LES MATINEES

CEMKAfé

Des bases de données en accès libre mais sans mode d'emploi : l'Open-Data *by* CEMKA

10/12/2020 : 8h30-10h

Sandrine BAFFERT, Marc Anthony CHECHOI, Cecile FABRON, Julien ROBERT, Clément TEISSIER

EN PRATIQUE

1. COUPEZ VOS MICROS

2. POSEZ VOS QUESTIONS **UNIQUEMENT PAR ÉCRIT** SUR L'OUTIL DE CHAT

MERCI

AGENDA

1. Contexte
2. Les bases spécifiques
3. L'Open DAMIR

Illustrées par des exemples de requêtes et d'utilisation à partir de notre outil d'Opendata

INTERVENANTS

Sandrine BAFFERT
Directeur HEOR
sandrine.baffert@cemka.fr

Clement TEISSIER
Biostatisticien
Clement.teissier@cemka.fr

Julien ROBERT
Directeur Adjoint HEOR
Modeleur HEOR
Julien.robert@cemka.fr

Cécile FABRON
Economiste de la santé
Cecile.fabron@cemka.fr

Marc Anthony CHECHOI
Modeleur HEOR
marc-anthony.chechoimienje@cemka.fr

Contexte

De quelles bases parle t-on ?

	Accès libre	Accès restreint
Bases hospitalières	▪ Statistique annuelle des établissements de santé ▪ Statistiques sur le nombre de séjours en fonction des codes diagnostiques ou des GHM ▪ ENCC MCO, HAD, SSR et EHPAD ▪ Référentiel de coût des unités d'oeuvres ▪ Open CCAM ▪ Montant des produits facturés en sus	PMSI
Bases de soins de ville générales	▪ Dépenses d'assurance maladie : Open data et Open DAMIR ▪ SNIR	EGB SNDS
Bases soins de ville spécifiques	▪ BioLAM, Open Bio ▪ MedicAM, Open Medic, Open PHMEV, RetrocedAM, Bdm_IT ▪ LPPAM, LPP ▪ CCAM	Bases Taxe sur les médicaments et sur les DM, ANSM Bases GERS Base IQVia Xpr-So-Openhealth Et autres bases de données privées

Quelle utilité de ces bases ?

- Pour cadrer des analyses et problématiques, et disposer de données agrégées sans avoir à passer par les bases exhaustives
- Pour avoir un accès « immédiat », sans temps réglementaire
- Pour disposer dans certains cas de données plus récentes que celles des bases reglementaires (SNDS)
- Pour alimenter les paramètres des modèles économiques et épidémiologiques
- Pour disposer d'une validité externe par rapport à des données de modélisation
- Pour choisir les outils de traitements des données (travail sous Python, R, …)

Les bases hospitalières

Les bases hospitalières

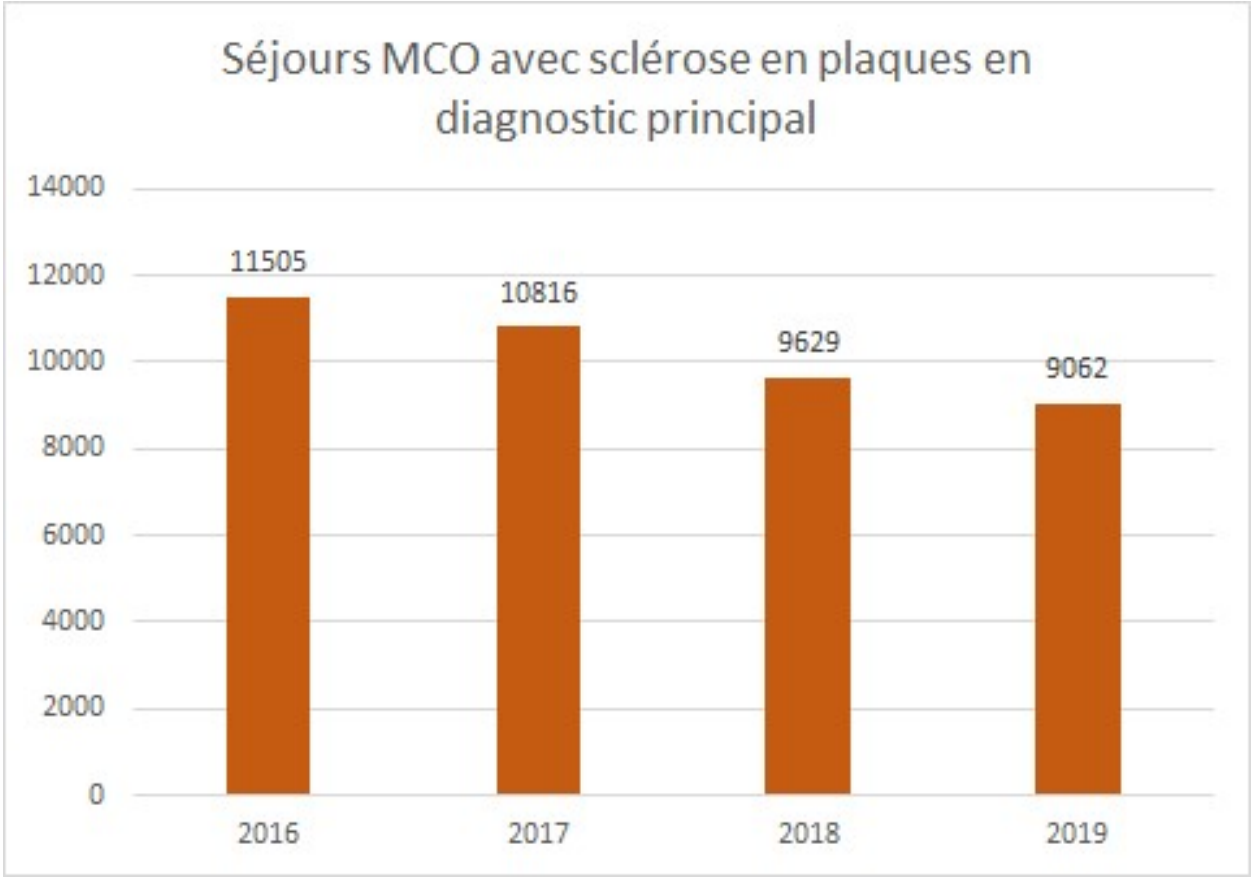
L'open data de l'ATIH : Scan Santé

- Etablissements de santé en MCO, HAD, SSR et PSY + EHPAD
- Contenu et historique variables selon l'activité considérée
- Informations en accès libre :
  - Des indicateurs synthétiques : indicateurs de performance Hospi-Diag, activités hospitalières par région,
  - Des données d'activité :
    - Statistiques médicales par groupe (GHM), diagnostic (CIM-10) ou acte (CCAM) ;
    - Consommations des médicaments et DMI en sus
  - Des données économiques :
    - Référentiel national de coûts ;
    - Référentiel de coût des unités d'œuvre (activités de soins et des fonctions support)

Les bases hospitalières

Un exemple MCO

L'open data de l'ATIH : Scan Santé



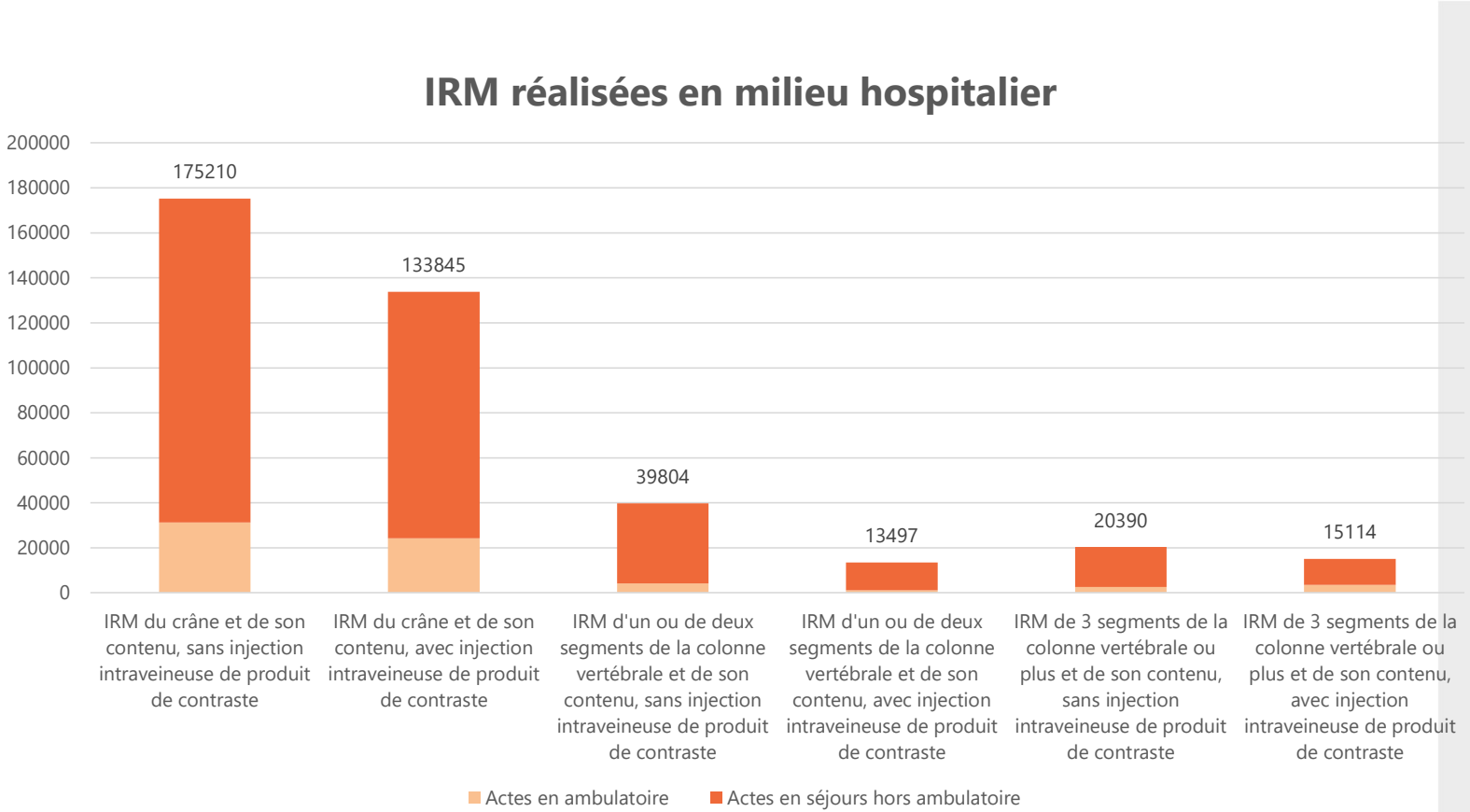
Les bases hospitalières

Open CCAM

- Constitué à partir du PMSI pour l'activité de MCO et odontologie des établissements de santé.
- **Actes réalisés par les établissements de santé.**
- Pour chaque ligne :
  - Identifiant de l'établissement de santé ;
  - Code de l'acte médical réalisé ;
  - Le nombre de réalisation de l'acte au cours de l'ensemble de séjours hospitaliers (à temps complet comme ambulatoires ou séances) ;
  - La durée moyenne des séjours hospitaliers ;
  - Le nombre de séjours ambulatoires.
- Données disponibles pour les années 2015 à 2019

Exemple sur SEP

OpenCCAM



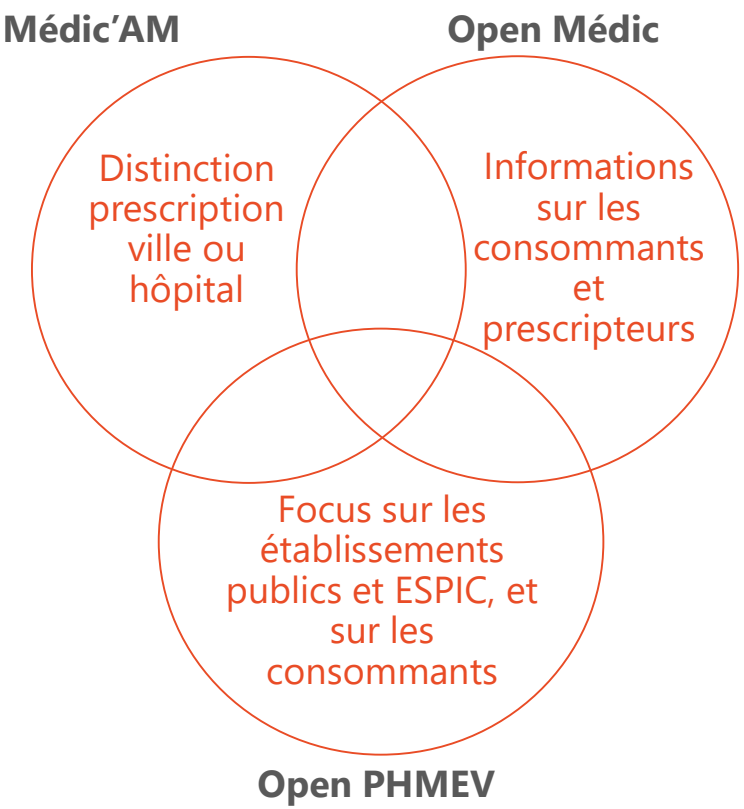
Des questions ?

Les bases spécifiques

Les bases spécifiques

Les remboursements de médicaments

Médicaments délivrés en ville



Bdpm : informations tarifaires sur les médicaments remboursés aux assurés sociaux et agréés aux collectivités

Médicaments délivrés en rétrocession

Rétrocéd'AM

Informations sur le nombre d'unités et les marges de rétrocession

Médicaments délivrés en sus du GHM

Liste en sus ATIH

Base historique. Informations sur les médicaments délivrés

Les bases spécifiques

Les remboursements de médicaments

Medic'AM

- **Nombre de boîtes remboursées** et montants correspondants par code CIP
- **France entière**, y compris DOM-TOM, sans détail géographique
- **Distinction entre les prescripteurs** « ville » et « hôpital »
- Format **annuel, ou mensuel** depuis 2015
- Mise-à-jour : *a priori*, décalage de 3 mois

Les bases spécifiques

Les remboursements de médicaments

Open MEDIC

- Format équivalent à Medic'AM mais détail en plus
  - **Dénombrement de consommants et leurs caractéristiques (tranche d'âge, sexe, région)**
  - **Spécialité de prescripteur**
- Contenu plus restreint : format **annuel** de 2014 à 2019 ;
- Mise à jour annuelle, au mois de mars

Les bases spécifiques

Les remboursements de médicaments

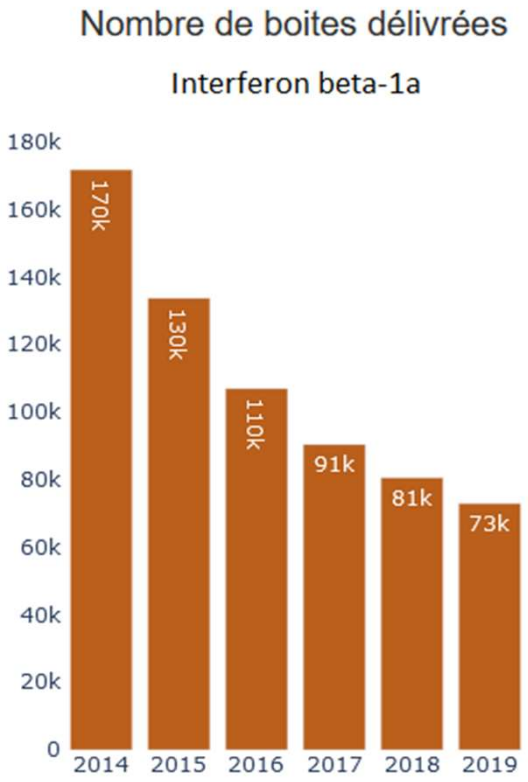
Medic'AM ou Open MEDIC ?

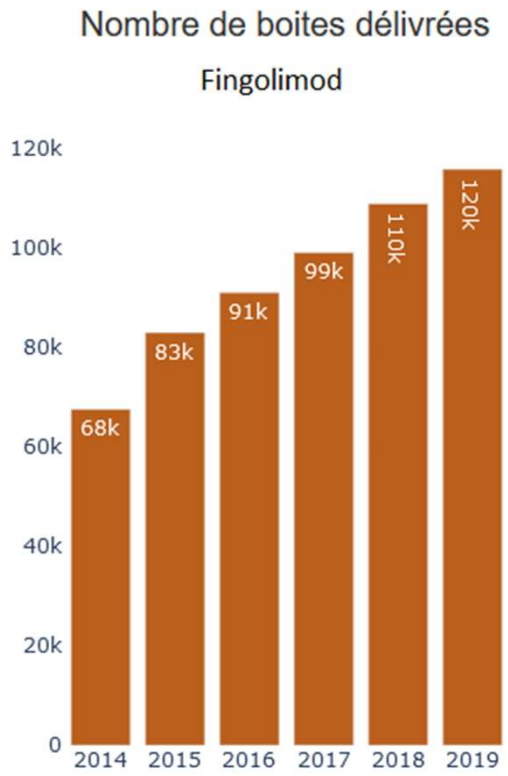
Contenu	Medic'AM	Open MEDIC
Nombre de consommants	X	√
Caractéristiques des consommants	X	√
Détail du prescripteur	Ville (sans spécialité) ou hôpital public	Ville (avec spécialité) ou hôpital public
Détail mensuel	√	X
Fréquence de mise à jour	Mensuelle	Annuelle
Décalage	3 mois	3ème mois de l'année suivante

Open MEDIC

Les bases spécifiques

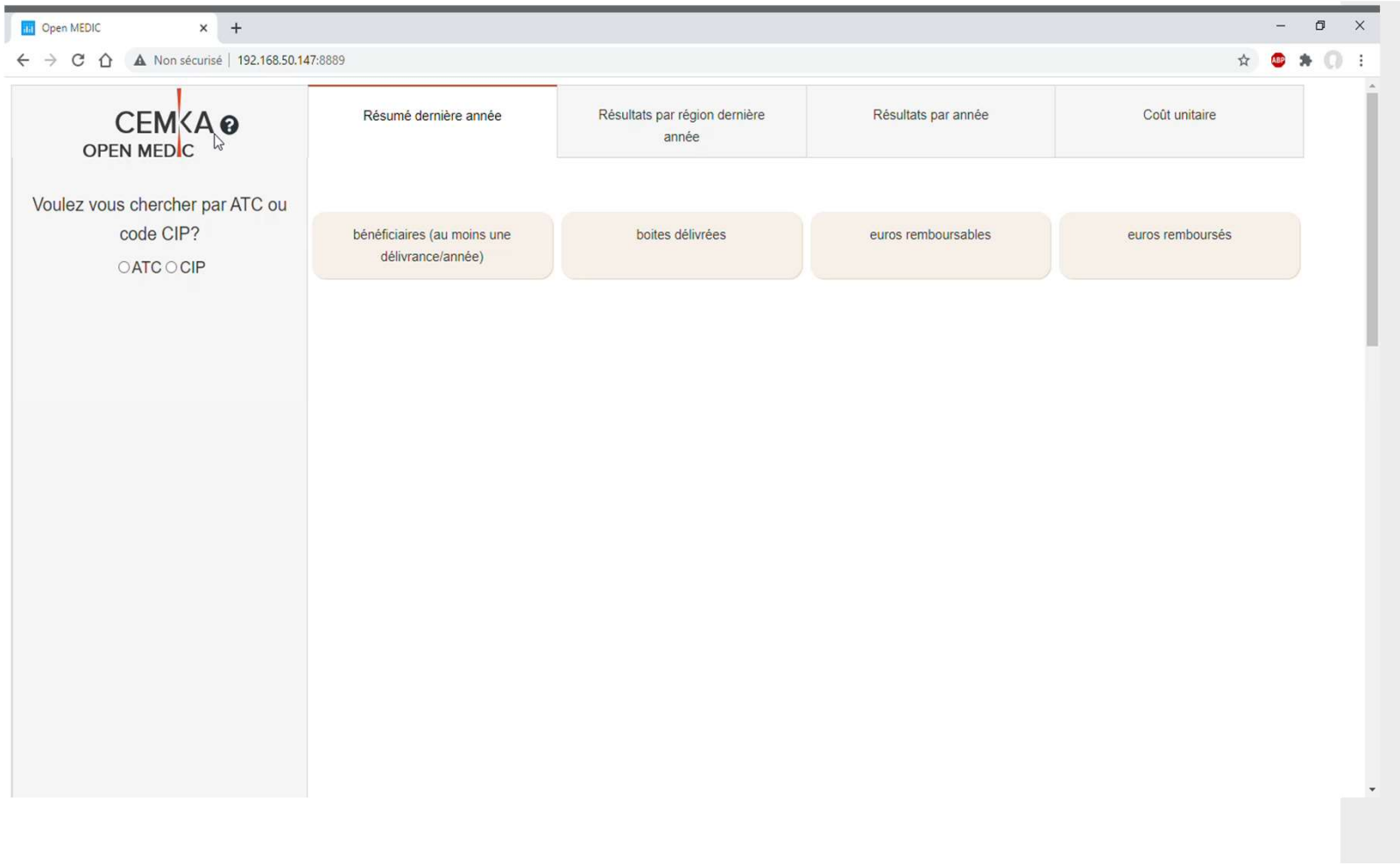
Les remboursements de médicaments





OPEN MEDIC

by CEMKA



Les bases spécifiques

Les remboursements de médicaments

Open PHMEV

- Contenu : du même format que Open Medic, mais **restreint aux prescripteurs hospitaliers** ;
- Détail
  - Un **établissement prescripteur public** (FINESS)
  - Une **tranche d'âge** donnée
  - Un **sexe** donné.
- Mise à jour annuelle, de 2014 à 2018
- Pas de dénombrement des consommants.

Les bases spécifiques

Les remboursements de médicaments

Retroced'AM

- 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.**
- Contenu :
  - Format annuel.
  - Dernières données dispo : 2019

Les bases spécifiques

Les remboursements de médicaments

Liste en sus

- Deux tableaux : **public ou privé** ;
- Données disponibles :
  - **Nombre d'UCD administrés** dans l'année ;
  - **Montant au prix hors remise**
- Dernière version disponible : Août 2020

Les bases spécifiques

Les autres bases

- **Biologie :** Biol'AM, Open Bio ;
- **Dispositifs médicaux :** LPP'AM ;
- **Causes de décès :** CépiDc ;
- **Caractéristiques des médecins :** RPPS, données de l'AM, SNIR.

Biologie : Biol'AM et Open Bio

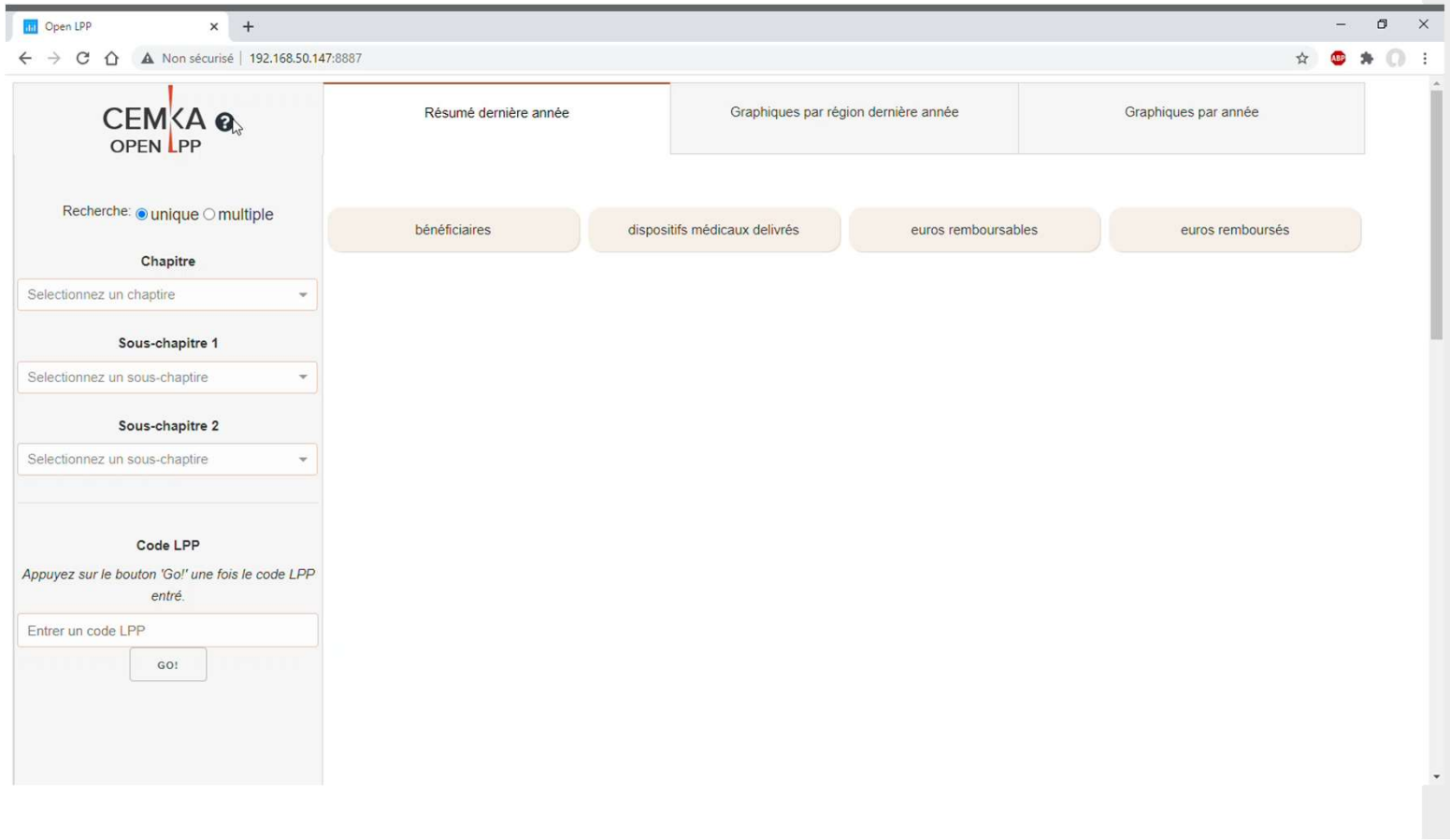
- Biol'AM
  - dénombrement annuel des actes de biologie médicale codés et remboursés par le RG et les SLM
  - Nb d'unité remboursée, montant remboursable cumulé, montant remboursé cumulé par acte de biologie
- Open Bio
  - Biol'AM + caractéristiques et dénombrement des consommants + région de résidence + spécialité prescripteur
- Couverture :
  - De 2014 à 2019 pour Open Bio ;
  - De 2013 à 2020 pour Biol'AM ;
  - Fréquence de mise à jour : annuelle.

Dispositifs médicaux : LPP'AM et OPEN LPP

- LPP'AM
  - Dénombrement annuel des dispositifs médicaux et des prestations inscrits à la liste des produits et prestations (LPP) et remboursés par le RG (hors SLM), par code LPP
  - Nombre d'unités remboursées, montant remboursable cumulé, montant remboursé cumulé
- Open LPP
  - LPP'AM + nombre de bénéficiaires, détail par âge, sexe, région de résidence, spécialité du prescripteur
- Couverture :
  - De 2006 à 2019 ;
  - Fréquence de mise à jour annuelle ;
  - Champ : régime général hors SLM en France métropolitaine.

OPEN LPP

by CEMKA



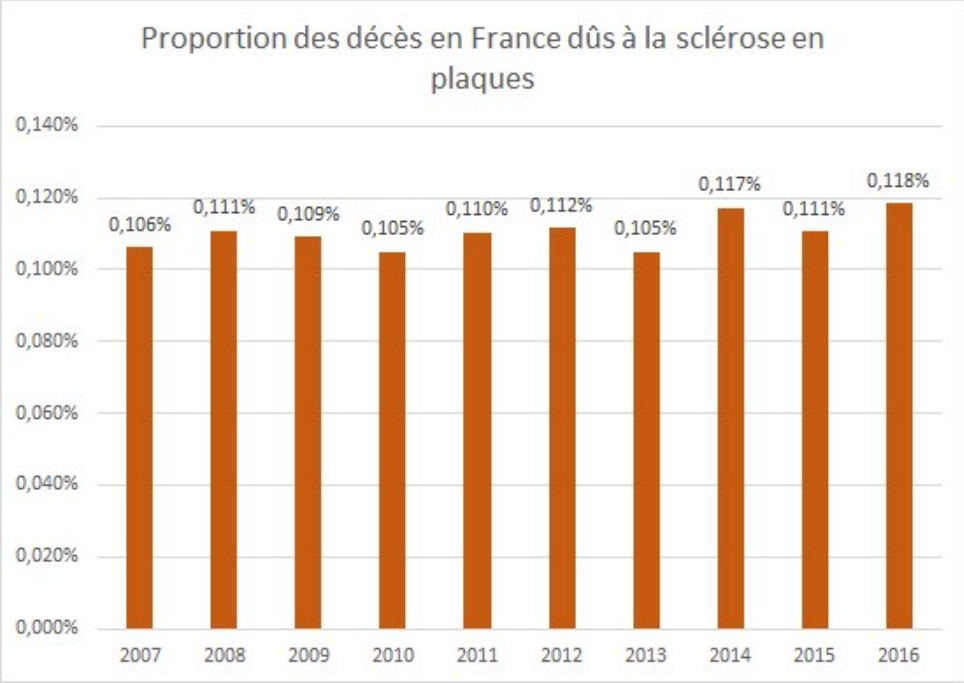
Causes de décès

- CépiDC : causes de décès de 1980 à 2016
- Indicateurs disponibles :
  - Effectifs de décès ou taux bruts de décès
  - Effectifs et taux standardisés par âge, part des décès, surmortalité masculine
- Détail géographique : des grandes villes à la France entière
- Détail de la cause de décès : liste abrégée ou CIM-10
- Pas de données récentes et seulement un outil d'interrogation en ligne

Les bases spécifiques

Les autres bases

Causes de décès – Code CIM-10 G35



Les bases spécifiques

Les autres bases

Caractéristiques des médecins

- L'outil d'interrogation mis en place par la DREES :
  - Données issues de l'exploitation des référentiels nationaux : Répertoire Partagé des Professionnels de Santé (RPPS), répertoire ADELI
  - PS : médecins, sages-femmes, chirurgiens dentistes, pharmaciens, MK pédicures-podologues, infirmiers, ergothérapeutes, audioprothésistes…
  - Données disponibles : effectifs des PS par année, spécialité, mode d'exercice, sexe, tranche d'âge, région d'obtention du diplôme, secteur d'activité, … / densité & âge moyen des PS ;
- Données statistiques de l'AM :
  - Professionnels de santé libéraux
  - Caractéristiques socio-démographiques, dénombrements d'actes, honoraires, région ou département

Des questions ?

La base générale : DAMIR

La base générale : DAMIR

Le contenu

Open DAMIR

- Damir = **D**épenses de l'**A**ssurance **M**aladie **I**nter-**R**égimes
- Contenu semblable à celui du SNIIRAM, mais restitué d'une manière différente → absence de détail individuel = pas de chaînage envisageable

- Tout tourne autour de la variable de prestations **PRS_NAT** :

PRS_NAT	Libellé
1111	CONSULTATION COTEE C
1112	CONSULTATION COTEE CS
1351	ACTE D'IMAGERIE (hors ECHOGRAPHIE) CCAM
1813	IK PIED SKI
6115	IJ MAJOREES -3MOIS
1145	REMUNERATION OBJECTIF MEDECIN

- Certaines prestations (CCAM, NGAP, médicaments, biologie, …) sont présentées sous forme de **variables agrégées**

La base générale : DAMIR

Le contenu

Format d'une ligne de la base

- Données correspondant à chaque PRS_NAT, avec les volumes totaux mensuels (correspondants à la périodicité de la base) pour :
  - Nombre de remboursements ;
  - Montant remboursable ;
  - Montant remboursé ;
  - Dépense.
- Selon le format de la base, chaque prestation peut être précisée selon :
  - Des caractéristiques du bénéficiaire : âge, sexe, …
  - Des caractéristiques sur le prescripteur et l'exécutant : nature, spécialité, statut juridique, mode de fixation des tarifs, …
  - Des caractéristiques géographiques : région/département de l'organisme de liquidation, région du bénéficiaire, région du prescripteur et de l'exécutant, …
  - Des caractéristiques administratives : taux de remboursement, qualificatif parcours de soins, nature de l'assurance, type d'enveloppe, T2A, …

  -> Au total, chaque ligne de prestation est décrite par **55 variables**

La base générale : DAMIR

	Base A	Base R	Base N
Données	Exhaustif	Hospitalier exclu	Hospitalier exclu
Champ	**Tous régimes**	Régime général	Régime général
Mise à jour	Annuelle		
Période	2009 - 2019		
Périodicité	Mensuelle		
Couverture	France entière		
Granularité	9 zones géo. de 2009 à 2014 13 régions à partir de 2015	Caisse primaire / **département**	Aucune
Taille de la base	Un mois = **4 Go**	Un mois = 280 Mo	Un mois = 60 Mo

La base générale : DAMIR

La base générale : DAMIR

Des exemples

L'avis ponctuel de consultant

Attention aux subtilités de la base…

- Avis ponctuel de consultant applicable seulement si le patient est reçu sur « demande explicite » de son médecin traitant ;
- Deux types de tarif : **C2 ou C3** ;
- Dans le DAMIR, **C2 = PRS_NAT 1117** ;
- En revanche, pas de PRS_NAT pour le C3 ;
- Donc **C3 codé comme « 3 fois C »** (PRS_NAT 1111) → non identifiable précisément et « perdu » dans les C ;
- Un exemple chez les spécialistes libéraux en oncologie médicale :

Type de consultation	PRS_NAT	Nombre de consultations	Nombre de coefficients	Coefficients par consultation
C	1111	13 839	21 079	**1,52**
CS	1112	326 542	326 319	1,00
C2	1117	53 929	107 642	2,00

Les forfaits techniques d'imagerie

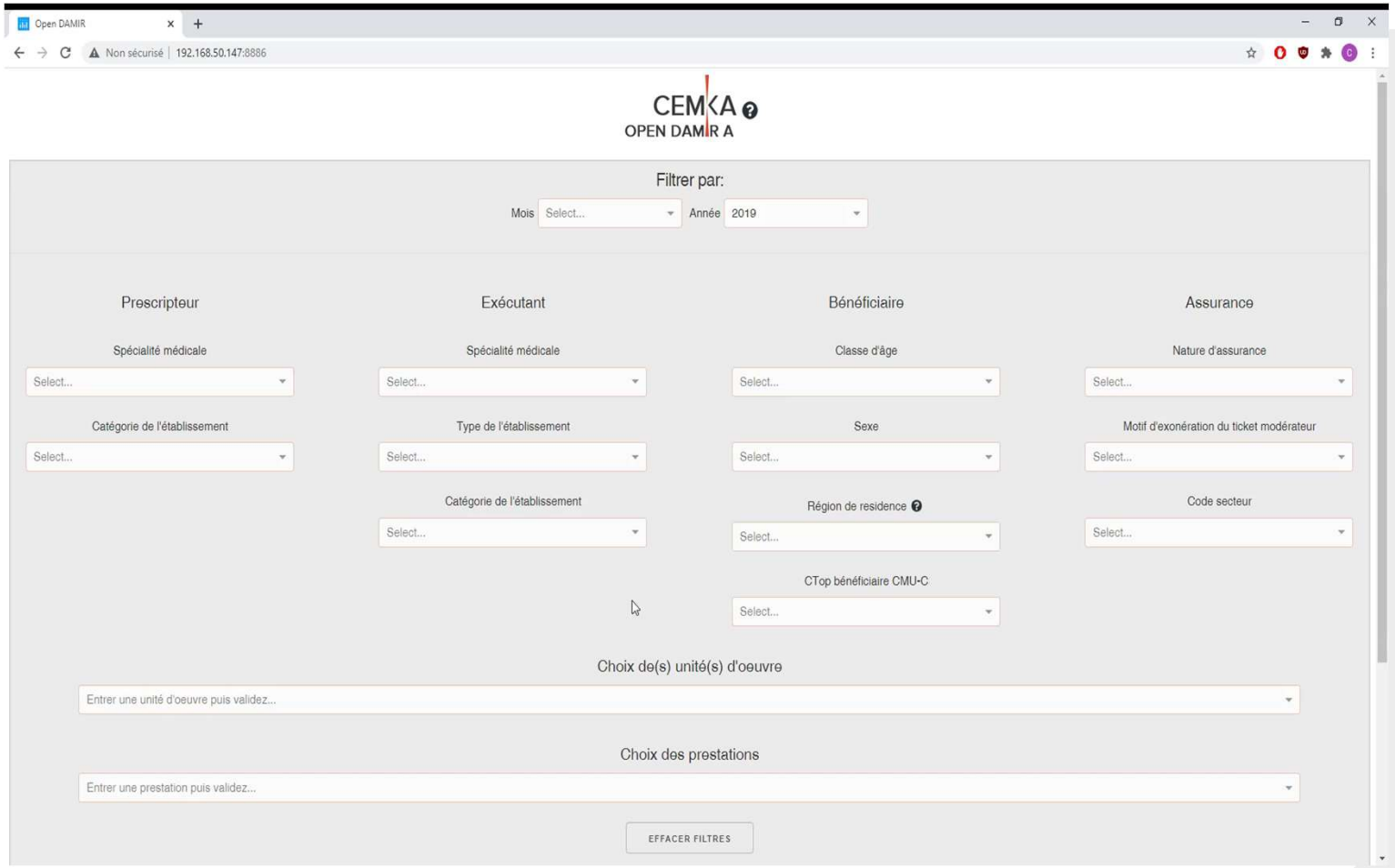
La base générale : DAMIR

Des exemples

- Pas assez de détail dans la variable PRS_NAT : forfait technique normal, réduit, réduit n°2 ou réduit n°3 ;
- Nécessité de coupler avec la série mensuelle : scanner ou IRM ;
- Résultat :

	Montants totaux	Nombre de forfaits	Montant moyen
SCANNER			
FORFAIT TECHNIQUE NORMAL IRMN -SCANNERS	239 398 973 €	2 949 118	81 €
FORFAIT TECHNIQUE REDUIT IRMN -SCANNERS	83 247 699 €	1 927 865	43 €
FORFAIT TECHNIQUE TARIF REDUIT N°2	5 492 883 €	167 815	33 €
FORFAIT TECHNIQUE TARIF REDUIT N°3	6 127 705 €	267 855	23 €
TOTAL	**334 267 261 €**	**5 312 653**	**63 €**
IRM			
FORFAIT TECHNIQUE NORMAL IRMN -SCANNERS	382 404 674 €	2 273 485	168 €
FORFAIT TECHNIQUE REDUIT IRMN -SCANNERS	88 230 873 €	1 292 518	68 €
FORFAIT TECHNIQUE TARIF REDUIT N°2	6 705 086 €	128 911	52 €
FORFAIT TECHNIQUE TARIF REDUIT N°3	407 265 €	15 530	26 €
TOTAL	**477 747 898 €**	**3 710 444**	**129 €**

OPEN DAMIR – DAMIR R



Des questions ?

CONCLUSION

De nombreuses possibilités…

- Suivi macro-économique :
  - Suivi de l'activité des médecins généralistes, mois par mois
  - Comparaisons de recours aux soins entre différents départements
  - Evolution du taux de remboursement pour une prestation
  - Profil de consommation par tranche d'âge
- Médicaments :
  - Historique des remboursements
  - Détail selon les caractéristiques du bénéficiaire
  - Comparaison entre zones géographique, entre établissements prescripteurs
  - Comparaison entre deux médicaments, deux classes

Limites

- **Liées aux données médico-administratives :**
  - Nécessite une veille scientifique
  - Peut manquer de détails, notamment sur le plan clinique ou socio-démographique
  - Couverture des données : régimes / panier remboursable par la CNAM
  - Difficultés pour différencier hôpital privé du cabinet de ville
  - Médecin non identifiable dans l'hôpital public
- **Liées à l'Open Data :**
  - Manque de profondeur sur les bases
  - Règle d'anonymisation : moins de 10 bénéficiaires = NA
  - Fréquence de mise à jour et décalage temporel
  - Impossibilité de faire une répartition des montants pour une prestation
  - Format prédéfini par l'Assurance Maladie
- **Techniques :**
  - Manque de définitions et de dictionnaire des données
  - Volumétrie de certaines bases
  - Méthodologie pour traiter certaines problématiques

Conclusion & discussion

L'Open Data, c'est bien !

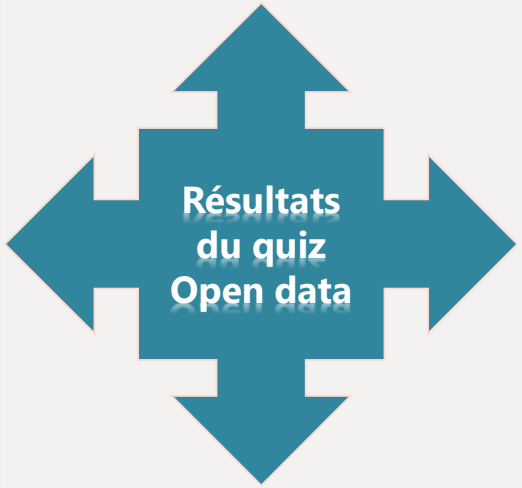
Les bases permettent de traiter un large panel de problématiques

Elles fournissent un apport complémentaire aux bases réglementaires, sans s'y substituer

L'avantage majeur de ces bases est leur accessibilité et leur libre exploitation

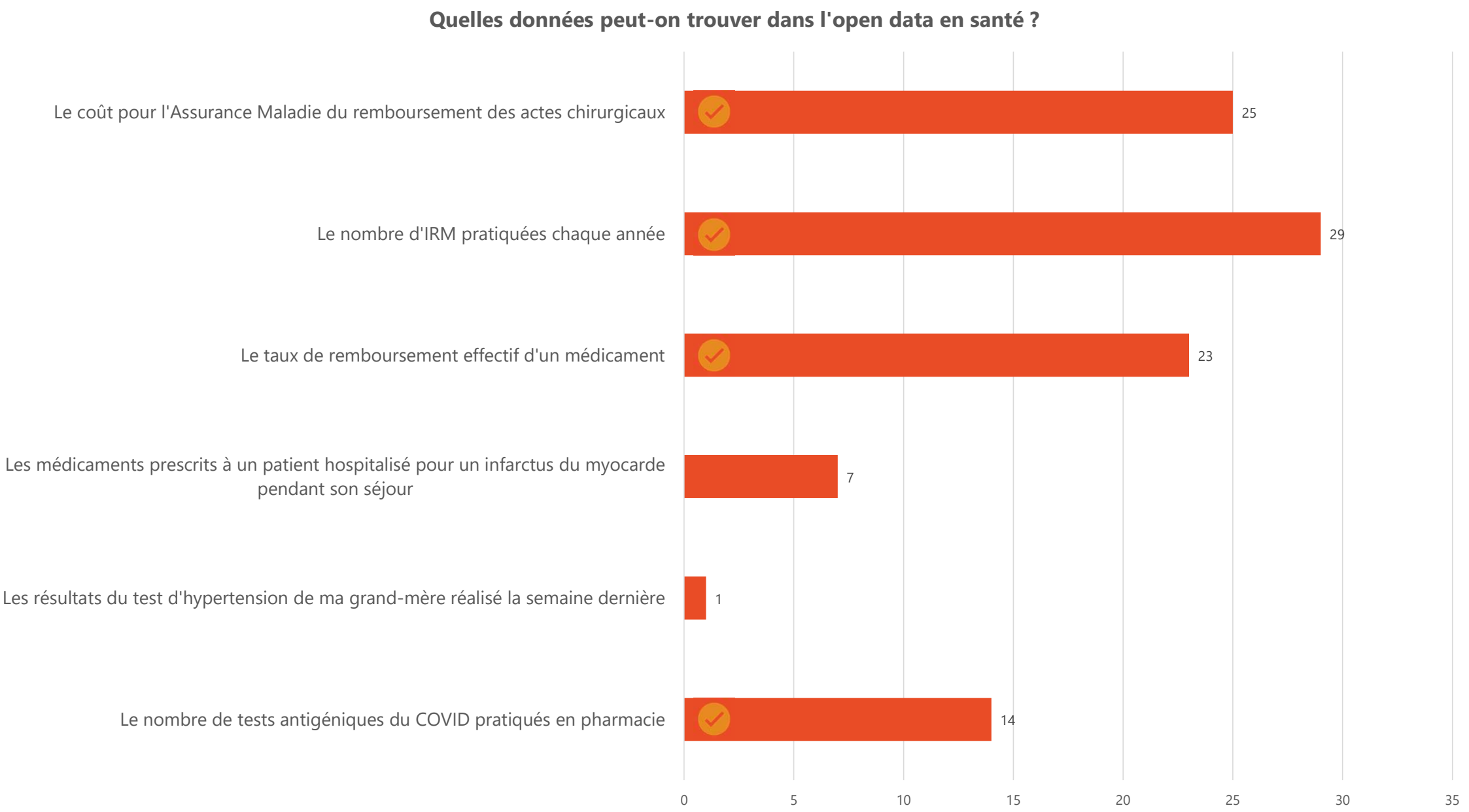
L'offre d'Open Data s'enrichit constamment.

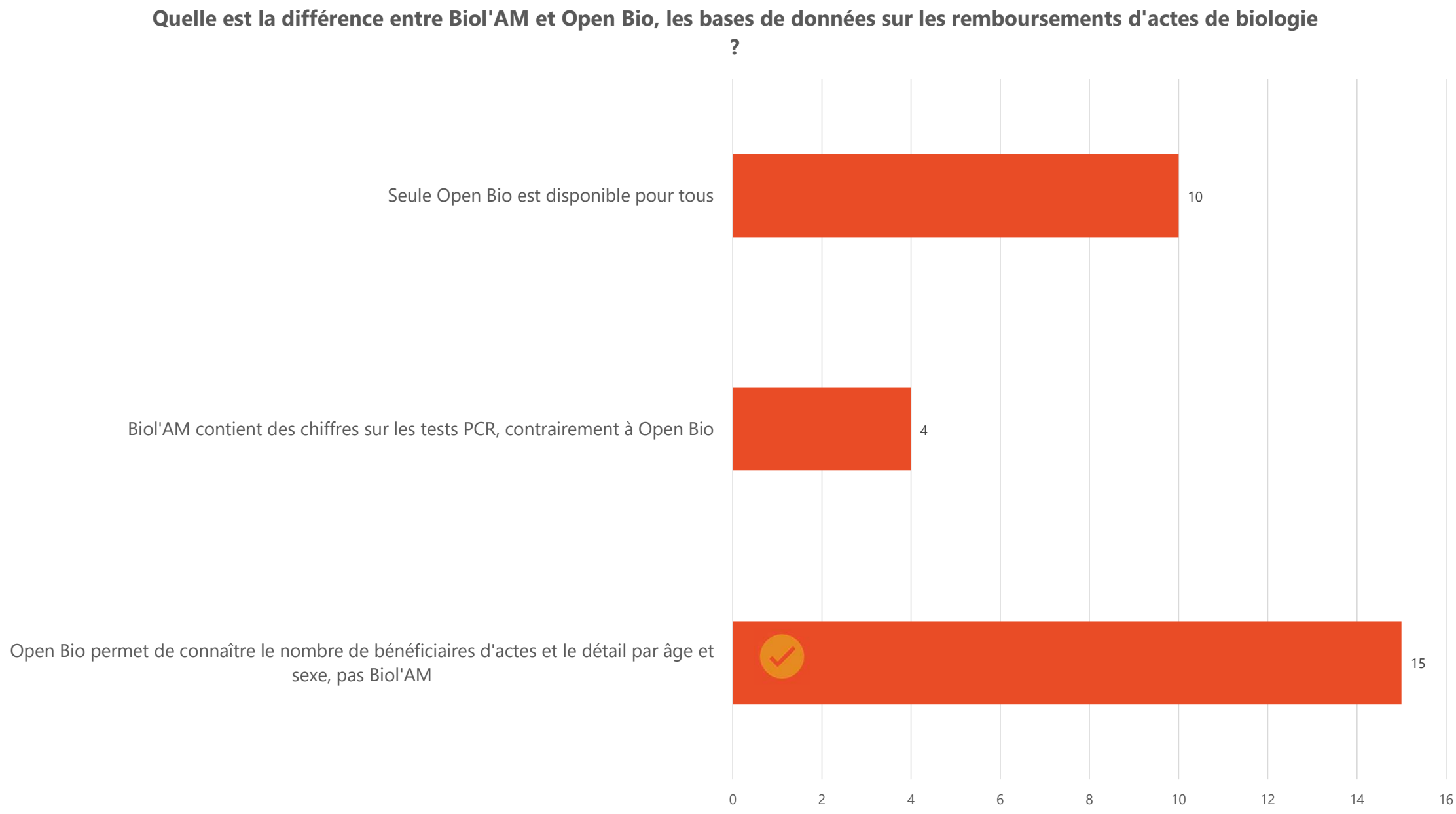
Merci de votre attention !

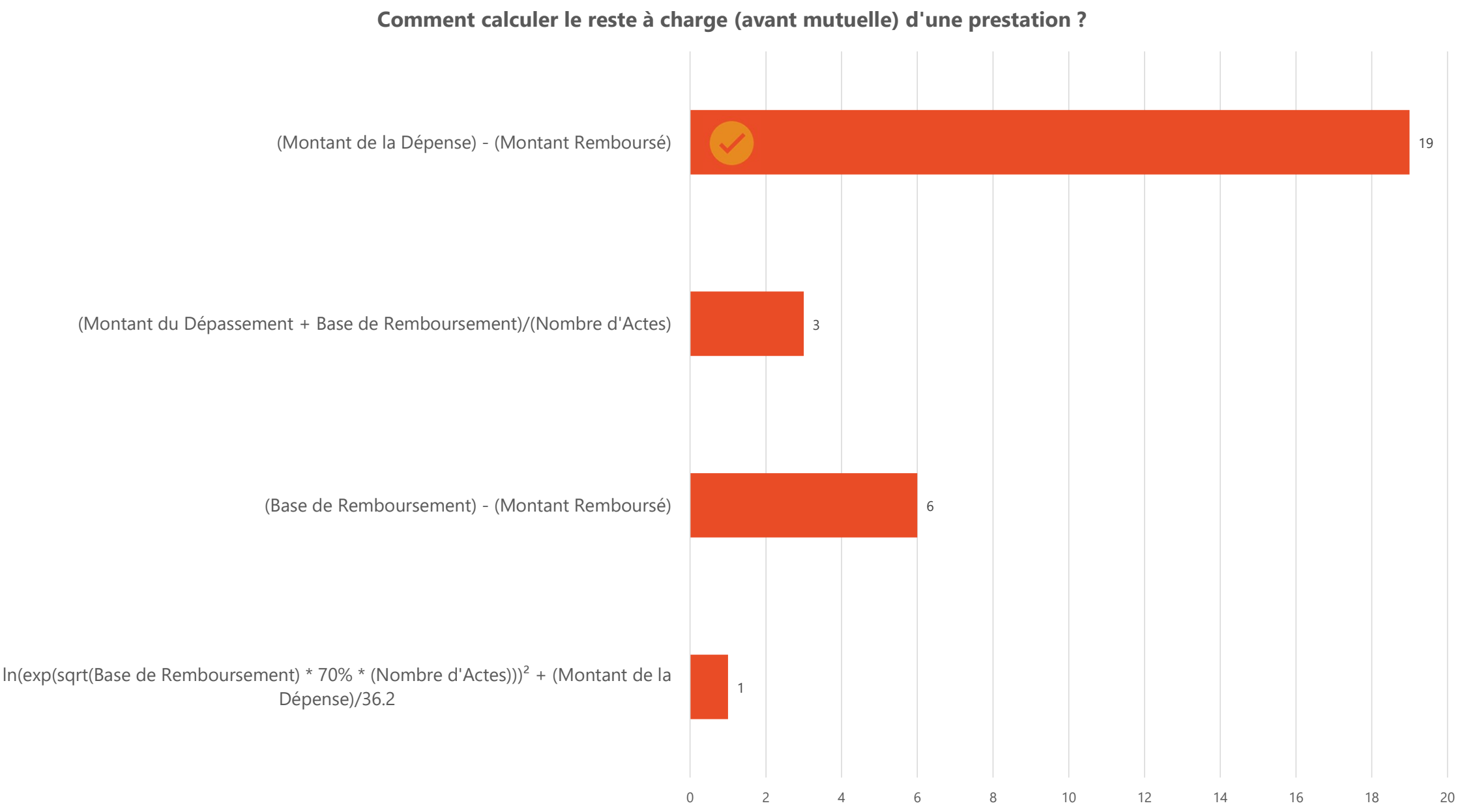


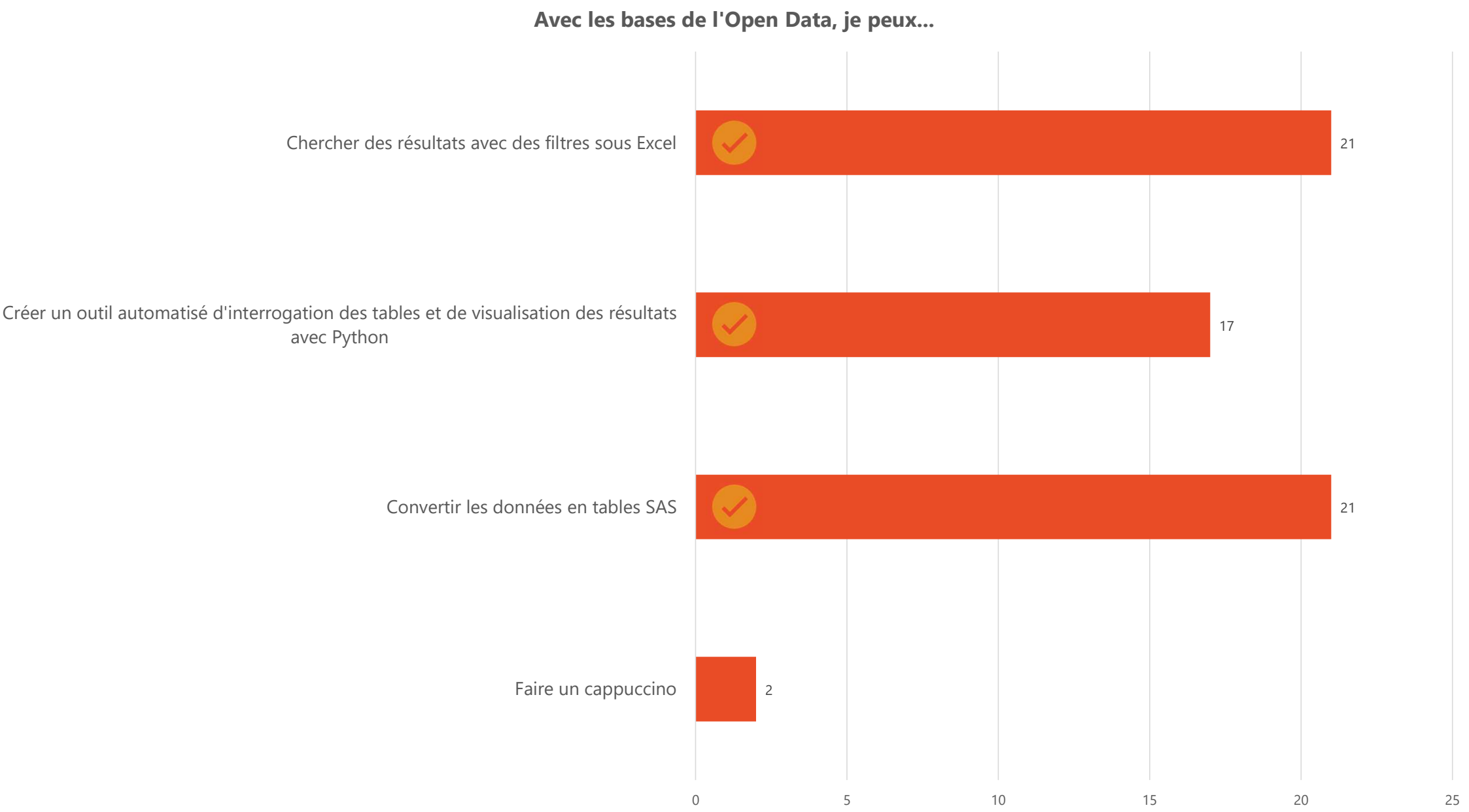
l'Open-Data *by* CEMKA

Résultats du quiz


Quelles données peut-on trouver dans l'open data en santé ?
Le coût pour l'Assurance Maladie du remboursement des actes chirurgicaux
25
Le nombre d'IRM pratiquées chaque année
29
Le taux de remboursement effectif d'un médicament
23
Les médicaments prescrits à un patient hospitalisé pour un infarctus du myocarde pendant son séjour
7
Les résultats du test d'hypertension de ma grand-mère réalisé la semaine dernière
1
Le nombre de tests antigéniques du COVID pratiqués en pharmacie
14
0
5
10
15
20
25
30
35


Quelle est la différence entre Biol'AM et Open Bio, les bases de données sur les remboursements d'actes de biologie ?
Seule Open Bio est disponible pour tous
10
Biol'AM contient des chiffres sur les tests PCR, contrairement à Open Bio
4
Open Bio permet de connaître le nombre de bénéficiaires d'actes et le détail par âge et sexe, pas Biol'AM
15
0
2
4
6
8
10
12
14
16


Comment calculer le reste à charge (avant mutuelle) d'une prestation ?
(Montant de la Dépense) - (Montant Remboursé)
19
(Montant du Dépassement + Base de Remboursement)/(Nombre d'Actes)
3
(Base de Remboursement) - (Montant Remboursé)
6
ln(exp(sqrt(Base de Remboursement) * 70% * (Nombre d'Actes)))² + (Montant de la Dépense)/36.2
1
0
2
4
6
8
10
12
14
16
18
20


Avec les bases de l'Open Data, je peux...
Chercher des résultats avec des filtres sous Excel
21
Créer un outil automatisé d'interrogation des tables et de visualisation des résultats avec Python
17
Convertir les données en tables SAS
21
Faire un cappuccino
2
0
5
10
15
20
25