



Choix de la population de référence dans l'estimation de taux de recours aux soins, de taux d'incidence et de prévalences à partir des données du SNDS

Ce travail a été réalisé dans le cadre du réseau ReDSiam par le Groupe de Travail « Méthodologie-Dénominateurs »

Marjorie Boussac-Zarebska (Santé publique France), Jérôme Brocca (ARS Centre Val de Loire), Marie Dalichampt (ORS Pays de Loire), Claire-Lise Dubost (DREES), Kristel Jacquier (DREES), Céline Leroy (ARS Normandie), Laurence Mandereau-Bruno (Santé publique France), Antoine Rachas (CNAM), Fanny Raguideau (HEVA), Sébastien Rivière (CNAM) pour le groupe de travail

Mars 2021

Table des matières

1. Considérations méthodologiques	3
1.1. Caractéristiques des personnes incluses au numérateur	3
1.2. Evolution des caractéristiques des personnes	3
1.3. Définition de l'âge	3
1.4. Définition du lieu de résidence	4
2. Sources d'information pour estimer le dénominateur	4
2.1. Choisir la source d'information selon l'objectif.....	4
2.2. Délai de disponibilité des données.....	5
2.3. La proportion de « consommateurs » varie en fonction du sexe et de la classe d'âge	5
2.4. La proportion de « consommateurs » dynamiques varie suivant les territoires.....	6
Annexe 1 – Populations de référence	7
1.1. Données Insee.....	7
1.1.1. Résultats complets du recensement de la population	7
1.1.2. Estimations localisées de population (ELP)	9
1.1.3. Projections de l'outil OMPHALE.....	11
1.2. Fichiers annuels de consommateurs de la Cnam.....	13
1.3. Fichiers des consommateurs sur une période supérieure à 1 an	15
1.4. Répertoire National Inter régimes des bénéficiaires de l'Assurance Maladie (RNIAM).....	16
1.5. Tables de dénombrement des bénéficiaires par maillage géographique	18
Annexe 2 - Méthode utilisant le référentiel des bénéficiaires.....	20
Annexe 3 –Comparaison des taux de recours annuel au médecin généraliste (MG) en 2017 en fonction du dénominateur Insee	22
Annexe 4 – Les identifiants dans le SNDS.....	23

Le choix du dénominateur dans le cadre d'estimation de taux de recours ou d'incidence semble à première vue un sujet assez trivial, qui ne laisse pas forcément présager de difficultés particulières. Or dans la pratique, l'exploitation des bases du Système National des Données de Santé (SNDS) fort de ces différents régimes et de règles de gestions complexes invite à repenser la notion même de dénominateur et à s'interroger sur l'impact du choix de ce dernier et l'interprétation des résultats obtenus.

Le document donne des recommandations générales pour le choix du dénominateur utilisé pour le calcul de taux de recours aux soins, de taux d'incidence ou de prévalence d'une pathologie à partir du SNDS (à l'exclusion de l'échantillon généraliste des bénéficiaires). Il recense en annexe des populations de référence possibles en précisant leurs limites et avantages.

Issu de l'expérience de ses auteurs, ce document se veut une aide pour les utilisateurs du SNDS. Il n'a pas vocation à être exhaustif. En revanche, il souligne le fait que le choix du dénominateur n'est pas une évidence et devrait systématiquement être précisé dans les rapports d'analyse, et ses limites (et avantages) discutés.

1. Considérations méthodologiques

1.1. Caractéristiques des personnes incluses au numérateur

Le champ couvert par la population utilisée pour estimer le dénominateur doit être le même ou à défaut le plus proche possible de celui de la population utilisée pour estimer le numérateur, en termes notamment de zone géographique (France entière ? métropolitaine ? Mayotte comprise ?), de période considérée, de régimes d'assurance maladie (régime général ? y compris sections locales mutualistes ? tous régimes ?) ou de caractéristiques des bénéficiaires (âge, sexe).

1.2. Evolution des caractéristiques des personnes

Lors du calcul d'un indicateur stratifié (par exemple par département de résidence), il est nécessaire de définir dans quel(s) groupe(s) considérer une personne dont la caractéristique d'intérêt évolue au cours de la période d'étude.

Deux approches sont possibles :

- Considérer la personne dans la strate correspondant à sa situation à une date donnée (le plus souvent soit en fin soit en début de période).
- Considérer la personne dans chaque strate qui la concerne.

1.3. Définition de l'âge

Que ce soit pour sélectionner la population ou pour stratifier l'analyse, si un critère d'âge est utilisé, il doit avoir la même définition dans les populations utilisées pour estimer le numérateur et le dénominateur.

Exemples :

- Age atteint dans l'année (= différence entre l'année considérée et l'année de naissance)
- Age révolu au 1er janvier de l'année (=âge au dernier anniversaire)
- Age révolu à la date de la prescription (variable BEN_AMA_COD)

1.4. Définition du lieu de résidence

L'analyse à un niveau géographique infranational nécessite la vérification des variables « département de résidence » et « commune de résidence » dans le SNDS (% de codes manquants, code postal ou code Insee, % de codes erronés) et si possible leur correction. La qualité du codage varie en fonction des régimes et pour un régime donné peut varier en fonction de l'année, en particulier le pourcentage de communes non renseignées.

Une fiche spécifique¹ et un programme² de correction des codes sont disponibles sur la documentation collaborative du SNDS du Health Data Hub pour aider et permettre aux utilisateurs de réaffecter certains codes erronés.

La liste des codes postaux en vigueur actuellement et leur correspondance avec les codes Insee est disponible à l'adresse : <https://www.data.gouv.fr/fr/datasets/base-officielle-des-codes-postaux/>.

La liste des communes existantes évolue chaque année. Il est donc possible de trouver dans les bases des codes INSEE de communes qui n'existent plus. Pour récupérer les correspondances éventuelles, AtlaSanté a créé des tables avec l'évolution des nomenclatures de 2010 à 2020. Ces tables renseignent également les correspondances entre le code INSEE et le code géographique du PMSI. Elles sont disponibles dans le répertoire RFCOMMUN du portail SNDS. La correspondance entre le nom de la commune et son code géographique INSEE est disponible dans la bibliothèque RFCOMMUN, dans les fichiers DATASANTE_T_GEO_COM et DATASANTE_T_CORRESP_COM.

Les tables d'AtlaSanté sont également disponible à l'adresse [Atlasanté - data.gouv.fr](https://atlasante.data.gouv.fr).

En outre, il est fortement recommandé de suivre les communiqués de la Cnam pour être alerté sur d'éventuelles anomalies (par exemple le communiqué de mars 2018 signalant un codage erroné du département pour la SLM étudiante 617).

A noter que lorsque le dénominateur est issu du SNDS (par exemple du fichier des consommateurs), le lieu de résidence a la même définition au numérateur (issu également du SNDS) et au dénominateur, à savoir le département inscrit dans les bases locales de l'Assurance Maladie. Il s'agit du département du destinataire du remboursement ou du bénéficiaire s'il est connu. Mais le dénominateur n'inclut alors pas l'ensemble de la population générale puisque seuls les consommateurs sont inclus.

Dans les données Insee, le département correspond au département de résidence au recensement.

2. Sources d'information pour estimer le dénominateur

2.1. Choisir la source d'information selon l'objectif

En l'absence d'une base de données unique répertoriant les effectifs de personnes selon les caractéristiques étudiées, et/ou la notion de couverture par un régime d'assurance maladie, il est nécessaire de choisir la source d'information répondant au mieux à l'objectif poursuivi.

¹ https://documentation-snds.health-data-hub.fr/fiches/localisation_geographique_beneficiaires.html#construction-de-la-localisation-dans-le-sniiram

² <https://gitlab.com/healthdatahub/programmes-snds/-/tree/master/ARS/Localisation%20B%C3%A9n%C3%A9ficiaire/SAS>

Les sources de données existantes, leurs avantages et limites ainsi que d'autres informations, sont précisées en Annexe 1 afin que le lecteur puisse choisir par lui-même la source la plus appropriée à son analyse. L'Annexe 2 présente une méthode faisant appel au référentiel des bénéficiaires.

Quelques suggestions :

- Si l'on s'intéresse aux bénéficiaires d'un régime en particulier, le choix pourra se porter sur les données du RNIAM ou sur les dénombrements des bénéficiaires du régime général hors sections locales mutualistes (cf. Annexes 1.4 et 1.5).
- Si l'on s'intéresse aux bénéficiaires de la CMU-c, on pourra utiliser les fichiers de consommateurs (cf. Annexes 1.2, 1.3).
- Si l'on s'intéresse à la population générale stratifiée par âge et sexe, les données fournies par l'Insee pourront être utilisées (cf. Annexe 1.1).

2.2. Délai de disponibilité des données

Le choix de la source d'information pour estimer le dénominateur peut également dépendre de la disponibilité des données pour le niveau géographique souhaité au moment où l'analyse est faite.

Si l'on s'intéresse aux niveaux national, régional ou départemental, les données de recensement de la population de l'année n ne sont disponibles que l'année n+3, alors que les estimations localisées de population (ELP) au niveau départemental peuvent l'être dorénavant le 1^{er} janvier de l'année n. On peut aussi avoir besoin de calculer la taille moyenne de la population de l'année n, mais alors il faut utiliser les ELP des années n et n+1. Les fichiers issus du SNDS (fichiers des consommateurs) de l'année n sont disponibles l'année n+1.

L'Annexe 3 compare les résultats observés en fonction de différents dénominateurs issus des données de l'Insee pour le recours au médecin généraliste en 2017, au niveau national, pour la région Pays-de-Loire et pour le département de Loire-Atlantique.

2.3. La proportion de « consommateurs » varie en fonction du sexe et de la classe d'âge

Tableau 1. Proportion de consommateurs selon le sexe et la classe d'âge en 2017³

	Hommes	Femmes	Ensemble
0-4 ans	1,02	1,01	1,02
5-14 ans	0,96	0,96	0,96
15-24 ans	0,87	0,95	0,91
25-34 ans	0,97	1,05	1,01
35-44 ans	0,96	1,00	0,98
45-54 ans	0,97	0,99	0,98
55-64 ans	0,97	0,97	0,97
64-74 ans	0,98	0,97	0,98
75-84 ans	0,98	0,96	0,97
85 ans et plus	0,97	0,96	0,96
Ensemble	0,96	0,98	0,97

³ Dénominateur : données du recensement 2017 (source : Insee) ; numérateur : fichier des consommateurs de 2017 (source : EXTRACTION_PATIENTS2017TR de la bibliothèque CONSOPAT sur le portail du SNDS)

En comparaison aux données du recensement de la population, pour la plupart des classes d'âge, excepté les classes 0-4 ans et 25-34 ans, la proportion de consommateurs est inférieure à 1 (Tableau 1). Elle est la plus basse chez les hommes de la classe d'âge 15-24 ans (0,87)

Une proportion de consommateurs inférieure à 1 entraîne une sur-estimation des indicateurs calculés lorsque le dénominateur est issu des bases de consommateurs en particulier pour les années antérieures à 2015. En effet depuis 2015, le fichier des consommateurs inclut tous les régimes présents dans la table des prestations, ce qui n'est pas le cas pour les années antérieures pour lesquelles les proportions de consommateurs par rapport à la population sont plus basses. Lorsque l'on considère une période supérieure à 1 an, le pourcentage de consommateurs augmente.

Les proportions supérieures à 1 pourraient s'expliquer par la présence de doublons, fréquents dans ces classes d'âge. Il est conseillé d'utiliser l'identifiant BEN_IDT_ANO (identifiant bénéficiaire) ou un identifiant basé sur la méthodologie utilisée pour la cartographie de la Cnam afin de réduire le nombre de doublons (cf. Annexe 4).

La présence de doublons au numérateur entraîne une surestimation des indicateurs si le dénominateur est issu de données sans doublons (Insee ou RNIAM). Lorsque le dénominateur est issu des fichiers de consommateurs, les doublons étant également présents au dénominateur, les doublons au numérateur ont moins d'impact sur l'estimation des indicateurs, notamment dans les classes d'âge les plus concernées.

Par ailleurs, selon les années, les prestations de certains régimes ne remontent pas dans le DCIR. L'utilisation des données issues de l'Insee peut alors entraîner une sous-estimation des indicateurs.

2.4. La proportion de « consommateurs » dynamiques varie suivant les territoires

Le nombre de consommateurs peut aussi être calculé de façon dynamique en prenant directement dans le DCIR/DCIRS l'ensemble des bénéficiaires qui ont consommé au moins un soin sur ce territoire sur une année écoulée par exemple. Là aussi, la proportion de consommateurs varie suivant les territoires : plus les territoires comportent une proportion d'étudiants importante comme ici l'Indre et Loire et plus la différence entre consommateurs et population résidente est forte, du fait des entrées-sorties au cours de l'année.

La colonne (a), correspond ici au nombre de patients qui ont consommé au moins un soin dans les tables de prestations du DCIRS en 2019. La colonne (b), à la population recensée au 1 janvier 2020.

Code département	département	Nb patients consommateurs 2019 (a)	Pop INSEE 2019 (b)	ratio
18	Cher	307 058	300 174	1,02
28	Eure-et-Loir	439 995	431 437	1,02
36	Indre	222 662	217 312	1,02
37	Indre-et-Loire	632 753	608 387	1,04
41	Loir-et-Cher	339 662	330 727	1,03
45	Loiret	693 903	678 722	1,02

Annexe 1 – Populations de référence

1.1. Données Insee

1.1.1. Résultats complets du recensement de la population

Source	INSEE
Références	Recensement de la population Conseil pour l'utilisation des résultats statistiques
Population couverte	Ensemble de la population résidant habituellement (en général plus de 6 mois) sur le territoire français au 1er janvier de l'année quelle que soit la nationalité. Logements (dont résidences hôtelières, habitations de fortune) + hôtels (résidents habituels) + « communautés » (cité universitaire, maison de retraite, internat, communauté religieuse, foyer de travailleurs...) + personnes sans abri + résidents d'une habitation mobile terrestre. Obligation théorique de répondre à l'enquête
Recueil des données	Depuis 2006, recensement par collecte annualisée par roulement sur un cycle de 5 ans (depuis 2004). - les communes de moins de 10000 habitants sont réparties en 5 groupes : recensement exhaustif d'un groupe chaque année - communes de 10000 habitants et plus : chaque année, enquête par sondage de 8% des logements de la commune A partir de 2008, les résultats complets du recensement sont identiques aux populations municipales ⁴ à tous les échelons communaux et supra-communaux. En 2006 et 2007, il peut exister quelques différences dues à des arrondis.
Zone géographique	France entière
Niveau géographique le plus fin commun au SNDS	Commune de résidence

⁴ « Comprendre la population légale de votre commune - Des enquêtes de recensement aux populations légales ». Cnis. Novembre 2010

(https://www.cnis.fr/wp-content/uploads/2018/01/DC_2011_15e_reunion_CNERP_comprendre_population.pdf)

Variables disponibles communes au SNDS	Année d'âge (âge en années révolues au 1 ^{er} janvier de l'année considérée, depuis le recensement de 2006). Sexe Commune/département/région de résidence
Disponibilité	Recensements disponibles pour chacune des années depuis 2006 jusqu'à l'année n-3 incluse Pour Mayotte : recensements en 2007, 2012 et 2017 Sur le site de l'Insee mise à jour en juin des résultats complets du recensement de l'année n-3.
Doublons	NON
Limites	Limites liées aux variables disponibles communes au SNDS (populations par âge/sexes/zone géographique) Ne peut pas être utilisée pour le DCIR pour les années antérieures à 2009 compte tenu de la couverture du DCIR (RG y compris SLM de 2006 à 2008 soit environ 85% des bénéficiaires) sans utilisation d'une pondération éventuelle. Attention : vérifier la version du code commune dans le SNDS et dans les données de recensement. Une table de passage d'une géographie à une autre peut être nécessaire. L'utilisation des données au niveau des communes nécessite de prendre des précautions, la commune présente dans les bases de l'AM ne correspondant pas toujours à la commune de résidence Insee (pour les étudiants par exemple).
Avantages	Population par sexe et année d'âge. Possibilité d'avoir dans le PMSI ou le DCIR des numérateurs correspondant au champ de la population INSEE : DCIR : depuis la création de la couverture maladie universelle en 1999 ⁵ la quasi-totalité des personnes résidant en France sont bénéficiaires de l'Assurance Maladie. A partir de 2009, les régimes inclus dans le DCIR permettent de couvrir plus de 98% des bénéficiaires d'un régime d'Assurance Maladie. Possibilité dans le DCIR et le PMSI de sélectionner les personnes résidant en France

⁵ Loi n°99-641 du 27 juillet 1999 portant création d'une couverture maladie universelle

1.1.2. Estimations localisées de population (ELP)

Source	INSEE
Référence	Estimations de population
Population couverte	Ensemble de la population résidant habituellement (en général plus de 6 mois) sur le territoire au 1er janvier de l'année quelle que soit la nationalité. Logements (dont résidences hôtelières, habitations de fortune) + hôtels (résidents habituels) + « communautés » (cité universitaire, maison de retraite, internat, communauté religieuse, foyer de travailleurs...) + personnes sans abri + résidents d'une habitation mobile terrestre.
Recueil des données	Le recensement de la population sert de base aux estimations annuelles de population. Il fixe les niveaux de référence. Entre deux recensements, des estimations provisoires sont réalisées grâce aux statistiques de l'état civil relatives aux naissances et aux décès par sexe et âge, et à une estimation du solde migratoire par sexe et âge. Les estimations deviennent définitives une fois les résultats d'un nouveau recensement connus, soit pour l'année n-3. Pour l'année n, les estimations sont basées sur les décès et les naissances de l'année n-1 et non celles de l'année n. Données non consolidées pour les années n, n-1 et n-2.
Zone géographique	France entière (Mayotte inclus depuis 2014)
Niveau géographique le plus fin commun au SNDS	Département de résidence
Variables disponibles communes au SNDS	Année d'âge (âge en années révolues au 1 ^{er} janvier de l'année considérée) Sexe Département/région de résidence
Disponibilité	sur demande, le réseau des directions régionales de l'Insee, les publics de type institutionnel, chercheur, chargé de mission de service public peuvent avoir accès gratuitement aux données et plus généralement tout public qui en a besoin pour un motif justifié.
Doublons	NON

<p>Limites</p>	<p>Limites liées aux variables disponibles (populations par âge/sexe/zone géographique)</p> <p>Ne peut pas être utilisée pour le DCIR pour les années antérieures à 2009 compte tenu de la couverture du DCIR (RG y compris SLM de 2006 à 2008 soit environ 85% des bénéficiaires) sans utilisation d'une pondération éventuelle</p> <p>Il s'agit d'estimations localisées de population, donc nécessairement affectées d'une marge d'incertitude (laquelle peut être sensible si les effectifs sont faibles : cas des âges élevés, sur les départements peu peuplés, par exemple). Il est donc toujours conseillé de travailler sur des tranches d'âges ad-hoc, plutôt que sur des âges détaillés isolés.</p>
<p>Avantages</p>	<p>Age en année d'âge</p> <p>Estimations disponibles pour l'année n.</p> <p>Possibilité d'avoir dans le PMSI ou le DCIR des numérateurs correspondant à la population INSEE :</p> <ul style="list-style-type: none"> • DCIR : depuis la création de la couverture maladie universelle en 1999⁶ la quasi-totalité des personnes résidant en France sont bénéficiaires de l'Assurance Maladie. A partir de 2009, les régimes inclus dans le DCIR permettent de couvrir plus de 98% des bénéficiaires d'un régime d'Assurance Maladie. • Possibilité dans le DCIR et le PMSI de sélectionner les personnes résidant en France

NB : les données ne peuvent pas être diffusées par âge détaillé mais seulement par regroupements d'âge d'au moins 5 ans ou correspondant à des effectifs d'au moins 2000 personnes, en mentionnant la source : Insee - estimations de population.

⁶ Loi n°99-641 du 27 juillet 1999 portant création d'une couverture maladie universelle
10

1.1.3. Projections de l'outil OMPHALE

Source	INSEE
Référence	Le modèle de projection démographique Omphale
Population couverte	Projections sur toute zone géographique de plus de 50 000 habitants. Ensemble de la population résidant habituellement (en général plus de 6 mois) sur la zone géographique considérée au 1er janvier de l'année quelle que soit la nationalité. Logements (dont résidences hôtelières, habitations de fortune) + hôtels (résidents habituels) + « communautés » (cité universitaire, maison de retraite, internat, communauté religieuse, foyer de travailleurs...) + personnes sans abri + résidents d'une habitation mobile terrestre.
Recueil des données	L'outil OMPHALE (outil méthodologique de projections d'habitants, d'actifs, de logements et d'élèves) élaboré par l'Insee à partir des données du recensement permet d'obtenir des <u>projections</u> de population aux niveaux départemental, régional et national. L'outil Omphale 2017 utilise les populations par sexe et âge au 1 ^{er} janvier 2013, issues du recensement de la population.
Zone géographique	France entière (y compris Mayotte)
Niveau géographique le plus fin commun au SNDS	Département de résidence
Variables disponibles communes au SNDS	Classes d'âge quinquennales (âge en années révolues au 1 ^{er} janvier de l'année considérée) Sexe Département/région de résidence
Disponibilité	Projections de population sur la période 2013-2050 aux niveaux départemental et régional et sur la période 2013-2070 au niveau national Sur le site de l'Insee
Doublons	NON

<p>Limites</p>	<p>Limites liées aux variables disponibles (populations par classes d'âge/sexe/zone géographique)</p> <p>Ne peut pas être utilisée pour le DCIR pour les années antérieures à 2009 compte tenu de la couverture du DCIR (RG y compris SLM de 2006 à 2008 soit environ 85% des bénéficiaires) sans utilisation d'une pondération éventuelle</p> <p>« Les projections départementales et régionales mises à disposition sur le site sont une ventilation des projections nationales. Elles reposent sur des hypothèses d'évolutions démographiques identiques et sur le maintien des tendances migratoires intérieures à la France observées entre 2012 et 2013. Comme toute projection, elles ne constituent pas une prévision mais une simulation de l'avenir dans le cadre théorique des hypothèses formulées. Toutes ces projections sont issues d'un scénario standardisé pour l'ensemble des territoires français qui ne tient compte ni des spécificités locales (en particulier les flux avec l'étranger) ni des ruptures de tendances démographiques les plus récentes. Des traitements spécifiques, avec des hypothèses différentes, peuvent être réalisés dans le cadre de partenariats avec les directions régionales de l'Insee »(source : https://www.insee.fr/fr/statistiques/2859843#consulter)</p>
<p>Avantages</p>	<p>Projections disponibles pour l'année n.</p> <p>Possibilité d'avoir dans le PMSI ou le DCIR des numérateurs correspondant à la population INSEE :</p> <p>DCIR : depuis la création de la couverture maladie universelle en 1999⁷ la quasi-totalité des personnes résidant en France sont bénéficiaires de l'Assurance Maladie. A partir de 2009, les régimes inclus dans le DCIR permettent de couvrir plus de 98% des bénéficiaires d'un régime d'Assurance Maladie.</p> <p>Possibilité dans le DCIR et le PMSI de sélectionner les personnes résidant en France</p>

⁷ Loi n°99-641 du 27 juillet 1999 portant création d'une couverture maladie universelle
12

1.2. Fichiers annuels de consommateurs de la Cnam

Source	Cnam
Référence	Documentation dans la rubrique Actualités du portail SNDS
Population couverte	Personnes ayant eu au moins une prestation entre le 1 ^{er} janvier et le 31 décembre de l'année aaaa (liquidation jusqu'au 30 juin de l'année suivante) : dans un régime donné* (EXTRACTION_PATIENTSaaaaTR) au RG y compris SLM (EXTRACTION_PATIENTSaaaa) *Régimes pour lesquels les prestations sont transmises au SNDS l'année n. Une ligne par régime d'affiliation dans l'année considérée.
Recueil des données	Bases générées par la Cnam à partir des données du DCIR (prestations et/ou référentiel des bénéficiaires). Pour rappel, la table des prestations ne comprend pas les prestations non remboursées ni les prestations remboursées non facturées directement à l'Assurance Maladie ; en particulier elle ne comprend pas l'ensemble des hospitalisations du secteur public. Création des bases l'année n+1 (septembre/octobre)
Zone géographique	France entière Mayotte depuis chargement d'août 2008 Collectivités d'Outre-mer
Niveau géographique le plus fin commun au SNDS	Commune de résidence
Variables disponibles communes au SNDS	Les principales variables communes sont : Identifiant SNDS Année/mois de naissance* Date de décès (jour*/mois/année) Sexe Couverture maladie universelle complémentaire (CMUC) Aide médicale d'état (AME) Présence d'une affection de longue durée Organisme d'affiliation Organisme de liquidation Régime Dernière date de soin dans le régime d'affiliation Département et commune* de résidence correspondant au dernier soin (ou dernier département connu) dans le régime d'affiliation Quintile de désavantage social (FDEP) à partir de 2013 Les variables sensibles (*) sont disponibles en fonction du profil de connexion Certaines des variables correspondent au dernier remboursement de l'année dans le régime, d'autres à l'ensemble des remboursements du régime (se référer au document de référence cité au-dessus).

Disponibilité	<p>Sur le portail SNDS, bibliothèque CONSOPAT, en fonction du profil de connexion à partir de l'année 2010.</p> <p>Pour les accès hors portail SNDS (bulle sécurisée), faire la demande auprès de la CNAM.</p>
Présence de doublons	<p>OUI</p> <p>Possibilité pour une personne d'avoir plusieurs identifiants SNDS {BEN_NIR_PSA, BEN_RNG_GEM} correspondant à plusieurs ouvreurs de droit.</p> <p>Cette table comprend une ligne par régime et bénéficiaire, un bénéficiaire ayant pu avoir plusieurs régimes au cours de l'année.</p>
Limites	<p>Disponibilité des données en fonction du profil de connexion</p> <p>Le codage de la variable commune de résidence doit être vérifié avant toute utilisation comme population de référence. En particulier, certaines communes ont un code postal et non un code géographique Insee.</p> <p>La Cnam recommande d'utiliser la variable FDEP présente dans les tables uniquement pour les régimes RG, MSA et RSI compte-tenu de la qualité de la variable commune de résidence.</p> <p>Non prise en compte des non consommateurs, avec risque pour certaines classes d'âge et sexe d'une sous-estimation du nombre de bénéficiaires</p> <p>Gestion des doublons car l'identifiant (couple ben_nir_psa, ben_rng_gem) peut ne pas être unique pour une personne. Il est fortement conseillé de prendre comme identifiant la variable ben_idt_ano ou un identifiant basé sur la méthodologie utilisée pour la cartographie de la Cnam au lieu du couple (ben_nir_psa,ben_rng_gem).</p> <p>Pour une utilisation pour laquelle on ne retient qu'un enregistrement par bénéficiaire, il est nécessaire d'avoir une règle pour choisir la commune de résidence de référence : par exemple, retenir la ligne correspondant à la dernière date de soin (variable dds) pour chaque ben_idt_ano. Dans ce cas, il peut-être pertinent de recalculer les top (ALD, C2S, etc...) par ben_idt_ano.</p>
Avantages	<p>Permet de faire des analyses par régime d'affiliation et sur des sous-groupes de population (par exemple en fonction de l'indice de désavantage social (FDEP) ou de la couverture maladie universelle complémentaire (CMU-C)</p> <p>Les numérateurs et les dénominateurs sont construits à partir de la même source de données, les éventuels biais se retrouvent donc au numérateur et au dénominateur.</p>

1.3. Fichiers des consommateurs sur une période supérieure à 1 an

Les fichiers annuels de consommateurs peuvent être combinés pour générer la population des personnes ayant eu au moins une prestation au cours de plusieurs années consécutives. Par exemple, la table des consommateurs de 3 ans de l'année 2019 comprend les bénéficiaires ayant eu au moins une consommation dans la période 01/01/2017 au 31/12/2019. Cette population peut également être générée directement à partir des tables de prestations du DCIRS ou DCIR. Ce procédé permet de prendre en compte au dénominateur les 'petits consommateurs' c'est-à-dire les personnes n'ayant pas eu de consommation dans l'année mais qui en ont eu une sur une période de plusieurs années consécutives.

Plusieurs étapes de data management sont nécessaires pour aboutir à une table ne contenant qu'une ligne par bénéficiaire. Les principales étapes consistent à récupérer dans le référentiel IR_BEN_R les identifiants uniques BEN_IDT_ANO mis à jour (leur valeur dans les bases Consopat étant obsolètes), de ne garder que la ligne associée à la dernière date de soin (variable DDS des tables annuelles de Consopat ou max(exe_soi_dtd) dans les tables de prestations NS_PRS_F ou ER_PRS_F et de supprimer de la table les bénéficiaires qui n'étaient plus affiliés au début de l'année (il s'agit principalement des personnes décédés et des NIR fictifs).

Intérêts de ces tables de consommateurs sur plusieurs années :

_ Permet d'avoir une base de dénombrement de bénéficiaires tous régimes, pour tous les croisements d'intérêt.

_ Permet de prendre en compte la situation de chaque bénéficiaire connue à la dernière date de soin de la période considérée (par exemple entre le 01/01/2017 et le 31/12/2019, et non à la date de l'extraction comme c'est le cas avec IR_BEN_R)

_ Les numérateurs et les dénominateurs sont construits à partir de la même source de données, les éventuels biais se retrouvent donc au numérateur et au dénominateur.

Inconvénients de ces tables de consommateurs sur plusieurs années :

_ coûteuses en temps de calcul pour leur création. Il peut être intéressant de les enregistrer (comme les tables de consommateurs annuelles de la bibliothèque 'consopat'). Il est conseillé d'enregistrer ces tables dans des dossiers partagés pour ne pas perdre trop d'espace de stockage.

_ conserve les défauts inhérents aux bases du SNDS : existence de doublons, non prise en compte des non consommateurs depuis plus de x ans, inclusion des faux non consommateurs (décédés ou plus affiliés à l'AM). La fiabilité des effectifs est donc moindre chez les jeunes adultes (plus de doublons mais aussi plus de non consommateurs) et chez les personnes âgées (défaut d'enregistrement du décès).

1.4. Répertoire National Inter régimes des bénéficiaires de l'Assurance Maladie (RNIAM)

Source	CNAV (Caisse nationale d'assurance vieillesse) Gestion pour le compte et sous le contrôle des organismes d'Assurance Maladie
Référence	Pour les tables RG+SLM disponibles sur le portail SNDS, documentation dans la rubrique 'SNIIRAM_Documentation SNIIRAM-SNDS\Actualités' sur le portail.
Population couverte	<p>Par régime, pour l'ensemble des régimes : personnes affiliées au régime de sécurité sociale, qu'elles résident ou non sur le territoire.</p> <p>Le RNIAM reprend la population des bénéficiaires des remboursements des frais de santé gérés par un organisme d'Assurance Maladie (en France). Seuls sont présents au RNIAM, les individus rattachés à un organisme d'assurance maladie qui disposent d'un NIR certifié (c'est-à-dire des personnes pour lesquelles l'identité a été validée (nom, prénoms, date et lieu de naissance). Ainsi, sont exclus du répertoire :</p> <ul style="list-style-type: none"> les individus disposant d'un NIR provisoire / NIA (Numéro d'Identifiant d'Attente) ne sont pas présents au RNIAM (personnes faisant l'objet d'une demande d'identification, ex immatriculation). les bénéficiaires de l'Aide Médicale d'Etat (pour le RG) les demandeurs d'asiles les migrants de passage les personnes non rattachées à un organisme d'assurance maladie par exemple : personnes ayant une pension d'invalidité ou d'une rente et vivant à l'étranger <p>Les mises à jour du RNIPP (Répertoire National d'Identification des Personnes Physiques), naissances et décès, sont transmises quotidiennement au SNGI (Système National de Gestion des Identifiants). Le RNIAM est adossé au SNGI</p>
Recueil des données	Les informations contenues dans le RNIAM sont confrontées à celles provenant respectivement, du Répertoire national d'identification des personnes physiques (RNIPP) tenu par l'INSEE, et des organismes gestionnaires des régimes d'assurance maladie (site Insee).
Zone géographique	France entière Mayotte depuis 2011
Niveau géographique le plus fin commun au SNDS	Niveau géographique de la caisse (le RNIAM ne comprend pas l'adresse de la personne)
Disponibilité	Tables SAS disponibles sur le portail (RGETSLM_JANVIER20XX) de 2007 à 2021 pour RG+SLM mais pas de procédure identifiée pour l'accès aux données des autres régimes. Disponibilité l'année n. Pour les accès hors portail SNDS (bulle sécurisée), faire la demande auprès de la CNAM.

Années disponibles	Mis en place en 1998
Variables disponibles communes au SNDS	<p>Dans le fichier géré par la CNAV (données individuelles) : Sexe Date de naissance Date de décès⁸ Identifiant de l'organisme d'AM servant les prestations et date de rattachement Le cas échéant identifiant de l'organisme ayant servi précédemment les prestations Eventuellement un numéro d'identification propre à l'organisme d'AM</p> <p>Dans les fichiers de dénombrement mis à disposition sur le portail dans le répertoire RFCOMMUN pour la population protégée RG+SLM Année de naissance Sexe Statut vital⁵ Code 'Grand régime' Caisse d'affiliation Compteur</p>
Présence de doublons	NON une personne n'est recensée que dans un régime
Limites	<p>Disponibilité des données (pas de procédure identifiée pour l'accès aux données autres que RG+SLM)</p> <p>Peu de variables disponibles dans les fichiers fournis sur le portail</p> <p>Les données relatives à une personne sont supprimées dans la deuxième année à compter du décès ou de la fin de droit de l'Assurance Maladie (27 mois).</p> <p>Le signalement de changement de régime à l'Assurance Maladie (mutations intra régime et inter régime assurance maladie) n'est pas systématique. Il est effectué par le nouvel organisme d'assurance maladie suite à une demande d'ouverture de droit. En revanche, les avis de fin de droit à l'Assurance Maladie sont rares hors cas du décès (exemple départ d'un assuré à l'étranger).</p>
Avantages	<p>Permet de faire des études par régime d'affiliation</p> <p>Par définition, il ne s'agit que de personnes protégées par un régime d'assurance maladie. Les taux peuvent être calculés par âge et sexe voire par région ou département en fonction du niveau géographique de la caisse d'affiliation.</p>

⁸Un organisme de l'Assurance Maladie peut avoir connaissance du décès de l'individu avant que celui-ci ne soit connu au SNGI et/ou RNIPP. Dans ce cas le décès sera « présumé » tant qu'il ne sera pas « certifié » par l'INSEE

1.5. Tables de dénombrement des bénéficiaires par maillage géographique

	Tables de dénombrement des bénéficiaires par maillage géographique
Source	Cnam - INSEE
Référence	Documentation dans la rubrique 'SNIIRAM_Documentation SNIIRAM-SNDS\Manuels utilisateurs\Dénombrements Bénéficiaires et PS'
Population couverte	Ensemble des personnes affiliées au régime général (RG) hors Sections Locales Mutualistes (SLM) au 1 ^{er} janvier de l'année
Recueil des données	Bénéficiaires du RG hors SLM présents dans les bases de données opérantes des caisses primaires d'Assurance Maladie et des caisses générales de Sécurité Sociale. La localisation géographique est assurée par l'INSEE à partir des adresses des bénéficiaires. Données disponibles en général l'année n+1
Zone géographique	France entière et collectivités d'Outre-mer de Saint-Barthélemy et de Saint-Martin
Niveau géographique le plus fin commun au SNDS	- Commune en France métropolitaine et dans les DOM - Commune pour les collectivités d'Outre-mer de Saint- Barthélemy et Saint-Martin
Disponibilité	Données disponibles depuis l'année 2014. Sur le portail SNDS : - Une table au format SAS pour chacun des 6 maillages géographiques dans la bibliothèque RFCOMMUN dans SAS (le maillage le plus fin correspond à l'IRIS) - Un fichier au format Excel pour chacune des 18 régions dans un répertoire sur BO Web Pour les accès hors portail SNDS (bulle sécurisée), faire la demande auprès de la CNAM.

Variables disponibles communes au SNDS	Données agrégées par : Classes d'âge Sexe Bénéfice ou non de la CMUC Qualité du bénéficiaire (assuré/ayant droit) et ventilées par région/département/commune/bassin de vie Remarque : la ventilation est aussi faite au niveau infra-communal : IRIS/ zones urbaines sensibles (ZUS) et dans la version 2016 par zones franches urbaines (ZFU) /quartiers prioritaires de la ville (QPV)
Présence de doublons	NON
Limites	Classes d'âge pré-définies (<20 ans ; 20-39 ; 40-59 ; 60-74 ; ≥75 ans) Modification de la définition des régions en 2016 Les tables mises à disposition ne constituent pas un référentiel de la population protégée au sens du RNIAM.
Avantages	Disponibilité d'effectifs au niveau des bassins de vie. Elles ne contiennent pas de doublons et les codes communes sont corrigés par l'INSEE. Elles permettent d'effectuer des analyses à un niveau fin (par sexe et classes d'âge).

Annexe 2 - Méthode utilisant le référentiel des bénéficiaires

Le référentiel des bénéficiaires ne constitue pas un référentiel de la population au sens de l'INSEE. En effet, ce référentiel n'est pas complètement exhaustif, on ne trouve pas tous les non-consommateurs (la MSA ne remonte pas ces non consommateurs ainsi que d'autres petits régimes).

Une étude interne de la Cnam (non publiée) portant sur la faisabilité d'inclure les « non consommateurs » dans la cartographie des pathologies et des dépenses de la Cnam retrouve dans le référentiel des « faux » non-consommateurs. Ces « faux non-consommateurs » sont des patients du Régime général et des organismes ou régimes infogérés, non consommateurs depuis le début de l'historique autorisé mais qui ont toujours des droits ouverts dans les bases locales.

On ne peut donc l'utiliser à proprement parler comme dénominateur. Toutefois, ce référentiel peut s'avérer précieux pour corriger des taux de recours « aberrants » sur un territoire donné.

Une méthode⁹ développée par l'ARS Normandie et l'ARS Centre-Val de Loire vise à proposer une correction des taux de recours en prenant en compte la dimension spatiale et dynamique de la consommation de soins au sein d'un territoire. En effet, au cours de l'année, la population de référence change et bouge géographiquement. De même, les référentiels bénéficiaires du SNDS ne permettent pas de recenser de manière exhaustive ces mouvements ni la situation exacte des bénéficiaires. Un individu peut consommer des soins sur un territoire tout au long de l'année ou sur plusieurs territoires. Les étudiants peuvent consommer du soin dans la région de leurs parents et gonfler les effectifs des résidents au sens de l'INSEE.

Cette méthode basée sur la décomposition du taux de recours se fait à partir d'hypothèses sur les tables des prestations et sur le référentiel des individus associé. Elle tient compte des contraintes inhérentes à chacune de ces deux bases et permet de s'affranchir des problèmes de variations de périmètres entre données dynamiques et statiques. En outre, elle permet une approche géographique fine et tous régimes du territoire.

Le facteur de non recours aux soins, même imparfaitement connu sur un territoire est néanmoins pris en compte.

La méthode consiste à décomposer ce taux en plusieurs facteurs mesurables à partir du SNDS de manière à les estimer sur des périmètres invariants.

L'estimation se fait à partir d'hypothèses sur les tables des prestations et sur le référentiel des individus associé, qui tiennent compte des contraintes inhérentes à chacune de ces deux bases.

Le taux de recours s'exprime de la façon suivante :

$$\text{Taux de recours} = \underbrace{\frac{\text{Nb de bénéf ayant recours à un soin sur la période d'intérêt}}{\text{Pop des consommateurs sur la période d'intérêt}}}_{\text{Calculé à partir de la table des prestations}} \times \underbrace{\frac{\text{Pop des consommateurs}}{\text{Pop des conso} + \text{pop des non conso}}}_{\text{Calculé à partir du référentiel des bénéficiaires}}$$

⁹ Brocca, Jérôme, et Céline Leroy. « Estimation du taux de recours aux soins au travers du SNDS : une approche spatiale et dynamique ». <https://www.ea-reperes.com/congres/session-poster>, EHESP, 25 Juin 2019. Web.

La méthode s'appuie sur l'information du Top consommant/non-consommant présent dans le référentiel des bénéficiaires. Il est donc nécessaire de préciser son périmètre, ses limites et les hypothèses qu'elle induit.

En effet, les informations ne sont disponibles que pour certains régimes : Régime général, SLM infogérées, Marins, Militaires, Cultes, Clercs et employés de notaire. Il faut donc faire l'hypothèse que cette estimation du pourcentage de non-consomphants est robuste, à savoir qu'on peut considérer que les régimes qui la renseignent sont suffisamment représentatifs du territoire et qu'elle est stable dans le temps. (Pour le premier point déjà l'intégration du RSI au Régime Général en 2020 permet de renforcer cette hypothèse).

De même, le référentiel ne permet pas d'estimer la part des consommateurs d'une année sur la population protégée de la même année. Par exemple si on s'intéresse à 2017, on aura un référentiel des individus ayant reçu au moins une prestation depuis le 1er janvier 2013.

Enfin, une étude de la CNAM a pu relever quelques imprécisions et erreurs sur cette variable.

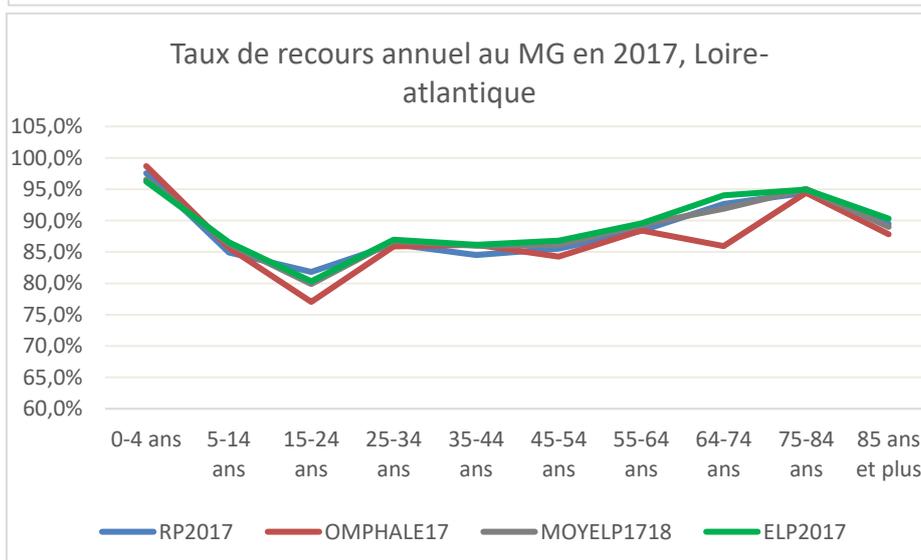
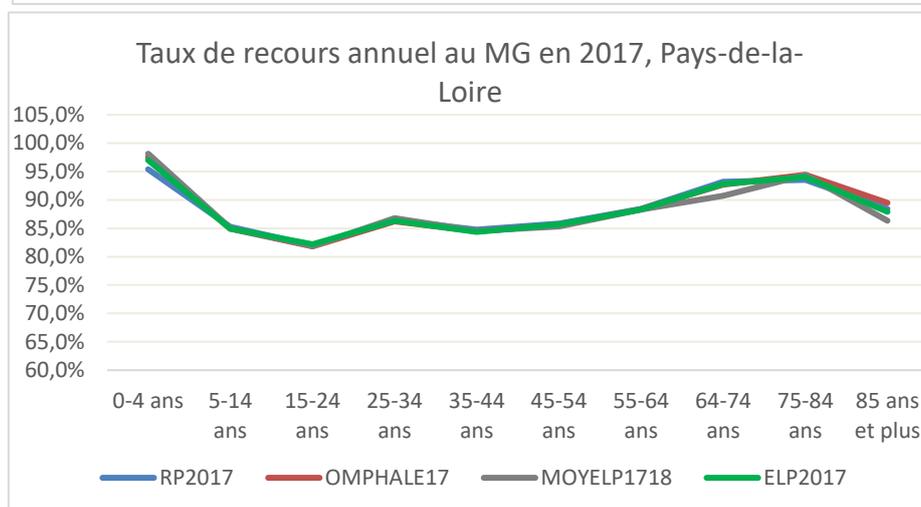
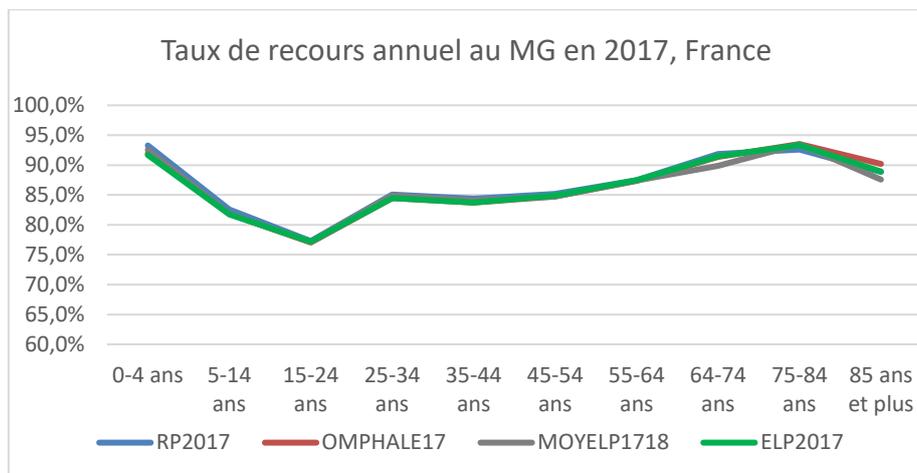
Toutefois, il convient ici de rappeler que la finalité première de cette méthode n'est pas d'estimer de manière précise cette part de non consommateurs ni son évolution dans le temps. Il s'agit surtout d'établir une correction et de considérer le taux de non consommation comme une pondération. En effet, pour la thématique étudiée l'information principale et prépondérante reste et demeure le taux de recours calculé sur les seuls « consommateurs, la deuxième composante ayant très largement un impact moindre.

D'autre part, les imprécisions et les biais se compensant les uns par rapport aux autres on peut raisonnablement s'appuyer sur cette méthode pour finaliser l'estimation d'un taux de recours. C'est particulièrement vrai sur des territoires infra départementaux comme des EPCI ou des communes où l'utilisation directe au dénominateur de référentiels classiques comme la population INSEE peuvent conduire à afficher des taux de recours supérieurs à 100%, et complexifie l'appropriation de ces résultats par les décideurs et les acteurs de terrain.

L'illustration de cette méthode s'appuie sur l'estimation du taux de recours aux médecins généralistes dans le DCIRS. Les exemples ci-dessous permettent de représenter l'apport de cette correction.

cod_comm	LIBELLE_DE_LA_COMMUNE	NB_CONSO_REF	NB_TOT_REF	nb_patient_conso	nb_patient_mg	pop_insee	est_txrecours	tx_insee
50430	Reigneville Bocage	26	26	129	34	39	26,4%	330,8%
50320	Le Mesnil Rogues	112	114	261	136	155	51,2%	168,4%
76620	St Martin L Hortier	196	197	937	248	275	26,3%	340,7%

Annexe 3 – Comparaison des taux de recours annuel au médecin généraliste (MG) en 2017 en fonction du dénominateur Insee



Légende :

- Données du recensement de la population en 2017
- Projection OMPHALE de la population en 2017
- Population moyenne en 2017 calculée à partir des estimations localisées de population en 2017 et 2018
- Estimations localisées de population en 2017

Annexe 4 – Les identifiants dans le SNDS

L'identifiant BEN_NIR_PSA est constitué, avant pseudonymisation, du n° de l'ouvreur de droit, du sexe et de la date de naissance du bénéficiaire. On accole souvent le rang gémellaire à cet identifiant pour différencier les jumeaux de même sexe. On obtient un identifiant {BEN_NIR_PSA ; BEN_RNG_GEM}. Si la personne a plusieurs ouvriers de droits, elle a plusieurs couples {BEN_NIR_PSA ; BEN_RNG_GEM} dans le référentiel IR_BEN_R. C'est le cas en particulier :

- des enfants affiliés simultanément à l'assurance maladie en tant qu'ayant droit de leurs deux parents,
- des personnes qui passent du statut d'ayant droit de leurs parents au statut d'assuré social (changement d'ouvreur de droits),
- des travailleurs étrangers affiliés à l'assurance maladie tout d'abord avec un n° de sécurité sociale provisoire puis, au bout de quelques mois, avec un n° définitif.

L'identifiant propre au bénéficiaire (NIR personnel unique tout au long de sa vie) est pseudonymisé et remonte, de façon inégale en fonction des années dans le SNDS, sous la variable BEN_NIR_ANO. En 2018, 98,5% des couples {ben_nir_psa ; ben_rng_gem} avaient un BEN_NIR_ANO renseigné (pourcentage variant de 93% à 100% en fonction du régime d'affiliation du bénéficiaire) versus 88% en 2011.

L'identifiant synthétique pseudonymisé du bénéficiaire BEN_IDT_ANO est égal :

- au NIR pseudonymisé du bénéficiaire ben_nir_ano lorsque ce NIR est renseigné (c'est de plus en plus le cas),
- à la concaténation de l'identifiant SNDS pseudonymisé ben_nir_psa et du rang de bénéficiaire ben_rng_gem lorsque le NIR du bénéficiaire ben_nir_ano n'est pas renseigné.

Cet identifiant ben_idt_ano est accompagné du top ben_idt_top qui vaut :

- 1 lorsque ben_idt_ano est égal à ben_nir_ano → 97% des cas en 2018
- 0 (zéro) sinon.

Dans le DCIRS ainsi que dans le référentiel des bénéficiaires qui lui est associé « ir_iba_r », les bénéficiaires sont identifiés par leur identifiant synthétique ben_idt_ano.

L'identifiant synthétique pseudonymisé du bénéficiaire utilisé pour la cartographie de la Cnam est égal :

- au NIR pseudonymisé du bénéficiaire ben_nir_ano lorsque ce NIR est renseigné,
- au numéro d'assuré, ASS_NIR_ANO, si la personne est son propre assuré (BEN_QAF_COD=10 ou BEN_QAF_COD=1),
- à la concaténation de l'identifiant SNDS pseudonymisé ben_nir_psa et du rang de bénéficiaire ben_rng_gem sinon.

L'utilisation de cet identifiant peut réduire le nombre de doublons par rapport à la variable BEN_IDT_ANO.

L'identifiant présent dans la cartographie, CTO_IDT_ANO, a été inclus dans le référentiel des bénéficiaires en mars 2021.

→ Afin de minimiser le nombre de doublons, il faut utiliser :

- **BEN_IDT_ANO** plutôt que le couple {ben_nir_psa ; ben_rng_gem}. La valeur 0 de la variable **BEN_IDT_TOP** indique le nombre de NIR parmi lesquels peuvent se trouver des doublons.

ou

- un identifiant généré en utilisant la méthodologie de la cartographie.

En outre, si l'on travaille sur le DCIR, il est recommandé de **retirer les NIR fictifs ou inconnus (BEN_CDI_NIR égal à '01', '05', '06', '08', '09', '11', '12', '13', '14', '15', '99' ou 'NN')**. Ce nettoyage a été effectué par la CNAM dans le DCIRS.

Par ailleurs, il peut y avoir dans les référentiels des personnes décédées dans les 3 ans. Il conviendra donc d'exclure les personnes décédées avant la période d'intérêt.

Dans le PMSI (MCO, SSR, HAD, RIM-P), les patients sont identifiés par la variable NIR_ANO_17, qui correspond à BEN_NIR_PSA.

Dans les tables des causes de décès, les personnes décédées sont identifiées par la variable DCD_IDT_ENC, qui correspond au numéro du certificat de décès pseudonymisé deux fois.

Lorsque l'appariement est effectif entre les données du DCIR et celles des causes de décès, le top DCD_IDT_TOP vaut 1 et les identifiants BEN_IDT_ANO et BEN_NIR_ANO apparaissent dans les deux tables relatives aux causes de décès.

Des informations complémentaires sont disponibles dans la fiche intitulée 'Identifiants des bénéficiaires' de la [documentation collaborative du Health Data Hub](#).