observationnelles » en santé, c'est-à-dire les données issues de la pratique courante et non d'études cliniques (1), est un véritable défi pour les chercheurs et les épidémiologistes, mais également pour les dirigeants d'établissements de santé ainsi que pour les personnes en charge des politiques de santé. Elle se heurte au coût de l'élaboration des outils de recueil, bien que celui-ci soit de moins en moins élevé, et à la complexité de leur exploitation. Le manque de standardisation et de normalisation des données collectées limite leur réutilisation. Les bases de données médico-administratives doivent avoir une interopérabilité sémantique et syntaxique qui nécessitent une coordination des différents éditeurs de solutions informatiques. Enfin, le plus souvent, le recueil est limité à une seule analyse ou du moins une seule finalité, et la réutilisation des données peut être complexe, voire biaisée du fait de l'ajustement de ce recueil à un objectif prédéfini.

Pourtant depuis plus de 20 ans, des études en santé portant sur la surveillance épidémiologique (2,3), le suivi de cohorte (4), la pharmaco-épidémiologie (5), les politiques de santé (6) etc. réutilisent des données de santé issues de Bases de données médico-administratives (BDMA) pour mener à bien leurs travaux. Ces travaux ont démontré qu'en utilisant des démarches et méthodes adaptées, les données stockées dans ces BDMA pouvaient faire l'objet d'une utilisation secondaire, définie par l'Agence du numérique en santé comme le « traitement ultérieur de données de santé électroniques collectées initialement à d'autres fins pour : des statistiques publiques dans le secteur de la santé ; des activités d'intérêt public dans le domaine de la santé, telles que la protection contre les menaces transfrontalières ou la surveillance de la santé publique ; la recherche scientifique ayant trait aux secteurs de la santé ; des activités de développement et d'innovation de produits ou services contribuant à la santé publique ou

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

à la sécurité des soins de santé, des médicaments ou des dispositifs médicaux ou des activités d'éducation dans le secteur de la santé » (7).

La définition des BDMA n'est pas consensuelle (8) : en France, un rapport du Haut Conseil de la santé publique publié en 2012 (9) les décrit comme des bases « centralisées, constituées et gérées par des organismes publics, couvrant de façon exhaustive et permanente l'ensemble de la population dans divers domaines stratégiques pour la santé publique et la recherche : recours aux soins, hospitalisation, handicaps, prestations et situation professionnelle, sociale et économique ». Pour cet article, nous retiendrons une définition plus étroite des BDMA : base constituée pour répondre à un besoin réglementaire ou légal avec une finalité principale généralement financière, permettant de faire valoir des droits liés à l'état de santé ou des prestations en nature ou en espèce.

Après une description synthétique des BDMA françaises, nous présenterons les BDMA du Ministère des armées et le concept de passage d'une donnée administrative vers une donnée de santé.

1 Les bases de données médico-administratives françaises

Les principales BDMA en France sont centralisées dans la plateforme des données de santé, appelée également *Health Data Hub* (HDH) dont l'hébergement de la base dite « principale », le Système national des données de santé (SNDS), dépend de l'assurance maladie. Le HDH est relativement récent (2019), même s'il est issu d'une histoire plus ancienne avec l'Institut des données de santé créé en 2007 puis l'Institut national des données de santé en 2017 (10). Le HDH est en évolution permanente avec l'insertion de sources de données qui sont généralement des registres ou des entrepôts de données de santé.

Certaines de ces bases sont restituées sous forme agrégée, non réidentifiable, en libre accès sur Internet : data pathologies, open medic, scansanté, Cepidc, data.gouv, etc.

Il existe également des travaux pour une mise en commun des informations au niveau européen (11).

Le SNDS est une base nationale comprenant des données relatives à 68,2 millions de personnes en 2022 (5). Elle contient des données sur l'ensemble des prestations remboursées de l'assurance maladie obligatoire issues du système national inter-régimes de l'assurance maladie, pour le régime général (87 % des assurés de la base), mais également pour les autres régimes, dont la Caisse nationale militaire de sécurité sociale (CNMSS). La base comportant ces prestations s'appelle le Datamart de consommation inter-régimes (DCIR). Elle est appariée au référentiel des bénéficiaires, qui comporte l'ensemble des assurés. Pour des raisons historiques, cet appariement est réalisé au moyen d'une combinaison du Numéro d'inscription au registre (NIR) de l'ouvreur de droit, de la date de naissance et du sexe du bénéficiaire. Le SNDS inclut une autre base, gérée par l'Agence technique d'information sur l'hospitalisation (ATIH), comportant les données du PMSI, qui est le recueil standardisé d'informations médicales permettant la tarification et la dotation globale de financement des établissements de santé dans les champs du court séjour, des soins médicaux et de réadaptation, de l'hospitalisation à domicile et de la psychiatrie. Le recueil des informations médicales

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

concernant les décès en France effectué par le Centre d'épidémiologie sur les causes médicales de Décès (CépiDc) est également présent dans le SNDS.
Pendant la crise Covid, des informations issues de la collecte spécifique pour cette maladie ont également été incorporées au SNDS. Enfin, le Code de la santé publique (article L. 1461-1) prévoit un élargissement du SNDS à d'autres bases telles que les données recueillies lors des visites médicales et de dépistages obligatoires, des visites en service de protection maternelle et infantile ou en santé au travail. Le comité stratégique des données de santé créé en 2021 propose des orientations et recommandations sur l'enrichissement des données du SNDS (12).

Un échantillon du SNDS (ESND), incluant 2 % de la population de ce dernier, permet une interrogation plus facile de la base en raison d'un volume moindre. D'autres bases de données médicalisées ont vocation à venir étoffer par la suite le SNDS : à court terme, une base issue des informations collectées par les maisons départementales des personnes handicapées, à plus long terme un échantillon des données issues des assurances maladie complémentaires et d'autres données prévues par le Code de la santé publique.

La couverture nationale des informations et la nature systématique et standardisée en termes de recueil font du SNDS un outil précieux pour les études en santé avec des finalités très variées : suivi de l'objectif national des dépenses de l'assurance maladie, études sur les effets des médicaments en vie réelle, c'est-à-dire sans intervention de l'investigateur, études sur les parcours de soins, études médico-économiques, études sur la qualité des soins. Le socle historique de cette base ne contient pas de résultats médicaux. Ainsi, la réalisation d'examens complémentaires comme les actes d'imagerie et de biologie sont tracés, mais leurs résultats ne sont pas disponibles. Son interrogation est complexe. La base étant pseudo-anonymisée, il n'est pas possible de revenir au dossier du patient. Les informations recueillies sont liées à des règlements et des lois qui évoluent plus ou moins rapidement et qu'il est nécessaire de connaître pour interpréter correctement les données. Enfin son volume, plusieurs milliards de lignes de remboursement, et son historique, avec des données depuis 2006 et exploitables depuis 2009, nécessitent des méthodes spécifiques et des capacités informatiques puissantes pour les calculs.

Les données du SNDS ne sont pas adaptées à une analyse et à un suivi en temps rapproché. Les données sont relativement stables au bout de 6 mois par rapport à la date d'exécution des soins pour la base DCIR. Pour les hospitalisations, le dépôt est annuel avec une disponibilité l'été suivant. Pour les décès, l'obtention des données est annuelle, avec une latence de deux ans.

L'accès à cette base nécessite une formation payante et des formalités administratives qui comprennent entre autres un respect des finalités qui sont définies par la loi (L. 1461-1 III du Code de la santé publique) et précisées par décret (R. 1461-1 du Code de la santé publique) (13), un passage devant un Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) qui a un avis consultatif et l'avis favorable de la Commission nationale informatique des libertés. Deux nouvelles méthodologies de référence (MR-007 et MR-008), adoptées par la CNIL en octobre 2023, simplifient l'accès à la base principale du SNDS (14).

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Le SNDS contenant des identifiants uniques pseudo-anonymisés, il est possible de chaîner les informations avec d'autres bases médicalisées selon une procédure d'appariement déterministe si la base importée contient le NIR, ou probabiliste dans le cas contraire, à l'aide de quelques variables simples : dates d'exécution des soins et numéro d'assurance maladie du praticien, par exemple.

D'autres bases médico-administratives existent en dehors du SNDS avec une interrogation possible pour des personnes extérieures aux services administrant ces bases. Il est cependant nécessaire de déposer des dossiers administratifs auprès de ces organismes pour en obtenir l'accès.

L'Échantillon démographique permanent (EDP) est une base détenue par l'Institut national de la statistique et des études économiques (INSEE) (15). Il a été créé en 1967. Il contient des données d'état civil, de fiscalité, des données du fichier électoral et un fichier individuel des déclarations administratives des employeurs. Il est apparié au SNDS.

L'INSEE établit également le fichier des personnes décédées depuis 1970 à partir des informations reçues des communes (16). Ces fichiers sont publiés mensuellement en open data en application des dispositions des articles L311-9 et L312-1-1 du code des relations entre le public et l'administration, et réutilisables dans les conditions prévues par le titre II de ce livre. Ce fichier contient les noms, prénoms, sexes, dates de naissance, codes et libellés du lieu de naissance, dates du décès, codes du lieu de décès et numéros de l'acte de décès et inclut les décès survenus à l'étranger.
Les informations sont croisées et mises en conformité avec les données du CépiDc.

Le Système national de gestion des carrières (SNGC) est une base permettant l'analyse sociale. Couplée au SNDS, elle apporte des éléments fondamentaux pour l'étude des liens entre santé et travail, ou sur les inégalités sociales. Cet appariement est réalisé par exemple dans la cohorte Constances (17). Le SNGC est géré par la caisse nationale d'assurance vieillesse. Il englobe les périodes de la vie active jusqu'au décès et parfois au-delà en cas de pension de réversion. Il est appelé à s'étendre à tous les régimes en devenant le Répertoire de gestions des carrières uniques (RGCU) (18).

Le système de surveillance syndromique SURSAUD n'est pas *stricto sensu* une BDMA selon notre définition. Il centralise des données hospitalières, d'associations SOS médecins, du CépiDc et de l'INSEE. Le caractère obligatoire ou conventionnel des bases de SURSAUD et l'intégration de la base « Organisation de la surveillance coordonnée des urgences » (OSCOUR), un des éléments centraux du système SURSAUD, dans le catalogue du SNDS en font un ensemble de bases similaires aux BDMA, en raison de leur standardisation, de leurs liens aux soins sans exclusion *a priori* et sans formulaire spécifique.

La base SI-DEP recueillait jusqu'en juillet 2023 les résultats des tests antigéniques et RT-PCR pour la Covid. Cette base a été alimentée dans le SNDS entre mai 2020 et fin juin 2023. En relais de SI-DEP, le système d'information LABOé-SI a été créé. Il est mis en œuvre sous la responsabilité de la Direction générale de santé (DGS).
Ce système permet une collecte exhaustive et pérenne des signalements et des transmissions de résultats émis par les laboratoires de biologie (19). Les résultats d'examens

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

biologiques positifs de dépistage pour l'ensemble des maladies à signalement obligatoire ainsi que les dépistages négatifs de la Covid sont téléversés dans ce système.

La base du Système d'information des victimes (SI-VIC) est une base administrative permettant le dénombrement, l'identification et le suivi préhospitalier, hospitalier et médico-psychologique des victimes dans le cas de situations sanitaires exceptionnelles. Il facilite également l'accompagnement étatique des victimes et de leur famille, la gestion de l'évènement par les autorités sanitaires et l'analyse des parcours de soins.
Une première version (V0) a été déployée en 2016, à l'occasion du Championnat d'Europe de football, mais présentait des limites importantes. Une nouvelle version (V1) a été conçue et déployée en 2017 à partir d'un outil, Terminal ORSAN, qui avait été utilisé en PACA pour le suivi des victimes de l'attentat de la promenade des Anglais. L'Agence du numérique en santé assure actuellement le pilotage du projet. La version V4 est attendue pour le troisième trimestre 2025. Il est disponible pour la DGS, les Agences régionales de santé, les établissements, les cellules d'urgence médico-psychologique et les SAMU (20,21).

2 Les bases de données médico-administratives dans les armées

Les armées possèdent également des BDMA. Les armées peuvent interroger *via* un accès permanent la base principale du SNDS. Le Centre d'épidémiologie et de santé publique des armées coordonne les études demandées et assure, avec le Représentant du responsable de traitement du SSA (RRT), correspondant du délégué à la protection des données du ministère des armées, le suivi réglementaire et légal des interrogations de ce système. Les données accessibles sont limitées à la population des affiliés de la Caisse nationale militaire de sécurité sociale (CNMSS), contenant les militaires en situation d'activité, leurs ayants-droits mais également les retraités bénéficiaires d'une pension militaire de retraite sans autre activité rémunérée et d'autres situations particulières. Si légalement la profondeur historique prévue est de 20 ans, elle débute de manière exhaustive en 2012, date du dépôt des données du PMSI par les hôpitaux d'instruction des armées.

La CNMSS est également détentrice d'un accès permanent au SNDS. Son accès inclut l'ensemble des affiliés de tous les régimes, lui permettant de faire des comparaisons adéquates. Son accès permet également l'identification du professionnel de santé, dont le médecin des armées *via* le numéro du Répertoire partagé des professionnels de santé. La déclaration d'étude est soumise à l'approbation du directeur de l'établissement et la validation réglementaire des informations demandées par le délégué à la protection des données de l'établissement.

La CNMSS a, en parallèle, un accès aux données natives des remboursements de ses affiliés : l'identification et l'adresse des personnes sont accessibles, sous couvert d'une autorisation respectant le cadre légal, permettant de réaliser des enquêtes nominatives (22) ou avec des résultats agrégés (23). Elle a la possibilité d'interroger également les données médicales détenues par le service médical de la CNMSS qui émet des avis médicaux pour certaines prestations (exonération du ticket modérateur, accord pour les transports, accord pour la délivrance de certains actes etc.). La profondeur historique des données de remboursement est de trois ans, celle sur les avis médicaux est d'au moins 20 ans. La CNMSS, par convention, gère les prestations de soins effectuées

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

en milieu civil par les militaires victimes d'accidents ou d'Affections présumées imputables au service (APIAS). Il est donc possible sous couvert d'un accord réglementaire d'utiliser ces données pour de nouvelles finalités (24).
De même, par délégation, la CNMSS gère le règlement des dossiers de soins médicaux et d'appareillage des titulaires d'une pension militaire d'invalidité. Des études en santé sur ce périmètre sont également possibles. Depuis 2016 pour les pensions militaires d'invalidité et depuis 2018 pour les APIAS, les prestations de ces missions déléguées sont également transmises de manière pseudo-anonymisée au SNDS.
Les données concernant l'accident ou l'affection ouvrant droit à des prestations ne sont pas disponibles dans cette base. Seules les prestations remboursées sont transmises dans le SNDS, elles sont identifiables par la modalité d'exonération affectée à cette dépense. Elles sont donc accessibles à tous sous réserve d'une demande conforme à la réglementation. Les modalités d'accessibilité sont disponibles sur le portail du HDH.
Les données des prestations prises en charge au titre du service de santé des armées ne sont pas disponibles ou elles ne permettent pas l'identification unique anonyme de l'individu dans le SNDS lorsqu'elles sont délivrées au sein des établissements appartenant au Service de santé des armées.

Le Système d'information des congés liés à l'état de santé (SI-CLES) est administré par le chef du bureau des instances médicales d'expertises des armées. Il contient l'ensemble des placements en Congés longue maladie (CLM) ou en Congé longue durée pour maladie (CLDM) de l'ensemble des militaires, gendarmes inclus, de manière numérisée et interrogeable informatiquement depuis 2017. Les affections sont codées selon la Classification internationale des maladies 10^{e} révision (CIM-10).
Les extractions sont disponibles pour la gestion du système sous forme de tableurs et sous forme de base de données disponibles *via* la Direction des systèmes d'information et du numérique en santé (DSIN-S) du service de santé des armées.
Le dossier patient informatisé Axone a pour finalités (25), outre la gestion du dossier médical des bénéficiaires des soins dispensés par le Service de santé des armées, la gestion de la médecine de prévention et de l'expertise médicale d'aptitude des personnels militaires, des candidats à l'engagement militaire et des civils.
Ce SI est administré par le bureau de l'information médicale et du numérique de la direction de la médecine des forces et compte plus d'un million de dossiers actifs. Il contient d'un point de vue médico-administratif l'ensemble des données d'aptitude (SIGYCOP, inaptitudes, éditions réglementaires), d'indisponibilités pour raison de santé (arrêt de travail, APIAS, CLM, CLDM) et d'expositions (fiche emploi nuisance, poste de travail et risques afférents, surveillance médicale). Les diagnostics et motifs médicaux sont codés selon la CIM-10. Des requêtes sont disponibles pour extraire différents indicateurs sous forme de base de données (par exemple, la liste des patients avec telle valeur pour tel sigle de SIGYCOP, la liste des CLM, CLDM et indisponibilités pour raison de santé). Des extractions plus complexes nécessitent une expression de besoins auprès de l'industriel.
Il existe également le SI PREVENTIEL qui permet le suivi en santé au travail des personnels civils du ministère des armées et contient également des données sur le poste de travail et les risques afférents, le type de surveillance médicale, les actions en milieu de travail, les maladies professionnelles, les arrêts et accidents de travail, les avis d'aptitudes. Ce SI est requêtable à partir d'un outil de *reporting* de type *Business Objects* (BO).

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Le service des pensions et des risques professionnels du ministère des Armées dispose d'une base médico-administrative contenant les informations nécessaires à l'instruction et au suivi des pensions militaires d'invalidité, cependant la finalité de ce traitement ne permet pas la réutilisation des données dans le cadre d'étude en santé.

3 De la base de données médico-administratives à l'étude en santé

Les diagnostics médicaux dans les BDMA présentées se limitent à des codes ou des regroupements de codes sans interprétation clinique ni règle formalisée de saisie, en dehors des résumés standardisés du PMSI et des certificats de décès.

Ces limitations n'excluent pas de déduire de ces combinaisons de codes des informations permettant de supposer des pathologies ou de modéliser des prises en charge.
Ces combinaisons avec leurs règles d'utilisation sont appelées des algorithmes.
La cartographie de la Caisse nationale d'assurance maladie (CNAM) permet d'avoir un outil d'analyse d'une cinquantaine de pathologies ou de prises en charge.
Pour exemple, dans cette cartographie, le diabète est défini par différentes combinaisons associant la consommation d'antidiabétiques oraux ou injectables spécifiques, les codes fournis par le PMSI ayant trait au diabète, la présence d'une exonération du ticket modérateur pour l'appartenance à une des pathologies de la liste etc. Si l'on peut distinguer la prise en charge d'un diabète par insulinothérapie ou par antidiabétiques oraux, on ne peut raisonnablement déduire de ces informations le type de diabète pris en charge. Pour donner un autre exemple d'interprétation de données médico-administratives, on peut supposer que le remboursement d'un test de dépistage du paludisme suivi d'une délivrance d'antiparasitaires spécifiques peut être le témoin de la présence d'un paludisme (26).

De nombreux algorithmes concernant ces pathologies ou prises en charge sont publiés. Le groupe informel « Réseau pour mieux utiliser le SNDS » (REDSIAM) recense beaucoup d'entre eux. Le site Internet du HDH en fournit également au sein de la bibliothèque ouverte d'algorithmes en santé (27), car l'utilisation des données du SNDS implique la transparence des études et donc la diffusion des résultats *via* une publication, bien que la nature de cette publication ne soit pas précisément définie.
L'appropriation de ces bases de données, ainsi que le nombre croissant de spécialistes de ces bases (de quelques centaines en 2006 à plusieurs milliers actuellement) permettent d'affiner la détection de différentes pathologies ou de décrire des parcours de soins. La grande difficulté est de valider l'algorithme avec un *gold* standard.
Les données étant pseudo-anonymisées, le retour au dossier clinique n'est pas autorisé et relativement difficile dans la plupart des cas. La question de la reproductibilité des études sur les BDMA se pose (28). Néanmoins, la simplification du recueil et de l'utilisation du NIR dans le cadre d'études en santé, permet le couplage individuel de ces bases à des études *ad hoc*, comblant ainsi en partie ce déficit et permettant la validation des algorithmes. Le HDH incite financièrement les équipes de recherche à valider leurs résultats avec le croisement de données cliniques et en contrepartie de diffuser les algorithmes leur permettant d'obtenir ces résultats.

Par ailleurs, certains facteurs de risque (tabagisme, poids, taille, exposition environnementale) et autres déterminants de santé ne sont pas détenus de manière simple dans ces bases, limitant les apports de celles-ci. Par exemple, les expériences des soins

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

perçus « Patient-Reported Experience Measures » (PREMs) et le résultat perçu par le patient « Patient-Reported Outcome Measures » (PROMs), outils de plus en plus utilisés dans les programmes de santé, ne sont pas disponibles.

Un exemple d'analyse des données du SNDS dans les armées est la description à coût réduit des pathologies non soumis à la surveillance épidémiologique dans les armées, à différents temps (par exemple, avant et après une intervention) sur la population entière des militaires d'active, ainsi que leur consommation et parcours de soins détaillés en milieu civil et hôpitaux des armées. Par ailleurs, les données des BDMA permettent de compléter les données de la Surveillance épidémiologique dans les armées (SEA). Par exemple, un travail (29) sur l'incidence et les causes médicales des décès parmi les militaires d'active dans le SNDS avait permis de montrer que le SNDS permet une meilleure identification des décès par pathologies chroniques car ces individus sont consommateurs de soins ; à l'inverse, des décès par suicide ou AVP, sont mieux détectés par la SEA.

Si l'utilisation d'autres BDMA que le SNDS est possible, il n'y a pas d'initiatives similaires à celles de la CNAM et du REDSIAM, complexifiant leur réutilisation de manière standardisée et reproductible. Par contre, dans le SSA, l'avantage est d'avoir un organisme centralisateur de tous les SI : la Direction des Systèmes d'Information et du Numérique en santé (DSIN-S), permettant d'avoir une vue globale des systèmes existants et d'avoir accès au catalogue et aux référentiels des données.

4 Perspectives dans le domaine de la réutilisation de données de santé, directes ou indirectes

4.1 Interopérabilité et mise en qualité

Pour faciliter l'exploitation des données et l'interopérabilité entre les différentes bases (qu'elles soient cliniques ou médico-administratives), une harmonisation des approches de mise en qualité et de standardisation est nécessaire. Des modèles standards de données comme le modèle OMOP-CDM (*Observational Medical Outcomes Partnership – Common Data Model*), soutenu par l'*Observational Health Data Sciences and Informatics* (OHDSI) permettent de normaliser la structure, le contenu et la sémantique des données et ainsi de rendre possible l'écriture de codes d'analyse statistique uniques réutilisables sur chaque base de données (30). Un échantillon du SNDS a été converti au format OMOP-CDM (31). Cette conversion consiste en une restructuration de la base initiale en tables et variables standard OMOP-CDM et une traduction des nomenclatures du SNDS vers les terminologies standards OMOP-CDM, par exemple la terminologie internationale SNOMED-CT (Systematized Nomenclature of Medicine Clinical Terms) pour le codage des diagnostics et des procédures et la terminologie RxNorm pour les médicaments. Ce travail a permis de conserver les informations les plus pertinentes dans le cadre d'études observationnelles longitudinales à grande échelle au travers d'une vingtaine de tables seulement centrées sur le patient et ses interactions avec le système de santé.

Le constat actuel d'un manque d'interopérabilité constitue un frein technique à la création d'entrepôt de données de santé (32). Un groupe de travail mandaté en janvier 2023 par le comité stratégique des données de santé a proposé un socle commun de 51 *items*

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

permettant la collecte, la mise en qualité et la standardisation sur le modèle OMOP-CDM d'un périmètre de données commun (33). Parmi ces *items* figurent les identifiants nationaux tels que le numéro d'identification au répertoire des personnes physiques (NIR) ou l'identité nationale de santé (INS) en tant que clés d'appariement direct entre les différentes bases.

Au niveau des armées, le projet fédérateur d'identité du SSA, brique du programme d'armement e-SAN permettra de fédérer les identités d'un même patient au sein des différents systèmes d'information métier pour relier ses différents épisodes de soins et collecter les données d'intérêt au sein d'une plateforme de partage.

4.2 Enrichissement des données médico-administratives par d'autres sources de données

Au-delà de l'enrichissement de la base principale du SNDS sur proposition du Comité stratégique des données de santé, les initiatives de chaînage, sur projet ou pérenne, entre des registres, des cohortes ou des données d'entrepôt de données de santé et la base principale du SNDS sont de plus en plus nombreuses. On peut citer, à titre d'exemple, la plateforme AGORIA santé dont l'un des projets est le chaînage des données collectées dans le cadre de l'accès précoce au tixagévimab/cilgavimab en prophylaxie préexposition de la Covid-19 aux données du SNDS, l'entrepôt de données de santé Cardiohub alimenté par différents registres de la société française de cardiologie et apparié au SNDS ou encore la plateforme de données en cancérologie gérée par l'Institut national du cancer (INCa). Au 18 octobre 2024, 100 entrepôts de données de santé sont répertoriés en France dont dix comportent des données du SNDS. Une cartographie de ces entrepôts de données de santé est publiée par la CNIL et disponible en *open source* sur Internet (34).

Au niveau des armées, le projet de système d'information CORTEX a vocation à intégrer et unifier les données de santé individuelles des militaires au sein d'un entrepôt de données pour une utilisation secondaire à des fins de recherche ou de pilotage opérationnel ou stratégique. Par exemple, dans les armées, la problématique de la détection précoce pour la surveillance épidémiologique *via* les systèmes d'information est un sujet de longue date (35,36). La réutilisation des données d'Axone permettrait une telle surveillance grâce à un regroupement des codes CIM-10 des conclusions diagnostiques des consultations. À terme, Cortex devrait être alimenté par les données du SNDS des militaires d'active. Porté par le CESPA, le projet de cohorte sur les militaires et leurs expositions (CoMitEX) prévoit d'inclure et suivre 20 000 militaires quant à leur état de santé et leurs expositions professionnelles au travers d'un questionnaire de suivi annuel et à terme d'un appariement aux données du SNDS afin de mieux analyser leur parcours de soins.

4.3 Intelligence artificielle et réutilisation secondaire des données

Ces données de santé de plus en plus massives permettent l'entraînement de modèles d'intelligence artificielle pour prédire par exemple la survenue d'un évènement de santé, d'un évènement indésirable ou d'une rupture dans le parcours de soins du patient. Par exemple, le projet Hydro (37), porté par Implicity, a utilisé les données du SNDS et celles issues des capteurs cardiaques connectés pour développer un algorithme de prédiction des hospitalisations pour insuffisance cardiaque.

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Dans plusieurs travaux comme celui de Guyomard *et al.* (38), des modèles de *deep learning* appliqués au langage naturel de type *transformers* ont permis de modéliser des séquences de soins afin de prédire la survenue d'une pathologie, comme l'on prédit la suite des mots dans un grand modèle de langage.

Les algorithmes de traitement du langage naturel sont aussi utilisés sur les données non structurées qui représentent à peu près 80 % des données hospitalières (comptes-rendus médicaux, résultats d'examens biologiques ou radiologiques, etc.) pour extraire les données de façon automatisée sous un format structuré. Par exemple, le projet TARPON (39) porté par le CHU de Bordeaux, le centre de recherche sur la santé des populations de Bordeaux et l'institut national de la santé et de la recherche médicale (Inserm) a développé des algorithmes de traitement automatique du langage naturel sur les dossiers médicaux électroniques des urgences afin d'extraire, entre autres, les circonstances et mécanismes des traumatismes afin d'enrichir les Résumés de passage aux urgences (RPU) exploités dans le système OSCOUR (40). L'objectif de ce projet est de permettre à terme l'établissement d'un observatoire national des traumatismes apparié aux données du SNDS.

Un autre exemple de cas d'usage est le développement de grand modèle de langage (LLM) pour automatiser le processus de codage dans des standards internationaux (CIM-10, RxNorm…) à partir de comptes-rendus médicaux. PLM-ICD et ChatICD en sont des exemples basés respectivement sur les architectures RoBERTa et ChatGPT avec une prédiction du code CIM-10 supérieur à 0,9 en termes d'AUC (41). Néanmoins, il est important de souligner que les LLM peuvent souffrir d'hallucinations et assigner un code non existant, rappelant qu'il est crucial de conserver un contrôle humain pour détecter et rectifier les erreurs.

Les algorithmes d'apprentissage automatique peuvent aussi être utilisés pour générer des données artificielles à partir des données réelles. Ces données artificielles peuvent être utilisées pour l'entraînement de modèles en limitant les contraintes réglementaires mais également en recherche clinique pour enrichir les essais cliniques et ainsi compenser les difficultés de recrutement des patients réels dans les bras de contrôle (concept de cohorte augmentée) (42).

L'interopérabilité, l'intégration de données provenant de multiples sources et leurs échanges devraient favoriser une collaboration accrue entre les différents acteurs.
Cela enrichit également les possibilités d'analyses multicentriques et multidisciplinaires. La variété des sources et des usages reflète une dynamique d'innovation et un effort collectif visant à maximiser l'exploitation des données de santé en vie réelle, dans le but d'améliorer la compréhension et la prise en charge des pathologies.
Cette dynamique s'inscrit dans un mouvement plus large, visant à capitaliser sur la richesse et la complémentarité des données disponibles pour faire avancer la recherche et l'innovation en santé.

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

Remerciements

Médecin en chef Claire BYLICKI et Médecin en chef Sieu-You UNG du Bureau de l'information médicale et du numérique (DMF / Division organisation / BIMN).

Remerciements au linguiste relecteur : Pierre-Alexandre Clerc.

Financements

Aucun

Contributions des auteurs

Rédaction : GD, JML, FD
Relecture : PG

Approbation éthique et consentement à la participation

Non applicable

Conflits d'intérêt

Guillaume Desjeux est président d'une société réalisant des études sur les bases de données en santé.

Références bibliographiques

1. Haute Autorité de la Santé. Données en santé et social : savoir les exploiter pour mieux suivre les pratiques [En ligne]. Saint-Denis (FR) : Haute Autorité de la Santé [cité le 26 mars 2025]. Disponible : https://www.has-sante.fr/jcms/p_3272150/fr/donnees-en-sante-et-social-savoir-les-exploiter-pour-mieux-suivre-les-pratiques

2. Bounoure F, Beaudeau P, Mouly D, Skiba M, Lahiani-Skiba M. Syndromic surveillance of acute gastroenteritis based on drug consumption. Epidemiology and Infection. 2011 ; 139(9) : 1388-95. doi : 10.1017/S095026881000261X

3. Lauzun V de, Pivette M, Nicolay N, Scanff A, Hubert B. Caractéristiques des hospitalisations avec diagnostic de grippe, France, 2012-2017. Saint-Maurice (FR) : Santé publique France ; 2019. Disponible : https://www.santepubliquefrance.fr/docs/caracteristiques-des-hospitalisations-avec-diagnostic-de-grippe-france-2012-2017

4. Schwoebel V. Utilisation des systèmes d'information sanitaire pour le bilan et l'aide à la décision après l'explosion de l'usine « AZF » à Toulouse le 21 septembre 2001. Bulletin épidémiologique hebdomadaire. 2004 ; (38-39) : 186. Disponible : https://www.santepubliquefrance.fr/regions/occitanie/documents/article/2004/utilisation-des-systemes-d-information-sanitaire-pour-le-bilan-et-l-aide-a-la-decision-apres-l-explosion-de-l-usine-azf-a-toulouse-le-21-septembr

5. Weill A, Païta M, Tuppin P, Ricordeau P, Allemand H. Benfluorex et valvulopathies cardiaques : rapport final du 09/11/2009 (correction du 10/11/2009). Paris (FR) : Caisse nationale de l'assurance maladie ; 2009. Disponible : https://www.assurance-maladie.ameli.fr/etudes-et-donnees/2009-2010-benfluorex-risques-valvulopathie-cardiaque-deces#download

6. Caisse nationale de l'assurance maladie. Méthode de la cartographie des pathologies et des dépenses de l'Assurance Maladie [En ligne]. Paris (FR) : Caisse nationale de l'assurance maladie ; [cité le 26 mars 2025]. Disponible : https://www.assurance-maladie.ameli.fr/etudes-et-donnees/par-theme/pathologies/cartographie-assurance-maladie/methode-cartographie-pathologies-depenses-assurance-maladie

7. Agence du numérique en santé. Utilisation secondaire des données de santé électroniques [En ligne]. Agence du numérique en santé [cité le 26 mars 2025]. Disponible : https://participez.esante.gouv.fr/project/chapitre-4-utilisation-secondaire-des-donnees-de-sante-electroniques/presentation/presentation

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

8. Bannay A. Bases médico-administratives : de l'épidémiologie à la pharmacovigilance. Nancy (FR) : Université de Lorraine ; 2021. Disponible : https://theses.fr/2021LORR0183
9. Goldberg M. Pour une meilleure utilisation des bases de données nationales pour la santé publique et la recherche. Paris (FR) : Haut Conseil de la santé publique ; 2012. Disponible : https://www.hcsp.fr/explore.cgi/avisrapportsdomaine?clefr=258
10. Système national des données de santé. [cité le 26 mars 2025]. Dans : snds.gouv.fr [En ligne]. Système national des données de santé [environ 2 écrans]. Disponible : https://www.snds.gouv.fr/SNDS/Actualites/Actu-1
11. Agence du numérique en santé. Espace européen des données de santé [En ligne]. Agence du numérique en santé [cité le 26 mars 2025]. Disponible : https://esante.gouv.fr/espace-europeen-donnees-sante
12. Arrêté du 29 juin 2021 portant création du comité stratégique des données de santé. Journal officiel de la République française. 2021.
13. Système national des données de santé. [cité le 26 mars 2025]. Dans : snds.gouv.fr [En ligne]. Système national des données de santé [environ 2 écrans]. Disponible : https://www.snds.gouv.fr/SNDS/Finalites-autorisees
14. Commission nationale de l'informatique et des libertés. Recherche en santé : la CNIL adopte deux méthodologies de référence permettant l'accès à la base principale du SNDS [En ligne]. Paris (FR) : CNIL [cité le 26 mars 2025]. Disponible : https://www.cnil.fr/fr/recherche-en-sante-la-cnil-adopte-deux-methodologies-de-reference-permettant-lacces-la-base
15. Robert-Bobée I, Gualbert N. En 50 ans, l'EDP a bien grandi ! Courrier des statistiques. 2011 ; (6) : 47-63. Disponible : https://www.insee.fr/fr/information/5398685?sommaire=5398695
16. Institut national de la statistique et des études économiques. Fichiers des personnes décédées depuis 1970 [En ligne]. Paris (FR) : INSEE [cité le 26 mars 2025]. Disponible : https://www.insee.fr/fr/information/4190491
17. Cohorte Constances. Données des bases administratives [En ligne]. Cohorte Constances [cité le 26 mars 2025]. Disponible : https://www.constances.fr/donnees-des-bases-administratives/
18. Sureau C, Merlen R. Le Répertoire de gestion de carrières unique (RGCU), un nouveau référentiel et des perspectives pour l'analyse sociale. Courrier des statistiques. 2011 ; (6) : 64-82. Disponible : https://www.ccomptes.fr/sites/default/files/2024-12/20241220-S2024-Repertoire-gestion-carriere-unique-RGCU.pdf
19. Arrêté n° 1413-58-1 du 31 juillet 2023 relatif à la création d'un système d'information dénommé "LABOé-SI". Journal officiel de la République française. 2023.
20. Agence du numérique en santé. Doctrine du numérique en santé 2025 [En ligne]. Agence du numérique en santé [cité le 26 mars 2025]. Disponible : https://participez.esante.gouv.fr/project/doctrine-du-numerique-en-sante-ok/presentation/presentation
21. Agence du numérique en santé. SI-VIC, le système d'information pour le suivi des victimes d'attentats et de situations sanitaires exceptionnelles [En ligne]. Agence du numérique en santé [cité le 26 mars 2025]. Disponible : https://esante.gouv.fr/SI-VIC
22. Desjeux G, Balaire C, Pommier de Santi V, Léon C, Aspar A-M, Deparis X, *et al.* Enquête préliminaire sur les besoins de prévention en santé des militaires d'active. Médecine et armées. 2009 ; 37(5) : 389-397.
23. Martinez-Lorenzi E, Desjeux G, Marsan P, Thévenin-Garron V. Incidence des ALD 30 chez les militaires d'active entre 2005 et 2011. Santé publique. 2013 ; 25(5) : 599-607.
24. Ressort T, Desjeux G, Marsan P, Thévenin-Garron V. Les affections en service liées aux sports chez les militaires français. Santé publique. 2013 ; 25(3) : 263-270. doi : 10.3917/spub.135.0599
25. Arrêté du 19 mai 2020 portant création d'un traitement automatisé de données à caractère personnel dénommé « CMA numérique – AXONE » relatif à la gestion des dossiers médicaux par le service de santé des armées et à la médecine de prévention pour les personnels civils et militaires. Journal officiel de la République française. 2020.
26. Delon F, Mayet A, Thellier M, Kendjo E, Michel R, Ollivier L, *et al.* Assessment of the French National Health Insurance Information System as a tool for epidemiological surveillance of malaria. Journal of the American Medical Informatics Association. 2017 ; 24(3) : 588-95. doi : 10.1093/jamia/ocw164
27. Health Data Hub. Bibliothèque ouverte d'algorithmes en santé [En ligne]. Health Data Hub [cité le 26 mars 2025]. Disponible : https://gitlab.com/healthdatahub/boas
28. Looten V. Les études portant sur les bases de données médico-administratives sont-elles reproductibles ? Médecine/sciences. 2019 ; 35(8-9) : 689-92.
29. Loreau JM, Mmadi-Mrenda B, Collange F, Marchi J, Mayet A. Décès dans les armées, résultats du système national des données de santé – 2013 à 2018. Marseille (FR) : Centre d'épidémiologie et de santé publique des armées ; 2019.

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816

30. Health Data Hub. Fiches descriptives des standards et terminologies nationaux et internationaux des données de santé [En ligne]. Health Data Hub [cité le 26 mars 2025]. Disponible : https://documentation-snds.health-data-hub.fr/standards/

31. Health Data Hub. Documentation de la standardisation de la base principale du SNDS au format OMOP [En ligne]. Health Data Hub [cité le 26 mars 2025]. Disponible : https://documentation-snds.health-data-hub.fr/omop/

32. FCRIN. Rapport pour l'utilisation des données de santé en vie réelle en recherche clinique du 16 mai 2024 [En ligne]. FCRIN ; 2024 [cité le 26 mars 2025]. environ 50 p. Disponible : https://www.fcrin.org/sites/default/files/brique/fichier/05-2024/Rapport_Pour\%20l\%E2\%80\%99utilisation\%20des\%20donn\%C3\%A9es\%20de\%20sant\%C3\%A9\%20en\%20vie\%20r\%C3\%A9elle\%20en\%20recherche\%20clinique_16\%20mai\%202024_0.pdf

33. Chêne G, Delpierre C, Abbara S, Fery-Lemonnier E, Garcelon N, Georgin-Lavialle S, *et al.* Note à l'attention du Comité stratégique du Système national des données de santé [En ligne]. Health Data Hub ; 2023 [cité le 26 mars 2025]. Disponible : https://www.health-data-hub.fr/sites/default/files/2023-11/20230927_GT\%20socle\%20de\%20donn\%C3\%A9es_Note\%20de\%20synth\%C3\%A8se\%20des\%20travaux_VF.pdf

34. Commission nationale de l'informatique et des libertés. Entrepôts de données de santé : cartographie des acteurs qui en sont responsables en France [En ligne]. Paris (FR) : CNIL ; [s.d.] [cité le 26 mars 2025]. Disponible : https://carto-eds.beta.cnil.fr/

35. Texier G, Chaudet H, Pellegrin L, Meynard J-B. Surveillance épidémiologique pour l'alerte précoce et systèmes d'information. Dans : Astagneau P, Crépy P, rédacteurs. Surveillance épidémiologique et veille sanitaire. Paris (FR) : Lavoisier ; 2023. p. 87-98.

36. Desjeux G, Matton T, Galoisy-Guibal L. Utilisation du PMSI à des fins épidémiologiques dans les armées. Médecine et armées. 2004 ; 32(4) : 352-9.

37. Health Data Hub. HYDRO [En ligne]. Health Data Hub [cité le 26 mars 2025]. Disponible : https://www.health-data-hub.fr/partenariats/hydro

38. Guyomard M, Bouhnik AD, Tassy L, Urena R. Encoding breast cancer patients' medical pathways from reimbursement data using representation learning : a benchmark for clustering tasks. Dans : 2024 IEEE 37th International Symposium on Computer-Based Medical Systems (CBMS) [En ligne]. Guadalajara (MX) : IEEE ; 2024 [cité le 26 mars 2025]. p. 120-125. Disponible : https://ieeexplore.ieee.org/document/10600940/

39. Health Data Hub. TARPON II : traitement automatique des résumés de passage aux urgences pour un observatoire national – Phase 2 [En ligne]. Health Data Hub [cité le 26 mars 2025]. Disponible : https://www.health-data-hub.fr/projets/tarpon-ii-traitement-automatique-des-resumes-de-passages-aux-urgences-pour-un-observatoire

40. Chenais G, Touchais H, Avalos M, Bourdois L, Revel P, Gil-Jardiné C, *et al.* Performance en classification de données textuelles des passages aux urgences des modèles BERT pour le français [En ligne]. 2021 [cité le 26 mars 2025]. Disponible : https://inria.hal.science/hal-03276129

41. Zhou H, Liu F, Gu B, Zou X, Huang J, Wu J, *et al.* A survey of large language models in medicine : progress, application, and challenge [En ligne]. arXiv ; 2023 [cité le 26 mars 2025]. Disponible : https://arxiv.org/abs/2311.05112

42. Allassonnière P, Fraysse D. Données de santé artificielles : analyse et pistes de réflexion [En ligne]. [cité le 26 mars 2025]. Disponible : https://static.botdesign.net/docs/VF_Livre_blanc_Donn\%C3\%A9es_de_sant\%C3\%A9_artificielles-250424.pdf

Published in : SERVICE DE SANTÉ DES ARMÉES, - (dir.), *Médecine et Armées Vol.51 No.2. Dossier Numérique en santé*, Editions des archives contemporaines, ISBN : 9782813005816, doi : 10.17184/eac.9782813005816