PLATEFORME DES DONNÉES DE SANTÉ
FRENCH HEALTH DATA HUB



Health Data Hub

Guide pédagogique

Savoir cibler sa population d'étude dans la base principale du Système National des Données de Santé (SNDS)

Remerciements

Ce guide pédagogique a été réalisé en partenariat avec la Direction du Numérique (DNUM) du Ministère de la Santé, des Familles, de l'Autonomie et des Personnes Handicapées et a bénéficié de l'expertise de **Jérôme Brocca**, Chef de projet Données de santé, en particulier pour la construction du jeu de données et du cas pratique.

Sommaire de la formation

1. Contexte introductif
    - Définitions des concepts abordés
    - Illustration : cas d'un bénéficiaire ayant plusieurs BEN_NIR_PSA
    - Glossaire
2. Identifier les critères de ciblage dans la base principale du SNDS
3. Exemples illustrés de ciblage d'individus
    - Ciblage d'individus selon leur âge au cours d'une période donnée, avec collecte de leurs données socio-démographiques
    - Ciblage d'individus atteints de la maladie de Crohn et de rectocolite hémorragique
4. Exercices pratiques



Références



Cette formation se base notamment sur :

- Les ressources documentaires de la Caisse Nationale d'Assurance Maladie (CNAM), partenaire institutionnel ;
- Le Module 1 du MOOC SNDS *"Présentation du SNDS, des identifiants et des variables sociodémographiques"* ;
- Les ressources pédagogiques du site sur la documentation collaborative du SNDS ;
- Les guides portant sur les nomenclatures présentes dans la base principale du SNDS ;
- Le *"Guide d'utilisation du Système National des Données de Santé pour la surveillance et l'observation épidémiologiques (juillet 2024)"* de Santé Publique France (SpF).

Présentation de l’objet du module

- L’objet de ce module est d’**introduire les notions de base nécessaires** aux utilisateurs de la base principale* du Système National des Données de Santé (SNDS) afin de les **accompagner dans le ciblage d’une population d’étude** et de **renforcer leurs compétences** en la matière. Il propose des outils permettant aux utilisateurs d'identifier une population au cours de leurs travaux.

- Ce module se divise en 2 volets :
  - **un guide** qui présente **la partie théorique** de ce module et **deux exemples illustrés** (avec le programme SAS associé, adapté pour des données restituées **dans le cadre d’un accès permanent****)
  - **deux exercices pratiques** (disponibles via ce lien) à réaliser de façon autonome et qui permettront aux utilisateurs de s’exercer à manipuler une base de données synthétiques de la base principale du SNDS, telle qu’elle est restituée **dans le cadre d’un accès permanent****, comprenant des données fictives du Datamart de Consommation Inter-Régimes (DCIR) et du Programme de Médicalisation du Système d’Information (PMSI) reproduisant fidèlement la structure et le réalisme des données.

Pour des données extraites **dans le cadre d’un accès sur projet**, c’est la CNAM qui réalise l’extraction des données et donc le ciblage

* La base principale du SNDS comprend les données du SNDS historique (données du DCIR, du PMSI, du CépiDC, données de handicap de la CNSA) et les données des bases exhaustives appariées à la base historique (données de dépistage COVID-19 SIDEP et de vaccination COVID-10 VAC-SI).

** Les noms des variables présentés dans les exercices illustrés et les exercices pratiques sont ceux figurant dans les données en accès permanent. Certaines différences peuvent exister avec le nom des variables dans le cadre d’un accès sur projet.



1. Contexte introductif

Les concepts abordés dans ce guide

Ciblage

- Le ciblage est une **méthode d'identification d'un ensemble d'individus présentant certaines caractéristiques**, dans le cadre de la réalisation d'une étude
- Les **caractéristiques** (ou critères de sélection) des individus **varient selon les projets et leurs objectifs**
- La **nature** des caractéristiques peut être :
    - **Socio-démographique :** âge, sexe, lieu de résidence ou des soins, régime d'assurance maladie…
    - **Liée à la présence de pathologies :** via les codes Classification Internationale des Maladies - via les codes de Classification Internationale des Maladies - 10ème révision (CIM-10) des diagnostics d'hospitalisations ou des ALD, ou si la pathologie est repérable via les délivrances de médicaments spécifiques de celle-ci avec les codes CIP (Code Identifiant de Présentation), UCD (Unité Commune de Dispensation) ou classe ATC (Anatomique, Thérapeutique et Chimique), etc…
    - **Liée à la délivrance/réalisation de soins :** via la nature de prestation des soins de ville, codes CIP, UCD ou classe ATC pour les délivrances de médicaments, code Classification Commune des Actes Médicaux (CCAM) pour les actes médicaux, code Nomenclature affinée des Actes de Biologie Médicale (NABM) pour les actes de biologies, code Liste des Produits et Prestations remboursables (LPP) pour les dispositifs médicaux, etc…
    - **Autres :** temporelle,…

Population d'étude

- **Ensemble d'individus sélectionnés selon des critères d'inclusion et d'exclusion préalablement définis,** et sur lesquels porteront les analyses dans le cadre d'un protocole de recherche.

Exemple *: Les sujets de la population d'étude devront présenter les critères d'inclusion suivants :*

- *Etre âgé de 18 ans ou plus au 1er janvier 2015 ;*
  ET
- *Présenter au moins une délivrance remboursée de médicaments pour la pathologie X (code ATC = …) entre le 1er janvier 2018 et le 31 décembre 2022 ;*
  OU
- *Présenter une Affection de longue durée (ALD) ou invalidité pour la pathologie X (code CIM-10 = …) entre le 1er janvier 2015 et le 31 décembre 2022 ;*
  OU
- *Présenter au moins une hospitalisation en MCO pour la pathologie X (code CIM-10 = … diagnostic principal, relié ou associé) entre le 1er janvier 2015 et le 31 décembre 2022.*

Et présenter le critère d'exclusion suivant :

- *Ne pas avoir eu de délivrance de médicaments pour la pathologie X (code ATC = …) avant le 1er janvier 2018.*

Focus - Notions autour du ciblage d'une population d'étude

Critère de ciblage

- **Règles d'inclusion (et d'exclusion) permettant de repérer les individus de la population d'étude** au sein de la population couverte par la base principale du SNDS
- **Nomenclatures et codes reliés à ces critères précisés le cas échéant**

Période de ciblage

- **Fenêtre temporelle** au cours de laquelle **les critères de ciblage sont appliqués** pour sélectionner la population d' étude
- Peut être **différente en fonction des critères** *(Ex : sélection des sujets avec diagnostic de maladie XX en 2020 et au moins 1 délivrance d'un médicament YY en 2022-2023)*

Articulation des critères

- **Règles d'imbrication des critères de ciblage**
- Typiquement s'il faut considérer un "ET" ou un "OU" entre les critères et éventuellement un ordre associé

Période d'extraction

- **Période sur laquelle seront extraites l'ensemble des données nécessaires** à la réalisation d'une étude pour la population préalablement ciblée

Extraction

- **Sous-ensemble de données**, à un instant *t* issu de la base principale du SNDS

Nomenclatures

- **Termes ou codes pour identifier, structurer et interpréter** de façon normalisée des **informations médicales** (diagnostics, procédures, traitements, médicaments ...) **ou médico-administratives** (type de soin, motif d'exonération du ticket modérateur ...)
- Facilitent la **communication**, la **documentation**, la **facturation** et l'**analyse des données de santé**

Conseil : Il est parfois nécessaire de tenir compte d'une **chronologie d'apparition entre les différents codes d'intérêt** ainsi que de **l'évolution des codes au cours de la période d' étude**

Focus - Notions autour du patient (1/2)

Numéro d'inscription au Répertoire (NIR)

Identification unique donnée à chaque personne à la naissance par l'INSEE sur la base d'éléments d'état civil transmis par les mairies, **et identique tout au long de la vie**, que la personne soit "assurée" ou "ayant-droit".

➔ Communément intitulé **"Numéro de sécurité sociale"**

➔ Le NIR est présent **sous forme pseudonymisée dans le SNDS (DCIR et CépiDc) dans la variable BEN_NIR_ANO**

Pseudo-NIR du bénéficiaire

Identifiant créé pour le SNIIRAM, à la fois dans les bases de l'Assurance Maladie et dans le PMSI, qui correspond à **la pseudonymisation de la concaténation du Numéro de l'assuré** (NIR), **de la date de naissance et du sexe du bénéficiaire** (variable dans DCIR : BEN_NIR_PSA et dans PMSI : NIR_ANO_17)

Assuré

Personne affiliée à un régime d'Assurance Maladie et **sous le nom duquel le remboursement du soin est effectué**, qu'elle soit la bénéficiaire de ce soin ou non

➔ Appelé aussi **"l'ouvreur de droit"** (variable dans DCIR : ASS_NIR_ANO)

Bénéficiaire

Personne qui reçoit le soin, qu'elle soit "assurée" ou "ayant-droit"

Ayant-droit

Personne qui n'a pas de droits ouverts à son nom, mais des droits ouverts sous le compte d'un assuré (appelé "ouvreur de droit")

Exemple : enfant ayant-droit de son père et/ou de sa mère

Focus - Notions autour du patient (2/2)

Pseudonymisation

C'est un **codage irréversible des identifiants des bénéficiaires** réalisé dans la base principale du SNDS par technique de hachage

C'est également un traitement de données à caractère personnel qui consiste à **remplacer les informations "directement" identifiantes** (comme le NIR, etc.) **par des identifiants artificiels**, de façon à rendre une personne non directement identifiable sans l'utilisation d'informations supplémentaires

Il s'agit d'un processus moins fort que le **processus d'anonymisation** qui vise à **empêcher totalement** la possibilité d'identifier une personne même avec des informations supplémentaires

Illustration : bénéficiaire ayant plusieurs BEN_NIR_PSA (1/3)

Rappel : exemples de contenu du NIR *(fictifs)*

2	98	01	31	155	158
Sexe (1 pour un homme, 2 pour une femme)	*Année de naissance*	*Mois de naissance*	*Département de naissance*	*Code commune de naissance*	*N° ordre de naissance (valeur numérique sur 3 chiffres)*
1	**84**	**12**	**75**	**123**	**267**
Sexe (1 pour un homme, 2 pour une femme)	*Année de naissance*	*Mois de naissance*	*Département de naissance*	*Code commune de naissance*	*N° ordre de naissance (valeur numérique sur 3 chiffres)*

Contextualisation du cas illustré



Adrien est un enfant né le 12 juin 2015 à Metz. Il est bénéficiaire de soins puisque c'est lui qui consulte et bénéficie des soins. Or, les enfants n'ont pas de droits ouverts à leur nom auprès de l'Assurance Maladie et ne disposent souvent pas de Carte vitale. Leurs droits sont généralement ouverts par l'intermédiaire d'un ou de leurs deux parents. L'enfant est donc un "ayant-droit" de ses parents (ou d'un de ses parents).

De ce fait, **une seule et même personne peut avoir plusieurs pseudo-NIR (BEN_NIR_PSA) différents** au cours de sa vie dans les données de la base principale du SNDS.

Module 1 du MOOC SNDS *"Présentation du SNDS, des identifiants et des variables sociodémographiques"*

Illustration : bénéficiaire ayant plusieurs BEN_NIR_PSA (2/3)

Père

→ Adrien est l'ayant-droit de son père, né le 3 avril 1975 à Strasbourg. Son père est l'ouvreur de droit (= assuré). Pour Adrien, le NIR de l'assuré ouvreur de droit sera celui de son père. Son pseudo-NIR prendra la forme suivante :

1 75 04 67 114 YYY	**20150612**	**1**
NIR du père d'Adrien	*Date de naissance d'Adrien (YYYYMMDD)*	*Code sexe d'Adrien*

= pseudo-NIR 1

Le 1er BEN_NIR_PSA d'Adrien sera composé de la concaténation du NIR de son père, de la date de naissance d'Adrien et de son sexe. La valeur sera pseudonymisée dans la base principale du SNDS, pas en clair *(Exemple (fictif) : BCoIiTHK2vh9VtpMU après hachage).*

Mère

→ Adrien peut être simultanément ayant-droit de sa mère, née à Metz le 25 décembre 1978. Sa mère est ouvreuse de droit (= assurée). Pour Adrien, le NIR de l'assuré ouvreur sera celui de sa mère. Son pseudo-NIR prendra la forme suivante :

2 78 12 57 387 ZZZ	**20150612**	**1**
NIR de la mère d'Adrien	*Date de naissance d'Adrien (YYYYMMDD)*	*Code sexe d'Adrien*

= pseudo-NIR 2

Le 2ème BEN_NIR_PSA d'Adrien sera composé de la concaténation du NIR de sa mère, de la date de naissance d'Adrien et de son sexe. La valeur sera pseudonymisée dans la base principale du SNDS, pas en clair *(Exemple (fictif) : CDmnrZVC8qj3MgdBD après hachage).*

Illustration : bénéficiaire ayant plusieurs BEN_NIR_PSA (3/3)

Adulte



→ **Quand Adrien devient adulte**, il devient son propre ouvreur de droit. Le NIR de l'assuré ouvreur de droit change et par conséquent, le BEN_NIR_PSA de la personne également. Il aura alors l'identifiant SNDS (pseudo NIR) issu de la concaténation des éléments suivants (qui seront pseudonymes dans la base principale du SNDS, pas en clair) :

1 15 06 57 555 XXX	**20150612**	**1**
NIR d'Adrien	*Date de naissance d'Adrien*	*Code sexe d'Adrien*

= pseudo-NIR 3

Exemple (fictif) : ZBtmoVFH6lp4AqoFZ, après hachage.

Dans la base principale du SNDS, **Adrien aura donc trois pseudo-NIR (BEN_NIR_PSA) différents** selon les périodes de sa vie :

- deux correspondant à son identité en tant qu'enfant (*BColiTHK2vh9VtpMU, CDmnrZVC8qj3MgdBD)*
- et un en tant qu'adulte *(ZBtmoVFH6lp4AqoFZ)*.

Il sera possible de les **relier grâce à un identifiant unique** qui sera le résultat de la pseudonymisation de son NIR dans la base principale du SNDS **(BEN_NIR_ANO)** ou via un **identifiant composite BEN_IDT_ANO** (*Exemple (fictif) : VIzqrYAN7bk2PdiRS, après hachage*) et **qui sera relié à tous ses BEN_NIR_PSA dans une table de correspondance**.

Focus - Les limites du pseudo-NIR

Les limites du pseudo-NIR (BEN_NIR_PSA)

1

Utilisé seul, il ne permet pas de distinguer les jumeaux de même sexe : dans les bases du DCIR, le rang de naissance du bénéficiaire (variable BEN_RNG_GEM) est ajouté, ce qui sert à distinguer deux jumeaux enfants de même sexe (Régime Général y compris SLM) ou des enfants distincts au sein d'une même famille (autres régimes) présentant une valeur de BEN_NIR_PSA identique.

Il est recommandé de dénombrer des patients **en concaténant les variables BEN_NIR_PSA || BEN_RNG_GEM**.

Attention : **dans le PMSI**, le rang de naissance n'est pas directement utilisable car moins fiable que dans le DCIR, l'identifiant NIR_ANO_17 seul (correspondant à BEN_NIR_PSA dans DCIR) est disponible. Le chaînage avec les données du PMSI se fait donc en utilisant uniquement NIR_ANO_17 <-> BEN_NIR_PSA.

2

Un même bénéficiaire peut présenter plusieurs pseudo-NIR différents au cours de sa vie :

- ➔ enfants présents sur la carte Vitale de leurs 2 parents
- ➔ bénéficiaire qui change de statut (enfant ayant-droit de son parent devenant étudiant assuré ouvreur de droit)
- ➔ étranger passant d'un NIR provisoire à un NIR définitif
- ➔ changement de statut marital en l'absence d'activité professionnelle (< 2016)
- ➔ décès du conjoint(e) qui couvrait son épouse/mari (< 2016)

Nécessité de relier tous les pseudo-NIR d'un bénéficiaire à son NIR unique BEN_NIR_ANO (ou à BEN_IDT_ANO)

Récupération des identifiants uniques (1/2)

- Le **NIR du bénéficiaire BEN_NIR_ANO** est bien remonté dans la base principale du SNDS
  - Mais **reste encore non alimenté exhaustivement dans les flux DCIR pour certains régimes d'Assurance Maladie***
  - Et **n'est pas encore disponible dans les bases du PMSI** (i.e. n'est pas disponible si le bénéficiaire n'est présent que dans le PMSI)
  - Cependant, à partir de 2016, il est renseigné à plus de 97% pour tous les régimes*.

  ➔ De ce fait, certains BEN_NIR_PSA **n'auront pas de valeur dans BEN_NIR_ANO.**

- Depuis 2016, un identifiant composite du bénéficiaire **BEN_IDT_ANO** est proposé dans IR_BEN_R. Celui-ci prend la valeur de :
  - BEN_NIR_ANO, **lorsque ce dernier est renseigné** (associé avec la variable top BEN_IDT_TOP = 1)
  - BEN_NIR_PSA II BEN_RNG_GEM, **lorsque BEN_NIR_ANO est vide** (associé avec la variable top BEN_IDT_TOP = 0).

  ➔ L'identifiant BEN_IDT_ANO est un identifiant créé pour « faciliter le travail » de récupération des identifiants uniques d'un même bénéficiaire étant donné que contrairement au BEN_NIR_ANO, il est renseigné pour tous les bénéficiaires. Il permet directement de regrouper les informations d'un même bénéficiaire - néanmoins, **il existera tout de même des BEN_IDT_ANO différents associés à une même personne** (lorsque BEN_NIR_ANO n'est pas renseigné pour cette dernière).

* Dans ce guide, la complétude de la variable est référencée par régime et année (page 56)

Récupération des identifiants uniques (2/2)



Dans le cadre du ciblage d'un groupe de bénéficiaires répondant aux critères de sélection définis, il faudra d'abord identifier les bénéficiaires à partir de leur pseudo-NIR BEN_NIR_PSA et du rang de naissance BEN_RNG_GEM associé, en lien avec ces critères, puis récupérer tous les autres pseudo-NIR et rang de naissance de ces individus afin de pouvoir reconstituer la totalité de leur consommation de soins sur la période d'étude considérée, y compris sur les périodes antérieures ou postérieures au ciblage initial.

1

Étape 1 : Ciblage initial

Critères de sélection : pathologie CIM-10, acte CCAM, médicament CIP…

↓

Sélection de bénéficiaires (DCIR et/ou PMSI)

↓

Liste d'identifiants NIR_ANO_17 *(si issus du PMSI)* et de **couples BEN_NIR_PSA||BEN_RNG_GEM** *(si issus de DCIR)* **liés aux critères**

2

Étape 2 : Rattachement identifiant unique

Relier **NIR_ANO_17** *(issu du PMSI)* à **BEN_NIR_PSA (et BEN_RNG_GEM)** via IR_BEN_R et IR_BEN_R_ARC *

↓

Puis, pour chaque couple BEN_NIR_PSA||BEN_RNG_GEM *(issus de DCIR et PMSI)* **:**

→ Rechercher BEN_IDT_ANO correspondant dans IR_BEN_R et IR_BEN_R_ARC

↓

1 BEN_NIR_PSA||BEN_RNG_GEM ↔ 1 BEN_IDT_ANO

** si un NIR_ANO_17 est lié à plusieurs BEN_RNG_GEM (= jumeaux), il vaut mieux supprimer ces sujets - au choix)*

3

Étape 3 : Extension à tous les pseudo NIR du bénéficiaire

Pour chaque BEN_IDT_ANO distinct retrouvé :

→ Récupérer tous les couples BEN_NIR_PSA||BEN_RNG_GEM associés dans IR_BEN_R et IR_BEN_R_ARC

4

Etape 4 : Reconstitution des consommations

Extraire dans DCIR / PMSI l'ensemble de leurs consommations de soins au cours de la période d'étude considérée pour tous les BEN_NIR_PSA||BEN_RNG_GEM associés.

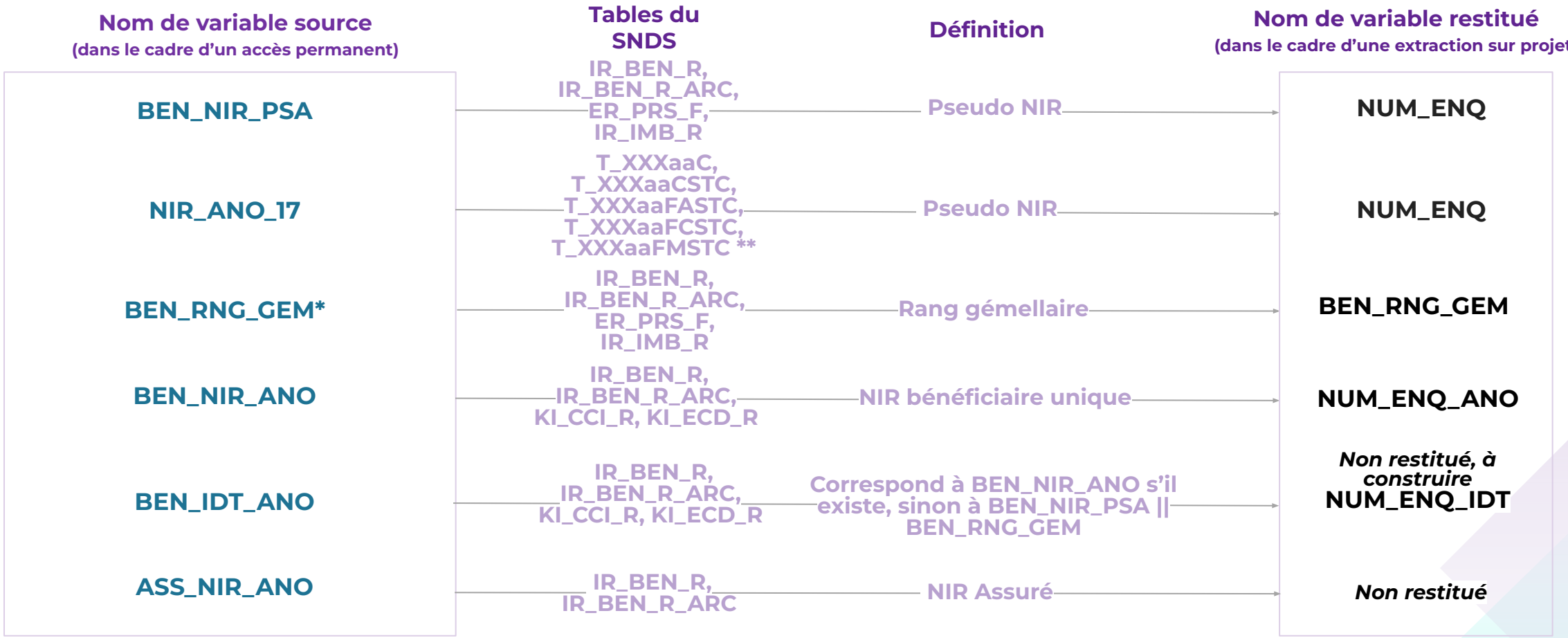
→ A partir des tables ER_PRS_F, T_MCOxxB…

→ Permet de reconstituer le parcours complet du bénéficiaire.

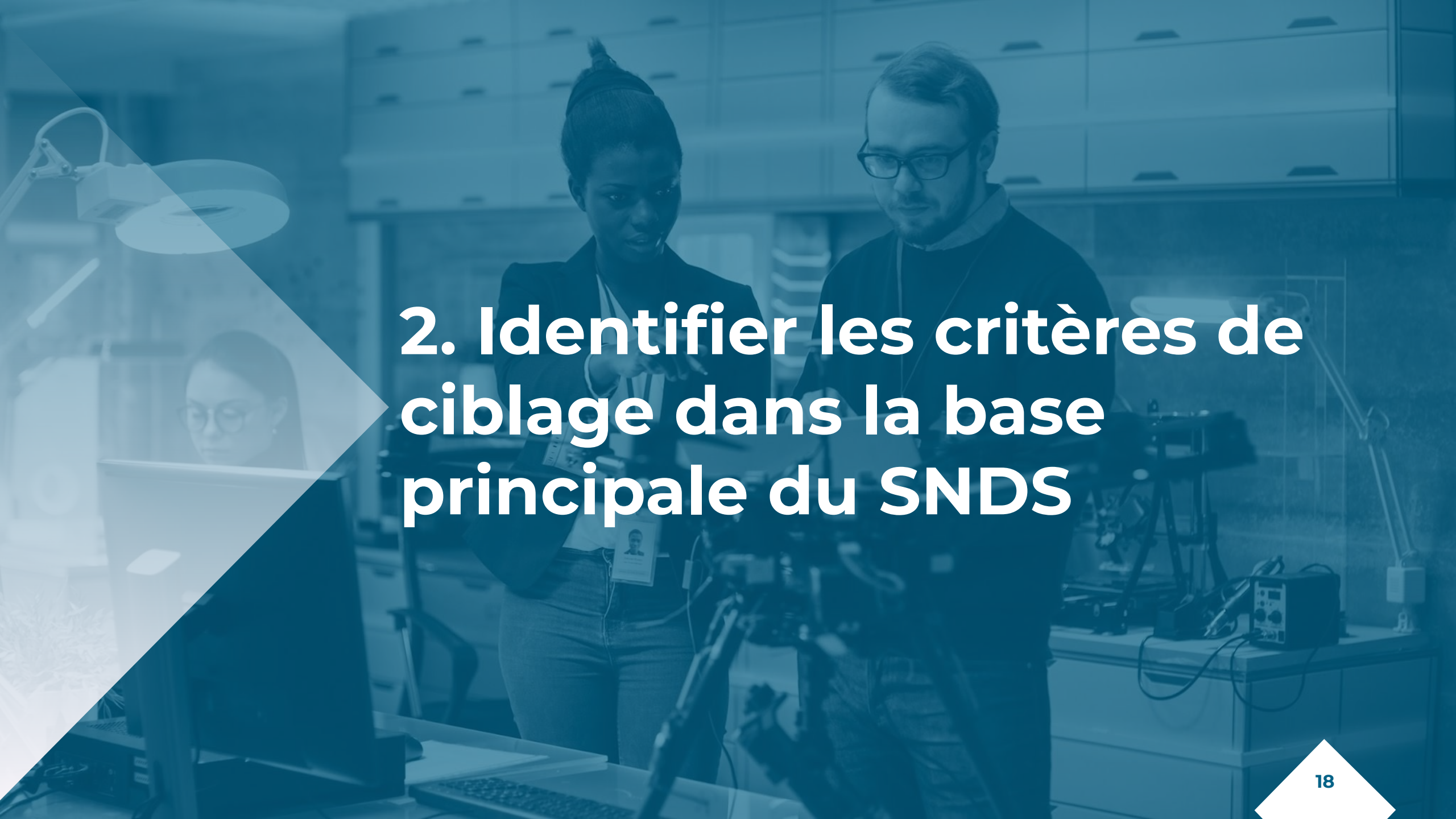


Vous trouverez plus de détails sur ces étapes (notamment techniques, avec les programmes associés) dans la présentation de la CNAM « les Clés du SNDS » sur : *"La méthodologie de ciblage de tous les identifiants patients d'une cohorte dans le SNDS : exemple d'une cohorte de patients diabétiques"*

Glossaire des noms des variables permettant de qualifier les bénéficiaires dans la base principale du SNDS



* Rang gémellaire pour le Régime Général (y compris SLM) OU, pour les autres régimes, le rang pour différencier les bénéficiaires rattachés au même ouvreur de droit.
** Tables du PMSI : *XXX* = MCO, SSR, HAD ou RIP (si disponible) ; *aa* = année de fin du séjour



2. Identifier les critères de ciblage dans la base principale du SNDS

Afin de cibler une population, il existe plusieurs étapes génériques à réaliser dans le cadre de son étude

1

Définir au préalable **les critères de ciblage :**

- Pour les accès sur projet : dans le formulaire d'expression des besoins (EDB) pour être pris en charge par le pôle DEMEX de la Cnam qui va livrer l'extraction associée ;
- Pour les accès permanents : le ciblage est réalisé directement sur les données exhaustives

Exemple : les sujets adultes diabétiques dans la base principale du SNDS en 2024, les enfants traités par double immunothérapie sur la période 2018-2024…

2

Identifier **les variables dans la base principale du SNDS** qui sont reliées à ces critères et leurs modalités d'intérêt pour pouvoir réaliser le ciblage

Exemple : calcul de l'âge via l'année de naissance BEN_NAI_ANN dans IR_BEN_R, OU sélection des diagnostics de diabète dans le PMSI et ALD via les codes CIM-10 E10 et E11, OU sélection des délivrances de médicaments antidiabétiques avec ATC = A10xx

3

Ordonner **la combinaison des critères de sélection** (ET, OU, ordre, …)

Exemple : les sujets doivent avoir 18 ans et plus en 2024 ET répondre à l'un des critères suivants : [présenter au moins 1 code CIM-10 (E 11 ou E10) via les diagnostics du PMSI ou d'ALD en 2024] OU [avoir eu au moins 3 délivrances de médicaments de la classe ATC A10xx en 2024]

4

Sélectionner **les observations dans la base de données**, en lien avec les critères de ciblage

5

Récupérer **les identifiants des sujets concernés** par les critères de ciblage dans la période de ciblage puis dans toute la période d'extraction d'intérêt

Critères présents dans la base principale du SNDS et documents associés



Socio-démographiques

Age, sexe, mois et année de naissance, date de décès, commune de résidence, régime d'assurance maladie, type de prestations de soins ...



Socio-économiques

Affiliation à la CSS (ou C2S, Complémentaire Santé Solidaire) ou à l'AME (Aide Médicale d'Etat), indice français de désavantage social (FDep - French Deprivation Index)...



Nomenclatures des codifications du DCIR

Documentation listant les codes des variables présentes dans le DCIR, liées aux organismes de prise en charge et d'affiliation, au décompte des soins, au bénéficiaire, au mode de prise en charge, à la prestation, au professionnel de santé

Exemple : nature de prestation, régime d'affiliation, motif d'exonération du ticket modérateur, spécialité du professionnel de santé exécutant / prescripteur, discipline de prestation ...



Nomenclatures médicales, classification GHM *(cf focus diapositive suivante)*

Nomenclature des pathologies *(Exemple : codes CIM-10 pour les diagnostics hospitaliers des maladies ou pour les ALD dans le cas de maladies chroniques à traitement prolongé)*

Nomenclature des délivrances et actes de soins *(Exemple : codes CIP, UCD ou ATC pour les médicaments, codes LPP pour les dispositifs médicaux, code CCAM pour les actes médicaux, code NABM pour les actes de biologie, etc.)*

Classification des séjours hospitaliers MCO *(groupage GHM)*

Focus - Nomenclatures médicales et classification des séjours MCO dans la base principale du SNDS

*La connaissance des nomenclatures est essentielle pour **caractériser une population au sein de la base principale du SNDS** ainsi que **d'exprimer ses besoins en termes de codes** lors des demandes de ciblage de population et d'extraction de données. Les principales nomenclatures et la classification des séjours hospitaliers MCO se déclinent comme suit :*

Nomenclatures des délivrances de soins

- **Classification des médicaments**
  - ATC *(Anatomique, thérapeutique et chimique)*
  - UCD *(Unité Commune de Dispensation)*
  - CIP *(Code Identifiant de Présentation)*
- **Dispositifs médicaux**
  - LPP *(Liste des produits et prestations)*

Nomenclatures des actes

- **Actes professionnels**
  - NGAP *(Nomenclature générale des actes professionnels)*
- **Actes médicaux**
  - CCAM *(Classification commune des actes médicaux)*
- **Actes de biologie médicale**
  - NABM *(Nomenclature affinée des Actes de biologie médicale*

Nomenclatures des pathologies

- **Maladies, symptômes, diagnostics**
  - CIM-10 *(Classification internationale des maladies, 10ème révision)*
- **Maladies chroniques à traitement prolongé**
  - N° ALD *(Affection de longue durée)*

Classification des séjours hospitaliers

- **PMSI Médecine, Chirurgie, Obstétrique (MCO)**
  - GHM *(Groupe homogène de malades)*



Pour chaque nomenclature, un guide présentant l'objet de celle-ci, sa structure, les variables en lien dans la base principale du SNDS ainsi que des cas d'usage, est disponible sur le site du HDH.



3. Exemples illustrés de ciblage d’individus

Exemple illustré 1 : Ciblage d'individus selon leur âge au cours d'une période donnée, avec collecte de leurs données sociodémographiques (1/4)

1

- A partir du référentiel des bénéficiaires IR_BEN_R, **sélectionner l'ensemble des bénéficiaires adultes en 2023** et présents dans le DCIR en 2023 (au moins un jour)

2

- **Extraire les données socio-démographiques** (âge, sexe, département, date de décès et organisme d'affiliation) **des bénéficiaires adultes en 2023**, par la constitution d'une table avec une seule ligne par bénéficiaire

3

- **Créer une table de correspondance entre les différents pseudo-NIR d'un bénéficiaire** au cours de sa vie (BEN_NIR_PSA+BEN_RNG_GEM) avec l'identifiant composite bénéficiaire (BEN_IDT_ANO)

Penser à enlever les patients non consommants à l'aide des variables MAX_TRT_DTD (date la plus récente des prestations perçues par un bénéficiaire) et BEN_DTE_INS (date d'insertion de l'individu dans le référentiel des bénéficiaires)

Exemple illustré 1 : Ciblage d'individus selon leur âge au cours d'une période donnée, avec collecte de leurs données socio-démographiques (2/4)

Le programme SAS associé est le suivant (adapté pour des données restituées dans le cadre d'un accès permanent) :

1

```
PROC SQL;
CREATE TABLE table_patients_tmp1 AS
SELECT *
FROM IR_BEN_R

/* Adultes en 2023 (exclusion des patients qui auraient plus de 115 ans en 2023) */
WHERE BEN_NAI_ANN <= '2005' AND BEN_NAI_ANN >= '1908'; /* AND BEN_NAI_ANN NOT IN ('0001', '1600') */

/* Retrait des bénéficiaires sans prestation de soins depuis 2023 */
AND (MAX_TRT_DTD >= '2023-01-01'

/* Retrait des non-consommants du RG insérés dans les 2 dernières années avant 2023*/
OR (MAX_TRT_DTD = '1600-01-01' AND BEN_DTE_INS >= '2021-01-01'))

/* Bénéficiaires en vie au début de la période d'étude */
AND (BEN_DCD_AME >= '202301' OR BEN_DCD_AME = '160001')

QUIT;
```

Exemple illustré 1 : Ciblage d'individus selon leur âge au cours d'une période donnée, avec collecte de leurs données socio-démographiques (3/4)

2

```
/* Sélection des variables socio-démographiques à retenir et tri sur les variables d'enregistrement et de mise à jour des données (les données les plus récentes seront classées en dernier) */
PROC SQL;
CREATE TABLE table_patients_tmp2 AS
SELECT BEN_IDT_ANO, BEN_NAI_ANN, (2023 - BEN_NAI_ANN) AS age_en_2023, BEN_SEX_COD, BEN_DCD_AME, BEN_RES_DPT, SUBSTR(ORG_AFF_BEN,1,3) AS code_regime, MAX_TRT_DTD
FROM table_patients_tmp1
ORDER BY BEN_IDT_ANO, MAX_TRT_DTD, BEN_DTE_MAJ;
/* Ajout d'un identifiant de ligne */
ALTER TABLE table_patients_tmp2
ADD COLUMN id INT AUTO_INCREMENT PRIMARY KEY;

/* Sélection de l'information la plus récente (i.e. de la dernière ligne, donc de l'identifiant de ligne le plus élevé) */
CREATE TABLE table_patients AS
SELECT *
FROM table_patients_tmp2
WHERE id IN
   (SELECT MAX(id)
   FROM table_patients_tmp2
   GROUP BY BEN_IDT_ANO);
QUIT;
```

Exemple illustré 1 : Ciblage d'individus selon leur âge au cours d'une période donnée, avec collecte de leurs données socio-démographiques (4/4)

3

```
/* Création d'une table de correspondance entre les différents pseudo-NIR des bénéficiaires adultes en 2023 au cours de leur vie (BEN_NIR_PSA+BEN_RNG_GEM) avec BEN_IDT_ANO pour relier plus facilement les patients à leur consommation de soins (non fait ici)*/
PROC SQL;
CREATE TABLE table_corresp_PSA_RNG_avec_ANO AS
SELECT DISTINCT BEN_IDT_ANO, BEN_NIR_PSA, BEN_RNG_GEM
FROM table_patients_tmp1;
QUIT;
```

Exemple illustré 2 : Ciblage d'individus atteints de la maladie de Crohn et de rectocolite hémorragique (1/2)

Afin de vous préparer au mieux à l'exercice pratique qui vous est proposé en volet 2 de ce module :

Nous vous suggérons **un nouvel exemple illustré portant sur un algorithme plus complexe** que celui présenté dans l'exemple illustré précédent, représentatif de ce que vous allez réaliser en volet 2.

Cet algorithme de ciblage figure dans la **Bibliothèque ouverte d'algorithmes en santé (BOAS)** du Health Data Hub. Il est programmé en code SAS et a pour objectif de **cibler** et de **caractériser une population atteinte des maladies inflammatoires chroniques de l'intestin (MICI)** dans la base principale du SNDS en 2021, pour obtenir un état des lieux des MICI et de leur prise en charge en France en 2021.

Vous trouverez l'énoncé de cet exemple illustré ici ainsi que le programme ici.

Exemple illustré 2 : Ciblage d'individus atteints de la maladie de Crohn et de rectocolite hémorragique (2/2)

1

Étape 1 : ALD

- Code CIM-10 **K50 ou K51**
- **ALD actives en 2021** (début d'ALD <= 2021 et fin d'ALD > 2020 ou en 1600
- **Suppression ALD trop anciennes** (date MAJ+5 ans avant 2021)
- Diagnostic **le plus récent**
- **Identifiants anonymisés** (BEN_NIR_PSA et CTO_IDT_ANO via IR_BEN_R)

+

2

Étape 2 : PMSI MCO

- Séjours **avec DP, DR ou DAS = K50 ou K51** en 2017-2021
- **Suppression séjours** en doublons ou erreur
- Diagnostic **le plus récent**
- **Identifiants anonymisés** (NIR_ANO_17 et CTO_IDT_ANO)

+

3

Étape 3 : PMSI SSR

- Séjours **avec MP, AE ou DAS = K50 ou K51** en 2021
- **Suppression séjours** en erreur
- Diagnostic **le plus récent**
- **Identifiants anonymisés** (NIR_ANO_17 et CTO_IDT_ANO)

4

Etape 4 : Fusion des 3 sources + récupération données

Pour chaque individu :

- **Diagnostic** le plus récent
- **Calcul date de référence globale** (la plus récente des 3 sources)
- Données **socio-démographiques**
- Données de **soins spécifiques** en 2021

Dans cet exemple, l'identifiant **CTO_IDT_ANO** utilisé n'a été présent que sur une durée restreinte dans les données, notamment de cartographie. **Nous vous conseillons de le remplacer par BEN_IDT_ANO.** Pour précision, CTO_IDT_ANO prend la valeur de :

- BEN_NIR_ANO II 'U' (lettre 'U' pour Unique), lorsque BEN_NIR_ANO est renseigné
- ASS_NIR_ANO II 'U', lorsque BEN_NIR_ANO est vide et la qualité d'affiliation du bénéficiaire est "Assuré" (BEN_QAF_COD = 1 ou 10)
- BEN_NIR_PSA II BEN_RNG_GEM, quand BEN_NIR_ANO est vide et la qualité d'affiliation du bénéficiaire n'est pas "Assuré" (BEN_QAF_COD ≠ 1 ou 10).



4. Exercices pratiques

Exercices pratiques - Introduction



- **Deux exercices pratiques** sont proposés :
  - Pour être réalisés de façon autonome ;
  - Afin de permettre aux utilisateurs de s'exercer à manipuler une base de données synthétiques de la base principale du SNDS.

- **Pour réaliser ces exercices**, un jeu de données synthétiques de la base principale du SNDS est mis à disposition sur Gitlab, restitué au format dans le cadre d'un accès permanent.

- **La correction de chaque exercice** est fournie pour permettre aux utilisateurs de s'auto-évaluer et de progresser en autonomie.



Vous pouvez poser vos questions sur le forum de la communauté d'entraide SNDS

Exercices pratiques - Le jeu de données synthétiques (1/2)

Le jeu de données synthétiques a été généré **à partir d'un programme SAS®** conçu pour **reproduire la structure et la logique des données issues de la base principale du SNDS**. Ce programme, développé par Jérôme Brocca à partir de son expertise métier, n'utilise aucune extraction des données source à partir de la base principale du SNDS. Les données synthétiques obtenues reprennent **la même organisation (librairies, tables, variables, codage ...) et ont la même cohérence que les données source** disponibles sur la plateforme SNDS dans le cadre d'un accès permanent.

Ce jeu de données contient :

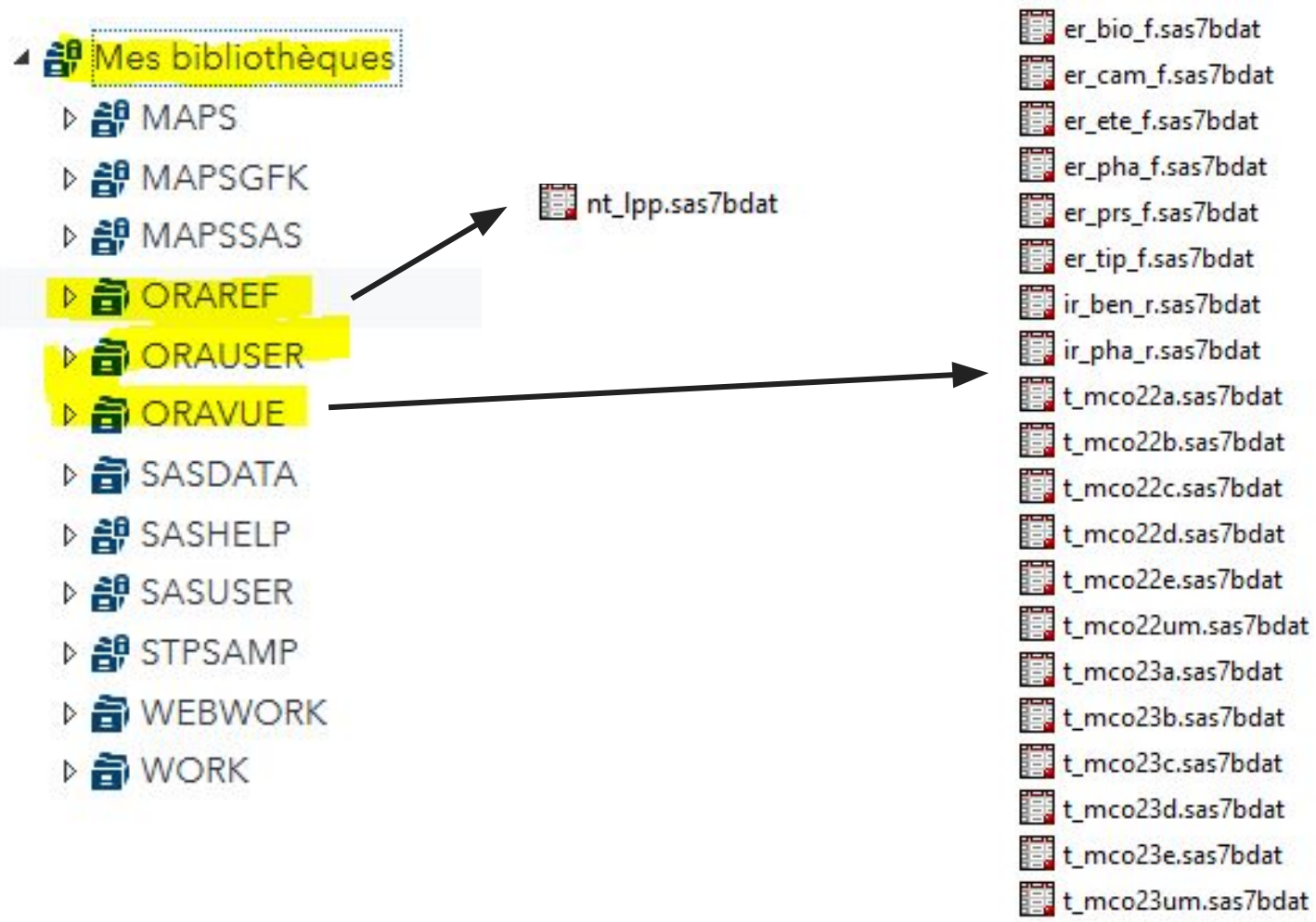
1 092 sujets distincts

Des données de remboursement de soins de ville DCIR (médicaments, consultations, procédures ...) sur la période 2023-2025

Des données de séjours hospitaliers du PMSI MCO sur la période 2022-2023

Exercices pratiques - Le jeu de données synthétiques (2/2)

Pour les deux exercices, les librairies et tables *.sas7bdat* suivantes ont été générées :



Jeu de données et exercices pratiques (1/2)

Les attendus

1

Exercice pratique n°1 :

Identifier les patients **ayant eu au moins une délivrance d'antibiotiques en 2023**

- ➔ Cibler les patients indépendamment de toute pathologie
- ➔ Calculer la répartition par sexe du nombre de consommants d'antibiotiques en ville pour l'année 2023

Indications



Antibiotiques : classe ATC = J01

2

Exercice pratique n°2 :

Cibler **les patients ayant eu en 2023 >= 1 délivrance d'antidiabétiques en ville et/ou** hospitalisés **avec présence d'un diagnostic de diabète en 2023**

- ➔ Rechercher les patients avec >= 1 séjour PMSI MCO avec présence d'un diagnostic de diabète (diagnostic principal (DP) ou diagnostic relié (DR) du séjour ou du Résumé de sortie d'Unité Médicale, ou diagnostic associé (DA)) et/ou ayant eu >= 1 délivrance de médicament antidiabétique dans les données de DCIR
- ➔ Déterminer le nombre total de patients distincts (BEN_IDT_ANO) sélectionnés ainsi que leur nombre selon les catégories suivantes :
  - patients présentant >= 1 délivrance de médicaments antidiabétiques en ville mais sans avoir eu d'hospitalisation avec diagnostic de diabète
  - patients présentant >= 1 délivrance de médicaments antidiabétiques en ville et/ou >= 1 séjour MCO avec diagnostic de diabète (DP, DR ou DA)
  - patients présentant >= 1 séjour MCO pour diabète (DP ou DR uniquement), qu'ils aient eu ou non >= 1 délivrance de médicaments antidiabétiques en ville

Indications



Médicaments du diabète : classe ATC = A10



Diagnostics du diabète : codes CIM-10 (E10 à E14)

Jeu de données et exercices pratiques (2/2)

Les résultats

1

Exercice pratique n°1 :

➔ Répartition par sexe du nombre de consommants d'antibiotiques en 2023

Code Sexe	Libellé Sexe	Nombre de consommants antibiotiques
1	Homme	212
2	Femme	201

2

Exercice pratique n°2 :

➔ Répartition des patients avec médicaments ou séjour MCO pour diabète

Nombre total de bénéficiaires	Nombre de bénéficiaires avec seulement médicaments pour diabète	Nombre de bénéficiaires avec médicament/hospitalisés pour diabète	Nombre de bénéficiaires hospitalisés en motif principal
925	436	400	298

PLATEFORME DES DONNÉES DE SANTÉ
FRENCH HEALTH DATA HUB

Suivez-nous sur les réseaux sociaux !