Une image contenant dessin

Description générée automatiquement

ETUDES EN VIE REELLE – OPEN DATA

Les études en vie réelle autour des médicaments constituent un enjeu pour les laboratoires pharmaceutiques. Alors que les projets se multiplient, l’accès aux données devient crucial. Voici les questions que nous avons abordées dans notre commission QUANTISTAT.

* Quelles sont les sources de données accessibles ?
* Comment les études en vie réelles impactent-elles notre quotidien ?
* Comment les études peuvent-elles s’approprier ces données pour apporter de la valeur ajoutée aux équipes internes ?
* Quelles utilisations ?

1. **Définition des données en vie réelles :**

Les données patients en vie réelles sont l’ensemble des données collectées pendant la durée de vie d’un produit.

D’une manière générale elles sont générées à l’occasion des soins réalisés en routine pour un patient, et reflètent donc a priori la pratique courante. De telles données peuvent provenir de multiples sources : elles peuvent être extraites des dossiers informatisés de patients, ou constituer un sous-produit des informations utilisées pour le remboursement des soins.

Elles peuvent être collectées de manière spécifique, par exemple dans le cadre de procédures de pharmacovigilance, ou pour constituer des registres ou des cohortes, ou plus ponctuellement dans le cadre d’études ad hoc.

On distingue :

* + Les données patients issues des panels tels que LDP, LTP, Patient Intelligence, Market Intelligence (voir le document sur les Panels Patients).
  + Les données patients issues de l’open data : PMSI, SNIIRAM, SNDS, EGB,…

Nous nous intéressons dans le présent document uniquement aux données issues de l’open data, le sujet des panels patients fait l’objet d’un document à part.

1. **Données issues des dépenses de l’Assurance Maladie**
2. **Open DAMIR** : Base complète sur les **D**épenses d'**A**ssurance **M**aladie **I**nter **R**égimes

Ce jeu de données est une extraction du Système National Inter Régimes d'Assurance Maladie (SNIIRAM) portant sur l'ensemble des remboursements de l'assurance maladie tous régimes confondus (y compris les prestations hospitalières facturées directement à l'Assurance Maladie pour l'ensemble des régimes.).

Les dépenses sont détaillées selon :

* 6 axes d'analyse (période, prestation, organisme de prise en charge, bénéficiaire des soins, professionnel de santé exécutant, professionnel de santé prescripteur)
* 7 indicateurs de montant (total de la dépense, base de remboursement, montant remboursé, dépassement) et de volume (dénombrement, quantité, coefficient).

Au total, chaque ligne de prestation est décrite par 55 variables. Il est indispensable de se référer au descriptif accompagnant le jeu de données.

Les données sont téléchargeables au format csv.

|  |  |
| --- | --- |
| VARIABLES – BASE DAMIR | |
| **PERIODE DE TRAITEMENT** | Année et Mois de Traitement |
| **PERIODE DE SOINS** | |  | | --- | | Année de Soins  Mois de Soins | |
| **PRESTATION** | |  | | --- | | Nature de Prestation | | Nature d'Assurance | | Nature de l'Accident du Travail | | Type d'Enveloppe | | Complément d'Acte | | Motif d'Exonération du Ticket Modérateur | | Taux de Remboursement | | Code Secteur Privé/Public | | Type de Prise en Charge Forfait Journalier | | Indicateur TAA Privé/Public | | Code Qualificatif Parcours de Soins (sortie) | | Nature du Destinataire de Règlement affiné | | Type de Remboursement | |
| **ORGANISME** | |  | | --- | | ZEAT de l'Organisme de Liquidation jusqu'à 2014 (tables P) | | Région de l'Organisme de Liquidation à partir de 2015 (tables A) | |
| **BENEFICIAIRE** | |  | | --- | | Sexe du Bénéficiaire | | Tranche d'Age Bénéficiaire au moment des soins | | Qualité du Bénéficiaire | | ZEAT de Résidence du Bénéficiaire jusqu'à 2014 (tables P) | | Région de Résidence du Bénéficiaire à partir de 2015 (tables A) | | Modulation du Ticket Modérateur | | Top Bénéficiaire CMU-C | |
| **EXECUTANT** | |  | | --- | | Catégorie de l' Exécutant | | Spécialité Médicale PS Exécutant | | Nature d'Activité PS Exécutant | | ZEAT du PS Exécutant jusqu'à 2014 (tables P) | | Région du PS Exécutant à partir de 2015 (tables A) | | Statut Juridique PS Exécutant | | Mode de Fixation des Tarifs Etb Exécutant | | ZEAT d'Implantation Etb Exécutant jusqu'à 2014 (tables P) | | Région d'Implantation Etb Exécutant à partir de 2015 (tables A) | | Type Etb Exécutant | | Catégorie Etb Exécutant | | Discipline de Prestation Etb Exécutant | | Mode de Traitement Etb Exécutant | |
| **PRESCRIPTEUR** | |  | | --- | | Catégorie du Prescripteur | | Spécialité Médicale PS Exécutant | | Nature d'Activité PS Prescripteur | | ZEAT du PS Prescripteur jusqu'à 2014 (tables P) | | Région du PS Prescripteur à partir de 2015 (tables A) | | Statut Juridique PS Prescripteur | | ZEAT d'Implantation Etb Prescripteur jusqu'à 2014 (tables P) | | Région d'Implantation Etb Prescripteur à partir de 2015 (tables A) | | Catégorie Etb Prescripteur | |
| VARIABLES – BASE DAMIR | |
| **INDICATEURS** | |  | | --- | | Coefficient Global de la Prestation Préfiltré | | Dénombrement de la Prestation Préfiltré | | Quantité de la Prestation Préfiltrée | | Montant du Dépassement de la Prestation Préfiltré | | Montant de la Dépense de la Prestation Préfiltrée | | Montant Versé/Remboursé Préfiltré | | Coefficient Global | | Dénombrement | | Quantité | | Montant du Dépassement | | Montant de la Dépense | | Montant Versé/Remboursé | | Base de Remboursement | |

1. **Dépenses d'assurance maladie hors prestations hospitalières en données nationales** :

Base de données regroupant l'ensemble des remboursements (hors prestations hospitalières) effectués par le régime général de l'Assurance Maladie en France.

Cette base de données contient l'ensemble des remboursements **mensuels** effectués par le régime général de l'Assurance Maladie (hors prestations hospitalières) par :

* type de prestations (soins et prestations en espèces),
* type d'exécutant (médecins par spécialité, chirurgiens dentistes, auxiliaires médicaux, laboratoires d'analyse, pharmaciens, ...)
* type de prescripteurs.

Les dépenses sont indiquées en montants remboursés et présentées au remboursement.

1. **Dépenses d'assurance maladie hors prestations hospitalières par caisse primaire/département :**

Base de données regroupant l'ensemble des remboursements (hors prestations hospitalières) effectués par le régime général de l'Assurance Maladie en France.

Cette base de données contient l'ensemble des remboursements mensuels effectués par le régime général de l'Assurance Maladie (hors prestations hospitalières) **par caisse primaire/département** par

* type de prestations (soins et prestations en espèces),
* type d'exécutant (médecins par spécialité, chirurgiens dentistes, auxiliaires médicaux, laboratoires d'analyse, pharmaciens, ...)
* par type de prescripteurs.

Les dépenses sont indiquées en montants remboursés et présentées au remboursement.

Les tables **mensuelles** au format csv sont regroupées dans des fichiers zip **annuels**.

Deux présentations sont disponibles :

* une version complète avec les libellés des variables et des regroupements pour certaines variables
* une version simplifiée, sans les libellés et sans les regroupements.

1. **Dépenses annuelles d'assurance maladie :**

Tableaux de bord portant sur l'ensemble des remboursements (y compris les prestations hospitalières) effectués par le régime général de l'Assurance Maladie en France métropolitaine.

Rétrospective annuelle de l'ensemble des remboursements du régime général de l'Assurance Maladie effectués en France métropolitaine, par :

* type de risque : maladie, maternité, invalidité et décès, accident du travail et maladie professionnelle.

Pour chaque type de risque, les dépenses sont présentées :

* par catégorie de professionnels de santé

Pour chaque catégorie, les dépenses sont présentées

* par acte ou par groupe d'actes.

1. **Données issues du remboursement des médicaments par l’Assurance Maladie**
2. **Medic'AM** : **médic**aments remboursés par l’**A**ssurance **M**aladie

Données **mensuelles** et **annuelles** sur les médicaments remboursés par l'Assurance Maladie.

Des informations détaillées sur les médicaments remboursés par l'ensemble des régimes d'assurance maladie, **en France entière**. Dernières données disponible : Année 2018.

Les données Medic'AM proposent, pour chaque médicament, par code CIP13 :

* la base de remboursement
* le montant remboursé
* le nombre de boites remboursées

Une ventilation selon le type de prescripteur (libéral ou salarié) est également proposée.

Remarque importante : la base de remboursement et le montant remboursé n'incluent pas les honoraires de dispensation au conditionnement perçus par les pharmaciens d'officine depuis le 1er janvier 2015.

1. **Open Medic** : Base complète sur les dépenses de **médic**aments inter régimes

L’offre de données Open Medic est constituée d’un ensemble de bases annuelles, portant sur l’usage du médicament, délivré en pharmacie de ville de 2014 à 2019. Elle fournit des informations complémentaires au fichier [**Medic’AM**](https://www.data.gouv.fr/fr/datasets/medicaments-rembourses-par-lassurance-maladie/).

Toutes les données sont extraites du Système National des Données de Santé (SNDS).

Les données sur le médicament sont restituées au travers de la classification ATC.

La classification ATC est utilisée pour classer les médicaments selon un principe hiérarchique. Les médicaments sont divisés en différents groupes selon l'organe ou le système sur lequel ils agissent et/ou leurs caractéristiques thérapeutiques et chimiques.

Cette offre s’articule autour de deux types de jeux de données :

1. **Base complète sur les dépenses de médicaments interrégimes**

Les premiers jeux de données Open\_Medic\_2014 à 2019 permettent d’étudier

* les dépenses de médicaments (montants remboursés et remboursables) et le nombre de boîtes délivrées,
* en fonction d’éléments descriptifs sur les bénéficiaires (tranche d’âge, sexe, région de résidence selon la nouvelle nomenclature Insee)
* ou de l’information sur la spécialité du prescripteur,
* pour une année.

1. **Bases complémentaires enrichies des dénombrements de consommants**

Les seconds jeux de données ont été constitués en complément des bases Open\_Medic\_2014 à 2019 pour restituer le nombre de personnes ayant consommé des médicaments à différents niveaux de regroupements.

Les indicateurs sur les montants et le nombre de boîtes sont également mis à disposition sur ces différents niveaux intermédiaires.

Ces fichiers sont préfixés par NB et l'année concernée. Puis le suffixe indique d’une part le niveau de la classification ATC étudié (ATC1, ATC2,.., CIP13), d’autre part les critères de ventilation complémentaires.

Quarante-huit bases ont été constituées afin de répondre au besoin et de garantir le non double-compte des consommants. Elles sont accessibles selon l'année considérée et les différents niveaux de la hiérarchie ATC .

Remarque importante : la base de remboursement et le montant remboursé n'incluent pas les honoraires de dispensation au conditionnement perçus par les pharmaciens d'officine depuis le 1er janvier 2015.

1. **Données spécifiques issues de l’hôpital**
2. **Open PHMEV** : bases sur les **p**rescriptions **h**ospitalières de **mé**dicaments délivrées en **v**ille

L’offre de données PHMEV est constituée d’un ensemble de bases annuelles inter régimes, portant sur les remboursements de médicaments prescrits par les établissements publics et ESPIC (établissements de santé privés d’intérêt collectif) et délivrés en officine de ville de 2014 à 2017.

Toutes les données sont extraites du système national inter régimes de l’Assurance Maladie (SNIIRAM).

Les données sur le médicament sont restituées au travers de la classification ATC.

Le bénéficiaire est caractérisé par sa classe d’âge et son sexe.

Cette offre s’articule autour de deux types de jeux de données :

**1. Base complète PHMEV :**

La base complète PHMEV comprend l’hôpital prescripteur décrit par :

* son numéro (géographique pour l’APHP et la Croix Rouge (ainsi que pour les hôpitaux militaires à partir de 2017) ou juridique pour les autres établissements),
* sa catégorie juridique
* ses éléments de géolocalisation.

**2. Base régionale PHMEV :**

De manière complémentaire, les prescriptions hospitalières délivrées en ville sont restituées, **à un niveau plus agrégé**, au travers de la région d’implantation des établissements (région géographique pour l’APHP et la Croix Rouge (hôpitaux militaires à partir de 2017), région gestionnaire pour les autres) et de leur catégorie juridique.

1. **Retroced'AM**

Des informations détaillées sur les médicaments remboursés dans le cadre de la rétrocession hospitalière par le régime général (y compris les sections locales mutualistes) de l'Assurance Maladie, en France métropolitaine, sont présentées pour les années 2010 à 2017.

Les données Retroced'AM proposent, pour chaque médicament, par code UCD :

* la base de remboursement ;
* la marge de rétrocession (montants inclus dans la base de remboursement présentée) ;
* le montant remboursé ;
* le nombre d'unités remboursées.

Pour les prestations ne donnant pas lieu à la remontée d'informations par code affiné UCD (médicament avec une autorisation temporaire d’utilisation, préparation magistrale hospitalière, etc.), la base de remboursement et le montant remboursé sont présentés de manière agrégée selon le code de nature de la prestation (PHU, PHM, PHP, PHI, MAR ou PHD).

1. **Accessibilité des données :**

Les données sont accessibles en open data sur le site <https://www.data.gouv.fr/>. Néanmoins pour pouvoir exploiter l’ensemble des données, il est nécessaire d’avoir une connaissance fine des indicateurs. De plus il semble important d’avoir un logiciel ou le matériel informatique adéquat (volumétrie importante des fichiers) pour réussir à exploiter les bases disponibles.

Les bases sont donc récupérables directement par les laboratoires, comme DAMIR, PHMEV (liste non exhaustive).

Néanmoins les laboratoires qui n’ont pas toujours les ressources internes (logiciel informatique, data scientists, temps homme,…) pour exploiter ces bases et peuvent faire appel à des tiers comme le gers, iqvia, real consulting data,…

A noter que pour certaines bases de données (comme le PMSI) il est important de définir le cadre juridique :

* + Besoin de définir la finalité des données : acces médical et non pas marketing
  + Aide à la réalisation des briefs pour coller au cadre juridique

1. **Limites / Avantages des données « Open Data » :**

|  |  |
| --- | --- |
| LIMITES | POINTS POSITIFS |
| Fréquence des données | Exhaustivité des données |
| Volumétrie des fichiers | Faible coût si traité en interne |
| Pas possible d’agréger / partitionner |  |
| Ne convient pas à tous les marchés – Pas de diagnostic |
| Données brutes |
| Appropriation/compréhension de la donnée |

1. **Exemples pratiques :**

Partage d’expérience d’un labo :

* Analyse de données dans le sevrage tabagique avant et après déremboursement
* Données DAMIR sur les pathologies hivernales

1. **Prestataires**

On distingue

* + des sociétés très spécialisées pour traiter la données de vie réelle comme par exemple :
    - Semca
    - Heva
    - Real Data Consulting
    - MEDIA CONSEIL Contexte
  + des sociétés plus généralistes qui ont également des panels
    - Cegedim
    - IQVIA

1. **Exemples de types d’études réalisées à partir des données Open Data**

* La définition d’une population cible
* La description de la prise en charge d’une maladie, en particulier les lignes de traitement, le parcours de soin des patients
* L’utilisation des médicaments en vie réelle: efficacité, effets indésirables (morbidité/mortalité), persistance
* Etude comparative d’efficacité, de tolérance entre deux médicaments ou deux dispositifs médicaux
* La pharmaco-vigilance
* Le coût de prise en charge d’une pathologie (études médico-économiques)

1. **Evolution métier**

On peut s’attendre à un besoin de personnes qui sont capables de traiter de plus en plus de données (volumétrie) tels que les Data scientist.

Nous avons observé au cours de cette année, l’émergence d’un nouveau métier data scientist, personne souvent dédiée mais rattachée dans un service autre que celui des études de marché.

**Néanmoins, aujourd’hui peu de laboratoires ont des Datascientists au niveau France, il semblerait que ce métier commence à émerger au niveau global ou dans les services Real World Data.**

1. **Sources :**

La liste des bases exposées dans ce document n’est pas exhaustive.

Le lien ci-dessous vous permet d’avoir accès **à la liste de l’ensemble des bases de données répertoriées** par la mission Etalab, publiée en 2014.

<https://www.data.gouv.fr/fr/datasets/cartographie-des-bases-de-donnees-publiques-en-sante/>

Vous pouvez télécharger la liste qui vous donne des informations précieuses sur les accès aux bases répertoriées.