

CNAM (DEMSI, DATAD) avec l'aide de Laurence De Roquefeuil

Avec l'aide de Grégoire Rey, Claire Imbaud, Diane Martin, Pierre Boulet et Elise Coudin
(INSERM – CépiDc)

Communication du 8 mars 2024

Documentation à l'attention des utilisateurs
concernant l'intégration des causes médicales de décès
dans le Système National des Données de Santé
Version 5

SOMMAIRE

Avant-propos	3
Introduction	4
I/ Constitution des informations sur les causes médicales de décès	5
II/ Appariement indirect entre les informations des causes de décès et les données de santé ...	8
III/ Comment relier les données de santé aux causes médicales de décès	12
1/ Recherche des informations dans le référentiel médicalisé IR_IMB_R et DCIR.....	13
2/ Recherche des informations dans le PMSI	14
IV/ Description des données	15
1/ La table des circonstances et de la cause initiale du décès KI_CCI_R.....	15
a/ Rappel sur les identifiants des bénéficiaires.....	18
b/ Version du certificat, statut de traitement, type de certificat, type de support, volet du certificat.....	19
c/ Département et commune de décès ou de résidence	20
d/ Date de décès.....	20
e/ Profession et catégorie socioprofessionnelle.....	20
f/ Cause initiale du décès.....	21
g/ Apgar à une minute	21
h/ Âge gestationnel et poids de naissance de l'enfant en grammes	21
i/ Type de Codage et score de confiance obtenu	22
2/ La table de l'ensemble des causes de décès KI_ECD_R	22
Précisions sur le numéro de ligne du certificat CER_LIG_NUM.....	24
3/ Les tables de valeur.....	25
Annexe 1 – fin de séjour hospitalier pour cause de décès dans le PMSI et DCIR.....	26
1/ Dans le PMSI annuel.....	26
2/ Dans DCIR.....	28
Annexe 2 – sélection des informations des bénéficiaires décédés dans le référentiel IR_BEN_R et le référentiel archivé IR_BEN_R_ARC	29
Annexe 3 - Volet médical du certificat général et néonatal pour la version 2017.....	30

Avant-propos

Par rapport à la version précédente, datée du 3 septembre 2020, cette version 5 est enrichie :

- des résultats de l'appariement des données de 2017 à 2021 selon l'appariement indirect avec les référentiels bénéficiaires (Cf. §II page 8),
- des nouvelles tables de valeurs, des tables de valeurs modifiées, des variables ajoutées ou supprimées suite au changement de la méthodologie du CépiDc pour le codage des causes de décès intégrant des modèles de machine learning (Cf. §IV page 15)
- d'une présentation rapide des campagnes de production 2018, 2019 et 2021, lesquelles reposent maintenant en partie sur l'usage de l'IA (réseaux de neurones) pour prédire les causes de décès de certains certificats, avec les références vers les documents détaillant ces approches (rapports de production de campagne)

Introduction

La mise en œuvre du Système National des Données de Santé (SNDS) est prévue par la loi N° 2016-41 et 2019-774 de modernisation de notre système de santé. Outre les données de remboursements des régimes obligatoires de sécurité sociale et les informations hospitalières du PMSI¹ qui composent le Sniiram², le SNDS doit être enrichi par les informations sur les causes médicales de décès, puis ultérieurement, par des données concernant les personnes handicapées.

Les données individuelles sur les causes médicales de décès sont élaborées par le centre d'épidémiologie sur les causes médicales de décès (CépiDc) qui est une unité de service de l'Inserm³, en collaboration avec l'Insee, à partir des certificats de décès remplis par les médecins constatant les décès. Il s'agit des décès survenus sur le territoire français, métropole et départements d'outre-mer.

Les informations sur les causes de décès disponibles au SNDS couvrent les années 2006 à 2021. Leurs mises à dispositions ont suivi le schéma suivant :

Année de décès concernée et date de mise à disposition des données dans le SNDS

2006 - 2012	Septembre 2020
2013 - 2015	Novembre 2017
2016	Septembre 2020
2017	Mai 2021 (provisaires) – Février 2022 (définitives)
2018 - 2019	Décembre 2023
2020	Février 2023
2021	Février 2024

Les données sur les causes médicales de décès intégrées au SNDS sont appariées aux référentiels des bénéficiaires du SNDS par une **méthode d'appariement indirect** décrite dans le deuxième chapitre de ce document.

Toutefois, les données de la totalité des causes de décès sont disponibles dans le SNDS, même si elles n'ont pas pu être appariées avec le référentiel IR_BEN_R. Vous pourrez donc établir des statistiques sur l'ensemble des décès, que les informations des personnes concernées soient ou non appariées avec le référentiel des bénéficiaires du SNDS.

Cet appariement indirect avec les informations des référentiels bénéficiaires dépend de la qualité des variables d'appariement et de la façon dont celles-ci remontent dans le SNDS (notamment le statut vital), variable selon régimes (voir chapitre 2). Les utilisateurs du SNDS doivent avoir en tête que **les seules causes de décès appariées ne sont pas représentatives de l'ensemble des causes de décès en population générale**. Pour avoir cette représentativité, il faut utiliser l'ensemble des décès (appariés ou non aux référentiels bénéficiaires du SNDS).

Cette documentation doit vous permettre d'exploiter les informations sur les causes médicales de décès intégrées au SNDS et sur les liens que vous pourrez établir entre ces informations et

¹ PMSI : programme médicalisé des systèmes d'informations, données produites par l'agence technique sur l'information hospitalière (ATIH).

² Sniiram : système national d'informations inter-régimes de l'assurance maladie.

³ Inserm : Institut national de la santé et de la recherche médicale, Insee : Institut national de la statistique et des études économiques.

les données de remboursements de soins et du PMSI. Vous pouvez rechercher des informations complémentaires sur le site Internet du CépiDc : <https://www.cepidc.inserm.fr/>

I/ Constitution des informations sur les causes médicales de décès

Ce paragraphe vous présente, de façon très résumée, le circuit des informations qui permettent au CépiDc de produire les données sur les causes médicales de décès. Ce circuit est plus développé dans le document de travail '[Les statistiques sur les causes médicales de décès de A à Z](#)' publiée sur le site du CépiDc.

Lors de la survenue d'un décès sur le territoire français, un certificat de décès est établi par un médecin. Ce certificat peut être établi sur **papier** ou de façon **électronique** sur un site Internet ou une application (CertDc) sécurisé mais les informations contenues dans ces deux types de certificats sont identiques.

Le certificat de décès est divisé en deux parties :

- une partie nominative, le volet administratif (VA), comportant notamment le nom et les prénoms de la personne décédée, son adresse, la date et l'heure de son décès ;
- une partie médicale, le volet médical (VM), non nominative, qui comporte les informations sur les causes médicales du décès. Les seules informations administratives qui figurent sur cette partie médicale sont la date de naissance, la date de décès, le sexe, le nom et le code postal de la commune de décès, le nom et le code postal de la commune de domicile.

Dans les certificats papier, la partie médicale est cachetée par le médecin afin d'en garantir la confidentialité.

- A noter qu'il existe aussi depuis 2018 et plus rarement un volet médical complémentaire (VMC), identique au volet médical (VM) du certificat et disponible en version électronique exclusivement. Il est renseigné quand les causes de décès n'étaient pas connues au moment de la première déclaration et que les résultats d'une recherche médicale ou médico-légale viennent préciser les causes du décès.

Si le VMC est plus complet que le premier VM transmis, on remplace le VM par le VMC. Si les deux volets se contredisent, seule l'information du VMC est retenue. Le CépiDc reçoit encore peu de volets médicaux complémentaires qui pourraient faire diminuer les décès de causes indéterminées ou d'intentions indéterminées dans les morts violentes.

Le circuit des informations dans le **cas d'un certificat papier** est alors le suivant :

- Le certificat de décès est envoyé à la commune de décès qui conserve la partie nominative ;
- Un bulletin de décès (B7) est établi par la mairie, il contient notamment les données d'état civil (sans le nom ni le prénom) de la personne décédée ainsi que sa date et son lieu de naissance, sa date et son lieu de décès, son département et sa commune de domicile, son état matrimonial, sa profession ou sa catégorie socioprofessionnelle ;
- Le bulletin de décès nominatif (B7bis), est envoyé à l'Insee, ce qui permet à l'Insee de rechercher le NIR de la personne décédée dans le répertoire national d'identification des personnes physiques (RNIPP), d'enrichir le RNIPP avec la date et la commune de décès

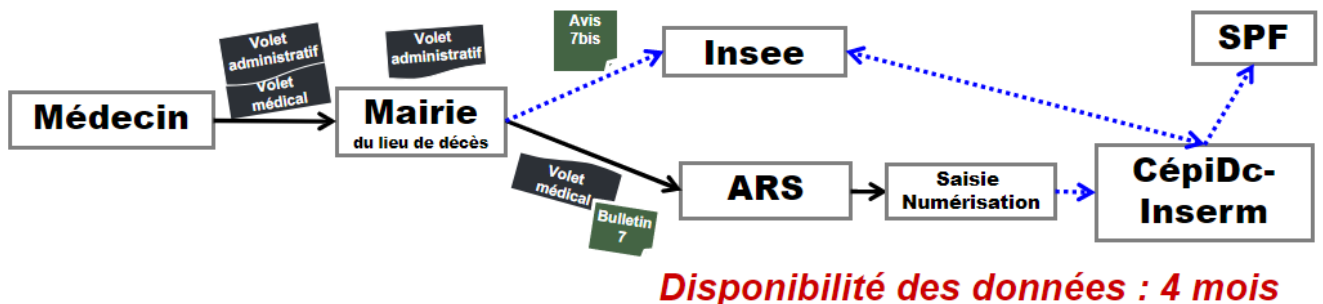
et de vérifier ou de corriger les informations administratives et socio-démographiques figurant sur le bulletin de décès ;

- Dans le même temps, la mairie adresse la partie médicale du certificat de décès et le bulletin de décès non nominatif (B7) à l'Agence Régionale de Santé (ARS) (ou sa délégation territoriale ou départementale). L'ARS ou la délégation territoriale envoie ensuite ces deux documents au CépiDc via un prestataire de saisie.

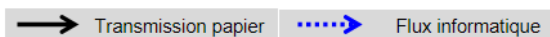
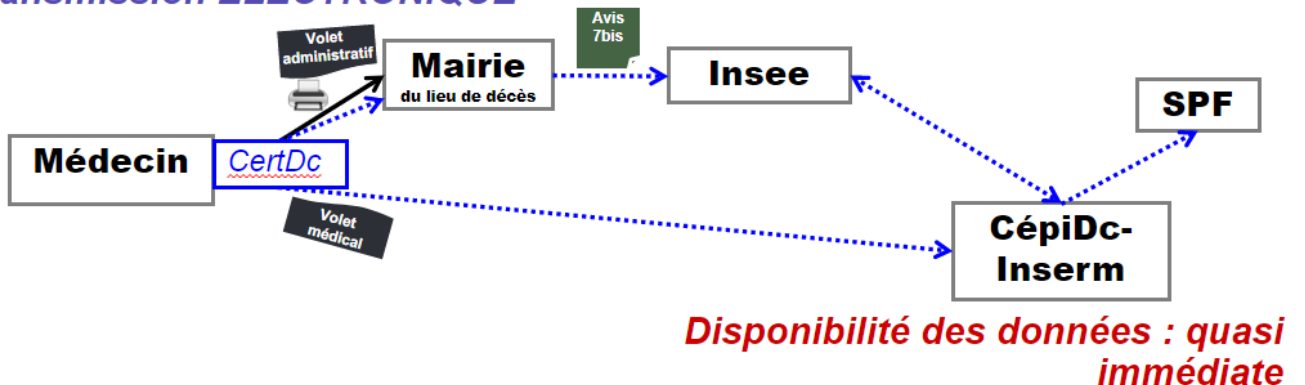
Le circuit des informations d'un **certificat électronique** est légèrement différent pour la partie administrative et complètement remanié pour le volet médical :

- Cette fois, seul le volet administratif est envoyé à la commune de décès, soit automatiquement et sous format dématérialisé pour les communes raccordées au Hub d'Échange de l'État (HubEE), soit sous format papier, mais imprimé par le médecin à partir de l'application CertDc, pour les autres communes ;
- L'étape d'envoi du bulletin de décès nominatif à l'Insee (B7bis) reste la même que pour le bulletin papier ;
- La mairie n'a plus à générer le bulletin de décès (B7) et n'intervient plus dans le circuit du volet médical. Les données du volet médical sont transférées quasiment instantanément depuis la plateforme CertDc directement vers les serveurs du CépiDc.

Transmission PAPIER



Transmission ELECTRONIQUE



1. Le CépiDc procède alors à deux opérations :

- il envoie à l’Insee les informations administratives dont il dispose, sans aucune donnée médicale, afin de permettre à l’Insee d’apparier ces informations avec les données des bulletins de décès et le RNIPP.
- il procède au codage des causes médicales de décès et à la détermination de la cause initiale du décès, cause à l’origine du processus morbide ayant conduit au décès selon la classification internationale des maladies et les recommandations de l’OMS.

Pour rattraper les années 2018/2019 et à partir de l’année 2021, le CépiDc a utilisé des algorithmes d’apprentissage profond (IA) pour coder une partie des causes de décès en complément du codage manuel et automatique (IRIS). Pour plus de détails, le site du CépiDc met à disposition [les rapports de productions](#)⁴ de chaque année et [les documents de travail sur les évolutions du codage](#)⁵.

2. L’Insee de son côté, envoie au CépiDc des informations administratives et socio-démographiques sur l’ensemble des décès déclarés à l’Insee qui permettent de dédoublonner, compléter et enrichir la base des causes médicales de décès. En particulier y sont ajoutées les caractéristiques des décès pour lesquels le CépiDc n’a pas reçu le volet médical (2% des décès).

Deux versions du certificat coexistent dans les déclarations des décès : celle de 1997 et celle de 2017 (entrée en vigueur le 1^{er} janvier 2018). L’application de certification électronique a mis en place la version 2017 du certificat le 02/01/2018 au matin. En papier, il persiste dans les faits une cohabitation des deux versions (1997/2017) pour des raisons pratiques et logistiques.

Le volet médical de la dernière version permet de renseigner plus de précisions sur les causes de décès, grâce à une partie consacrée aux informations complémentaires plus détaillée (voir annexe 3). Pour les certificats 1997, les variables liées aux informations complémentaires de 2017 sont juste vides dans les tables CépiDc du SNDS.

En plus des deux versions du certificat, il existe aussi deux types de certificat en fonction de l’âge du défunt : le certificat général et le certificat néonatal (voir annexe 3).

- Le type de **certificat néonatal** est spécifique pour les décès des enfants nés vivants (à plus de 22 semaines d'aménorrhée ou pesant plus de 500 g) et décédés au plus tard le 27^{ème} jour de leur vie. Son volet médical contient des informations complémentaires ciblées sur les parents et l’accouchement.
- Le **certificat général** (pour les décès à 28 jours et plus) a quant à lui des informations complémentaires sur l’activité professionnelle ou la grossesse au moment du décès et des précisions en cas de mort violente.

Le type de certificat utilisé pour constater le décès n’est pas une information suffisante pour repérer les décès néonataux à partir des certificats au format papier. En effet, un médecin peut

⁴ [Hebbache Z., Robert A., Clanché F., Coudin E., Martin D. - Rapport de production 2018/2019 – Données définitives – document de travail du CépiDc n3/2023](#)

⁵ [E. Zambetta, N. Razakamanana, A. Robert, F. Clanché, D. Martin, Z. Hebbache, R. Flicoteaux, and E Coudin. Codage des causes de décès de 2018 et 2019 en CIM 10 : Approche combinant deep learning, système expert et codage manuel ciblé - Combining a transformer-based approach, rule-based expert system, and targeted manual coding for icd-10 cause of death coding of french death certificates in 2018 and 2019. Technical report, CEPIDC Working Papers, n2, 2023, also accessible on researchgate DOI: 10.13140/RG.2.2.36127.33443/1](#)

remplir un certificat néonatal pour le décès d'un adulte, ou inversement, s'il n'a pas le bon type de certificat papier à sa disposition au moment où il rédige le certificat. Il est conseillé de **toujours vérifier l'âge du défunt en plus du type de certificat.**

Comme pour la version des certificats, les informations spécifiques au certificat néonatal sont manquantes pour les décès renseignés par le certificat général et inversement pour les informations spécifiques au certificat général.

II/ Appariement indirect entre les informations des causes de décès et les données de santé

Les informations de chaque décès en provenance du CépiDc sont caractérisées par un identifiant non nominatif nommé DCD_IDT_ENC. Il s'agit du numéro de certificat de décès pseudonymisé deux fois, une première fois par le CépiDc, une deuxième fois lors de l'arrivée des informations à la Cnam.

Dans le référentiel des bénéficiaires du SNDS, IR_BEN_R, un individu est caractérisé par son identifiant synthétique BEN_IDT_ANO. Cet identifiant est égal à :

- son NIR pseudonymisé BEN_NIR_ANO lorsqu'il est alimenté dans le SNDS,
- la concaténation de son identifiant SNIIRAM pseudonymisé BEN_NIR_PSA et de son rang de bénéficiaire BEN_RNG_GEM lorsque son NIR pseudonymisé BEN_NIR_ANO n'est pas alimenté.

En décembre 2023 (*pour les informations du référentiel IR_BEN_R dont la date maximale de traitement est supérieure ou égale au 01/01/2023*), le NIR pseudonymisé BEN_NIR_ANO est alimenté dans 98 % des enregistrements du référentiel IR_BEN_R.

Rappel :

L'identifiant SNIIRAM en clair d'un individu est égal à la concaténation du numéro de sécurité sociale de son ouvreur de droits, de sa propre date de naissance et de son propre code sexe. Cet identifiant est pseudonymisé dans le SNDS. Il est complété par le rang de naissance du bénéficiaire BEN_RNG_GEM qui permet de distinguer les personnes ayant le même identifiant SNIIRAM (jumeaux, triplés ... de même sexe en particulier). Un même individu a autant d'identifiants SNIIRAM qu'il a d'ouvreurs de droits.

En revanche, le NIR bénéficiaire est attribué par l'INSEE à chaque individu dès sa naissance et reste identique tout au long de sa vie. Un même individu a donc un seul NIR pseudonymisé BEN_NIR_ANO dans le SNDS.

Pour associer la consommation de soins des personnes décédées à leurs caractéristiques socio-démographiques et à leurs causes médicales de décès, on recourt actuellement à un **appariement indirect** pour faire correspondre un seul identifiant BEN_IDT_ANO des données de santé à un seul identifiant DCD_IDT_ENC du CépiDc.

La méthode d'appariement indirect entre les données de décès et d'autres systèmes d'informations a été mise au point par le CépiDc, et adaptée au SNDS en collaboration entre la Cnam et le CépiDc.

Les données sur lesquelles porte l'appariement sont l'année de décès, le mois de décès, le jour de décès, le code sexe, l'année de naissance, le mois de naissance, le département de résidence et la commune de résidence.

L'appariement utilise un algorithme qui se déroule en deux temps, en fonction des données disponibles. Dans son principe :

- il consiste à réaliser un appariement entre les données de décès et le référentiel IR_BEN_R sur tous les critères retenus, on parle d'appariement total sur les critères donnés en entrée ou, plus simplement, d'appariement total ;
- ensuite, pour les identifiants DCD_IDT_ENC de l'Inserm pour lesquels aucune correspondance avec un identifiant BEN_IDT_ANO du SNDS n'a été trouvée, un appariement est réalisé tour à tour sur tous les critères sauf un, à l'exception de l'année de décès qui doit toujours coïncider dans les deux sources de données (tous les critères sauf le code sexe, tous les critères sauf le département de résidence, ...). Ces appariements sont dits partiels ;
- lorsque l'appariement total ou les appariements partiels ont permis d'associer *un seul* identifiant BEN_IDT_ANO à *un seul* identifiant DCD_IDT_ENC, on considère que les informations sur les causes de décès peuvent être rapprochées des données de consommation de soins et on introduit l'identifiant BEN_IDT_ANO correspondant dans les tables des causes médicales de décès du SNDS.

L'algorithme décrit ci-dessus est utilisé lorsque **la date de décès du patient est renseignée dans le référentiel bénéficiaires IR_BEN_R ou IR_BEN_R_ARC** ⁶:

- pour les individus du référentiel IR_BEN_R et IR_BEN_R_ARC dont la date de décès est renseignée, on réalise d'abord un appariement total puis des appariements partiels sur l'année⁷, le mois, le jour de décès, le code sexe, l'année de naissance, le mois de naissance, le département de résidence et, dans certains cas, la commune de résidence.

Taux d'appariement indirect avec les référentiels bénéficiaires IR BEN R et IR BEN ARC pour les décès de l'année N :

2006 : 34,27%	2007 : 39,61%
2008 : 48,75%	2009 : 57,95%
2010 : 62,74%	2011 : 68,34%
2012 : 76,79%	2013 : 80,25%
2014 : 82,41%	2015 : 82,94%
2016 : 83,29%	2017 : 84,10%
2018 : 87,34%	2019 : 87,22%
2020 : 87,17%	2021 : 87,80%

Les taux d'appariement sont bien plus faibles pour les années les plus anciennes car les données sont moins fiables (ceci est dû en partie aux processus d'alimentation de la date de décès qui sont propres à chaque régime et qui ont évolué au cours du temps).

Il est possible d'améliorer le taux d'appariement en complétant les informations contenues dans le référentiel bénéficiaires IR_BEN_R et IR_BEN_R_ARC par :

⁶ Table contenant des données antérieures à celles de la table IR_BEN_R.

⁷ Rappel : l'année de décès figure dans tous les appariements partiels, variable bloquante.

- la date de fin de séjour hospitalier pour motif de décès
- le département de l'établissement géographique où a eu lieu ce séjour, provenant du PMSI ou le département des remboursements de soins de DCIR⁸.

A titre d'information, cet appariement supplémentaire a été fait sur les décès des années 2013, 2014 et 2015 et avait permis d'atteindre des taux d'appariement de l'ordre de 90%. (voir encadré n°1).

Mais, pour des raisons d'automatisation, l'appariement indirect des bases CépiDc utilisé pour le SNDS n'a été réalisé que sur les référentiels IR_BEN_R et IR_BEN_R_ARC, sans appel au DCIR ni au PMSI.

Encadré n°1 : Méthode d'appariement indirect des données avec le référentiel bénéficiaires IR_BEN_R et IR_BEN_R_ARC, DCIR et PMSI

Pour réaliser cet appariement, les informations du référentiel des bénéficiaires du SNDS IR_BEN_R et IR_BEN_R_ARC ont été enrichies par des informations en provenance :

- du PMSI : la date de fin de séjour hospitalier pour motif de décès et le département de l'établissement géographique où a eu lieu ce séjour,
- du DCIR : le département des remboursements de soins.

Le procédé utilisé pour rechercher ces informations supplémentaires est décrit en annexe 1 (page 26).

Cela permet d'ajouter des dates et département de décès dans l'appariement indirect pour les individus dont la date de décès n'est pas renseignée dans le référentiel bénéficiaires mais dont la date de fin de séjour hospitalier pour motif de décès est renseignée dans le PMSI.

On ajoute alors aux données sur lesquelles porte l'appariement :

- la date de décès correspondant à la date de fin de séjour hospitalier, lorsque le patient est décédé au cours d'un séjour hospitalier,
- et, lorsqu'il existe, le département de l'établissement hospitalier dans lequel le patient est décédé, en provenance du PMSI ou du DCIR, qui est comparé au département de décès contenu dans les données du CépiDc.

L'algorithme d'appariement total puis partiel est alors utilisé dans les deux étapes suivantes :

1^{ère} étape - la date de décès du patient est renseignée dans le référentiel IR_BEN_R :

- lorsque la date de fin de séjour pour cause de décès s'écarte d'au plus un jour de la date de décès, pour les individus du référentiel IR_BEN_R dont la date de décès est renseignée, on réalise un appariement total puis des appariements partiels sur l'année⁹, le mois, le jour de décès, le code sexe, l'année de naissance, le mois de naissance, le département de résidence et, dans certains cas, de la commune de résidence et le département de l'établissement où a eu lieu le décès (comparé au département de décès des données du CépiDc) ;
- pour les décès où cet appariement n'a pas réussi ainsi que pour les autres individus du référentiel IR_BEN_R dont la date de décès est renseignée (personnes décédées hors d'un établissement hospitalier ou dont la date de fin de séjour pour cause de décès n'a pas pu être trouvée), on réalise les appariements total puis partiel sur tous ces critères, à l'exception du département de décès et, dans certains cas, de la commune de résidence.

2^{ème} étape – la date de décès n'est pas renseignée dans le référentiel IR_BEN_R mais la date de fin de séjour pour cause de décès est connue :

⁸ DCIR : datamart de consommation inter-régimes, contenant les remboursements des régimes obligatoires d'assurance maladie, à l'exception des séjours hospitaliers publics.

⁹ Rappel : l'année de décès figure dans tous les appariements partiels, variable bloquante.

La date de décès n'est pas toujours renseignée dans la table IR_BEN_R parce que, pour des raisons d'ordre technique, certains régimes ou organismes de sécurité sociale ne l'alimentent pas dans le SNDS. Dans un tel cas, elle est forcée au 1^{er} janvier 1600. Par contre, en raison des résultats exposés en préambule, si une personne est décédée au cours d'un séjour hospitalier, on peut supposer que sa date de fin de séjour s'écarte d'au plus un jour de sa date de décès dans plus de 99 % des cas.

On va alors remplacer la date de décès de la table IR_BEN_R par la date de fin de séjour hospitalier, puis réaliser les appariements total puis partiels sur l'année, le mois, le jour de décès, le code sexe, l'année de naissance, le mois de naissance, le département de résidence, la commune de résidence et le département de l'établissement hospitalier (comparé au département de décès des données du CépiDc).

Attention :

Les données qui ont été appariées au cours de la seconde étape correspondent à des personnes décédées à l'hôpital et dont la date de décès n'est pas alimentée dans le référentiel IR_BEN_R. La raison principale de cette absence d'alimentation dans IR_BEN_R est que, pour des raisons de norme d'échange des données, le régime ou l'organisme de sécurité sociale de ces personnes ne transmet pas la date de décès au SNDS.

Si vous réalisez des études sur ces personnes, vous devez garder présent à l'esprit que la totalité des décès provenant de leur régime ou organisme de sécurité sociale n'a pas pu être appariée avec les données du CépiDc, d'une part parce qu'elles ne sont pas toutes décédées à l'hôpital, d'autre part parce que la date de fin d'hospitalisation pour cause de décès n'a pas toujours pu être trouvée (Cf. annexe 1 page 26).

A titre d'information, pour les décès survenus en 2013, sur l'ensemble des deux étapes d'appariement, le taux d'appariement obtenu par cette méthode monte à 93,8 % des bénéficiaires éligibles du référentiel du SNDS, contre 88,7 % des décès recensés par le CépiDc avec l'appariement réalisé sur le référentiel des bénéficiaires du SNDS uniquement. Les proportions correspondantes sont respectivement de 93,9 % contre 90,1 % pour 2014, et de 94 % contre 90,1 % pour 2015.

III/ Comment relier les données de santé aux causes médicales de décès

Les informations sur les causes médicales de décès sont restituées dans deux tables du répertoire « ORAVUE » du SNDS, elles sont accessibles via SAS Enterprise Guide :

- la table des circonstances et de la cause initiale du décès « KI_CCI_R »,
- la table de l'ensemble des causes de décès « KI_ECD_R ».

Les données contenues dans chaque table sont présentées dans le §IV (page 15).

Lorsqu'un décès, identifié par son code DCD_IDT_ENC, a pu être apparié avec les données de santé, les tables KI_CCI_R et KI_ECD_R contiennent :

- l'identifiant BEN_IDT_ANO du bénéficiaire,
- le top associé à cet identifiant BEN_IDT_TOP, qui vaut 1 si l'identifiant BEN_IDT_ANO est égal au NIR pseudonymisé du patient BEN_NIR_ANO, 0 sinon,
- le NIR pseudonymisé du bénéficiaire BEN_NIR_ANO, renseigné uniquement quand BEN_IDT_TOP = 1.

En revanche, lorsqu'un décès n'a pas pu être apparié avec les données de consommation de soins, les champs BEN_IDT_ANO, BEN_IDT_TOP et BEN_NIR_ANO ne sont pas renseignés dans les deux tables des causes de décès.

Un top a été alimenté pour vous permettre de distinguer facilement les décès appariés avec les données de santé des décès non appariés. Ce top, noté « DCD_IDT_TOP », vaut 1 lorsque le décès est apparié, 0 sinon.

Le fait de disposer des informations de l'ensemble des décès vous permet d'une part, de calculer des statistiques globales sur les causes médicales de décès (âge, sexe, pathologie, ...) et d'autre part, de mesurer les biais engendrés par l'utilisation d'une méthode d'appariement indirect pour relier les causes médicales de décès aux données de consommation de soins et par la non exhaustivité de l'alimentation de la date de décès dans la table IR_BEN_R. Vous devrez **tenir compte de ces biais** et chercher à les corriger lorsque vous ferez des études sur le parcours et la consommation de soins des personnes décédées.

Le champ BEN_IDT_ANO sert à identifier les patients dans l'application DCIR simplifié¹⁰ (tables NS_XXX_F) et le référentiel associé IR_IBA_R. Toutefois, pour l'instant, l'historique des soins retracés dans cette application démarre au 1^{er} janvier 2016, ce qui ne vous permet pas d'étudier la consommation de soins des personnes décédées avant cette date.

Le champ BEN_IDT_ANO figure également dans le référentiel des bénéficiaires du SNDS IR_BEN_R et dans le référentiel archivé IR_BEN_R_ARC, ces deux tables vous permettent d'associer à l'identifiant BEN_IDT_ANO d'un patient, tous les couples [identifiant Sniiram BEN_NIR_PSA ; rang de bénéficiaire BEN_RNG_GEM} associés à ses différents ouvreurs de droits¹¹. En effet, une même personne, caractérisée par son identifiant BEN_IDT_ANO, a autant de couples {identifiant Sniiram ; rang} qu'elle a d'ouvreurs de droits (Cf. rappel en page 8).

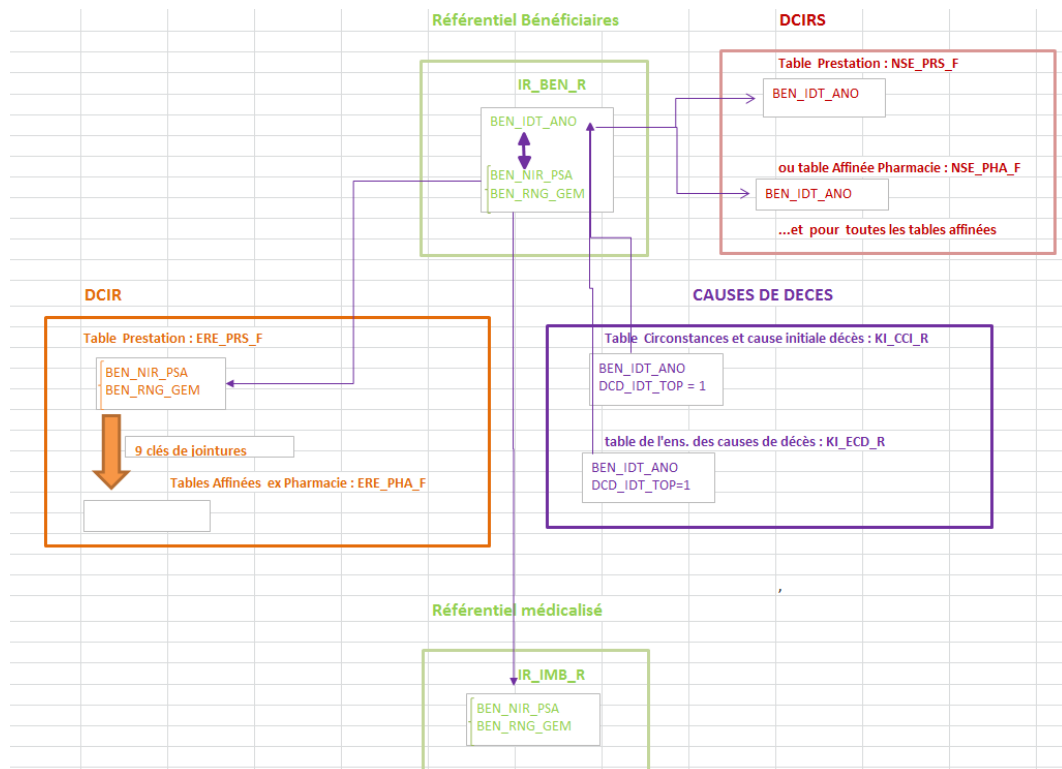
¹⁰ Vous pouvez vous reporter à la liste des tables et des variables ainsi qu'au support de formation de l'application DCIR simplifié (DCIRS), en ligne sur le portail Sniiram.

¹¹ Pour plus d'informations sur le référentiel IR_BEN_R, vous pouvez vous reporter à la documentation destinée aux utilisateurs du référentiel des bénéficiaires, en ligne sur le portail du Sniiram.

Les couples {BEN_NIR_PSA ; BEN_RNG_GEM} du référentiel IR_BEN_R et, le cas échéant, du référentiel archivé IR_BEN_R_ARC, vont vous permettre de rechercher les informations médicalisées sur les pathologies ayant donné lieu à une exonération des patients pour affection de longue durée (ALD) contenues dans le référentiel médicalisé IR_IMB_R, ainsi que les données de consommation de soins contenues dans les remboursements de soins de DCIR. L'identifiant Sniiram BEN_NIR_PSA vous permet de faire le chaînage avec les informations hospitalières du PMSI.

1/ Recherche des informations dans le référentiel médicalisé IR_IMB_R et DCIR

Schéma 1 – recherche des informations dans DCIR et le référentiel IR_IMB_R



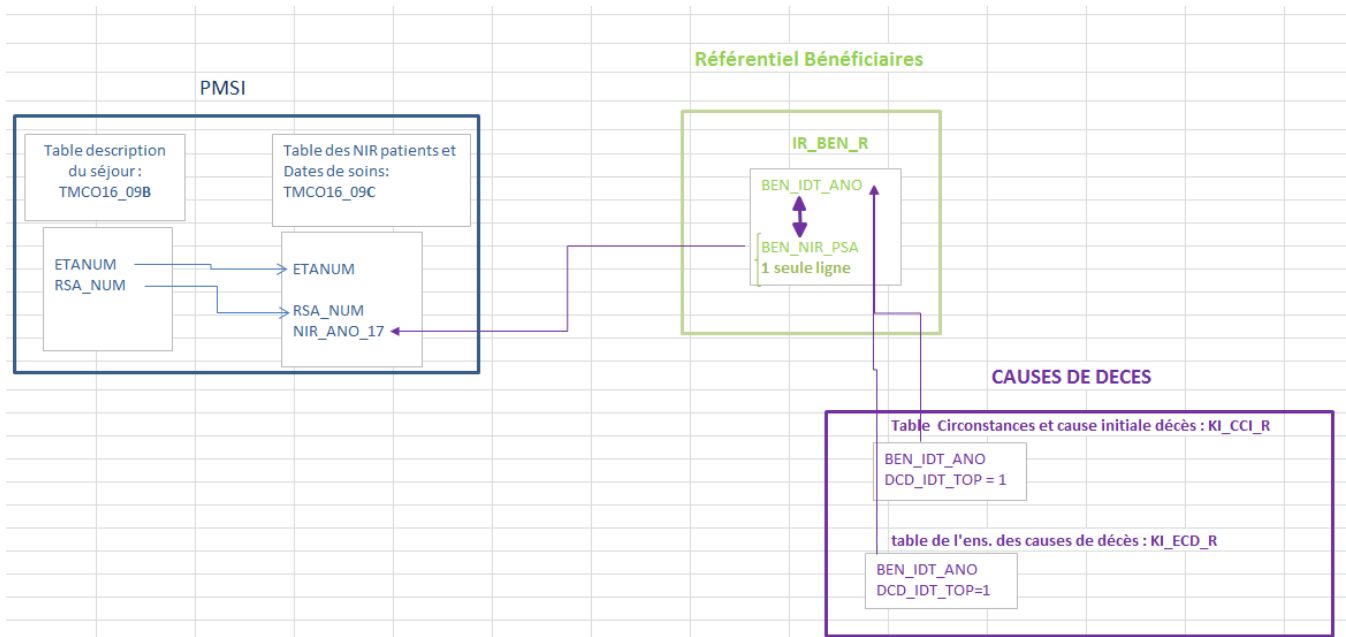
- Pour les décès appariés avec le SNDS (DCD_IDT_TOP = 1 dans la table des circonstances et de la cause initiale du décès KI_CCI_R¹²), vous recherchez tous les couples {BEN_NIR_PSA ; BEN_RNG_GEM} associés à leur identifiant BEN_IDT_ANO dans le référentiel IR_BEN_R. Si le champ BEN_IDT_ANO ne figure pas dans IR_BEN_R, vous le recherchez dans le référentiel archivé IR_BEN_R_ARC (Cf. annexe 2 page 29). Vous appariez donc la table KI_CCI_R et le référentiel IR_BEN_R + IR_BEN_R_ARC à l'aide de l'identifiant BEN_IDT_ANO et vous constituez une table de correspondance entre les champs BEN_IDT_ANO, BEN_NIR_PSA et BEN_RNG_GEM. Vous devez veiller à ce que cette table de correspondance ne contienne qu'une seule ligne par couple {BEN_NIR_PSA ; BEN_RNG_GEM}, afin de ne pas dupliquer les données de consommation de soins ou les données médicalisées que vous allez rechercher ensuite ;

¹² Le champ DCD_IDT_TOP figure également dans la table de l'ensemble des causes de décès KI_ECD_R mais cette table peut contenir plusieurs lignes pour un même décès. En revanche, la table KI_CCI_R contient une seule ligne par décès, ce qui simplifie les choses.

- pour les couples [BEN_NIR_PSA ; BEN_RNG_GEM] de votre table de correspondance, vous allez rechercher les informations que vous souhaitez dans le référentiel IR_IMB_R ou DCIR ;
- vous pouvez ensuite rassembler les informations que vous avez sélectionnées pour chaque identifiant BEN_IDT_ANO et les rapprocher des données des causes médicales de décès au moyen de ce champ BEN_IDT_ANO.

2/ Recherche des informations dans le PMSI

Schéma 2 – recherche des informations dans le PMSI



Vous devez garder présent à l'esprit que le rang de bénéficiaire n'est pas toujours alimenté dans le PMSI, surtout avant 2015. Il vous est donc impossible de distinguer les jumeaux, triplés, ... de même sexe dans le PMSI car ils ont même identifiant Sniiram mais vous ne connaissez pas leur rang :

- vous constituez une table à partir du référentiel IR_BEN_R + IR_BEN_R_ARC en sélectionnant les lignes pour lesquelles un même identifiant BEN_NIR_PSA n'y figure qu'une seule fois (exclusion des jumeaux, triplés, ...) ;
- pour les décès appariés avec les données de santé (DCD_IDT_TOP = 1 dans la table KI_CCI_R), vous constituez une table de correspondance entre l'identifiant BEN_IDT_ANO de la table KI_CCI_R et l'identifiant SNIIRAM BEN_NIR_PSA de la table que vous avez extraite du référentiel IR_BEN_R + IR_BEN_R_ARC ;
- vous recherchez ensuite dans le PMSI, les données correspondant aux identifiants BEN_NIR_PSA que vous avez sélectionnés (le champ BEN_NIR_PSA des tables IR_BEN_R et IR_BEN_R_ARC se nomme NIR_ANO_17 dans le PMSI) ;
- vous rassemblez ensuite les informations recueillies pour chaque identifiant BEN_IDT_ANO et vous les rapprochez des données sur les causes médicales de décès.

IV/ Description des données

A partir des éléments fournis par le médecin qui a délivré le certificat de décès, le CépiDc détermine la cause initiale du décès. Cette cause initiale est renseignée dans la table KI_CCI_R, accompagnée d'informations socio-démographiques et médicales. La cause initiale du décès ainsi que toutes les autres causes, s'il en existe, sont restituées dans la table de l'ensemble des causes de décès KI_ECD_R.

Ces deux tables sont disponibles dans le répertoire « ORAVUE » du SNDS, vous pouvez les exploiter via SAS Enterprise Guide. Le document « Causes Médicales de décès - liste des profils d'accès et des identifiants potentiels - 14 novembre 2023.pdf » mis en ligne sur Ameli Réseau et dans le portail Sniiram vous donne la liste des identifiants potentiels ainsi que les profils du SNDS sous lesquels vous pouvez exploiter les données sur les causes de décès. Selon le profil, certains identifiants potentiels ne seront pas accessibles (commune de décès, dates précise des soins, commune de résidence, mois de naissance, date précise du décès).

Afin de vous aider à exploiter ces deux tables, des tables de valeurs sont à votre disposition dans le répertoire « ORAVAL ». Elles contiennent les valeurs et les libellés des variables correspondantes. Les tables spécifiques aux causes médicales de décès commencent par « DS », les autres tables commencent par « IR » ou « MS » et sont également utilisées pour les données de consommation de soins.

1/ La table des circonstances et de la cause initiale du décès KI_CCI_R

La table des circonstances et de la cause initiale de décès KI_CCI_R contient une ligne par décès, une ligne par identifiant DCD_IDT_ENC. Lorsque les décès sont appariés avec les données de santé (DCD_IDT_TOP = 1), cette table contient une ligne par identifiant BEN_IDT_ANO.

Pour plus de détails sur les spécificités du traitement d'une année donnée, le lecteur peut se référer aux rapports de production du CépiDc – [Par exemple pour 2018 et 2019](#)¹³,

¹³ [Hebbache Z., Robert A., Clanché F., Coudin E., Martin D. - Rapport de production 2018/2019 – Données définitives – document de travail du CépiDc n3/2023](#)

Tableau 1 - Liste des informations de la table KI_CCI_R

	Code	Libellé	Format	Tables de valeurs	Changement à la mise à disposition 2023
Identifiants Bénéficiaire	DCD_IDT_ENC	Identifiant décès encodé	car(64)	-	
	DCD_IDT_TOP	Top apparié avec IR_BEN_R	num(1)	-	
	BEN_IDT_ANO	Identifiant synthétique SNDS	car(18)	-	
	BEN_IDT_TOP	Top identifiant synthétique SNDS	num(1)	-	
	BEN_NIR_ANO	NIR pseudonymisé du bénéficiaire	car(17)	-	
Tout certificat	FLX_PER_ANN	Année de décès	car(4)	-	
	CER_VER_NUM	Version du certificat	num(1)	DS_VER_V	
	TRT_STA_COD	Statut de traitement	num(1)	DS_STA_V	
	TYP_CER_COD	Type de certificat	num(1)	DS_CER_V	
	CER_SUP_TYP	Type de support	num(1)	DS_SUP_V	
	TYP_VOL_COD	Type de volet	num(1)	DS_VOL_V	
	DCD_DPT_COD	Département de décès	car(3)	IR_DPT_V	
	DCD_COM_COD	Commune de décès	car(3)	IR_GEO_V	
	BEN_RES_DPT	Département de domicile	car(3)	IR_DPT_V	
	BEN_RES_COM	Commune de domicile	car(3)	IR_GEO_V	
	BEN_DCD_DTE	Date du décès	date		
	BEN_DCD_AME	Année et mois du décès	car(6)	AAAAMM	
	DCD_LIE_COD	Lieu du décès	num(1)	DS_LIE_V	
	BEN_NAI_ANN	Année de naissance	car(4)		
	BEN_NAI_MOI	Mois de naissance	car(2)		
BEN_SEX_COD	Sexe	num(1)	IR_SEX_V		

	PFV_ACP_COD	Activité professionnelle	num(1)	DS_ACP_V	
	CAT_PCS_COD	Profession et catégorie socioprofessionnelle	car(4)	DS_PC1_V, DS_PC2_V, DS_PC3_V	
	ETA_MAR_COD	Etat matrimonial	num(1)	DS_MAR_V	
	DCD_CIM_COD	Cause initiale du décès	car(4)	IR_CCL_V, IR_CIM_V, MS_CIM_V	
	DCD_CAU_COD	Recherche de la cause de décès	num(1)	DS_CAU_V	
	DCD_CIM_VER	Version CIM utilisée pour coder les causes médicales de décès	num(1)	DS_CIM_V	
	DCD_CIR_COD	Circonstance apparente du décès	num(1)	DS_CIR_V	
	DCD_MSU_COD	Mort subite	num(1)	DS_MSU_V	
	CER_PRD_TYP	Type de codage utilisé	num(1)	DS_PRD_V	Variable Nouvelle pour 2018, 2019 et 2021
	PRD_SCR_COF	Score de confiance obtenu	num(5)		Variable Nouvelle pour 2018, 2019 et 2021
Certificat Adulte	DCD_GRS_COD	Contribution de la grossesse au décès	num(1)	DS_GRS_V	
	DCD_GRS_DEL	Délai entre fin de grossesse et décès	car(4)		
	DCD_ATT_COD	Accident du travail ou Activité professionnelle	num(1)	DS_ATT_V	
	DCD_ENC_COD	Statut de la grossesse si femme décédée	num(1)	DS_ENC_V	
	DCD_FIA_COD	Fiabilité de la date de décès	num(1)	DS_FIA_V	
	LIE_LMV_COD	Codification du lieu de décès si mort violente	num(1)	DS_LMV_V	
Certificat Néonatal	NEO_APG_SCO	Apgar à une minute	num(2)		
	NEO_GES_AGE	Âge gestationnel en semaines révolues d'aménorrhée	num(2)		
	NEO_NAI_POI	Poids de naissance en grammes	num(4,0)		
	RNG_NAI_TYP	Type de naissance	num(1)	DS_NAI_V	
	GRS_ORD_NUM	N° d'ordre de l'enfant si grossesse multiple	num(1)		

ACC_LIA_COD	Lieu d'accouchement	num(1)	DS_LIA_V	
ACC_PST_COD	Présentation de l'enfant	num(1)	DS_PST_V	
ACC_DEB_COD	Début du travail	num(1)	DS_DEB_V	
MOD_ACC_COD	Mode d'accouchement	num(1)	DS_ACC_V	
HOS_TRF_TOP	Transfert ou hospitalisation particulière de l'enfant	num(1)	DS_TRF_V	
MER_NAI_ANN	Année de naissance de la mère	car(4)		
MER_PFS_COD	Activité professionnelle de la mère	num(1)	DS_PFS_V	
MER_PFG_LIB	Profession de la mère exercée pendant la grossesse	car(50)		
MER_MAR_COD	Etat matrimonial de la mère	num(1)	DS_MAR_V	
MER_SIT_TOP	Top mère en couple	num(1)	DS_SIT_V	
TOT_GRS_NBR	Nombre total de grossesses	num(2)		
TOT_ACC_NBR	Nombre total d'accouchements	num(2)		
PER_PFS_COD	Activité professionnelle du père	num(1)	DS_PFS_V	
PER_PFG_LIB	Profession du père exercée pendant la grossesse	car(50)		
DCD_NEO_AGE	Mortalité néonatale groupe d'âge	num(1)	DS_AGE_V	
DCD_NRS_COD	Mort inattendue du nourrisson	num(1)	DS_NRS_V	

a/ Rappel sur les identifiants des bénéficiaires

- le champ DCD_IDT_ENC permet d'identifier chaque décès envoyé par l'Inserm, qu'il ait ou non pu être apparié avec les données de santé. Il correspond au numéro de certificat de décès pseudonymisé ;
- le top DCD_IDT_TOP vaut 1 si le décès a été apparié avec le référentiel IR_BEN_R / IR_BEN_R_ARC, 0 sinon ;
- l'identifiant synthétique du bénéficiaire dans le SNDS BEN_IDT_ANO est renseigné dans la table KI_CCI_R lorsque le décès a pu être apparié avec le référentiel des bénéficiaires IR_BEN_R / IR_BEN_R_ARC (DCD_IDT_TOP = 1).

Il correspond :

- au NIR pseudonymisé du bénéficiaire BEN_NIR_ANO lorsqu'il est alimenté dans le référentiel des bénéficiaires,
- à la concaténation de l'identifiant Sniiram pseudonymisé du bénéficiaire BEN_NIR_PSA et du rang de bénéficiaire BEN_RNG_GEM lorsque le NIR pseudonymisé BEN_NIR_ANO n'est pas alimenté dans le référentiel des bénéficiaires (Cf. §II page 9) ;
- le top BEN_IDT_TOP associé à l'identifiant synthétique BEN_IDT_ANO vaut 1 lorsque BEN_IDT_ANO = BEN_NIR_ANO, 0 sinon ;
- le NIR pseudonymisé du bénéficiaire BEN_NIR_ANO est alimenté pour les décès appariés avec les données de santé, lorsqu'il est renseigné dans le référentiel IR_BEN_R / IR_BEN_R_ARC (DCD_IDT_TOP = 1 et BEN_IDT_TOP = 1).

b/ Version du certificat, statut de traitement, type de certificat, type de support, volet du certificat

- pour les décès antérieurs à 2017, la version de tous les certificats CER_VER_NUM est celle de 1997. Un nouveau certificat de décès est disponible à partir de l'année 2017 pour 6 communes faisant partie de l'expérimentation de dématérialisation totale du certificat de décès. Le nouveau modèle de certificat est entré en vigueur à partir du 1^{er} janvier 2018 et a progressivement remplacé le certificat de 1997. Sur support électronique, la bascule a été effective au 2 janvier 2018. Le support papier peut quant à lui encore contenir des anciens modèles, de moins en moins au fur et à mesure des années.
- pour l'instant, les informations de décès intégrées au SNDS sont des données annuelles consolidées par l'Inserm et l'Insee, le statut de traitement TRT_STA_COD est toujours final.
- le type de certificat de décès TYP_CER_COD est destiné à distinguer les décès d'adultes des décès néonataux. Les décès néonataux concernent les enfants nés vivants et décédés au plus tard le 28^{ème} jour de leur vie.

En réalité, lorsque le médecin qui rédige le certificat de décès ne dispose que d'une version sur papier, il ne dispose pas toujours de la version adéquate du certificat (adulte ou néonatal). C'est pourquoi, certains décès considérés comme des décès d'adultes (TYP_CER_COD = 1 dans KI_CCI_R) sont en réalité des décès néonataux et certains décès considérés comme néonataux (TYP_CER_COD = 2) sont des décès d'adultes. Par conséquent, si vous souhaitez réaliser des études spécifiques aux décès néonataux, veuillez-vous rapprocher du CépiDc.

- le type de support CER_SUP_TYP permet de distinguer les certificats sur papier des certificats électroniques. La proportion de certificats électroniques augmente au cours du temps (de 4,9 % en 2013 à 43% fin 2023), cette tendance devrait continuer dans les années postérieures ;
- la variable TYP_VOL_COD concernant le type de volet du certificat de décès n'a été introduite qu'en 2017 dans les bases de données du CépiDc, ce qui explique que cette variable est très rarement renseignée pour l'instant. Tant que le nouveau certificat de décès n'est pas utilisé, le type de volet du certificat TYP_VOL_COD ne peut prendre que la valeur 1 (initial). Un volet complémentaire (TYP_VOL_COD = 2) pourra être utilisé dans les nouveaux certificats, mis en place progressivement à partir de 2018.

c/ Département et commune de décès ou de résidence

- La table de valeurs du code des départements IR_DPT_V prévoit qu'un département est codé sur trois positions : 001, 002 par exemple. Pour la Corse, la codification adoptée dans cette table est 201 pour la Corse du Sud et 202 pour la Haute Corse, au lieu de 02A et 02B. Pour les départements d'outre-mer (DOM), le code du département est 097.

La codification de la table IR_DPT_V a été appliquée au département de décès DCD_DPT_COD et au département de résidence BEN_RES_DPT de la table KI_CCI_R. Par conséquent, les départements de Corse sont codés 201 et 202 et les DOM sont codés 097.

- Le libellé des communes se trouve dans la table de valeurs IR_GEO_V mais nous vous rappelons que, pour trouver le nom d'une commune, il faut toujours indiquer le code du département dans lequel elle se situe. Le code du département figure également dans la table IR_GEO_V.

d/ Date de décès

La date BEN_DCD_DTE est la date complète du décès, année, mois et jour. Pour les utilisateurs qui ne sont pas autorisés à croiser les identifiants potentiels, le champ BEN_DCD_AME leur fournit l'année et le mois du décès quel que soit le profil utilisé.

e/ Profession et catégorie socioprofessionnelle

Lorsqu'il est renseigné, le code CAT_PCS_COD de la profession ou catégorie socioprofessionnelle (PCS) se rapporte à l'activité d'une personne au moment de son décès. Le code de la PCS n'est jamais renseigné pour les retraités ou les autres inactifs, il commence toujours par 1, 2, ... 6.

La codification des PCS est définie par l'Insee, c'est également l'Insee qui saisit ce code dans les données des causes médicales de décès.

La codification des PCS à deux caractères est hiérarchisée :

- le premier caractère de gauche du champ CAT_PCS_COD correspond à la codification en 8 postes, listée dans la table de valeurs DS_PC1_V. Si vous souhaitez connaître le libellé de ces PCS pour les personnes de la table KI_CCI_R, vous appariez le premier caractère du code CAT_PCS_COD (substr(CAT_PCS_COD,1,1) avec le code CAT_PC1_CH1 de la table de valeurs DS_PC1_V et vous listez le libellé correspondant CH1_PC1_LIB ;
- les PCS de la table KI_CCI_R sont saisis sur 2 caractères, suivant la nomenclature de l'INSEE en 42 postes listée dans la table DS_PC3_V. Si vous voulez connaître le libellé de ces PCS, vous pouvez donc apparier le code CAT_PCS_COD de la table KI_CCI_R avec le code CAT_PC3_CH3 de la table de valeurs DS_PC3_V et éditer les libellés correspondants CH3_PC3_LIB.

La table de valeurs DS_PC2_V correspond à la codification des PCS suivant la nomenclature de l'INSEE en 24 postes (codes saisis sur 2 caractères). Cette nomenclature n'a pas été utilisée par l'INSEE pour les causes de décès mais la table de valeurs DS_PC2_V pourra peut-être servir pour les décès d'autres années.

N.B. : le format du code de la PCS a été corrigé dans les tables de valeurs DS_PC1_V, DS_PC2_V et DS_PC3_V : il est maintenant alphanumérique, comme cela est également le cas dans la table KI_CCI_R. L'appariement entre la table KI_CCI_R et les tables de valeurs DS_PC1_V ... est donc simplifié.

f/ Cause initiale du décès

La cause initiale de décès est définie par l'OMS comme « a) la maladie ou le traumatisme qui a déclenché l'évolution morbide conduisant directement au décès, ou b) les circonstances de l'accident ou de la violence qui ont entraîné le traumatisme mortel ». Le code de la cause initiale du décès DCD_CIM_COD se conforme à la 10ème version révision de la classification internationale des maladies (CIM10) réalisée maintenue par l'organisation mondiale de la santé (OMS).

Ce code est hiérarchisé :

- les trois premiers caractères de gauche du code DCD_CIM_COD correspondent à la classe de la pathologie au sens de l'OMS. En appariant ces trois premiers caractères de gauche au code CAT_CIM_COD de la table de valeurs des classes de pathologies IR_CCI_V [substr(DCD_CIM_COD,1,3) = CAT_CIM_COD], vous pouvez connaître le libellé correspondant de la classe de pathologie CAT_CIM_LIB ;
- pour connaître le libellé de la pathologie elle-même DCD_CIM_COD, vous appariez ce champ avec le code CIM_COD de la table de valeurs des pathologies IR_CIM_V. Vous pouvez également rechercher ces libellés dans la table MS_CIM_V qui contient la codification CIM10 des diagnostics du PMSI.

La valeur « R99 » du code de la cause initiale de décès DCD_CIM_COD ne figure ni dans la table IR_CIM_V, ni dans la table MS_CIM_V. Elle correspond à une pathologie inconnue, ce qui se produit notamment lorsque le CépiDc ne reçoit pas le volet médical du certificat de décès ou que les informations fournies par le médecin ne permettent pas d'identifier une cause initiale de décès.

g/ Apgar à une minute

Le score d'Apgar mesure la vitalité d'un nouveau-né au moment de sa naissance. Ce score varie de 0 à 10, plus il est élevé, meilleur est l'état de santé du bébé. Ce score résulte de l'évaluation du rythme cardiaque, de la respiration, du tonus, de la couleur de la peau et de la réactivité de l'enfant. Pour plus de détails sur ce score, veuillez-vous rapprocher d'un médecin.

h/ Âge gestationnel et poids de naissance de l'enfant en grammes

Selon les recommandations de l'OMS, les enfants nés après 22 semaines révolues d'aménorrhée ou pesant au moins 500 grammes sont présumés viables, indépendamment de tout autre critère.

Dans la table KI_CCI_R, lorsque l'âge gestationnel NEO_GES_AGE est d'au moins 22 semaines ou lorsque le poids de naissance de l'enfant NEO_NAI_POI est d'au moins 500 grammes, on peut considérer qu'il s'agit d'un décès néonatal. Au-dessous de ces deux seuils, on considère qu'il s'agit d'un enfant mort-né mais pas d'un décès néonatal. Si vous souhaitez plus d'information, vous pouvez activer le lien suivant sur le site du Sénat :

<https://www.senat.fr/lc/lc184/lc1840.html>

i/ Type de Codage et score de confiance obtenu

La variable CER_PRD_TYP permet de connaître la **méthode de codage utilisée** pour chacun des décès. Elle est **renseignée à partir de l'année 2018 (hors 2020)** et prend les **valeurs de 1 à 6**, modalités contenues dans la table des valeurs DS_PRD_V disponible dans ORAVAL.

Pour les années 2018, 2019 et à partir de 2021 le CépiDc a **utilisé des prédictions par des algorithmes d'apprentissage profond (codage IA) en combinaison avec les traditionnels codages manuel et automatique** du logiciel IRIS des autres années. Le type de codage impliquant de l'IA prend les valeurs de 3 à 6 en fonction des différents modèles IA utilisés et retenus.

Les répartitions des modes de codage et les spécificités des algorithmes utilisés sont décrits dans les rapports de production disponibles :

<https://www.cepidc.inserm.fr/documentation/rapport-de-production-annees-de-deces-2018-et-2019-donnees-definitives-document-de-travail-du-cepidc-n32023>

A noter que le type de codage CER_PRD_TYP n'est pas renseigné dans le cas des décès pour lesquels le CépiDc n'a pas reçu le volet médical du certificat de décès. Dans ce cas, la cause initiale de décès DCD_CIM_COD est renseignée à 'R99' dans la table KI_CCI_R.

Pour les types de codage impliquant l'IA, un **score de confiance est calculé pour indiquer le niveau de confiance que l'on a dans la prédiction de la cause initiale** prédite par l'IA. Ce score de confiance est stocké dans la variable PRD_SCR_COF, en numérique et sur 5 caractères, **allant de 0.000 à 1.000**.

Plus cette valeur se rapproche de 1, plus on considère que la prédiction de la cause initiale a des chances d'être la même que celle obtenue par application du logiciel Iris /Muse ou par codage manuel.

Plus de détails sur ce score de confiance et plus généralement sur l'utilisation de réseaux de neurones pour prédire une partie des causes de décès se trouvent dans le document de travail : Codage des causes de décès de 2018 et 2019 en CIM 10 : Approche combinant deep learning, système expert et codage manuel ciblé¹⁴.

2/ La table de l'ensemble des causes de décès KI_ECD_R

La table de l'ensemble des causes de décès KI_ECD_R se rapporte aux mêmes identifiants de décès DCD_IDT_ENC que ceux de la table des circonstances et de la cause initiale du décès

¹⁴ [E. Zambetta, N. Razakamanana, A. Robert, F. Clanché, D. Martin, Z. Hebbache, R. Flicoteaux, and E Coudin. Codage des causes de décès de 2018 et 2019 en CIM 10 : Approche combinant deep learning, système expert et codage manuel ciblé - Combining a transformer-based approach, rule-based expert system, and targeted manual coding for icd-10 cause of death coding of french death certificates in 2018 and 2019. Technical report, CEPIDC Working Papers, n2, 2023, also accessible on researchgate DOI: 10.13140/RG.2.2.36127.33443/1](#)

KI_CCI_R. Tout identifiant DCD_IDT_ENC de l'une des deux tables se retrouve également dans l'autre.

Contrairement à la table KI_CCI_R qui ne contient qu'une seule ligne par décès, la table KI_ECD_R peut contenir plusieurs lignes pour un même identifiant de décès DCD_IDT_ENC. En effet, elle comporte une ligne pour chaque cause de décès déclarée par le médecin qui a rempli le certificat de décès. En pratique, la table KI_ECD_R contient une ligne par triplet {identifiant de décès encodé DCD_IDT_ENC ; N° de ligne du certificat de décès CER_LIG_NUM ; rang de la cause de décès ECD_CAU_RNG}.

Le N° de ligne du certificat de décès CER_LIG_NUM peut aller de 1 à 6, là où la limite définie pour le rang de la cause ECD_CAU_RNG peut monter jusqu'à 40. Pour les certificats codés avec l'IA (type_codage >= 3), le texte de la ligne correspondra au texte brut présent sur le certificat sans découpage par rang, mais sera répété par autant de rang repéré par l'IA. Il y aura toujours un code par rang mais le texte brut des causes sera identique sur l'équivalent d'une ligne.

Lorsque le décès a pu être apparié avec les données de santé, le top DCD_IDT_TOP prend la valeur 1, les champs BEN_IDT_ANO, BEN_IDT_TOP et s'il existe BEN_NIR_ANO sont alimentés.

Attention :

Nous vous déconseillons fortement d'apparier directement la table KI_ECD_R avec les données de santé car vous risquez de recopier les données de santé autant de fois qu'il y a de lignes pour un même décès dans la table KI_ECD_R.

Nous vous conseillons donc de calculer d'abord les données de santé pour chaque identifiant BEN_IDT_ANO correspondant à un décès apparié avec les données de santé, puis de rapprocher ces données de la table KI_ECD_R (Cf. paragraphe III page 12).

Il convient de noter que pour un même identifiant DCD_IDT_ENC, la cause initiale de décès présente dans la table KI_CCI_R ne correspond pas systématiquement à l'une des causes mentionnées dans la table KI_ECD_R. Dans certains cas, la cause initiale résulte d'une combinaison de causes de la table KI_ECD_R.

Le code CIM10 ECD_CIM_COD et le libellé ECD_CAU_LIB de toutes les causes de décès d'un individu figurent dans la table KI_CCI_R. Pour connaître les libellés attribués par l'OMS aux codes ECD_CIM_COD attribués à chaque décès, vous pouvez également apparier tout ou partie du code ECD_CIM_COD aux tables de valeurs IR_CCI_V, IR_CIM_V ou MS_CIM_V du SNDS (Cf. §IV-1-f page 15).

Tableau 2 – liste des informations de la table KI_ECD_R

	Code	Libellé	Format	Table de valeurs	Changement à la mise à disposition 2023
Identifiants Bénéficiaire	DCD_IDT_ENC	Identifiant décès encodé	car(64)		
	DCD_IDT_TOP	Top Apparié avec IR_BEN_R	num(1)		
	BEN_IDT_ANO	Identifiant synthétique SNDS	car(18)		
	BEN_IDT_TOP	Top identifiant synthétique SNDS	num(1)		
	BEN_NIR_ANO	NIR pseudonymisé du bénéficiaire	car(17)		
Ensemble des causes du décès	FLX_PER_ANN	Année du décès	car(4)		
	CER_LIG_NUM	N° de ligne du certificat de décès	num(1)		
	ECD_CAU_RNG	Rang de la cause	num(3)		
	ECD_CAU_LIB	Libellé de la cause	car(400)		
	ECD_CIM_COD	Code de la cause	car(4)	IR_CCI_V, IR_CIM_V, MS_CIM_V	
	BEN_DCD_DTE	Date du décès	Date		
	BEN_DCD_AME	Année et mois du décès	car(6)	AAAAMM	

Précisions sur le numéro de ligne du certificat CER_LIG_NUM

Lorsqu'un médecin remplit un certificat de décès, il doit décrire la ou les maladies ou affections qui ont directement provoqué le décès mais il peut également préciser les états morbides qui ont contribué au décès.

Dans la table KI_ECD_R, les lignes CER_LIG_NUM comprises entre 1 et 5 correspondent aux maladies ou affections qui ont directement provoqué le décès, la ligne N° 6 concerne les états morbides qui ont contribué au décès. Bien entendu, les 5 premières lignes ne sont pas systématiquement remplies pour tous les décès, cela dépend du nombre d'affections qui ont provoqué le décès.

C'est pourquoi, dans la table KI_ECD_R, vous pourrez trouver des décès (identifiés par le champ DCD_IDT_ENC), pour lesquels les lignes CER_LIG_NUM portent les numéros 1 et 6, 1, 2 et 6 par exemple. Vous pourrez également trouver des décès qui ne comportent pas de ligne N° 6.

La variable ECD_CAU_LIB est le descriptif de la cause tel qu'il a été rédigé par le médecin sur le certificat pour le codage IA et après traitement automatique ou manuel spécifique pour lui attribuer un code pour le codage manuel et automatique.

Pour un décès, il peut arriver qu'il y ait un décalage entre ECD_CAU_LIB et le code attribué ECD_CIM_COD, par exemple lorsqu'une entité nosologique génère deux codes distincts. Cela peut aussi être la conséquence de deux entités nosologiques sur une même ligne dont le code attribué est la même : seul le code de la première entité nosologique est apposé.

3/ Les tables de valeur

Les tables de valeur associées à chaque variable des tables KI_CCI_R et KI_ECD_R sont disponibles dans le fichier [« V0 documentation variables CepiDc SNDS.xlsx »](#).

Annexe 1 – Fin de séjour hospitalier pour cause de décès dans le PMSI et DCIR

La date de fin de séjour pour cause de décès et le département de l'établissement de décès sont recherchés d'abord dans le PMSI qui couvre les séjours des établissements publics et privés, puis dans DCIR qui couvre essentiellement les séjours privés. Dans le PMSI, on recherche ces informations d'abord dans le champ MCO¹⁵, puis dans le champ HAD, puis dans le champ SSR. On n'a pas recherché les données dans le PMSI de psychiatrie car les fins de séjours hospitaliers pour cause de décès y sont très peu nombreuses.

1/ Dans le PMSI annuel

Quel que soit le champ du PMSI (MCO, HAD ou SSR) :

- les séjours qui se terminent pour cause de décès sont caractérisés par la valeur 9 du mode de sortie SOR_MOD. Cette variable se trouve dans la table « B » des séjours hospitaliers ;
- la date de fin de séjour se trouve dans la table « C » du chaînage des patients, il s'agit du champ EXE_SOI_DTF.

On recherche les séjours dans le PMSI annuel (tables T_MCOAAx, T_HADAAx, T_SSRAAx, AA représentant le millésime de l'année) car les données du PMSI annuel sont scellées par l'ATIH¹⁶ et on dispose de la totalité des séjours d'une année donnée.

a/ PMSI MCO

- On commence par exclure les séjours des établissements géographiques de l'APHP¹⁷, de l'APHM et des HCL car ils sont en doublons avec ceux des établissements juridiques correspondants. Pour cela, on sélectionne les séjours pour lesquels les établissements ETA_NUM de la table « C » se trouvent également dans la table des établissements « E » ;
- on se limite aux séjours {ETA_NUM ; RSA_NUM} pour lesquels on est assuré de la qualité des données, en particulier sur les éléments de l'identifiant SNIIRAM des individus NIR_ANO_17. Pour cela, on sélectionne les séjours de la table « C » du chaînage des patients pour lesquels les 9 codes retours NIR_RET, NAI_RET, SEX_RET, SEJ_RET, PMS_RET, FHO_RET, DAT_RET, COH_NAI_RET et COH_SEX_RET sont égaux à 0 (zéro) ;
- pour les séjours ainsi sélectionnés, on recherche les séjours {ETA_NUM ; RSA_NUM} de la table « B » pour lesquels le mode de sortie SOR_MOD est égal à 9 ;
- pour ces séjours {ETA_NUM ; RSA_NUM}, on recherche l'établissement géographique ETA_NUM_GEO de la table « UM » des résumés d'unités médicales correspondant au numéro d'ordre de RUM UM_ORD_NUM le plus élevé. Les deux premiers caractères du code ETA_NUM_GEO correspondent au département de cet

¹⁵ MCO : médecine, chirurgie, obstétrique et odontologie, HAD : hospitalisation à domicile, SSR : soins de suite et de réadaptation.

¹⁶ ATIH : agence technique de l'information hospitalière, agence qui recueille les données du PMSI.

¹⁷ APHP : assistance publique des hôpitaux de Paris, APHM : assistance publique des hôpitaux de Marseille, HCL : hospices civils de Lyon.

établissement, département que l'on va assimiler au département de décès. On le fera précéder d'un 0 (zéro) avant de le comparer à la table des circonstances et de la cause initiale du décès KI_CCI_R, excepté pour la Corse que l'on codera 201 ou 202 (Cf. paragraphe IV-1-c page 16) ;

- s'il existe plusieurs séjours pour un même individu identifié par son identifiant SNIIRAM NIR_ANO_17, on retient celui dont la date de fin EXE_SOI_DTF est la plus récente ;
- on apparie les données avec le référentiel IR_BEN_R afin de ne conserver que les séjours des individus qui n'ont pas de jumeau, triplé, ... dans ce référentiel (il faut donc se limiter aux cas où il n'y a qu'une seule ligne pour un même identifiant SNIIRAM BEN_NIR_PSA¹⁸ dans la table IR_BEN_R) ;
- enfin, on exclut les séjours pour lesquels le département de l'établissement de décès est inconnu (exclusion des séjours pour lesquels les deux premiers caractères de l'établissement géographique ETA_NUM_GEO est égal à 0).

b/ PMSI HAD

La démarche est analogue à celle du PMSI MCO : on recherche soit le département de l'EHPAD¹⁹ du patient s'il est hospitalisé en EHPAD, soit son code postal s'il n'est pas hospitalisé en EHPAD et on en déduit son département de décès.

Le numéro d'EHPAD ETA_NUM_EHPA et le code postal du patient BDI_COD se trouvent dans la table « B » des séjours en HAD. Pour un même séjour {ETA_NUM_EPMSI ; RHAD_NUM}, on sélectionne :

- les informations correspondant à la dernière sous-séquence de chaque séjour (variable SSEQ_SEJ_DER égale à 1 dans la table « B ») et pour lesquels le mode de sortie SOR_MOD est égal à 9 (décès) ;
- les séjours pour lesquels la date de fin de séjour EXE_SOI_DTF est renseignée dans la table « C » et dont l'année de fin de séjour correspond à l'année des décès que l'on veut étudier (le PMSI de 2014 contient des séjours qui se terminent en 2014 ou en 2015, on s'est limité à ceux de 2014) ;
- s'il existe plusieurs séjours pour un même patient identifié par le champ NIR_ANO_17, on ne conserve que le séjour correspondant à la date de fin la plus récente ;
- on ne conserve que les séjours des patients pour lesquels il n'y a ni jumeau, ni triplé, ... dans le référentiel IR_BEN_R et pour lesquels le département de décès issu du code postal ou de l'EHPAD est renseigné.

N.B. : comme pour le PMSI MCO, dans la table du chaînage des patients « C » du PMSI HAD, vous devez sélectionner les séjours pour lesquels les 9 codes retours NIR_RET, ... COH_SEX_RET sont égaux à 0 (zéro).

¹⁸ Rappel : l'identifiant SNIIRAM BEN_NIR_PSA du référentiel IR_BEN_R correspond à l'identifiant NIR_ANO_17 du PMSI.

¹⁹ EHPAD : établissement hospitalier pour personne âgée dépendante.

c/ PMSI SSR

Même démarche que pour le PMSI MCO et HAD, ne pas oublier de sélectionner les séjours pour lesquels les 9 codes retours NIR_RET, ... COH_SEX_RET de la table « C » du chaînage des patients sont égaux à 0 :

- pour un même séjour {ETA_NUM ;RHA_NUM}, les informations à sélectionner correspondent à la valeur la plus élevée du numéro séquentiel de résumé hebdomadaire RHS_NUM de la table « B », avec un mode de sortie SOR_MOD égal à 9 (décès). L'établissement géographique de décès ETA_NUM_GEO se trouve également dans la table « B ». On s'est limité aux décès survenus en 2014 (année de la date de fin de séjour EXE_SOI_DTF de la table « C » égale à 2014) ;
- s'il y a plusieurs séjours pour un même patient, on ne conserve que celui dont la date de fin est la plus élevée ;
- On se limite aux séjours dont le département de l'établissement géographique de décès ETA_NUM_GEO est connu et pour lesquels il n'y a ni jumeau, ni triplé, ... dans le référentiel IR_BEN_R.

2/ Dans DCIR

Le numéro de l'établissement géographique du décès ETB_EXE_FIN et le motif de fin de séjour ETE_NAT_FSJ se trouvent dans la table des soins réalisés en établissement ER_ETE_F. Lorsque le code ETE_NAT_FSJ vaut « D », cela signifie que le séjour s'est terminé pour cause de décès du patient.

Pour sélectionner les séjours se terminant au cours d'une année donnée pour cause de décès, vous devez relier la table des prestations ER_PRS_F et la table des soins réalisés en établissement ER_ETE_F, puisque la date de fin de séjour EXE_SOI_DTF se trouve dans la table ER_PRS_F, le numéro de l'établissement et le motif de fin de séjour pour décès se trouve dans la table ER_ETE_F.

Pour rechercher les séjours se terminant au cours d'une année donnée pour cause de décès, nous vous conseillons d'effectuer une sélection sur 18 mois de traitement. Pour l'année 2014 par exemple, les séjours se terminant en 2014 pour cause de décès (année de la date EXE_SOI_DTF = 2014) ont été recherchés pour une date de traitement FLX_TRT_DTD allant du 1^{er} janvier 2014 au 30 juin 2015 (condition équivalente à la condition sur la date de mise à disposition des données FLX_DIS_DTD : FLX_DIS_DTD comprise entre le 1^{er} février 2014 et le 1^{er} juillet 2015).

S'il existe plusieurs séjours pour un même patient, identifié par son couple {identifiant SNIIRAM BEN_NIR_PSA ; rang de bénéficiaire BEN_RNG_GEM}, on ne conserve que le séjour dont la date de fin EXE_SOI_DTF est la plus récente.

Annexe 2 – Sélection des informations des bénéficiaires décédés dans le référentiel IR_BEN_R et le référentiel archivé IR_BEN_R_ARC

Cette annexe complète le §III (page 13) dans lequel on vous explique comment relier les causes médicales de décès aux données de santé (remboursements de soins et PMSI).

Dans les tables des causes médicales de décès, lorsqu'un bénéficiaire est apparié avec les données de santé, son top d'appariement DCD_IDT_TOP vaut 1. Vous disposez alors dans ces tables de l'identifiant synthétique du bénéficiaire BEN_IDT_ANO et, s'il existe, de son NIR pseudonymisé BEN_NIR_ANO.

Les identifiants BEN_IDT_ANO se trouvent dans le référentiel des bénéficiaires du SNIIRAM IR_BEN_R et pour certaines personnes décédées, dans le référentiel archivé IR_BEN_R_ARC. Cette table archivée contient notamment tous les couples {identifiant SNIIRAM BEN_NIR_PSA ; rang de bénéficiaire BEN_RNG_GEM} pour lesquels aucun remboursement de soins n'a eu lieu depuis le 1^{er} janvier 2013.

Comme il vous est conseillé dans le §III, lorsque vous souhaitez faire une étude où vous reliez les causes médicales de décès aux données de santé :

- dans la table IR_BEN_R, vous commencez par rechercher tous les couples {identifiant SNIIRAM BEN_NIR_PSA ; rang de bénéficiaire BEN_RNG_GEM} associés aux identifiants BEN_IDT_ANO de la table des circonstances et de la cause initiale du décès KI_CCI_R ;
- lorsque l'identifiant BEN_IDT_ANO de la table KI_CCI_R ne se trouve pas dans le référentiel IR_BEN_R, il se trouve dans le référentiel archivé IR_BEN_R_ARC. Vous sélectionnez alors les couples {BEN_NIR_PSA ; BEN_RNG_GEM} correspondants dans cette table IR_BEN_R_ARC.

A titre d'exemple, pour les causes de décès de 2013, 12 930 identifiants BEN_IDT_ANO proviennent de la table archivée IR_BEN_R_ARC, contre 202 en 2014 et seulement 6 en 2015. Plus vous remonterez dans le temps, plus le nombre de bénéficiaires dont les identifiants se trouvent dans la table archivée IR_BEN_R_ARC sera important.

