

MAI 2017

MALADIES NEURO-
DÉGÉNÉRATIVES

ÉTUDES ET ENQUÊTES

LA BANQUE NATIONALE ALZHEIMER,
UN OUTIL POUR LA SURVEILLANCE
ÉPIDÉMIOLOGIQUE DES DÉMENCES
EN FRANCE ?

Résumé

Compte tenu de l'augmentation attendue du nombre de malades d'Alzheimer ou autre démence (MAAD) en France, la mise en place d'un système de surveillance épidémiologique opérationnel est indispensable. En complémentarité des données de cohortes et des bases médico-administratives, la Banque nationale Alzheimer (BNA) qui collecte de manière systématique les données des patients consultant un centre mémoire (CM) ou un centre de mémoire de ressources et de recherche (CMRR), présente un potentiel intéressant pour le suivi des MAAD.

Nous avons analysé les forces et les faiblesses de la BNA pour la production de données pertinentes de surveillance de la MAAD. À partir d'une extraction regroupant les données d'environ 560 000 patients inclus entre janvier 2010 et décembre 2014, nous avons analysé la qualité des données recueillies puis évalué la possibilité de construire des indicateurs de fréquence et de pronostic de la maladie. Si en théorie la BNA inclut tous les patients consultant un CM/CMRR en métropole, il n'a toutefois pas été possible de vérifier l'exhaustivité des cas présents dans la base. De plus, nous avons identifiés certains défauts dans la qualité des données recueillies (valeurs manquantes, incohérences, erreurs de codage...). S'ajoute à ces manques l'absence d'information concernant la représentativité des patients consultant les CM/CMRR par rapport à la population générale et l'absence de suivi longitudinal pour tous les patients. Tous ces éléments constituent un frein majeur à l'utilisation de la BNA, en sa forme actuelle, pour la production d'indicateurs de surveillance de la MAAD. Néanmoins, le nombre important de cas de MAAD identifiés à l'aide d'une définition de qualité homogénéisée entre les centres et l'identification fine des différents sous-types de démence, rarement étudiés dans les cohortes, sont des arguments majeurs pour entreprendre son amélioration.

Plusieurs actions doivent être menées pour pouvoir envisager l'intégration de la BNA dans le système de surveillance des MAAD : 1. assurer la remontée exhaustive des informations pour tous les patients consultant un CM/CMRR; 2. améliorer la qualité des données recueillies ; 3. évaluer la représentativité des cas de la BNA par rapport à ceux de la population générale ; 4. chaîner la BNA aux bases médico-administratives pour permettre le suivi longitudinal de cas bien validés de démence.

Ces points fondamentaux sont discutés au sein de groupes de travail dans le cadre de la mesure 16 « Pérenniser et améliorer la BNA » du plan « Maladies neurodégénératives 2014-2019 », groupes auxquels participe Santé publique France.

MOTS CLÉS : BANQUE NATIONALE ALZHEIMER, MALADIE D'ALZHEIMER, ET AUTRES DÉMENCES, SURVEILLANCE, QUALITÉ DES DONNÉES

Abstract

Given the expected increase in the number of Alzheimer's and other dementias (ADOD) in France, the implementation of an operational epidemiological surveillance system is essential. To complement cohort and medico-administrative database, the French Alzheimer Database (BNA), which systematically collects data from patients visiting a memory center (MC) or a resource and research memory center (RRMC), has an interesting potential for monitoring ADOD.

We analyzed the strengths and weaknesses of the BNA for the production of relevant data on ADOD surveillance. Using a data extraction of approximately 560,000 patients enrolled between January 2010 and December 2014, we analyzed the quality of the data collected and evaluated the possibility of constructing indicators of the frequency and prognosis of the disease. While in theory the BNA includes all patients visiting a MC/RRMC in mainland France, it was not possible to check the exhaustiveness of the cases present in the database. In addition, we identified some gaps in the quality of the data collected (missing values, inconsistencies, coding errors ...). In addition to these gaps, there is a lack of information about the representativeness of patients visiting MCs/RRMCs in relation to the general population and a lack of longitudinal follow-up for all patients. All these elements represent a major obstacle to the use of the BNA, in its present form, for the production of surveillance indicators on ADOD. Nevertheless, the large number of ADOD cases identified using a definition of quality and homogenized between the centers, and the fine identification of the different subtypes of dementia, rarely studied in the cohorts, are major arguments for working towards its improvement.

Several actions need to be taken to consider the integration of the BNA into the french surveillance system of ADOD :1- Provide comprehensive feedback for all patients attending a MC/RRMC;2- Improve the quality of the data collected;3- Assess the representativeness of the BNA cases compared to those in the general population;4- Chain the BNA to the medico-administrative databases to allow longitudinal follow-up of well-validated cases of dementia; These critical points are discussed in working groups within the framework of measure 16 "Sustain and improve the BNA" of the governmental plan "Neurodegenerative diseases Plan 2014-2019", in which Santé publique France is involved.

KEY WORDS: NATIONAL ALZHEIMER'S BANK, ALZHEIMER DISEASE AND OTHER DEMENTIAS, SURVEILLANCE, DATA QUALITY RESISTANCE

Ont participé à la réalisation de ce rapport technique

Auteurs

Laure Carcaillon-Bentata, Anne Thuret, Nathalie Beltzer (Santé publique France)
Karim Trifatène, Philippe Robert, (CHU de Nice)

Relecteur

Jean-François Dartigues (Inserm U1219, ISPED, Université Victor Segalen Bordeaux2, Bordeaux, France)

Sommaire

1	INTRODUCTION	8
2	OBJECTIFS DU RAPPORT	9
3	PRÉSENTATION GÉNÉRALE DE LA BNA	10
3.1	Les consultations spécialisées	10
3.2	Contexte et objectifs de la BNA	10
3.3	Gouvernance et Comitologie	10
3.4	Population concernée	11
3.5	Informations recueillies dans les CM/CMRR et intégrées dans la BNA	12
3.6	Le corpus minimum d'information Alzheimer : le CIMA	13
3.7	Modalités d'intégration et de circulation des données	17
3.8	Diagnostics de démence : démarche de standardisation	17
4	QUALITÉ DES DONNÉES RECUEILLIES DANS LA BNA	19
4.1	Méthodes	19
4.1.1	Base de données et documents transmis par le CHU	19
4.1.2	Évaluation de la qualité des données	19
4.2	Résultats	20
4.2.1	Prise en main des données et compréhension des documents	20
4.2.2	Évaluation du processus d'intégration et de contrôle de la qualité des données	21
4.2.2.1	Constitution de la base de données	21
4.2.2.2	Comparaison de la structure de la base de données avec les documents	25
4.2.2.3	Qualité des données de la BNA	29
4.2.2.3.1	Vérification de la non-existence de doublons	29
4.2.2.3.2	L'identifiant unique, PatientId, est-il effectivement unique ?	30
4.2.2.3.3	Valeurs manquantes	32
4.2.2.3.4	Incohérences observées	36
4.3	Synthèse des observations et propositions d'actions pour l'amélioration de la BNA	39
4.3.1	Le circuit des données n'en garantit pas l'exhaustivité et la qualité	39
4.3.2	Une appropriation peu aisée de la documentation et de la base	40
4.3.3	Comment améliorer la qualité des données ?	40
5	ÉLÉMENTS DE VALIDITÉ EXTERNE DES DONNÉES DE LA BNA	42
6	BNA : QUEL POTENTIEL POUR LA SURVEILLANCE ÉPIDÉMIOLOGIQUE DES MAAD ? ...	43
6.1	Définition des cas de MAAD dans la BNA	43
6.2	Représentativité des patients inclus	44
6.3	Quels indicateurs de surveillance en l'état actuel de la BNA ?	45
6.4	Autres indicateurs en l'état actuel de la BNA	47
6.5	Perspectives d'amélioration	48
6.5.1	Étudier la représentativité des patients inclus dans la BNA	48
6.5.2	Chainage des données de la BNA avec les bases de données médico-administratives	48
7	PERTINENCE DE LA BNA EN REGARD DU SYSTÈME DE SURVEILLANCE EXISTANT	49
7.1	Sources de données de surveillance actuelles	49
7.2	Apport de la BNA	50
8	CONCLUSION DU RAPPORT	52

Références bibliographiques	54
Annexes	57
Annexe 1. Diagnostic d'une MAAD dans la BNA	57
Annexe 2. Variables dans les bases de données transmises.....	60
Base principale	60
Description des 30 fichiers accompagnants la base principale.....	61
Annexe 3. Comparaison de la structure de la base de données avec les documents	76
Annexe 4. Items retenus et cotations de ces items pour le score IADL.....	79
Annexe 5. Description des modalités et de leur répartition, 2010-2014	80
Variables qualitatives.....	80
Variables quantitatives	85

Abréviations

ALD	Affection longue durée
ARS	Agence régionale de santé
BNA	Banque nationale Alzheimer
CépiDc	Centre épidémiologique des causes médicales de décès
CHU	Centre hospitalo-universitaire
CIM	Classification internationale des maladies
CIMA	Corpus d'information minimum maladie d'Alzheimer
CM	Centre mémoire
CMRR	Centre mémoire de ressources et de recherche
CNIL	Commission nationale informatique et liberté
DCI	Dénomination commune internationale
DGOS	Direction générale de l'offre de soins
DOM	Départements d'outre-mer
EGB	Échantillon généraliste des bénéficiaires
Ehpad	Établissement hébergeant des personnes âgées dépendantes
IADL	Instrumental activity of daily living
IRM	Imagerie par résonance magnétique
MAAD	Maladie d'Alzheimer et autres démences
MCI	Mild cognitive impairment (troubles cognitifs légers)
MMSE	Mini mental status examination
NMDA	N-méthyl-D-aspartate
OMS	Organisation mondiale de la santé
PMND	Plan des maladies neurodégénératives
SIC	Spécifications de l'interface de collecte
Sniiram	Système national d'information inter-régimes de l'Assurance maladie

Table des illustrations

Tableau 1. Informations du Corpus minimum d'information Alzheimer (CIMA), recueillies dans les CM/CMRR.....	14
Tableau 2. Description des principales variables dans les fichiers transmis pour intégrer la BNA.....	15
Tableau 3. Description des difficultés identifiées pour leur prise en main des 24 variables présélectionnées et étudiées.	28
Tableau 4. Pourcentage de valeurs manquantes pour les 24 variables, 2010-2014 (556 069 PatientID# présumée unique).....	34
Tableau 5. Répartition des individus pour la variable IADLs en fonction de leur statut de démence au 31/12/2015, 2010-2014	38
Tableau 6. Liste des variables (par ordre d'ordonnement) et paramètres (type de variable et longueur du champ) dans la base transmise.	60
Figure 1. Schéma simplifié du flux des données pour la constitution de la base BNA et risques identifiés pour garantir au mieux l'exhaustivité des cas et de la qualité des données	20
Figure 2. Consolidation de la base transmise	28
Figure 3. Carte présentant la répartition des CM et CMRR sur le territoire à la date du 07 juin 2010.....	40

1 INTRODUCTION

Il est communément admis que la surveillance épidémiologique se définit comme un « processus de collecte, d'analyse et d'interprétation systématique et pérenne de données standardisées et validées sur un ou des événements de santé afin de planifier, mettre en œuvre ou évaluer des pratiques en santé publique. Elle est étroitement associée à leur juste diffusion à ceux qui ont besoin d'être informés » [1]. Les objectifs d'un système de surveillance ne sont donc pas seulement de collecter de données mais également de pouvoir décrire l'événement d'intérêt selon différentes dimensions (temps, espace, caractéristiques sociodémographiques...). Il s'agit très souvent de décrire la fréquence de la maladie à l'aide d'indicateurs tels que la prévalence (nombre de cas vivants à un moment précis), l'incidence (nombre de nouveaux cas au cours d'une période donnée) ou encore le nombre de décès lié à la maladie. De plus, des indicateurs permettant de décrire le pronostic des malades (survie, hospitalisations, comorbidités) ou les facteurs de risque de la maladie ont également toute leur importance.

En France, les principales données de surveillance épidémiologique concernant la maladie d'Alzheimer et autres démences (MAAD) proviennent d'études de cohortes [2-5]. En particulier la cohorte Paquid permet de donner des estimations de fréquence de la MAAD chez les sujets âgés de plus de 65 ans : environ 996 000 personnes concernées en 2010 (9,8 % des plus de 65 ans) et plus de 1,5 million attendues en 2030 [6]. Les principales limites de ces données de cohortes concernent l'impossibilité de décliner les indicateurs de fréquence à des niveaux géographiques fins et d'étudier la démence avant 65 ans du fait que seuls des sujets plus âgés sont inclus dans ces études, alors qu'il est bien établi que les processus impliqués dans le développement de cette maladie sont lents et évolutifs. La récolte et l'utilisation de données complémentaires pour la surveillance des MAAD en France est aujourd'hui indiscutable.

La Banque nationale Alzheimer (BNA) a été mise en place dans le cadre du 3^e plan présidentiel « Plan Alzheimer et maladies apparentées 2008-2012 »ⁱ pour répondre aux objectifs de la mesure 34 : « permettre de contribuer à adapter l'offre de soins spécialisés mémoire à travers les schémas régionaux d'organisation des soins et d'apprécier la qualité de leur fonctionnement » et « disposer de données épidémiologiques, [...et] de données d'activités des centres spécialisés ». Cette base de données repose sur la remontée systématique des informations recueillies pour l'ensemble des patients vus au sein des centres de consultations spécialisées MAAD en France entière. Le rapport d'évaluation du « Plan Alzheimer et maladies apparentées 2008-2012 » [7] a souligné le potentiel important pour la surveillance des MAAD de cette remontée systématique, exhaustive et standardisée de données et a ainsi préconisé d'en poursuivre le recueil.

En vue d'évaluer les potentialités de la BNA pour intégrer ou compléter le système de surveillance des MAAD, Santé publique France collabore avec le CHU de Nice, hébergeur de la BNA et principal utilisateur de ces données. Le présent rapport contribue également à répondre à la mesure 16 du nouveau « Plan maladies neurodégénératives 2014-2019 » (PMND, [8]) visant à « pérenniser et améliorer la BNA » dans l'objectif d'en faire un outil épidémiologique « utilisable par tous ».

ⁱ <http://www.plan-alzheimer.gouv.fr/>.

2 OBJECTIFS DU RAPPORT

Ce rapport a pour objectif principal d'évaluer les potentialités de la BNA pour la surveillance épidémiologique des MAAD. Aussi, après avoir présenté la BNA dans sa forme, sa structuration et son contenu, nous étudierons la pertinence et la qualité des informations recueillies dans la base et nous évaluerons la possibilité de produire des données de suivi épidémiologique de la MAAD.

En particulier, nous nous attacherons à évaluer la possibilité de construire des indicateurs de fréquence de la maladie et de pronostic pour les malades. À chaque étape de ce rapport, nous formulerons, lorsque nécessaire, des recommandations pour que la qualité des données de la BNA soit améliorée et que ces données puissent être opérationnelles dans la construction d'indicateurs de surveillance.

Si l'objectif principal est effectivement d'évaluer les potentialités de la BNA à la surveillance des MAAD, il est tout d'abord nécessaire de s'assurer de la qualité des données. À chaque étape de notre analyse, nous nous sommes attachés à indiquer ce que nous avons pu constater. Puis, nous nous sommes autorisés au lieu de faire uniquement un état des lieux de la qualité des données remontées au sein de la BNA, à proposer lorsque nécessaire quelques préconisations pour améliorer les données ainsi produites.

Ces préconisations sont signalées dans le rapport par le symbole ☹. Elles sont très variées (de processus, de vérifications,), et n'ont pas du tout la prétention d'être exhaustives. Elles résultent bien souvent de réflexions menées au sein de Santé publique France et partagées avec le CHU de Nice et les autres partenaires au cours de réunions portant sur le suivi de la mesure 16 du PMND 2014-2019. Nos interrogations qui demeurent en suspens sont signalées par le symbole ◇.

Il est à noter que ce rapport est un rapport technique permettant de décrire notre compréhension du fonctionnement de la BNA, ses apports pour l'étude des MAAD et ses potentialités pour leur surveillance. Une synthèse accompagne ce rapport.

3 PRÉSENTATION GÉNÉRALE DE LA BNA

3.1 Les consultations spécialisées

Les centres mémoire (CM) et les centres mémoire de ressources et de recherche (CMRR) ont été mis en place au début des années 2000 dans le cadre du 1^{er} plan maladie d'Alzheimer et maladies apparentées pour évaluer les troubles de mémoires et des autres fonctions cognitives et comportementales. Ils assurent, au-delà des consultations de proximité, une mission de recours spécialisée au niveau régional ainsi qu'une mission de recherche. Les consultations mémoires sont labélisées par l'agence régionale de santé (ARS). Les plans suivants ont permis de renforcer ce dispositif afin d'obtenir un maillage national suffisant et pluridisciplinaire.

Les CM constituent en France un deuxième niveau de soin, c'est-à-dire que la consultation peut être initiée par un médecin généraliste ou spécialiste, ou par une structure, comme l'hôpital pour les patients hospitalisés ou vus en service d'urgence (dont l'hôpital pouvant héberger le CM), les établissements hébergeant des personnes âgées dépendantes (Ehpad), les maisons de retraite ou les services sociaux. Enfin, cette consultation peut également être motivée par le patient lui-même ou son entourage « inquiété » par la présence de troubles cognitifs. Ces centres disposent de compétences médicales et paramédicales spécialisées, d'un accès aux plateaux techniques (examens biologiques et radiologiques) et d'une offre de prises en charges spécifiques variables selon les centres (ateliers mémoires, prises en charge et orientation des aidants de personnes atteintes de démences...).

Les CM sont donc des structures de diagnostic et de prise en charges des patients atteints de troubles de la mémoire et cognitifs, les CMRR assurant en plus l'animation de réseau de ces CM, ainsi que des missions d'enseignement et de recherche.

3.2 Contexte et objectifs de la BNA

En février 2008, les pouvoirs publics français ont lancé le 3^e plan Alzheimer 2008-2012 qui avait pour objectif de fournir un effort en termes de structuration de la recherche, de favoriser un diagnostic plus précoce et de mieux prendre en charge les malades et leurs aidants. Il comportait 44 mesures. Parmi elles, la mesure 34 avait pour objectif la mise en place au niveau national d'un système de recueil d'activité et de suivi épidémiologique de la démence à partir des centres de consultations spécialisées (CM et CMRR) et des centres libéraux volontaires. Cette mesure s'est matérialisée à travers la construction d'un système d'information, la BNA.

3.3 Gouvernance et Comitologie

La BNA fait l'objet d'un financement de la DGOS Le budget alloué a été de 0,8 M € lors de son installation et de 1,2 M € de fonctionnement de 2008 à 2012. Depuis 2012, le budget de fonctionnement est de 0,3 M € par an.

L'organisation fonctionnelle de la BNA fait appel à plusieurs comités et structures avec :

- le comité de pilotage de la BNA est composé par la DGOS, le CHU de Nice, Santé publique France, de représentants de la fédération des CMRR, de la fondation de coopération scientifique, de la mission de pilotage du plan Alzheimer et d'un neurologue libéral participant au recueil des données. Il a pour mission de valider l'organisation générale, la répartition et l'évolution des travaux ainsi que les cahiers des charges d'appels d'offre relatifs aux consultations nécessitées par la mesure du plan et de suivre l'exécution des travaux prévus dans les marchés ou délégués directement au CHU de Nice.
- un Comité scientifique a également été créé en septembre 2014. Il a pour rôle de définir les modalités permettant d'améliorer la qualité des données colligées dans la BNA et d'émettre un avis sur les demandes extérieures d'exploitation et d'utilisation de la banque de données. Ce comité scientifique est composé de représentants des CM, des CMRR, de la DGOS, de l'Inserm, de Santé publique France et du CHU de Nice. Enfin, trois structures du CHU de Nice assurent son fonctionnement quotidien. Le département de Santé publique (un ingénieur), le département du Système d'information (présence de plusieurs techniciens) et le CMRR qui assure la coordination de la BNA (un coordinateur).

3.4 Population concernée

La totalité des consultations mémoire labellisées ont l'obligation de documenter la BNA et un financement a été octroyé aux ARS depuis 2009 pour accompagner ce recueil de données. L'ensemble des CMRR et des CM alimentent donc la BNA. Certains médecins libéraux participent également au recueil, sur la base du volontariat, de manière indépendante ou dans le cadre de réseaux labellisés par les ARS (par exemple le réseau de consultation mémoire « de ville » Aloïs Alzheimer).

Il n'y a pas de critères d'inclusion, de non inclusion ou d'exclusion des patients dans la BNA. En effet, la BNA a été conçue non pas comme une étude mais pour favoriser le recueil le plus exhaustif possible de l'activité des centres. En particulier il n'y a pas de critères d'inclusion basés sur l'âge ou l'état de santé de la personne qui consulte, ni de critères d'exclusion concernant par exemple la participation à un programme de recherche.

Au total, l'ensemble des patients ayant recours à un CMRR ou un CM sont invités à participer et les données des patients volontaires sont intégrées dans la BNA, certains patients pouvant être inclus via la participation de médecins libéraux.

En fonction du contexte clinique et social du patient, les structures participant au recueil peuvent être amenées à réaliser des consultations (ou actes) uniques pour un patient donné ou à suivre le patient dans le temps.

◇ Combien de cas refusent que leurs données soient intégrées dans la base pour transmission à la BNA ? Est-ce que les accords Cnil ou CCTIRS/CCPP permettent de recueillir des informations sur ces patients afin de pouvoir les décrire ?

3.5 Informations recueillies dans les CM/CMRR et intégrées dans la BNA

Les données recueillies dans la BNA ont été définies par consensus. À chaque consultation, un enregistrement est généré, transféré puis implémenté dans la BNA. Ainsi, un même patient peut avoir un ou plusieurs enregistrements (lignes) dans la BNA correspondant à une ou plusieurs consultations. Une clef d'appariement unique et anonyme (cf. § 3.7) est générée pour chaque patient et permet de repérer l'ensemble des enregistrements possibles relatifs à une même personne. Ceci permet le suivi longitudinal du patient dans la BNA.

On peut distinguer deux grandes catégories de données saisies dans la base initiale, l'une fait référence aux données renseignant sur l'activité des centres, l'autre contient les informations liées à la maladie d'Alzheimer (Corpus d'information sur la maladie d'Alzheimer, CIMA). Certaines variables sont « obligatoires » et doivent être complétées pour que le fichier de données puisse être intégré dans la BNA. D'autres sont « facultatives ». Dans ce cas, il n'y a pas de blocage de la transmission à la BNA par le logiciel utilisé par le centre ou par l'application BNA du CHU de Nice. Parallèlement, certaines variables sont dites « fixes » et d'autres « réactualisables » par l'opérateur de saisie du dossier dans le centre participant. Les données « fixes », telles que le sexe par exemple, sont saisies lors de la première consultation et sont reportées automatiquement sur le formulaire de saisie à chaque nouvelle visite ; les données « réactualisables » peuvent être modifiées à chaque nouvelle visite si besoin.

Trois documents sont mis à disposition des utilisateurs de la BNA pour les aider à prendre en main la base de données. Ceux à notre disposition, fournis au moment de la remise de la base, sont :

- le corpus d'information minimal maladie d'Alzheimer (CIMA), version du 4 mai 2013, formulaire qui indique les noms des variables et leurs modalités pour les informations liées au patient à l'acte ;
- le glossaire du CIMA, version du 3 octobre 2013, qui donne des précisions sur le caractère « fixe » ou « réactualisable » de la variable et quelques indications pour le remplissage de cette variable pour les données du centre, du patient et de l'acte ;
- les spécifications d'interface de collecte (SIC), version 5.8 datant du 01/02/2012, qui indiquent les formats (numérique, alphanumérique, texte ou date) des variables de la base, le caractère « obligatoire » ou « facultatif » de celles-ci ainsi que les modalités définies pour la dernière révision.

La dernière version de ces documents est disponible sur le site du Centre mémoire de ressources et de recherche de Nice (www.cmrr-nice.fr/?p=plan-alzheimer, consulté le 17/02/2017).

Des informations sur les centres participants sont collectées dans la BNA. Lors de l'inclusion d'un centre dans la BNA et selon un rythme d'actualisation annuel, 49 variables renseignent sur l'identification du centre (adresse, type de centre (CM, CMRR, spécialiste libéral), nom du responsable du centre...), sa structuration (multi-site, inclusion dans un réseau de soin, hôpital de jour, type d'application utilisé pour la transmission des données BNA...), son personnel (qualifications et nombre d'équivalent temps plein par type de personnel) et ses activités (recherche, enseignement, soutien des aidants...). Seules, les données concernant l'identification du centre sont « obligatoires ».

3.6 Le corpus minimum d'information Alzheimer : le CIMA

Le CIMA, extrait de la base initiale enregistrée au niveau de chaque centre, est soumis à deux contraintes très générales : il doit contenir des informations pertinentes pour permettre des traitements statistiques de qualité et il doit être suffisamment court pour ne pas représenter une contrainte trop importante pour les centres. Ce CIMA a été défini en concertation avec les CM et CMRR français et validé par la DGOS en 2008 avant la mise en place effective du recueil. Un cahier des charges a été soigneusement rédigé afin de définir les modalités de fonctionnement de la base de données à mettre en place ainsi que le type d'information qui y est recueilli. Ainsi, à l'issue de chaque consultation le praticien complète le CIMA dans le système informatique de son centre et celui-ci est transmis et intégré de façon anonyme à la BNA. Le CIMA est composé de 33 groupes d'informations systématiquement transmises lorsqu'un acte (consultation) est réalisé (cf. tableau 1).

I TABLEAU 1 I

Informations du Corpus minimum d'information Alzheimer (CIMA), recueillies dans les CM/CMRR

Libellé de la variable	Type de variable	Commentaires
DONNÉES LIÉES AU PATIENT		
1	Nom d'usage	Ces 6 variables servent à l'anonymisation et à la création de la clef d'appariement unique par patient, elles ne sont pas transmises sous leur forme initiale à la BNA
2	Nom de naissance	
3	Prénom	
4	Date de naissance	
5	Département de naissance	
6	Code de la commune de naissance	
7	Sexe	DF
8	Nouveau patient	DF
9	Mode d'adressage	DF
10	Caractéristiques de l'accompagnant	DR
11	Mode de vie actuel	DR
12	Situation géographique des patients par rapport à la localisation du centre	DR
13	Niveau d'études	DF
14	APA	DR
15	ALD	DR
16	Mesure de protection	DR
17	Diagnostic Niveau 1	DR
18	Diagnostic Niveau 2	DR
19	Profession (ou dernier profession exercée)	DF
DONNÉES LIÉES À L'ACTE		
20	Date de l'acte	DR
21	Type d'acte	DR
22	L'acte est réalisé dans le cadre de la mission de recours (pour CMRR uniquement)	DR
23	si oui : Type de recours	DR
24	L'acte en cours est réalisé dans le cadre d'un protocole de recherche:	DR
25	si oui : Type de protocole de recherche	DR
26	Score du Mini Mental Test	DR
27	Score des IADL	DR
28	Traitements pharmacologiques	DR
29	Effet indésirable grave	DR
30	Interventions psycho sociales et de réhabilitations	DR
31	État du suivi	DR
32	Date d'entrée en institution (si entrée)	DR
33	Date de décès, si décès	DF

DF : donnée fixe.

DR : donnée réactualisable.

APA : Allocation personnalisée d'autonomie.

ALD : Affection longue durée.

IADL : Instrumental Activities of Daily Living (échelle d'activités instrumentales de la vie courante ou échelle de Lawton).

Parmi les informations contenues dans le CIMA, celles relatives au diagnostic de la démence sont centrales. Au moment de la saisie de ces informations, deux niveaux de diagnostic doivent être indiqués. Au niveau 1, il est indiqué si le patient présente 1) une absence de trouble, 2) un trouble cognitif, 3) une démence, ou 4) s'il s'agit d'un autre cas. Au niveau 2, le diagnostic nosologique est précisé (ex : plainte mnésique, maladie d'Alzheimer, épisode dépressif...). Il est à noter qu'en l'attente d'éléments contributifs (examens neuropsychologiques, biologiques ou imagerie) permettant d'établir un diagnostic précis, la transmission du CIMA peut se faire avec un diagnostic de niveau 2 codé en « diagnostic en attente ». Une fois le diagnostic établi et le patient revu, un nouveau CIMA sera rempli avec le diagnostic de niveau 2 renseignant l'état du patient.

Les informations recueillies sur les traitements concernent la présence, au moment de la consultation, 1) d'une prescription ou non d'anti-démence avec la distinction entre inhibiteur d'acétylcholinestérase et bloqueur de récepteurs NMDA, 2) d'une prescription de psychotrope avec les distinctions suivantes : antidépresseur, anxiolytique, hypnotique, antipsychotique, agent nootrope. Il est important de souligner qu'aucune information relative à la dénomination commune internationale (DCI) ou à la posologie ne sont recueillies dans la BNA.

Après extraction des CIMA et anonymisation de la BNA, les fichiers transmis comportent 118 modalités de variables qui sont de nature démographique, diagnostique, clinique et de prise en charge (cf. tableau 2). La plupart des variables du CIMA sont « obligatoires », seules deux variables sont « facultatives » : le score de MMSE (Mini-Mental State Examination) et les scores aux IADL (Instrumental Activities of Daily Living ou échelle de Lawton).

I TABLEAU 2 I

Description des principales variables dans les fichiers transmis pour intégrer la BNA

Nom de la variable <i>Libellé supposé de la variable</i>	Type de variable <i>Longueur ou format</i>	Statut de la variable <i>obligatoire ou facultatif</i> fixe ou réactualisable	Informations complémentaires (nombre de modalités)
Variables géographiques			
CentreRegion <i>Région du centre</i>	Num. ¹ 8	<i>obligatoire ou facultatif, non précisé</i> fixe ou réactualisable, <i>non précisé</i>	1 modalité pour chacune des 22 régions et 1 pour l'ensemble des DOM
CentreDepartement <i>Département du centre</i>	Texte 4	<i>obligatoire ou facultatif, non précisé</i> fixe ou réactualisable, <i>non précisé</i>	1 modalité pour chacune des 105 départements dont les 7 DOM-TOM
Variables « patient »			
PatientId <i>Identifiant unique du patient</i>	Texte 64 ou 300	obligatoire fixe	
PatientSexe <i>Sexe du patient</i>	Texte 1 ou 3	obligatoire fixe	2 modalités (M/F ; Femme, Homme)
PatientAnneeNaissance <i>Année de naissance du patient</i>	Num. ¹ 4 ou 8	obligatoire fixe	
DclAnneePremierDiag <i>Année de première consultation</i>	Num. ¹ 4 ou 8	obligatoire fixe	
DclALD <i>ALD</i>	Num. ¹ 2 ou 8	obligatoire réactualisable	2 ou 3 modalités (Oui, Non ± sans information)
DclAPA <i>APA</i>	Num. ¹ 2 ou 8	obligatoire réactualisable	2 ou 3 modalités (Oui, Non ± sans information)

DclMesureProtection <i>Mesure de protection</i>	Num. ¹ 8 ou 9	obligatoire réactualisable	4 modalités
DclModeDeVie <i>Mode de vie actuel</i>	Num. ¹ 8 ou 10	obligatoire réactualisable	12 modalités
PatientAccompagnant <i>Accompagnant du patient / Caractéristiques de l'accompagnant</i>	Num. ¹ 8 ou 9	obligatoire réactualisable	10 modalités
PatientEnvoyePar <i>Patient envoyé par / mode d'adressage</i>	Num. ¹ 6 ou 8	obligatoire fixe	14 modalités
PatientNiveauEtudes <i>Niveau d'étude</i>	Num. ¹ 6 ou 8	obligatoire	6 modalités
PatientProfession <i>Profession – dernier métier exercé</i>	Num. ¹ 8 ou 10	obligatoire fixe	10 ou 12 modalités
Variables « acte »			
DclDateActe <i>Date de l'acte</i>	Date 8	obligatoire fixe ou réactualisable, <i>non précisé</i>	
DclTypeActe <i>Type d'acte</i>	Num. ¹ 8	obligatoire réactualisable	14 modalités
DclEtatDemence <i>Diagnostic N1</i>	Num. ¹ 8	obligatoire réactualisable	3 modalités
DclDiagnostic <i>Diagnostic N2</i>	Num. ¹ 6 ou 8	obligatoire réactualisable	38 modalités
DclIADLDetection <i>IADL score détection</i>	Num. ¹ 2 ou 8	facultatif réactualisable	score sur une échelle de 4
DclMMSE <i>Score au Mini Mental test</i>	Num. ¹ 2 ou 8	facultatif réactualisable	score sur une échelle de 30
DclPCAucun <i>Examens dont les résultats ont été consultés le jour de l'acte</i>	Num. ¹ 1 ou 8	obligatoire réactualisable	2 modalités
DclTPAucun <i>Traitements pharmacologiques</i>	Num. ¹ 1 ou 8	obligatoire réactualisable	2 modalités
DclPSRAucun <i>Traitements non pharmacologiques</i>	Num. ¹ 1 ou 8	obligatoire réactualisable	2 modalités
DclFinSuivi <i>Type de fin de suivi</i>	Num. ¹ 7 ou 8	obligatoire réactualisable	8 modalités

¹ Num., numérique

* Modalité précisée dans le SIC

Des révisions sont périodiquement réalisées après proposition et validation par le conseil scientifique de la BNA afin d'adapter le contenu de cette banque de données à l'évolution des pratiques médicales et aux exigences de monitoring de l'activité des centres par les tutelles administratives.

Les révisions 2, 3 et 4 ont porté sur l'ajout de nouvelles modalités pour des variables existantes ou sur l'ajout de nouvelles variables. La révision 5 a eu lieu en 2016. Elle a essentiellement consisté en une modification dans les choix diagnostiques qui sont proposés et pour lesquels 3 dimensions au lieu de 2 sont maintenant décrites (le stade, le type de syndrome cognitif et comportemental et le type de lésion). Des précisions concernant la démarche ayant conduit à ces choix est détaillé dans le paragraphe suivant.

3.7 Modalités d'intégration et de circulation des données

Concernant le logiciel permettant la saisie des informations contenues dans la BNA, le comité de pilotage de la BNA a fait le choix en 2008 de s'appuyer sur les systèmes de recueil de données existant au sein des CM et CMRR et de ne pas mettre en place de dispositif unique de saisie des données pour la BNA dans tous les centres participants. Ce choix a été fait car la quasi-totalité des centres disposait déjà en 2008 de systèmes de recueil des données patients (notamment à des fins de dossier patient). Même si plusieurs logiciels différents étaient utilisés (12 sont répertoriés en 2014), il est apparu plus raisonnable de ne pas imposer un nouvel outil pour alimenter une base de données naissante mais de s'appuyer sur les systèmes existants. Les spécifications de la BNA (variables, modalités de réponses, caractère obligatoire ou optionnel) ont donc été transmises aux différents fournisseurs de logiciels. En parallèle, une interface web commune a été créée par le CHU de Nice pour la transmission des données après anonymisation. Le serveur qui héberge la BNA est localisé dans les locaux du CHU de Nice sous le contrôle de la direction des systèmes d'Information, agréée pour l'hébergement de données de santé à caractère personnel, elle assure ainsi la protection nécessaire à ce type d'information.

En pratique, le médecin ou l'utilisateur saisit les données normalement pendant la consultation ou après celle-ci et valide la transmission des données à la BNA. Deux flux de données sont organisés : celui pour tout nouvel acte enregistré dans le logiciel du centre et celui pour la mise à jour du suivi d'un patient en cas de fin de suivi, de décès ou d'entrée en institution. Cinq étapes se succèdent alors avant que les données ne soient définitivement intégrées à la BNA (cf. § 4.2.2) : extraction des informations, anonymisation des fichiers d'actes, transmission des données anonymisées, intégration dans la BNA puis retour d'information. Un rapport de transmission est adressé lors du retour d'information contenant soit une information de bonne transmission soit un rapport d'erreur précisant les raisons de l'échec d'importation des données dans la BNA.

3.8 Diagnostics de démence : démarche de standardisation

Il est attendu de la BNA qu'elle reflète l'activité des centres participant. Or, le processus diagnostic et les modes de prise en charge évoluent avec les connaissances médicales et les consensus, et ces évolutions ne sont pas sans impact sur l'activité des CM et CMRR. C'est probablement sur le plan du diagnostic que ces évolutions sont les plus marquées avec des regroupements ou des distinctions diagnostiques ou encore l'identification de nouvelles entités. Dans ce cadre, le Comité scientifique de la BNA s'est saisi de la question de l'harmonisation diagnostique dans la BNA. Cette démarche est également inscrite dans le cadre de la mesure 16 du nouveau Plan maladies neurodégénératives 2014-2019. L'enjeu ici est double. Il s'agit d'une part de « coller » le plus possible à la nosologie faisant le plus largement consensus dans la communauté scientifique mais également de faire en sorte que l'ensemble des centres participant au recueil saisisse les diagnostics de la même façon : un même patient qui serait vu à différents endroits devrait avoir le même diagnostic dans la BNA.

La démarche active mise en œuvre dans la BNA repose ainsi sur une procédure d'identification simple, rapide, standardisée des cas faisant consensus en s'appuyant sur des définitions communes telles que celles du DSM-IV (ou DSM-V plus récemment), tout en

tenant compte de l'évolution des processus de diagnostic et de prise en chargeⁱⁱ. Ces efforts permettent d'optimiser la qualité et l'homogénéité des cas de MAAD présents dans la BNA.

Après plusieurs révisions de cette variable, le diagnostic repose depuis 2016 (et donc non intégrée dans la base transmise par le CHU de Nice) sur trois dimensions définies par le comité scientifique de la BNA (cf. annexe 1) :

- 1- le stade : absence de trouble mental, plainte cognitive isolée, trouble mineur, trouble cognitif majeur, non applicable ;
- 2- le type de syndrome cognitif et comportemental : présentation clinique à dominante cognitive (à dominante amnésique, à dominante non amnésique ou cognitive « multi-domaines ») ; présentation clinique à dominante comportementale ou psychologique (syndrome dysexécutif comportemental, troubles émotionnels ou affectifs, troubles productifs, ou autres) ; ou présentation clinique sans dominante cliniquement significative) ; ou non applicable ;
- 3- et le type de lésion ou diagnostic étiologique : maladie d'Alzheimer, maladie à corps de Lewy ...

Un avantage de la BNA pour le suivi épidémiologique des MAAD réside dans l'utilisation combinée d'un examen clinique, guidée par une méthodologie standardisée, et la réalisation d'examens complémentaires (IRM, prise de sang) pour confirmer le diagnostic et en préciser l'étiologie.

ⁱⁱ La classification internationale des maladies (CIM10) initialement retenue pour coder les cas de démences a été abandonnée car elle ne permettait pas de prendre en compte l'évolution des connaissances scientifiques.

4 QUALITÉ DES DONNÉES RECUEILLIES DANS LA BNA

4.1 Méthodes

4.1.1 Base de données et documents transmis par le CHU

Diverses variables sont colligées dans la BNA à partir des différents CM et CMRR et, comme nous l'avons vu, leur recueil ou leur format évoluent avec le temps. S'interroger sur la possibilité d'utiliser la BNA pour la surveillance épidémiologique des MAAD suppose au préalable de s'assurer de la qualité de ces données.

En vue d'évaluer les potentialités de la BNA pour la surveillance épidémiologique des MAAD, le CHU de Nice a transmis à Santé publique France une extraction de la base de données BNA contenant l'ensemble des données enregistrées depuis la création de la BNA jusqu'au 31 décembre 2014. En pratique, la base de données a été transmise sous la forme d'une table SAS accompagnée de 30 autres tables supposées correspondre aux libellés des variables de la base principale. Une description de la structure de ces différentes tables est disponible en annexe (cf. annexe 2).

Comme nous l'avons vu précédemment, trois documents sont disponibles pour nous aider à prendre en main la base de données : le CIMA, le glossaire du CIMA et le SIC (cf. § 3.5). De plus, deux autres documents sont mis à disposition concernant la méthode d'anonymisation réalisée par les centres :

- le manuel d'utilisation de la librairie anonymat version 2.2 – Spécifications fonctionnelles et techniques ;
- le programme Anonymisation version 1.0.0 – Spécifications fonctionnelles et techniques.

La base transmise en janvier 2015 comportait 33 variables dichotomisées en 118 modalités. Parmi ces variables, **24 variables** ont été finalement sélectionnées *a priori* pour « évaluer » la faisabilité d'une surveillance épidémiologique en population générale à partir de la BNA. Ces 24 variables présélectionnées sont décrites dans le tableau 2 et peuvent être classées en trois catégories : les variables géographiques, les variables permettant de caractériser le patient et celles permettant de caractériser l'acte. Pour ce rapport, l'analyse porte sur les actes enregistrés entre le 1^{er} janvier 2010 et le 31 décembre 2014.

4.1.2 Évaluation de la qualité des données

L'évaluation de la qualité des données repose sur plusieurs étapes. Avant d'aborder la question de la qualité des données dans la base, nous avons étudié la qualité du processus mis en œuvre par l'infrastructure BNA pour intégrer les données des centres dans la base et pour garantir la qualité de ces données. Ainsi, nous avons tout d'abord procédé à l'analyse des documents fournis pour appréhender les données de la BNA.

Dans un second temps, nous avons procédé à une analyse qualitative des fichiers. Nous avons notamment vérifié l'absence de doublon d'acte (un même acte enregistré plusieurs

fois) et vérifié que le numéro d'identifiant du patient - la variable « **PatientID** »- issu du processus d'anonymisation correspondait bien à un seul et même patient. Pour rappel, dans la base de données une ligne correspond à un acte et un même patient peut donc être enregistré autant de fois qu'un acte le concernant est réalisé. Nous avons déterminé que l'import d'un acte peut être identifié par la combinaison de 5 variables dénommées par la suite les variables identifiantes d'actesⁱⁱⁱ. Un patient peut être identifié à l'aide d'un identifiant unique grâce à la variable **PatientID**.

Enfin, en faisant l'hypothèse que la base de données était vierge de doublons, nous avons recherché pour les 24 variables sélectionnées la présence de données manquantes (informations non renseignées) et de données aberrantes (information renseignée mais incohérentes avec d'autres informations de la base ou avec des référentiels). Par exemple, le nombre de modalités annoncé dans le glossaire du CIMA ou le SIC pour une variable a été comparé avec celui retrouvé dans les résultats du tri à plat de cette variable ; la distribution des échelles du MMSE et des IADL ont été analysées en regard des résultats obtenus dans les études de référence sur le sujet ; ou encore la cohérence dans la chronologie des dates a été recherchée.

Les analyses ont été réalisées à l'aide du logiciel SAS Enterprise Guide français (France) version 4.3.0.0. L'analyse de l'existence de doublon d'acte a été réalisée à l'aide de la procédure PROC SORT en utilisant simultanément les options « SORTDUP=LOGICAL » et « NODUPRECS ».

4.2 Résultats

4.2.1 Prise en main des données et compréhension des documents

Plusieurs éléments ont rendu la prise en main des données et la compréhension des documents difficiles. Parmi les 5 documents (cf. § 3.5 et § 4.1.1) décrivant la base elle-même (avec le SIC), les items de la base (CIMA et glossaire du CIMA), ou le processus d'anonymisation (le manuel d'utilisation de la librairie anonymat et le programme Anonymisation) à notre disposition, des informations incohérentes ont été retrouvées entre les documents. De plus, les informations sur les variables fournies dans les documents pouvaient pour certaines ne pas correspondre aux variables trouvées dans la base.

La première difficulté tient au fait qu'il n'y a pas de correspondance claire entre le libellé et le nom des variables dans les différents documents. Le libellé des variables n'est pas indiqué dans le SIC et les noms des variables ne sont pas précisés dans le CIMA ou dans le glossaire du CIMA. Cela rend difficile la recherche d'information pour certaines variables lorsque le nom de la variable n'est pas explicite.

La seconde difficulté réside dans le fait qu'il est difficile de retracer l'historique des évolutions du CIMA. Les dates de mises à jour des différentes versions des documents ne sont pas homogènes (le SIC date de 2012 alors que le CIMA et le glossaire datent de 2013). De plus, il n'est pas précisé dans les documents à quelle date une modification a été introduite dans la base. Le SIC n'indique pas les modifications apportées avant la version 5.1. Pour les modifications apportées à partir de la version 5.1, la date de mise en œuvre de modifications pour chaque révision n'est pas systématiquement précisée. Concrètement, le manque d'information sur l'historisation des variables ou des modalités des variables ne permet pas de savoir s'il est cohérent d'observer une valeur manquante pour une variable ou non : à la

ⁱⁱⁱ Les variables identifiantes d'actes sont : les variables intitulées **ID**, **NumFic**, **TsCreation**, **DateFinPeriod** et **CentreNom**. La variable **CentreNom** a été ajoutée en raison de la non-homogénéité du libellé d'un même centre au cours du temps.

date de l'acte, cette variable existait-elle ou pas ? Ces interrogations sont valables aussi pour la non-complétude d'une modalité d'une variable. Ainsi, par exemple la variable **PatientProfession** a été ajoutée avec la version 5.5, soit à partir du 1^{er} février 2011, dans tous les centres. Or elle est retrouvée complétée dès 2010 avec l'utilisation de la modalité « sans information », alors que cette dernière est censée n'apparaître qu'avec la version 5.7. Enfin, lorsque des dates sont disponibles, elles sont parfois incohérentes. Ainsi, pour les modifications apportées à l'occasion de la mise en place de la version 5.5, il est précisé dans le SIC que cette nouvelle version est disponible à partir du 01 janvier 2011 et qu'elle sera effective dans tous les centres à partir du 01 février 2011. Or, la version 2 a été mise en place le 06 janvier 2011 selon ce qui est mentionné dans le glossaire du CIMA. Ainsi, la chronologie de ces versions semble contradictoire.

La troisième difficulté concerne des différences observées entre les documents au sujet de certaines procédures, comme celle de l'anonymisation. La procédure d'anonymisation référencée dans le SIC est celle du hachage SHA-2^{iv} alors que dans les spécifications fonctionnelles et techniques - Librairie anonymat, manuel d'utilisation, version 2.2, il s'agit du hachage SHA-1 puis SHA-256.

Pour rendre plus aisé et facile l'appropriation de la BNA, il est nécessaire :

➤ de limiter le nombre de documents pour pouvoir assurer une mise à jour cohérente et aisée pour les équipes qui les réalisent et compréhensible pour les utilisateurs. La multiplication de documents rend en effet peu aisée la compréhension de la structure de la base et de ses évolutions.

➤ de synthétiser au sein d'un même document l'intitulé de la question dans le CIMA, le nom de la variable et ses caractéristiques (obligatoire ou facultative, format, longueur, modalités). Ce document devrait comporter aussi toutes les évolutions de chacune des variables au cours du temps avec les évolutions des règles de complétude, le cas échéant.

4.2.2 Évaluation du processus d'intégration et de contrôle de la qualité des données

4.2.2.1. Constitution de la base de données

Lors de la constitution d'une base de données, il est décrit plus particulièrement dans le SIC que des contrôles de qualité sont mis en place à différents moments du processus, l'objectif étant de s'assurer de la qualité des données colligées et de leur exhaustivité. Que ce soit pour un fichier d'acte ou un fichier de mise à jour patient, les cinq étapes de constitution sont détaillées et analysées ci-dessous :

1. **Extraction des informations** pour constituer le fichier d'acte ou de mise à jour patient. Le fichier d'acte est constitué de l'ensemble des variables (cf. SIC, le § « Exemple de fichier d'acte »). Le fichier de mise à jour patient est constitué de l'ensemble des variables mais où seules 9 d'entre-elles contiennent une information, les 92 restantes étant vides (cf. SIC, le § « Contenu du fichier de mise à jour patient » et « Exemple de fichier de mise à jour patient »).

^{iv} Les algorithmes SHA-1 puis SHA-2 (ou SHA-256) ont été préconisés par l'Agence nationale de la sécurité des systèmes d'informations suite au « risque de collision » rencontré récemment avec SHA-1. Avec SHA-1, il est très difficile, mais possible, d'obtenir le même numéro d'identification à partir de deux individus différents.

◇ Dans les documents mis à notre disposition, il n'est pas indiqué si les applications disponibles dans les centres intègrent des contrôles en cours de saisie lors de la collecte des informations. Or, ces contrôles de saisie permettent de limiter directement au moment de la saisie des données les valeurs manquantes et les incohérences. En l'absence de cette information, nous n'avons pas pu vérifier que les différents éditeurs de logiciels détectent les incohérences et les erreurs avec le même succès.

2. **Anonymisation** : avant envoi, il est mentionné dans le SIC (cf. SIC, le § « Principe de l'interface ») que chaque fichier est anonymisé, que ce soit un fichier d'acte ou de mise à jour patient. La procédure d'anonymisation décrite dans ce document et le document intitulé « programme anonymisation – Spécifications fonctionnelles et techniques (version 1.0.0, disponible en téléchargement sur le site) est faite sur 6 variables d'identification du patient, disponibles dans la base du centre. Ces 6 variables d'identification du patient sont : la date de naissance, le nom de naissance, le prénom, le code département, le code commune et le nom de famille (ou nom usuel).

◇ Est-on sûr que les fichiers mis à jour sont anonymisés de telle sorte qu'ils correspondent au fichier patient initial ?

◇ Après l'anonymisation, il peut être indiqué si une erreur est détectée, sans que les documents ne précisent de quels types d'erreurs il peut s'agir. Il n'est pas précisé, dans les différents documents mis à notre disposition, si le centre conserve sous un format quelconque une correspondance entre l'identité du patient et le numéro d'anonymat. Or, ceci devrait permettre de s'assurer qu'un même numéro lui est attribué lors de la transmission ultérieure d'un acte réalisé au sein du même centre. De même, il n'est pas expliqué si avant l'envoi le centre vérifie que les numéros d'anonymat produits ne sont pas déjà utilisés pour un autre patient du centre ou d'un autre centre. Inversement, il n'est pas expliqué si, avant envoi, le centre vérifie que le numéro d'anonymat est correctement réattribué à chaque nouvel acte du patient. Dans tous les cas, il est nécessaire de vérifier qu'il s'agit de la même personne. L'étude de l'attribution par erreur du même numéro d'anonymat sera étudiée plus en détail par la suite (cf. § 4.2.2.3.2 L'identifiant unique, PatientId, est-il effectivement unique ?).

3. **Transmission** : les fichiers sont envoyés de façon sécurisée, via un « tunnel sécurisé ».

4. **Intégration** : les informations décrites pour comprendre comment les données sont collectées et sont intégrées dans la base BNA sont dispersées dans les différents documents à notre disposition et même parfois disséminées au sein d'un même document ce qui en rend l'appropriation difficile. Certaines informations ne sont pas accessibles (ex : la procédure « M34_procédure_mise_en_place_révisions », la description des tests d'intégration). Or, il est utile que toutes les informations sur un point particulier puissent être décrites au moins une fois en détail dans un document synthétique dans une seule et même partie. Comme cela est évoqué dans le paragraphe « Déroulement d'une intégration » du SIC, chaque fichier reçu est intégré dans la BNA s'il répond à la procédure « M34_procédure_mise_en_place_révisions » et les tests d'intégration sont réalisés.

◇ La procédure « M34_procédure_mise_en_place_révisions » et les tests d'intégration réalisés ne sont pas décrits dans les documents transmis ou mis à disposition sur le site. Nous ne pouvons pas savoir quelle est cette procédure et quels sont les tests utilisés, ni quel est leur impact sur la qualité des données de la BNA.

☞ Il est proposé d'expliciter les tests d'intégration réalisés pour l'import des fichiers dans la base mais qui ne sont pas décrits dans les documents. Ces tests d'intégration ne sont décrits ni dans le SIC ni dans un autre document disponible sur le site internet du CHU de Nice. Comme nous le verrons dans la partie sur la qualité des données, les tests d'intégration ne semblent pas intégrer de vérifications logiques sur les variables (cf. §4.2.2.3.4 Incohérences observées).

5. **Retour d'information** : après l'intégration d'un lot de données, un message est transmis au centre ; il y est précisé si l'intégration du fichier est effective ou non avec un compte-rendu permettant d'identifier le cas échéant les erreurs (cf. § « Principe de l'interface » du SIC). Ce compte-rendu liste l'ensemble des fichiers reçus. Le paragraphe « Structure des comptes-rendus quotidiens d'intégration » du SIC précise que le fichier reçu peut ne pas avoir été intégré dans la BNA. Dans ce cas, le numéro de ligne où un problème a été identifié est précisé ainsi qu'une description du problème. La nature des erreurs détectées n'est pas précisée dans les documents et les informations disponibles ne précisent pas :

◆ i/ Quel est le délai entre la réception du fichier et le renvoi du compte-rendu d'intégration ? Après interrogation du CHU de Nice, il semble que : « *Le compte-rendu d'intégration est envoyé dans la nuit ou le jour suivant le jour de réception du fichier* ».

ii/ Quelles sont les erreurs détectées par les tests d'intégration ?

iii/ Qu'est-ce qui doit être entrepris par le centre si des erreurs d'intégration lui sont retournées et quel est le délai pour que ce fichier soit finalement intégré ou non dans la BNA ?

iv/ Si le centre ne corrige pas les erreurs identifiées, le fichier est-il malgré tout intégré dans la base BNA ? Dans ce dernier cas de figure, est-ce que ces erreurs sont indiquées d'une quelconque façon dans la base (un code erreur spécifique au niveau de la variable comportant une erreur identifiée ou bien une variable de la base permettrait d'identifier les lignes d'observation avec une ou des erreurs) ?

◆ Si pour un patient toutes les informations « obligatoires » ne sont pas renseignées, le fichier est-il tout de même transmis ? Si ce n'est pas le cas, cela est-il connu du centre et du CHU de Nice ? Cette information permettrait de suivre la mise à jour du fichier et qu'il soit véritablement intégré après correction.

La réponse apportée par le CHU de Nice : « *Les contrôles sont normalement effectués aux niveaux des systèmes de recueil de données qui ne doivent pas envoyer le fichier si toutes les informations obligatoires ne sont pas renseignées. Si un fichier comportant une donnée obligatoire non renseignée est quand même envoyé, alors il est rejeté au niveau de la BNA et la nature du rejet est précisée dans le compte rendu d'intégration retourné au centre. S'il n'est pas envoyé, la BNA n'a aucun moyen de le savoir. Il est très difficile pour un centre d'effectuer une correction du fait de l'anonymisation. En effet, le système est conçu pour que l'on ne puisse pas identifier le patient dans la BNA. La BNA ne peut donc pas transmettre une identification du patient au centre dans le message d'erreur. Le centre devra revoir toutes les saisies qu'il a faites dans la journée pour identifier le patient dont le corpus n'a pas été intégré.* »

◆ Dans les documents transmis, il n'est pas indiqué avec quelle périodicité les données sont transmises au CHU de Nice par les centres. Les comptes-rendus d'intégration y sont quotidiens. Le sont-ils pour les derniers fichiers transmis ou intègrent-ils aussi l'historique des fichiers où une erreur a été retrouvée et n'a pas été corrigée ?

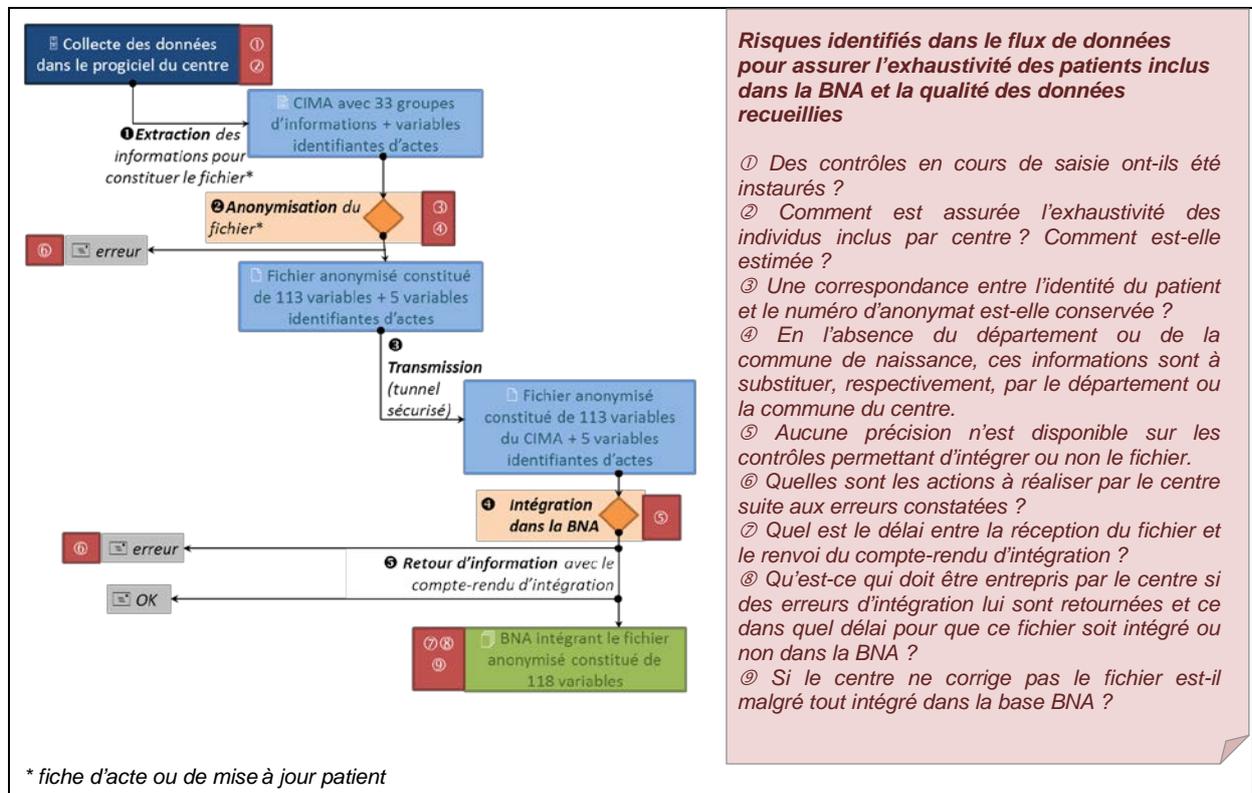
La réponse apportée par le CHU de Nice : « Les comptes-rendus ne concernent que les derniers fichiers transmis. »

❖ Il est important de pouvoir corriger les erreurs observées dans les fichiers pour assurer une bonne qualité des données et l'exhaustivité des individus inclus dans la BNA par rapport aux individus se présentant dans les centres mémoires. L'identification des lignes d'actes n'ayant pas été transmises ou transmises avec une erreur est indispensable pour estimer la qualité et l'exhaustivité des actes transmis pour chacun des individus au niveau central du CHU de Nice et pour chaque centre. Or, il n'est pas certain que cette exhaustivité des transmissions pour chaque centre puisse être estimée compte tenu des informations dont nous disposons.

La réponse apportée par le CHU de Nice : « La BNA stocke des informations de tous les fichiers qui lui sont envoyés avec s'il y en a un, le motif du rejet. Comme un fichier correspond à un acte d'un patient, [il est possible d'] estimer quel est le pourcentage de rejet d'actes par rapport aux envois, éventuellement par année de réception et/ou d'acte et par centre. L'identifiant du patient est récupérable, il est donc possible d'estimer un nombre de rejet patients par année et par centre. Par contre, [il n'est pas possible] d'estimer si le centre envoie systématiquement un corpus pour chaque acte effectué et donc pour chaque patient se présentant dans la CM ».

I FIGURE 1 I

Schéma simplifié du flux des données pour la constitution de la base BNA et risques identifiés pour garantir au mieux l'exhaustivité des cas et de la qualité des données



Dans les différents documents transmis, deux étapes de la constitution des données de la BNA (cf. § 3.7) font l'objet d'un contrôle qualité : il s'agit de l'anonymisation des identités des patients et de l'intégration des données dans la BNA. Dans les documents il est retrouvé qu'après l'anonymisation, il peut être indiqué si une erreur est détectée. De même, après l'intégration d'un fichier, un message est transmis au centre où il est précisé si l'intégration du fichier est effective ou non. Pour ces deux étapes, peu voire aucune précision sont apportées sur les contrôles qualité effectivement réalisés sur les informations saisies.

Nous n'avons pas connaissance d'autres processus mis en place pour garantir la qualité des données, soit lors de la collecte, soit au cours des différentes étapes du flux de données (cf. figure 1). Il n'a pas été retrouvé de documents, publiés ou non, sur la qualité du codage des variables : double saisie d'un même dossier, tirage au sort de dossiers pour recodage, vérification des incohérences avant transmission...

Pour les points relatifs à cette partie, nos recommandations s'attachent à indiquer qu'il serait nécessaire :

- que les lignes comportant des erreurs puissent être identifiées facilement par les personnes qui sont amenées à utiliser la base.
- de « geler » la base lorsqu'il est estimé que toutes les données d'une période donnée sont transmises et intégrées. Ainsi, la base de données est réputée consolidée et elle peut être analysée. Il n'est pas précisé dans les différents documents transmis ou sur le site internet si la BNA est gelée à certaines périodes.
- que le principe des contrôles qualité mis en place soient décrits dans les documents transmis afin d'indiquer aux équipes des centres mémoire le degré de précision demandée pour chaque variable.
- que les contrôles de cohérences puissent être réalisés en cours de saisie lors d'un enregistrement avec l'implémentation de masques de saisie sur le logiciel de saisie ; une uniformisation des contrôles entre les différents logiciels sera à vérifier. Des contrôles de cohérences pourraient être conduits aussi sur les enregistrements d'un centre, d'un seul patient ayant été dans plusieurs centres ou non, mais aussi sur l'ensemble de la base.
- que des bilans sur la qualité des données puissent être produits par centre périodiquement afin de vérifier que les consignes de collectes sont assimilées et appliquées par les équipes sollicitées pour la collecte des informations.
- que des méthodes ou processus appropriés soient mis en place (double saisie d'un même dossier, tirage au sort de dossiers pour recodage par exemple) pour évaluer la qualité des données saisies.
- que soit précisé lors de l'introduction d'une nouvelle révision, comment cette information est diffusée dans les centres et dans quel délai le CHU de Nice est assuré que cette révision est effective et maîtrisée dans les centres.

4.2.2.2. Comparaison de la structure de la base de données avec les documents

Pour notre objectif d'évaluation de la faisabilité d'une surveillance épidémiologique en population générale à partir de la BNA, nous avons examiné les documents et la base pour sélectionner les variables d'intérêt pour nos analyses.

Le fichier de la base de données BNA transmis par le CHU de Nice est accompagné de 30 fichiers (cf. Annexe 2. Variables dans les bases de données transmises). Aucun document n'indiquait leur utilité ni quelles variables étaient rattachées à chacun de ces 30 fichiers. Pour certains d'entre eux, il n'est pas possible de savoir si un fichier est rattaché à une ou plusieurs variables. Après lecture, il s'avère que ces 30 fichiers permettaient d'attribuer les libellés de certaines modalités pour des variables de la base de données BNA.

➔ Pour la diffusion de la base il est nécessaire de décrire tous les fichiers transmis en plus de la base afin de préciser si ce sont des fichiers à joindre à la base totale et d'indiquer leur utilité.

Des incohérences ont également pu être relevées entre la structure de la base de données, le SIC, le CIMA et le glossaire du CIMA. Il est important de noter que les libellés des variables ne sont pas indiqués dans le SIC et les noms des variables ne sont pas précisés dans le CIMA ou dans le glossaire du CIMA. Nous avons donc procédé à une correspondance *a priori* entre les informations contenues dans le SIC et celles du CIMA.

En raison du fait que le libellé en clair de certaines variables n'est pas connu, qu'il n'existe pas de tableau de correspondance entre les informations du CIMA et les variables des fichiers et que les intitulés des variables lorsqu'il sont connus sont très long (ex : « Liste diagnostics fédération des CMRR » pour **DcIDiagnostic** ou « Examens dont les résultats ont été consultés le jour de l'acte » pour **DcIPCAucun**), nous avons fait le choix pour ne pas alourdir le texte de présenter les variables par leur nom retrouvé dans la base. Elles sont mentionnées en gras dans le corps du rapport.

Globalement deux types d'incohérence peuvent être répertoriés lorsque l'on compare les éléments de la base de données avec les documents en notre possession :

1. Certaines variables existant dans la base ne sont pas retrouvées dans les documents ou inversement. Ainsi par exemple :

- Six variables apparaissent dans la base de données alors qu'elles ne figurent pas dans le SIC. Il s'agit des variables intitulées **CentreDepartement**, **CentreRegion**, **DateFinPeriod**, **ID**, **NumFic** et **TsCreation**. Les deux premières correspondent très certainement au département du centre et à la région du centre.
- La variable concernant la date de fin de suivi (**DcIDateFinSuivi** selon le SIC) est dans la base de données mais n'est pas mentionnée dans le CIMA ni dans le CIMA glossaire.
- Trois variables prévues dans le CIMA ne sont retrouvées ni dans la base ni dans les documents (ex : les variables **IADLConsequence**^v, **PatientNouveau**^{vi} et **DcIFinSuivi**). Deux de ces variables (**IADLConsequence**^{vii} et **PatientNouveau**) n'ont pas pu être étudiées alors qu'elles avaient été retenues pour notre analyse.

2. Les modalités ou format des variables diffèrent entre les documents et ce qui est mentionné dans la base (cf. tableau 3 et l'annexe 3. Comparaison de la structure de la base de données avec les documents) pour le détail sur les 24 variables présélectionnées). Plus précisément :

^v Cette variable est retrouvée dans le SIC et la base mais pas dans le CIMA ou le glossaire du CIMA ; elle n'a pu être incluse dans notre sélection.

^{vi} Cette variable **PatientNouveau**, créée pour la version 5.5 disparaît avec la version 5.6, n'a pu être présélectionnée puisqu'elle a disparu dans la base transmise.

^{vii} Cette variable est retrouvée dans le SIC et la base mais pas dans le CIMA ou le glossaire du CIMA ; elle n'a pu être incluse dans notre sélection.

- On observe des incohérences entre les formats numérique ou alphanumérique (i.e. texte) de certaines variables entre la base et les documents (e.g., les variables **PatientSexe** ou **DclDateEntreInstitution**) et de longueur (e.g., les variables **PatientID**, **PatientSexe**, **PatientAnneeNaissance**...);
- Les modalités indiquées ne sont pas les mêmes entre les documents (SIC, CIMA ou glossaire du CIMA) et la base pour certaines variables (e.g., **DclPCAucun**, **DclTPAucun**, **DclIPSAucun**, **DclTypeActe**, et **PatientProfession**);

*ex : Cette dernière variable (**PatientProfession**) permet d'enregistrer selon le CIMA et le glossaire du CIMA, la dernière profession exercée. Cette variable n'est pas précisée dans le SIC. Dans le CIMA, la variable dispose de 12 modalités dont 2 correspondent à des lieux de résidence ; alors que dans le glossaire du CIMA, le SIC et la base, cette variable est déclinée en 10^{viii} modalités. Il peut être noté que le lieu de résidence fait l'objet d'un recueil spécifique avec la variable **DclModeDeVie** qui correspond au « mode de vie actuel » et dont les modalités correspondent à des lieux de résidence. Le mode de vie peut aussi faire référence au statut marital.*

- Les informations données au sein d'un même document peuvent aussi être discordantes ; ex : selon le glossaire du CIMA, le score aux IADL a été introduit à partir de la révision 2 (appliquée le 06/02/2011). Or, il est indiqué aussi que quatre autres modalités, utilisées précédemment, ne le sont plus à partir de la révision 2. Ces informations incohérentes laissent à penser que le score aux IADL a été introduit avant la révision 2 avec d'autres modalités en plus de celles conservées avec la révision 2.
- le SIC indique que seules 9 variables sont remplies lorsqu'il s'agit de la mise à jour d'un fichier patient (**CentreMail**, **CentreNum**, **PatientId**, **DclTypeFinSuivi**, **DclDateFinSuivi**, **DclDateEntreInstitution**, **DclDateDeces**, **RetourMail**, et **Application**). Or, il est observé dans la base que cette règle n'est pas systématiquement suivie.

Nous avons sélectionné au total 24 variables, certaines variables n'ont pas été sélectionnées en raison soit de leur manque d'intérêt pour nos analyses (ex : les variables relatives à l'activité du centre) soit de la difficulté à les interpréter et des incohérences constatées. Ces variables sont décrites dans le tableau 3 et l'Annexe 3. Comparaison de la structure de la base de données avec les documents.

➡ Ces mêmes analyses pourraient être menées pour toutes les variables de la base afin d'actualiser les documents d'informations.

^{viii} Les modalités utilisées pour la profession exercée reprennent partiellement le niveau 1 de la PCS de l'Insee (http://www.insee.fr/fr/methodes/default.asp?page=nomenclatures/pcs2003/liste_n1.htm).

I TABLEAU 3 I

Description des difficultés identifiées pour la prise en main des 24 variables présélectionnées et étudiées

Nom de la Variable <i>Libellé supposé de la variable</i>	Difficultés retrouvées pour la prise en main de la base et des différents documents fournis
Variabiles géographiques	
CentreRegion <i>Région du centre</i>	Variable non décrite dans les documents fournis. Caractère obligatoire ou facultatif, non disponible. Date de création non connue Rattachée <i>a priori</i> à la table PBNAREGIONS
CentreDepartement <i>Département du centre</i>	Variable non décrite dans les documents fournis. Caractère obligatoire ou facultatif, non disponible. Date de création non connue Rattachée <i>a priori</i> à la table PBNADepARTEMENTS
Variabiles « patient »	
PatientId <i>Identifiant unique du patient</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Obtenue avec l'algorithme d'anonymisation (SHA-1 ou SHA-2 selon les documents). Variable de 64 caractères selon le SIC et de 300 dans la base
PatientSexe <i>Sexe du patient</i>	Modalités Femme/Homme selon le SIC et F/H selon le CIMA. Format numérique selon le SIC et alphanumérique dans la base. Longueur de la variable : 1 ou 3 caractères selon la source.
PatientAnneeNaissance <i>Année de naissance du patient</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Longueur de la variable : 4 ou 8 chiffres selon les documents et la base.
DclAnneePremierDiag <i>Année de première consultation</i>	Date de création non connue. Variable devenue obligatoire avec la version 5.5 ; le libellé est modifié avec la version 5.6. Les modifications avant la version 5.1 ne sont pas connues. Longueur de la variable : 4 ou 8 chiffres selon les documents et la base.
DclALD <i>ALD</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAALDS. Longueur de la variable : 2 ou 8 chiffres selon les documents et la base.
DclAPA <i>APA</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAAPAS. Longueur de la variable : 2 ou 8 chiffres selon les documents et la base.
PatientAccompagnant <i>Caractéristiques de l'accompagnant</i>	Date de création supposée le 01/01/2011. Création avec la version 5.5. Variable modifiée avec la 5.7 (ajout de modalités*). Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAACCOMPAGNANTS. Longueur de la variable : 8 ou 9 chiffres selon les documents et la base.
PatientEnvoyePar <i>Patient envoyé par/mode d'adressage</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAPROVSPAT. Libellé incertain : « Patient envoyé par » dans le CIMA et glossaire du CIMA ; « mode d'adressage » dans le tableau 1. Longueur de la variable : 6 ou 8 chiffres selon les documents et la base.
PatientNiveauEtudes <i>Niveau d'étude</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNANIVEAUXET. Longueur de la variable : 6 ou 8 chiffres selon les documents et la base.
DclMesureProtection <i>Mesure de protection</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAMESPROT. Longueur de la variable : 8 ou 9 chiffres selon les documents et la base.
DclModeDeVie <i>Mode de vie actuel</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAMODESVIE. Longueur de la variable : 8 ou 10 chiffres selon les documents et la base.
PatientProfession <i>Profession – dernier métier exercé / Profession – dernière profession exercée</i>	Date de création non connue. Création avec la version 5.5. Variable modifiée avec la 5.7 et 5.7 (ajout d'une modalité*). Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAPROFESSIONS. Le CIMA indique 12 modalités contre 10 dans la base ; celles relatives au lieu de résidence ne sont pas retrouvées dans la base. Longueur de la variable : 8 ou 10 chiffres selon les documents et la base. Libellé incertain : « Profession – dernier métier exercé » dans le CIMA et glossaire du CIMA ; « Profession – dernière profession exercée » dans le tableau 1.
Variabiles « acte »	
DclDateActe <i>Date de l'acte</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues.

Nom de la Variable <i>Libellé supposé de la variable</i>	Difficultés retrouvées pour la prise en main de la base et des différents documents fournis
DclTypeActe <i>Type d'acte</i>	Date de création non connue. Variable modifiée avec les versions 5.2 (ajout d'une modalité*), 5.5 (ajout de modalités, sans précision) et 5.8 (ajout d'une modalité* qui existe dans la base depuis 2010). Apparition de la modalité « intervention Assistance sociale » dès 2011. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNATYPESACTE.
DclEtatDemence <i>Diagnostic niveau 1 / Diagnostic N1</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Libellé incertain : « Diagnostic niveau 1 » dans le CIMA et glossaire du CIMA ; « Diagnostic N1 » dans le Erreur ! Résultat incorrect pour une table.
DclDiagnostic <i>Diagnostic niveau 2 / Diagnostic N2</i>	Date de création non connue. Variable modifiée avec les versions 5.1 (suppression du diagnostic G30.1, lésion vasculaire ou AVC sont codés F01.81) et 5.4 (suppression du diagnostic G30.8). Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNADIAGS. Longueur de la variable : 6 ou 8 chiffres selon les documents et la base. Libellé incertain : « Diagnostic niveau 2 » dans le CIMA et glossaire du CIMA ; « Diagnostic N2 » dans le Erreur ! Résultat incorrect pour une table.
DclIADLDetection <i>IADL score détection / IADL / Score des IADL</i>	Date de création non connue. Cette variable a été révisée avec la version 2 selon le glossaire du CIMA. Les modifications avant la version 5.1 ne sont pas connues dans le SIC. Longueur de la variable : 2 ou 8 chiffres selon les documents et la base. Libellé incertain : « IADL score détection » dans le CIMA ; « IADL » dans le glossaire du CIMA ; et, « Score des IADL » dans le Erreur ! Résultat incorrect pour une table. Une autre variable existe dans la base DclIADLConsequence, qui n'est pas définie ailleurs.
DclIMMSE <i>Score au Mini Mental test</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Longueur de la variable : 2 ou 8 chiffres selon les documents et la base.
DclPCAucun <i>Examens dont les résultats ont été consultés le jour de l'acte</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Modalités en « Oui/Non » dans le SIC et en « 0/1 » dans la base. Longueur de la variable : 1 ou 8 chiffres selon les documents et la base.
DclTPAucun <i>Traitements pharmacologiques</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Modalités en « Oui/Non » dans le SIC et en « 0/1 » dans la base. Longueur de la variable : 1 ou 8 chiffres selon les documents et la base.
DclFinSuivi <i>Type de fin de suivi / État du suivi</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattachée <i>a priori</i> à la table PBNAFINSSUIV. Libellé incertain : « Type de fin de suivi » dans le glossaire du CIMA et dans le CIMA ; et, « État du suivi » dans le tableau 1. Variable décrite dans le SIC et la base ; deux variables peuvent y être rattachées dans le CIMA et dans le glossaire du CIMA (variables « Type de fin de suivi » avec 8 modalités dont « En cours » et « État du suivi »). Longueur de la variable : 7 ou 8 chiffres selon les documents et la base.
DclIPSRAucun <i>Traitements non pharmacologiques</i>	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Modalités en « Oui/Non » dans le SIC et en « 0/1 » dans la base. Longueur de la variable : 1 ou 8 chiffres selon les documents et la base.

¹ Num., numérique

* modalité précisée dans le SIC

4.2.2.3 Qualité des données de la BNA

4.2.2.3.1. Vérification de la non-existence de doublons

Comme nous l'avons précédemment indiqué, la base de données est constituée d'une ligne pour chaque acte, plusieurs actes pouvant être enregistrés pour un même individu. Or, au cours de l'analyse de la base, nous avons observé que des actes strictement identiques étaient enregistrés plusieurs fois pour un même patient. Des lignes d'actes strictement identiques sur l'ensemble des variables semblent ainsi remonter plusieurs fois pour le même

centre à des dates différentes et sont intégrées dans la base avec des variables identifiantes d'acte (i.e., **ID**, **NumFic** et **TsCreation**, **DateFinPeriod** et **CentreNom**^{ix}) différentes.

❖ Nous ne pouvons expliquer cette présence de doublons d'actes. Ce que nous observons c'est que des observations identiques sur l'ensemble des variables remontent plusieurs fois pour le même centre à des dates différentes et sont intégrées dans la base avec des variables identifiantes d'acte différentes. Il pourrait peut-être s'agir soit d'actes enregistrés plusieurs fois, soit d'actes corrigés suite à l'identification d'erreur, sans que la ligne d'origine ne soit supprimée. Mais pour le confirmer, il nous aurait fallu vérifier in situ comment les centres envoient leurs fichiers et comment ils sont intégrés dans la base par le CHU de Nice.

Au total, **33 178 doublons** d'actes ont ainsi été mis en évidence et supprimés pour la suite des analyses, nous conduisant à retenir **1 429 270** observations (cf. figure 2).

➡ Afin d'éviter les multiples enregistrements d'un même acte, des procédures de vérification peuvent être mises en place en amont, et les doublons doivent être éliminés des bases transmises pour analyse. Cette vérification ne semble pas faire partie des tests d'intégration réalisés jusqu'alors ou les tests mis en place ne semblent pas parer à de telles éventualités. L'intégration des doublons engendre des effectifs erronés pour les estimations faites à partir d'une base non nettoyée des doublons.

4.2.2.3.2 L'identifiant unique, **PatientId**, est-il effectivement unique ?

La procédure d'anonymisation retenue (hachage) utilisée pour la BNA doit permettre d'attribuer un numéro d'identifiant unique et anonyme (**PatientId**) à un même patient quel que soit le centre où l'acte est réalisé. Les variables utilisées pour l'anonymisation sont le nom de naissance, le nom de famille le prénom, la date de naissance, le département et la commune d'habitation (cf. tableau 1).

Théoriquement le numéro d'identifiant anonyme peut ne pas toujours être unique, puisque la variable **PatientId** dépend :

- de la bonne orthographe faite sur les chaînes de caractère, du respect entre nom usuel et de naissance ainsi que du même nombre de prénoms retenu pour le dossier patient ;
- de la substitution du département et/ou de la commune de naissance par le département et/ou la commune du centre lorsque celles-ci sont inconnues ;
- des paramétrages utilisés pour procéder à l'anonymisation.

Il est ainsi possible que plusieurs numéros d'identifiant unique et anonyme **PatientId** soient attribués à un même sujet. Au sein d'un même centre, cela semble néanmoins peu vraisemblable. En effet, le numéro d'anonymat doit être conservé pour chaque patient et repris pour chaque nouvel acte de ce patient. En revanche, cela pourrait potentiellement survenir si un patient suivi dans un centre se présentait dans un autre centre. En effet, s'il n'y a pas une communication synchronisée entre les deux centres, le même patient pourrait se voir attribuer un nouveau numéro d'identifiant unique et anonyme **PatientID**. Cette erreur n'est toutefois pas quantifiable à ce jour et ne devrait concerner que très peu de cas.

L'Agence nationale de la sécurité des systèmes d'informations a relayé le « risque de collision », i.e., le risque d'obtenir le même numéro d'identification pour deux individus différents avec l'algorithme utilisé pour l'anonymisation dans la BNA (algorithme de hachage

^{ix} La variable **CentreNom** a été ajoutée en raison de la non-homogénéité du libellé d'un même centre au cours du temps.

SHA-1^x). Théoriquement, le risque de collision pour cet algorithme initialement estimé à 10^{-31} a récemment été réévalué à la baisse. D'autres algorithmes de hachage sont recommandés depuis octobre 2015 (SHA-2, encore appelé SHA-256, ou SHA-3^{xi})

Dans la BNA, nous avons observé que 88 (soit 0,016 %) des 556 069 **PatientID** ont un même numéro d'identifiant unique et anonyme (**PatientID**) alors que le sexe et ou l'année de naissance changent d'un acte à l'autre au cours du suivi. Le risque de collision observé à partir de ces trois variables dans la BNA est ainsi de l'ordre de 10^{-4} ; bien au-delà du risque de 10^{-31} théoriquement attendu. Mais cette analyse devrait être menée sur l'ensemble des variables pour obtenir l'effectif exact du nombre de numéro d'identification non unique par patient.

◇ Nous ne pouvons expliquer ce risque de collision. Des différences ont été constatées selon les documents auxquels on se réfère (soit, hachage SHA-1, soit hachage SHA-1 puis SHA-256, ou encore hachage SHA-256) sur la procédure d'anonymisation. L'attribution du numéro d'identifiant unique et anonyme du patient **PatientID** pourrait-elle être différente d'un centre à l'autre ou par un même centre au cours du temps selon la procédure suivie ? Certains centres ont-ils toujours appliqué uniquement le hachage SHA-1 ou alternent-ils indifféremment SHA-1 et SHA-256 ?

◇ À l'inverse, il n'a pas été possible, compte tenu des données dont nous disposons, de confirmer ou infirmer si un même patient pouvait avoir deux ou plusieurs numéros d'anonymisation différents, au cours du temps dans un même centre ou dans un autre centre.

Au total, la base est constituée de 556 069 numéros d'identifiant **PatientID** présumés anonymes et uniques après suppression des doublons d'actes sans qu'il nous soit possible de vérifier qu'il s'agit bien de 556 069 patients différents (cf. figure 2).

➔ Il est essentiel de :

- rechercher pourquoi l'algorithme permettant de générer un identifiant unique pour chaque patient a une si grande probabilité de collision (observée dans la BNA à 10^{-4} contre 10^{-31} estimée *a priori*).
- vérifier comment l'algorithme de hachage est utilisé par les centres ainsi que le paramétrage utilisé par chacune des applications ou des centres.
- savoir si le numéro d'anonymat est conservé par le centre afin de pouvoir le réattribuer au même patient lors du prochain export ? Au besoin, le dossier Cnil, voire CCTIRS/PPP, nécessitera des modifications pour le permettre.
- vérifier *a posteriori* qu'un même numéro d'anonymat est réattribué au même patient et pas à un autre patient sur l'ensemble des centres avant exploitation et diffusion de la base de données. Au besoin, le dossier Cnil nécessitera des modifications en ce sens.
- ne pas substituer (contrairement à la consigne prescrite dans le glossaire du CIMA) le département et/ou de la commune de naissance par le département

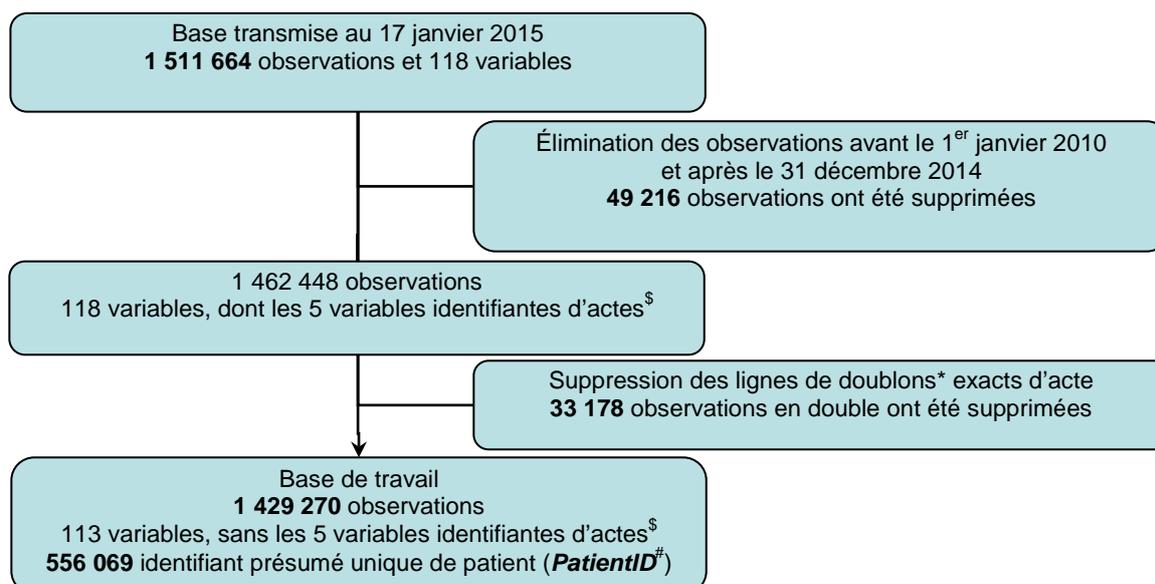
^x Agence nationale de la sécurité des systèmes d'informations. Bulletin d'actualité CERTFR-2015-ACT-042. 2015. Disponible sur : <http://www.cert.ssi.gouv.fr/site/CERTFR-2015-ACT-042/CERTFR-2015-ACT-042.html>, consulté le 17/02/2017.

^{xi} Agence nationale de la sécurité des systèmes d'information. Annexe B1 sur les mécanismes cryptographiques du référentiel général de sécurité 2.0 (version 2.03 du 21 février 2014, 63 p.). Disponible sur : http://www.ssi.gouv.fr/uploads/2014/11/RGS_v-2-0_B1.pdf, consulté le 17/02/2017.

et/ou la commune du centre lorsque celles-ci sont inconnues. Ces substitutions augmentent le risque de ne pas attribuer le même numéro d'identifiant unique et anonyme **PatientID** au même patient, en particulier si au cours d'une seconde visite ces données étaient actualisées ou si le patient allait consulter dans un nouveau centre.

I FIGURE 2 I

Consolidation de la base transmise



* *Doublon = lignes présentant la même valeur pour toutes les variables de la base hormis pour les variables identifiantes d'actes[§]. Une seule ligne est conservée dans la base de travail.*

[§] *Les 5 variables identifiantes d'actes sont : ID NumFic TsCreation DateFinPeriod et CentreNom*

[#] *sous l'hypothèse que le numéro d'identification anonyme PatientID est unique. Or, nous avons mis en évidence que cette hypothèse n'est pas vérifiée (cf. 4.2.2.3.2).*

4.2.2.3.3 Valeurs manquantes

Les analyses pour les valeurs manquantes sur les variables de la BNA ont été réalisées sur l'hypothèse qu'un seul numéro d'anonymisation **PatientID** est attribué par patient. Nous avons mis en évidence précédemment que cette hypothèse n'est pas forcément vérifiée (cf. § 4.2.2.3.1) pouvant entraîner des **effectifs observés qui ne correspondent pas au nombre de patients réellement inclus dans la base**. Aucune correction de la base n'est possible, aussi, les analyses suivantes sont-elles réalisées en considérant qu'un seul numéro d'identification est attribué par patient. **Les effectifs observés pour ce paragraphe et les suivants sont donc potentiellement inexacts.**

La BNA contient peu de données manquantes sur les 24 variables étudiées : 18 variables n'ont aucune donnée manquante. On observe néanmoins une proportion importante de valeurs manquantes pour trois variables particulièrement importantes pour l'analyse de la démence (cf. tableau 4) : le niveau d'études n'est pas renseigné pour 10 % des patients (**PatientNiveauEtudes**, variable fixe et obligatoire), le score MMSE (**DcIMMSE**, variable facultative et réactualisable) pour 15 % et le score aux IADL (**DcIIADLdetection**, variable facultative et réactualisable) pour 47 %, le remplissage de ces deux dernières variables étant

facultatif. Ce pourcentage de données manquantes évolue au cours du suivi du patient. Il diminue au cours du temps pour certaines variables, comme pour le niveau d'études (e.g., **PatientNiveauEtudes**) ; ce qui peut être interprété comme une meilleure qualité de remplissage au cours du suivi. Il peut aussi augmenter (e.g., **DcIMMSE**, **DcIIADLDetection**, **PatientAccompagnant**, **PatientProfession**) au cours du temps ou rester stable (e.g., **DcIFinSuivi** et les 18 autres variables, ces dernières restants à zéro) ; ceci indique qu'il n'y a pas d'amélioration de la qualité des données pour les variables mentionnées. Ces variations sont indiquées dans le tableau 4, où est documenté le pourcentage de données manquantes selon que le 1^{er} acte ou le dernier acte de chaque PatientID est pris en compte.

Pour des variables comme les scores aux IADL (**DcIIADLDetection**) ou le MMSE (**DcIMMSE**), deux variables facultatives qui supposent une mise à jour fréquente (tous les six mois), la proportion de données manquantes est plus faible dès qu'est pris en compte la première fois où le score est renseigné.

Parallèlement à ces données manquantes, certaines variables présentent des proportions élevées de modalités « sans information ». Ainsi, trois d'entre-elles : *Accompagnant du patient* (**PatientAccompagnant**), *Dernière profession du patient* (**PatientProfession**) et *Mesure de protection* (**DcIMesureProtection**) ont un pourcentage pour la modalité « sans information » qui varie de 16 à 68 % selon que l'on considère les variables enregistrées lors du premier enregistrement ou lors du dernier enregistrement (i.e., la dernière consultation) de chaque PatientID. Pour les variables *Accompagnant du patient* (**PatientAccompagnant**) et *Dernière profession du patient* (**PatientProfession**), l'analyse par année met en évidence que ces deux variables, créées en 2011 sont renseignées dès 2010, avec alors une forte proportion de « sans information » (plus de 97 %).

I TABLEAU 4 I

Pourcentage de valeurs manquantes pour les 24 variables, 2010-2014 (556 069 PatientID# présumée unique)

Libellé de la variable Nom de la variable étudiée (obligatoire/facultatif fixe/réactualisable)	Pourcentage de valeurs manquantes		Remarque
	Au premier acte pour chaque PatientID#	Au dernier acte pour chaque PatientID#	
Numéro d'identification PatientID	0	0	
Sexe PatientSexe	0	0	L'analyse met en évidence que pour certains patients, le sexe est recodé (parfois plusieurs fois) entre le premier et le dernier acte : 348 375 femmes et 207 694 hommes sont enregistrés au premier acte de chaque PatientID# contre 348 381 femmes et 207 688 hommes au dernier acte pour ces mêmes PatientID#. Le sexe d'un patient change aussi au cours des actes.
Âge PatientAnneeNaissance	0	0	
Date de l'acte DclDateActe	0	0	
Diagnostic niveau 1 DclEtatDemence	0	0	
Diagnostic niveau 2 DclDiagnostic	0	0	8 % de valeurs manquantes sur l'ensemble des enregistrements d'un PatientID# >20 % diagnostic en attente au premier acte et 34 % au dernier acte
Année de la première consultation DclAnneePremierDiag	0	0	
Score MMSE DclMMSE	21	32	15 % de valeurs manquantes sur l'ensemble des enregistrements
Score IADL Détection DclIADLDetection	55	59	47 % de valeurs manquantes sur l'ensemble des enregistrements
Patient envoyé par PatientEnvoyePar	0	0	
Accompagnant du patient PatientAccompagnant	0,1	0,3	65 % sans information Variable ajoutée avec la version 5.5, soit à partir 01/02/2011, dans tous les centres. Or elle est retrouvée complétée dès 2010 avec l'utilisation de la modalité « sans information », alors que cette modalité n'a été ajoutée qu'avec la version 5.7
Niveau d'études PatientNiveauEtudes	13	12	

Libellé de la variable Nom de la variable étudiée (obligatoire/facultatif fixe/réactualisable)	Pourcentage de valeurs manquantes		Remarque
	Au premier acte pour chaque PatientID [#]	Au dernier acte pour chaque PatientID [#]	
Profession du patient PatientProfession	0,1	0,3	68 % sans information, au premier acte ou au dernier acte Ajoutée avec la version 5.5, soit à partir 01/02/2011, dans tous les centres. Or elle est retrouvée complétée dès 2010 avec l'utilisation de la modalité « sans information », alors que cette modalité a été ajoutée avec la version 5.7
Patient avec une ALD DclALD	0	0	Près de 30 % sans information
Bénéficiaire de l'APA DclAPA	0	0	32 % sans information
Aucun traitement pharmaceutique DclTPAucun	0	0	
Fin de Suivi DclFinSuivi	2	2	
Mesure de protection DclMesureProtection	0	0	18 % sans information au premier acte de chaque PatientID [#] 16 % sans information sur l'ensemble des enregistrements de chaque PatientID [#]
Mode de vie DclModeDeVie	0	0	
Patient n'a aucune prise en charge DclPCAucun	0	0	
Région du centre CentreRegion ^α	0	0	
Département du centre CentreDepartement ^α	0	0	Pas de données pour les départements d'Eure-et-Loir et de la Creuse
Type d'acte DclTypeActe	0	0	
Type de centre CentreType	0	0	

^α Le statut réactualisable ou non est non précisé.

[#] Pour cette analyse, il est fait l'hypothèse que le numéro d'anonymisation PatientID est unique. Or, nous avons mis en évidence que cette hypothèse n'est pas nécessairement vérifiée.

4.2.2.3.4 Incohérences observées

Les analyses de l'incohérence entre les variables de la BNA ont été réalisées, comme pour les valeurs manquantes, en faisant l'hypothèse qu'un seul numéro d'anonymisation PatientID est attribué par patient.

Sur les 24 variables, 7 présentent des valeurs qui peuvent sembler aberrantes. Il s'agit de :

- l'année de naissance (variable **PatientAnneeNaissance**) : 763 (soit 0,1 %) des patients ont moins de 15 ans ;
- ➔ Pour cette observation, soit il s'agit soit d'erreurs de saisie, soit les populations couvertes diffèrent entre les centres.
- l'état de démence, diagnostic de niveau 1 (variable **DclEtatDemence**) : 3 cas (soit 0,5 pour 100 000 cas) sont dans une modalité non définie.

❖ Ceci nous amène à nous interroger sur les tests d'intégration qui sont réalisés et qui ne semblent pas intégrer la vérification des modalités autorisées pour toutes les variables.

➔ La base devrait être nettoyée avant transmission aux utilisateurs.

- l'état de démence, diagnostic de niveau 2 (variable **DclDiagnostic**) :

❖ Les codes G30.8 et G30.1, utilisés respectivement jusqu'à la version 5.4 et 5.1, ne sont pas retrouvés dans la base. Or, les documents disponibles ne permettent pas de savoir si ces modalités ont été recodées, respectivement, en F00.2 et F01.81, ou si elles ont été supprimées.

➔ Après « nettoyage » de la base, il est souvent recommandé de laisser en l'état les données collectées et corrigées. Ainsi, une base est réputée « gelée » à un temps donné sur une période définie. Toutes modifications dans les CM/CMRR ultérieures au gel de la base par le CHU de Nice ne seront prises en compte qu'à la mise à la mise à jour suivante.

- l'année de la première consultation (variable **DclAnneePremierDiag**)
Cette variable obligatoire a pour objectif d'identifier les patients n'ayant jamais consulté avant pour des troubles cognitifs. Elle nous intéresse car elle renseigne approximativement la date de début des troubles et pourrait peut-être nous permettre d'identifier les nouveaux cas de démence inclus dans la base.

Sa date de création est discordante entre le SIC (qui indique le 01/01/2011) et le glossaire du CIMA (qui mentionne juillet 2012). Par ailleurs, dans le CIMA, les instructions pour le remplissage de cette variable peuvent porter à confusion : « indiquer l'année de la première consultation pour plainte cognitive et/ou comportementale quel que soit le diagnostic et quel que soit le lieu de la consultation ». Il n'est pas clair s'il s'agit d'indiquer la date de la première consultation ayant permis un diagnostic de démence ou bien seulement la première consultation pour plainte sans que cela n'ait conduit à un diagnostic de démence. De plus, les diverses modifications de son intitulé au cours du temps rendent difficile l'utilisation de cette variable.

Sur l'ensemble de la période d'étude, quelques patients (337, soit 0,6 %) ont une année de 1^{re} consultation postérieure à l'année de l'acte.

➤ Malgré l'intérêt que pourrait avoir cette variable, la confusion sur la définition et les diverses modifications que cette variable a rencontrées en limite son utilisation.

- le score IADL (variable **DclIADLDetection**)

Le score IADL (Instrumental Activity of Daily Living) permet de mesurer le degré d'autonomie ou plus généralement de dépendance de la personne pour les activités de la vie quotidienne. Plus le score IADL est élevé, plus le patient est confronté à des déficits fonctionnels.

Nous avons déjà observé précédemment des discordances dans le glossaire du CIMA concernant la date de l'introduction du score aux IADL. Il n'est pas non-plus indiqué comment remplir cette variable dans la documentation. D'après les informations disponibles, il semble que la valeur du score corresponde au nombre d'atteintes déclarées et est attribuée de la façon suivante : score de 0 pour 4 atteintes ; score de 1 pour 3 atteintes ; score de 2 pour 2 atteintes ; score de 3 pour 1 atteinte ; et, score de 4 pour aucune atteinte. Plus la valeur de score est faible, plus le degré de dépendance du patient est élevé, donc contraire au sens de lecture habituellement trouvé dans la littérature et plutôt peu intuitif.

Et par ailleurs la présentation de l'échelle du score IADL dans le CIMA est ambiguë : pour les 4 items, chaque modalité est entourée de deux échelles de cotations disposée de part et d'autres du texte (cf. annexe 4). Enfin, le score final est demandé après le 3^e item et non le 4^e ce qui suggère de faire la somme sur les trois premiers items et non les quatre.

Nous observons ainsi que, sur la période 2011-2014, le score aux IADL varie entre 0 et 52 : 800 PatientID, soit 0,1 % des PatientID, ont un score aux IADL qui dépasse la valeur de 4. La distribution du score aux IADL sur l'ensemble des PatientID a un profil en forme de U : près de 30 % des PatientID ont 4 atteintes et autant n'ont aucune atteinte alors que chacune des autres modalités (1 à 3 atteintes) ne dépassent pas 15 %. Cette observation n'est *a priori* pas attendue et pourrait refléter le manque de compréhension des échelles de cotation. De plus, près de 22 % des PatientID ont un score aux IADL en 2010 alors que cette variable a été introduite en 2011.

Pour aller plus loin dans cette analyse, nous avons décliné le score aux IADL selon le diagnostic de démence et l'âge au diagnostic. La distribution en U est observée que ce soit parmi les déments et les non-déments. Les PatientID sont retrouvés davantage au niveau des modalités extrêmes (aucune ou 4 atteintes) à la fois pour les PatientID diagnostiqués déments et non-déments : 28 % des diagnostiqués déments et 24 % des non-déments ont 4 atteintes, 28 % des déments et 41 % des non-déments n'ont aucune atteinte. Avec un focus sur le score aux IADL parmi les non-déments, il est observé que contrairement à ce qui est attendu la proportion de patients avec 4 atteintes aux IADL n'augmente pas avec l'âge alors que la proportion des patients sans incapacité enregistrée sur cette échelle décroît avec l'augmentation de l'âge.

◇ Cette distribution du score aux IADL au-delà des valeurs du score nous suggère que les vérifications sur les modalités autorisées ne sont pas réalisées ; et si elles sont détectées, elles ne sont pas corrigées dans la version finale de la base.

◇ La distribution du score aux IADL en forme de U avec une plus grande part aux modalités extrêmes, nous interroge sur le fait que les équipes des centres se soient approprié correctement la cotation proposée.

➤ Il est recommandé de : 1/ reporter correctement l'historisation des versions du score aux IADL dans le SIC ; 2/ de remettre dans le CIMA l'échelle de score avec une seule cotation et un score après le 4^e item ; 3/ de faire un renvoi vers la page où se situe l'échelle ou de mettre l'échelle en dessous de cette variable dans le CIMA ; et 4/ de supprimer les scores enregistrés avant la date de mise en œuvre de la variable.

➤ Malgré l'intérêt de cette variable pour documenter l'état fonctionnel des individus de la BNA, les incohérences mises en évidence nous conduisent à préconiser de ne pas utiliser cette variable pour le moment.

I TABLEAU 5 I

Répartition des individus pour la variable IADL en fonction de leur statut de démence au 31/12/2015, 2010-2014

Score IADL valeur	Ensemble		Déments		Non-déments	
	n	%	n	%	n	%
manquant	259 077	46,6	113 856	44,9	145 221	48,0
0 : 4 atteintes	79 162	26,7	36 082	28,2	37 877	24,1
1 : 3 atteintes	43 969	14,8	27 155	21,3	16 283	10,3
2 : 2 atteintes	39 886	13,4	21 384	16,7	17 227	10,9
3 : 1 atteinte	41 130	13,8	18 471	14,5	21 212	13,5
4 : 0 atteinte	92 052	31,0	24 599	19,3	64 538	41,0
de 5 à 52	793	0,1	473	0,2	320	0,1

[#] Pour cette analyse, il est fait l'hypothèse que le numéro d'anonymisation PatientID est unique. Or, nous avons mis en évidence que cette hypothèse n'est nécessairement pas vérifiée.

- la mise en ALD : 197 247 (35,5 %) des patients ont une ALD

Il n'est pas précisé pour quelle pathologie (numéro d'ALD) l'ALD est attribuée au patient. La consigne de codage demande de « coter oui quel que soit le type de l'ALD (par ex pour [MAAD] ou pour une autre pathologie) ».

➤ Il est recommandé de faire préciser le numéro d'ALD si cette variable continue à être recueillie.

- la variable « Fin de suivi »

Pour près de 95 % des patients le suivi est en cours lors du dernier acte. Or, il s'avère que cette modalité est la modalité sélectionnée par défaut lors de la saisie. Par ailleurs, nous avons constaté que près d'un tiers des cas décédés ont des visites enregistrées après la date de décès. Parmi l'ensemble des cas décédés, certains sont indiqués décédés à tous les actes, que les actes aient eus lieu avant ou après le décès. Cette modification rétrospective n'est pas cohérente avec la transmission d'une mise à jour de patient.

❖ Pour la variable « Fin de suivi » (**DclFinSuivi**), la modalité de réponse possible « Choix du patient » pour préciser les motifs de fin de suivi n'est pas explicite. S'agit-il du choix du patient à ne plus se faire suivre ou à ce que ses données soient enregistrées dans la base ? De plus, des actes sont enregistrés après que la modalité « choix du patient » a été sélectionnée. L'usage de cette modalité serait peut-être à préciser.

➤ Malgré l'intérêt que pourrait avoir cette variable, il ne nous est pas possible de juger de la pertinence de son utilisation en l'état actuel de la qualité des données.

➤ Plus globalement, il est recommandé de vérifier la cohérence entre les dates, notamment la date de décès et la date de l'acte. Au besoin, des contrôles de cohérences pourraient être réalisés au cours de la saisie des informations.

➤ Si des règles de gestion de la base existent (génération de rapports d'erreur suite aux tests d'intégration comme il est indiqué dans le SIC), il est indispensable de les expliciter. S'il s'agit d'incohérences, il faut les corriger avant l'utilisation de la base.

Les résultats de l'analyse descriptive des 24 variables sont détaillés en annexe (cf. l'Annexe 5. Description des modalités et de leur répartition, 2010-2014).

4.3 Synthèse des observations et propositions d'actions pour l'amélioration de la BNA

4.3.1 Le circuit des données n'en garantit pas l'exhaustivité et la qualité

L'analyse du circuit des données semble montrer des faiblesses tant au niveau de la collecte que de l'intégration des données dans la base. Les faiblesses identifiées remettent en cause l'exhaustivité des cas contenus dans la base et la qualité des données. L'existence de ces faiblesses est confirmée par l'analyse de la qualité des données que nous avons réalisée par la suite.

L'exhaustivité des cas de MAAD ne peut être garanti dans la BNA en raison du caractère passif du recrutement des cas de MAAD d'une part (cf. § 6.2). Seuls les cas de MAAD se rendant dans un CM/CMRR ou cabinet libéraux participant à la BNA peuvent y figurer. De plus, lorsque les données pour un cas sont incomplètes ou erronées, il n'est pas indiqué si ce cas est finalement intégré dans la base. Ainsi, l'exhaustivité des cas suivis dans les centres ne peut pas non-plus être garantie dans la base.

L'utilisation de la BNA repose sur l'usage exclusif d'un numéro d'identification unique par patient, et ce quel que soit le centre. Or, nous avons mis en évidence que cette hypothèse n'était pas vérifiée (cf. § 4.2.2.3.1). Nous avons observé dans la base que le numéro d'anonymisation PatientID pouvait être attribué à plusieurs patients. De plus, l'hypothèse qu'un même patient dispose uniquement d'un seul et unique numéro d'anonymisation PatientID ne peut être confirmé ou infirmé.

➤ Il semble donc y avoir une défaillance dans l'application de l'algorithme d'anonymisation. Il est nécessaire d'en retrouver l'origine afin de comprendre la probabilité élevée de collision observée pour cette base estimé à 10^{-4} alors qu'il est estimé à 10^{-31} . D'autres variables que le sexe ou l'année de naissance pourrait également être analysées pour détecter les doublons de numéro d'identification.

➤ Substituer le département et/ou de la commune de naissance par le département et/ou la commune du centre lorsque celles-ci sont inconnues augmente le risque de ne pas attribuer le même numéro d'anonymisation PatientID au même patient.

➤ Les tests d'intégration réalisés lors de l'import des fichiers en provenance des centres dans la base ne sont pas explicités et ne permettent pas d'identifier ni les lignes comportant des erreurs, ni les variables concernées. Ces lignes d'enregistrement pourraient être

supprimées des analyses en fonction des erreurs identifiées pour les analyses, comme par exemple les doublons.

4.3.2 Une appropriation peu aisée de la documentation et de la base

Plusieurs documents sont disponibles pour nous aider à nous approprier la base. Ces documents comportent des informations importantes et utiles. Cependant, des améliorations sont nécessaires afin d'en faciliter l'appropriation par des nouveaux utilisateurs. À l'usage, cette diversité de documents a rendu difficile la compréhension des données contenues dans la base. Enfin, nous ne disposons pas de la procédure d'intégration ni d'un document explicitant les tests d'intégration.

Tout d'abord, il s'avère qu'aucun document ne fait un lien entre le nom de la variable dans la base (ou le SIC) et son libellé dans le CIMA ou dans le glossaire du CIMA. Une autre difficulté réside dans l'impossibilité de reconstituer correctement l'historique d'une variable ou de ses modalités afin de réaliser les sélections adéquates en fonction de la date de l'acte. Enfin, nous avons trouvé des incohérences entre les documents eux-mêmes ou entre les documents et la base indiquant que ces documents ne sont pas à jour.

➤ Pour la diffusion de la base, il serait nécessaire de décrire tous les fichiers transmis ainsi que les articulations qui existent entre eux avec un/des documents cohérents et mis à jour.

➤ Il est préconisé de limiter le nombre de documents pour garantir la cohérence des informations et d'en assurer plus aisément la mise à jour. Il est notamment indispensable de disposer d'un document mentionnant les variables et leurs caractéristiques ainsi que les libellés en miroir du CIMA.

➤ Il est fondamental également que l'historique précis (avec des dates) soit disponible pour chaque variable afin de documenter l'évolution des variables au cours du temps sur tout l'historique de la base. Ceci est primordial pour pouvoir sélectionner les périodes d'études pertinentes. La seule mention de la version ne permet pas de dater correctement les évolutions successives.

➤ Lors de la création ou de la suppression de variables, il est préférable de ne pas lui attribuer le nom d'une autre variable, déjà utilisé. Ainsi, lors de la fin de l'utilisation d'une variable, il est souhaitable de laisser la variable dans la base et de mentionner sa suppression dans les documents d'information. Si la modalité d'une variable est supprimée la modalité doit être gardée dans les enregistrements antérieurs avec une explication claire dans le document d'information comment recoder cette modalité au besoin pour une analyse historique. Dans tous les cas, il est nécessaire et indispensable d'expliquer les choix faits par l'équipe en terme de recodage ou de gestion de la base afin d'accompagner au mieux les utilisateurs qui pourront recoder (ou supprimer) les variables ou modalités qui ne leur permettent pas de faire les analyses.

4.3.3 Comment améliorer la qualité des données ?

Dans la partie traitant de l'analyse du circuit des données (cf. § 4.2.2), nous nous sommes interrogés sur l'existence de contrôles en cours de saisie pour limiter les valeurs manquantes ou les incohérences. Notre réflexion sur le processus mis en œuvre pour garantir la qualité des données et notre analyse de la qualité des données (valeurs manquantes, incohérences), nous amènent à nous interroger sur les efforts à faire pour garantir la qualité des données collectées.

➤ Malgré les limites à utiliser la base en raison d'un numéro d'anonymisation qui n'est pas unique par patient, les analyses réalisées sur les quelques variables étudiées ont mis en évidence que peu de données étaient manquantes. Néanmoins, quelques variables (**PatientAccompagnant**, **PatientProfession**, **PatientNiveauEtudes**, **DclMesureProtection**), souvent très utiles dans les analyses, ont une proportion élevée de données manquantes. Il pourrait être intéressant de savoir pourquoi elles ne peuvent être obtenues du patient : Comment est formulée la question ? Est-ce une question ouverte ou fermée ? À qui est posée la question ? À quel moment ? Est-ce le moment le plus judicieux pour obtenir un tel renseignement ? Qui d'autre dans l'entourage pourrait y répondre ? Faut-il poser cette question à chaque passage ? Certaines questions sont-elles à poser dès le premier acte ou lors d'un acte ultérieur, à chaque acte ?

➤ Un certain nombre d'incohérences pourraient également être évitées en instaurant des contrôles de cohérences sur les masques de saisies. Certains d'entre eux relèvent de la logique, notamment sur les dates. Une uniformisation des contrôles entre les différents logiciels utilisés par les CMRR sera alors à vérifier. D'autres analyses pourraient être réalisées sur l'ensemble de la base afin de définir des contrôles de cohérences internes pour toutes les variables.

Cela nécessiterait peut-être d'établir une programmation de la transmission et des contrôles, tant internes qu'externes, des données transmises. La périodicité serait à organiser selon les centres en fonction de leur patientèle. La faisabilité d'un contrôle de qualité interne par les centres pourrait être étudiée pour leur permettre de reprendre les données avant envoi. Après transmission de la base, les données pourraient être analysées avec les mêmes contrôles de cohérences internes mais aussi avec des contrôles de cohérences entre les différents enregistrements au niveau central de la BNA. Les incohérences retrouvées seraient retransmises au centre pour confirmation/infirmation et correction le cas échéant. Après validation, la base serait « gelée » pour la période considérée, c'est-à-dire qu'il n'y aurait plus de modifications sur les données antérieures. Un code erreur resterait pour les variables non corrigées ; les utilisateurs pourraient alors choisir de supprimer ou non la ligne si cette erreur perturbe ou non leurs analyses.

L'ensemble de ces éléments de réflexion pourrait être porté par le groupe de travail sur la « qualité des données de la BNA » mis en place dans le cadre de la mesure 16 du Plan maladies neurodégénératives 2014-2019. Des analyses pourraient être entreprises afin de vérifier que les consignes de collectes sont assimilées et appliquées par les équipes sollicitées pour la collecte des informations. De même, ce groupe de travail pourrait se prononcer sur la pertinence de réaliser des travaux sur les qualités des données recueillies (double saisie d'un même dossier, tirage au sort de dossiers pour recodage) ou toutes autres méthodes appropriées. De même, des visites en centre ou encore des « séminaires » pour la qualité des données de la base avec les personnes chargées de la saisie et de la qualité au sein des centres pourraient être organisés.

5 ÉLÉMENTS DE VALIDITÉ EXTERNE DES DONNÉES DE LA BNA

Plusieurs travaux ont permis de décrire la BNA [9, 11] et de fournir des éléments concernant la validité externe de la base de données, c'est-à-dire de reproduire des résultats issus d'autres sources de données ou des évidences scientifiques à haut niveau de preuve.

Un premier travail sur la conformité des prescriptions d'anti-démence aux recommandations de la Haute Autorité de santé émise en 2008 [12], a permis de décrire la proportion de patients atteints de maladie d'Alzheimer traité par l'un ou les deux types d'anti-démence disponibles sur le marché français (inhibiteur d'acétylcholinestérase et bloqueur de récepteurs NMDA) dans la BNA. Dans cette étude un traitement anti-Alzheimer était prescrit chez 76,9 % au cours de l'année 2010. Selon une étude basée sur le panel Thalès¹², 64,5 % des patients diagnostiqués comme atteints de maladie d'Alzheimer étaient traités par des médicaments anti-Alzheimer en 2007. Cette dernière étude n'étant pas spécifique des centres spécialisés comme l'est la BNA on peut penser que l'écart d'environ 10 % de plus dans la BNA s'explique par la spécialisation du réseau BNA.

Plus récemment, les taux de progression des patients diagnostiqués comme MCI (Mild Cognitive Impairment, troubles cognitifs légers) ont été décrits dans la BNA [13]. Cette première étude utilisant des données longitudinales de la BNA a analysé le devenir de 23 676 patients diagnostiqués comme MCI sur une durée médiane de suivi de 406 jours (intervalle 1 – 1 882 jours). Le taux de progression dans la BNA était de 13,7 % personne-années (intervalle de confiance à 95 % : 13,5 %-13,9 %). Ce taux déduit de données de prise en charge en vie réelle était comparable aux taux de progression en milieux spécialisés décrits dans la littérature en conditions expérimentales. Outre le fait de démontrer le caractère opérant en pratique du concept de diagnostic de MCI, cette étude a fourni des éléments de validité externe sur les données de la BNA et le suivi longitudinal dans la BNA. Par ailleurs, cette étude s'est intéressée aux facteurs associés à la progression et a permis de retrouver des facteurs connus comme l'âge, le niveau de MMSE initial ou le sous type MCI confirmant encore la validité de ces données dans ce type de diagnostics.

D'autres travaux sont en cours et s'inscrivent dans le même objectif de fournir des éléments de validité externe à la BNA. Un travail est en cours avec l'unité de pharmaco-épidémiologie du département hospitalo-universitaire de pharmacologie de Bordeaux consistant en une comparaison longitudinale des données de prescription des anti-Alzheimer dans la BNA à celles des patients repérés à partir de l'EGB (échantillon généraliste des bénéficiaires) de l'assurance maladie.

¹² Données sur l'utilisation et la prescription en France des médicaments indiqués dans la maladie d'Alzheimer. Annexe du dossier de réévaluation des médicaments indiqués dans la maladie d'Alzheimer. 2007. www.has-sante.fr.

6 BNA : QUEL POTENTIEL POUR LA SURVEILLANCE ÉPIDÉMIOLOGIQUE DES MAAD ?

Un système de surveillance nécessite une bonne définition de l'événement étudié ainsi que l'exhaustivité ou la représentativité de la population couverte par le système. Dès lors que ces conditions sont remplies et que la qualité des données est assurée, l'estimation d'indicateurs pertinents de surveillance peut être réalisée. Ces indicateurs ont pour objectifs d'estimer la fréquence et la gravité de la maladie étudiée, de suivre les tendances temporelles et spatiales, de décrire les caractéristiques démographiques des malades ainsi que leurs facteurs de risque ou pronostic. Ces indicateurs peuvent également permettre d'évaluer les actions de santé publique menées, voire de déclencher des alertes sanitaires.

6.1 Définition des cas de MAAD dans la BNA

Le diagnostic de MAAD est difficile à poser et son étiologie ne peut être déterminée avec certitude. Il n'existe pas de standardisation du diagnostic pour les MAAD comme cela peut exister pour d'autres maladies (maladies cardiovasculaires, infectieuses, diabète ou cancers) que ce soit au niveau national ou international. Or, l'existence d'une définition claire et partagée est le prérequis à la constitution d'indicateurs de surveillance.

Selon l'OMS, « La démence est un syndrome, généralement chronique ou évolutif, dans lequel on observe une altération de la fonction cognitive (capacité d'effectuer des opérations de pensée), plus importante que celle que l'on pourrait attendre du vieillissement normal. Elle affecte la mémoire, le raisonnement, le comportement et l'aptitude à réaliser les activités quotidiennes. La conscience n'est pas touchée. Une détérioration du contrôle émotionnel, du comportement social ou de la motivation accompagne souvent, et parfois précède, les troubles de la fonction cognitive ». En pratique, les critères couramment utilisés pour établir un diagnostic sont ceux du DSM-IV [14]. Le diagnostic est conduit en évaluant si le patient a développé des déficits cognitifs multiples, à l'origine d'une altération significative du fonctionnement social ou professionnel et significatifs par rapport au niveau de fonctionnement antérieur, non-expliqués par une autre pathologie. Néanmoins, si les critères et les échelles sont bien identifiés, il n'existe pas de seuil précis pour guider le diagnostic. Ainsi, l'appréciation du trouble cognitif et de son retentissement sur les activités de la vie quotidienne dépend de l'appréciation du clinicien et de son expérience. En complément de cet examen clinique, les récentes techniques d'imagerie et de biomarqueurs sont couramment utilisées pour exclure des diagnostics différentiels et objectiver une atteinte hippocampique.

Depuis la création de la BNA, une démarche active pour standardiser l'évaluation des troubles cognitifs en s'appuyant sur des définitions communes telles que celles du DSM-IV (ou DSM-V plus récemment) est mise en œuvre dans l'ensemble des CM/CMRR de France. Ces efforts permettent d'optimiser la qualité et l'homogénéité des cas de MAAD présents dans la BNA. Un avantage de la BNA pour le suivi épidémiologique des MAAD par rapport aux données issues de cohorte réside dans l'utilisation combinée d'un examen clinique, guidée par une méthodologie standardisée, et la réalisation d'examens complémentaires (IRM, prise de sang) pour confirmer le diagnostic et en préciser l'étiologie. En effet, dans les

cohortes, cette deuxième étape n'est généralement pas possible étant donné les coûts liés à ce type d'examen.

Ainsi, le nombre important de cas de MAAD identifiés dans la BNA à l'aide d'une définition de qualité et homogénéisée entre les centres est un atout majeur de la BNA pour envisager un suivi épidémiologique de la MAAD. De plus, la connaissance de l'étiologie précise de chaque cas inclus dans la BNA (maladie d'Alzheimer, démence vasculaire, corps de Lewy...) pourra également permettre le suivi de maladies rares pour lesquelles la mise en place d'études de taille suffisante est très difficile à envisager.

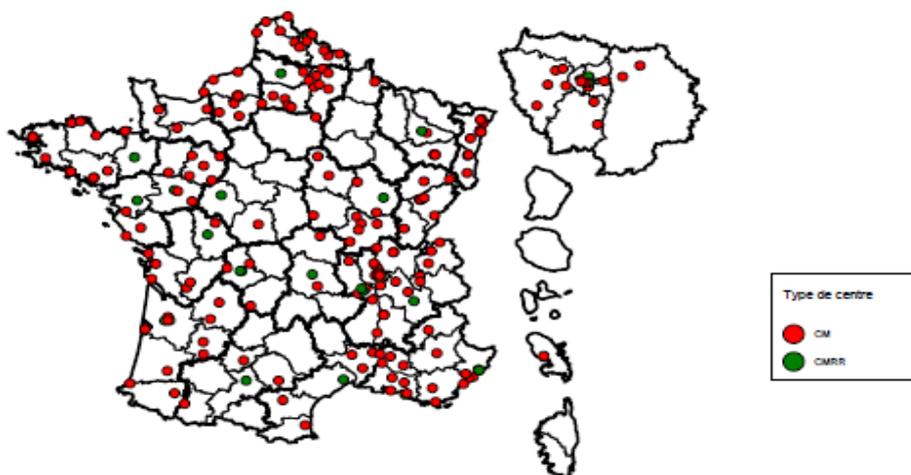
6.2 Représentativité des patients inclus

La BNA est constituée des données de tous les patients vus dans toutes les CM/CMRR labellisées de France. En 2010, 28 CMRR et plus de 350 CM, réparties de manière homogène sur l'ensemble de la France remontent les données de tous leurs patients. Malgré un bon maillage territorial des CM/CMRR en France (cf. figure 3), le recours à ces centres spécialisés n'est pas organisé de manière systématique dans le parcours de soins des personnes âgées. En effet, la consultation dans ces centres intervient en second niveau de soin, elle peut être initiée par un médecin généraliste ou spécialiste mais peut également être motivée par le patient lui-même ou son entourage « inquiété » par la présence de troubles cognitifs. La population incluse dans la BNA est donc issue d'un recrutement non-aléatoire et non-standardisé ne laissant supposer une quelconque représentativité de cette population vis-à-vis de la population française et *a fortiori* des cas de MAAD par rapport à l'ensemble des cas en France. De plus, aucune information concernant la remontée effective des données de tous les patients vus dans tous les centres n'est disponible.

I FIGURE 3 I

Carte présentant la répartition des CM et CMRR sur le territoire à la date du 07 juin 2010

CENTRE BNA 07/06/2010



Même s'il est raisonnable de considérer que la BNA est quasi-exhaustive de l'ensemble des patients ayant recours à une CM/CMRR en France, elle ne l'est pas de l'ensemble des cas de MAAD en France. À ce jour, aucune information n'est disponible concernant la représentativité des cas de la BNA vis-à-vis l'ensemble des cas de MAAD en France. Plus largement, il n'existe pas d'information concernant la représentativité des patients inclus dans la BNA vis-à-vis de la population française.

6.3 Quels indicateurs de surveillance en l'état actuel de la BNA ?

Nous présentons dans cette partie les indicateurs qui pourraient théoriquement être estimés sous la condition que les faiblesses retrouvées précédemment dans la BNA soient corrigées.

Prévalence de la MAAD à la première consultation

À partir des données de la BNA, il est possible d'identifier un nombre de cas prévalents de MAAD à la première consultation. En effet, au cours de la première visite d'un patient, une évaluation complète de la mémoire est réalisée à l'issue de laquelle la présence d'une MAAD est écartée ou suspectée. Dans le second cas, soit le diagnostic peut être directement établi lors de la première visite, soit des examens complémentaires sont nécessaires et le diagnostic sera confirmé lors d'une seconde visite. Techniquement, sous condition que le numéro d'identifiant anonyme soit unique, un nombre de cas de MAAD prévalents « à la première consultation » pourrait donc être calculé. De plus, le taux de prévalence de MAAD à la première consultation pourrait également être estimé, le nombre de cas prévalents étant rapporté au nombre de patients vus pour la première fois en consultation.

Prévalence annuelle de la MAAD

Les données de la BNA devraient permettre de calculer un nombre de cas de MAAD vus en CM/CMRR une année étudiée « N », qu'ils soient vus pour la première fois ou non si le numéro d'identifiant anonyme est unique. Toutefois, le calcul d'un taux de prévalence annuel de la MAAD sur la population totale n'est quant à lui pas envisageable car aucun dénominateur n'est à mettre en regard du nombre de cas annuel de MAAD identifiés. En effet, le nombre de patients vus l'année « N » est constitué des patients vus pour la première fois cette année étudiée et de patients revus issus des années précédentes. De plus, dans la mesure où le suivi dans les CM/CMRR n'est pas obligatoire, tous les anciens patients qu'ils aient été identifiés MAAD ou non, ne sont pas revus. Et pour ces patients non revus, nous ne connaissons pas les raisons de cette absence de suivi, notamment s'ils sont décédés ou non, si le diagnostic a évolué depuis la dernière consultation au sein du centre. De plus, nous pouvons supposer un biais différentiel entre les cas de MAAD et les « non-cas » : la probabilité pour un cas identifié l'année « N-1 » d'être revu l'année suivante est certainement supérieure à celle d'un « non-cas » de l'année « N-1 ».

Prévalence des sous-types de MAAD

Les données de la BNA permettent d'identifier les différents sous-types de MAAD. L'estimation de la fréquence de chaque sous-type présente un intérêt majeur pour l'épidémiologie des démences. Si cet indicateur ne peut pas être utilisé comme un indicateur de suivi de la fréquence des différents sous-types de démence, sa variation pourrait constituer une « alerte ». À ce jour, aucune autre source de données n'est disponible pour étudier ces évolutions et potentiellement détecter un changement dans la prévalence des différents sous-types de démence, en particulier lorsqu'il s'agit d'un sous-type de démence rare.

Incidence de la MAAD à la première consultation dans la BNA

En 2011 ou 2012, une nouvelle variable, intitulée « année de 1^{ère} consultation » (DclAnneePremierDiag) a été introduite dans la BNA dans l'objectif d'identifier les individus venant consulter pour la 1^{re} fois pour leurs troubles cognitifs. La question posée est la suivante : «indiquer l'année de la première consultation pour plainte cognitive et/ou comportementale quel que soit le diagnostic et quel que soit le lieu de la consultation». Cette variable a pour objectif de dater approximativement le début des troubles et de construire un algorithme permettant de classer les nouveaux patients diagnostiqués dément lors de la première consultation, en « cas prévalent » ou « cas incident » à la première consultation dans la BNA. Un exemple d'algorithme pourrait être : si l'année de première consultation est l'année en cours ou l'année précédente, le cas est considéré « incident », sinon il est considéré « prévalent ». Ainsi, le nombre de cas incidents à la première consultation peut être calculé et le taux d'incidence à la première consultation également (nombre de cas incidents divisé par le nombre de patients vus pour la première fois en consultation). Ceci serait possible sous la condition que le numéro d'identifiant anonyme soit unique.

Incidence de la MAAD au cours du suivi

En parallèle des cas incidents à la première consultation, il est possible d'identifier des cas incidents au cours de leur suivi. Il s'agit de patients non diagnostiqué « dément » lors de leur première consultation, mais qui le sont lors d'une consultation ultérieure. Comme c'était le cas pour la prévalence annuelle, le calcul du taux d'incidence au cours du suivi ne peut pas être envisagé étant donné le manque de dénominateur pertinent. Le nombre de cas incidents au cours du suivi devrait être rapporté au nombre de patients revus. Néanmoins, comme nous l'avons vu précédemment, le suivi des patients de la BNA n'étant pas obligatoire, un patient non dément à la première consultation et qui revient consulter aura probablement un risque plus élevé d'être dément que les patients qui ne reviendront pas. Ainsi, le taux d'incidence mesuré à partir de ces données serait fortement surestimé.

Avantage et limites de ces indicateurs

Comme nous venons de le détailler, un centre pourrait mesurer le nombre de cas de MAAD, prévalent ou incident, et la part de ces cas identifiés lors d'une première consultation parmi l'ensemble des patients vus pour la première fois au sein du centre. Le suivi de l'évolution de ces indicateurs au cours du temps serait alors possible. De plus, il serait possible pour chaque CM/CMRR de comparer, toutes choses égales par ailleurs, ces indicateurs et leur évolution par rapport aux données issues des autres CM/CMRR. Mais pour ce faire, il est indispensable de disposer de l'exhaustivité des cas ayant recours à toutes les CM/CMRR et que le numéro d'identifiant anonyme soit unique. Or, nous ne savons pas si l'intégralité des cas sont véritablement dans la BNA et nous avons mis en évidence que la variable **PatientID** ne correspond pas à un numéro d'identifiant anonyme unique.

À l'image de la prévalence de cas de démence vus à la première consultation, la prévalence annuelle des cas de démence dans la BNA pourrait être utile au suivi de l'activité de ces centres mémoire. Néanmoins, le calcul d'un taux de prévalence n'étant pas possible cela limite les comparaisons possibles au cours du temps ou entre les centres.

Si l'intérêt de ces indicateurs pour le suivi et la planification de l'activité des CM/CMRR est indéniable, il présente un intérêt limité pour le suivi épidémiologique des MAAD au niveau national. Il serait erroné de considérer qu'une augmentation (diminution) du nombre ou du taux de prévalence ou d'incidence des cas de MAAD vus en première consultation en CM/CMRR reflète une augmentation (diminution) réelle de la prévalence ou de l'incidence de la MAAD en France. En effet, une telle augmentation (diminution) pourrait être le reflet d'une modification dans les pratiques de recours aux centres (envoi plus important de patients de

la part des collègues praticiens dû à une action de prévention, par exemple). En parallèle, une augmentation (diminution) réelle de la prévalence ou de l'incidence de la MAAD devrait se traduire par une augmentation (diminution) du nombre global de sujets présentant des troubles cognitifs et donc adressés aux CM/CMRR sans pour autant qu'une proportion plus ou moins importante de cas avérés ne soient observés parmi tous ces patients. En l'absence d'uniformité concernant le recours des patients aux CM/CMRR et d'information concernant la représentativité de la population incluse dans la BNA par rapport à la population française, ces indicateurs ne pourront pas être utilisés pour suivre l'évolution du nombre de MAAD en France. Concernant l'évolution du nombre de cas incidents de MAAD au cours du suivi dans la BNA, l'absence d'information concernant la représentativité de ces cas par rapport à la population française, ne nous permet pas d'utiliser cet indicateur pour le suivi du nombre de cas incidents de MAAD en France.

6.4 Autres indicateurs en l'état actuel de la BNA

Suivi annuel des performances cognitives au moment du diagnostic

Un suivi du score au MMSE au moment du diagnostic est un élément intéressant pour apprécier l'évolution de la sévérité des cas de démence nouvellement diagnostiqués dans les CM/CMRR, travail déjà réalisé et publié par l'équipe du CHU de Nice, en collaboration avec l'Inserm [11]. Comme pour les indicateurs de fréquence, ce score renseignerait sur l'activité des CM/CMRR mais, en l'absence d'information concernant la représentativité des cas de la BNA, il ne pourra pas être généralisé à l'ensemble des cas de MAAD en France.

Suivi annuel du déclin fonctionnel au moment du diagnostic

Selon les mêmes modalités que le suivi du MMSE au diagnostic, le suivi annuel des capacités fonctionnelles au moment du diagnostic pourrait être intéressant pour apprécier l'évolution de la sévérité de la maladie au moment du diagnostic. Néanmoins, comme nous l'avons vu précédemment, la variable disponible pour évaluer les capacités fonctionnelles des patients (variable relative aux activités instrumentales de la vie quotidienne, IADL) est de mauvaise qualité et ne peut être utilisée pour le moment.

Suivi du pronostic des patients

Dans le cadre de la surveillance épidémiologique des MAAD, les informations sur le pronostic des malades en termes de morbi-mortalité et sur l'évolution des facteurs de risque de la MAAD sont essentielles. L'absence de suivi actif des patients de la BNA ne permet pas de récupérer des données relatives aux hospitalisations, à l'évolution précise des capacités cognitives ou fonctionnelles, à la survenue d'autres pathologies ou encore au décès. Concernant l'évolution des facteurs de risque, en plus du non suivi obligatoire des patients, peu d'information sur les comorbidités et les comportements de santé des patients inclus dans la BNA est recueillie. L'étude des facteurs de risque de MAAD à partir de ces données n'est donc pas possible en l'état actuel des choses.

6.5 Perspectives d'amélioration

6.5.1 Étudier la représentativité des patients inclus dans la BNA

Mettre en place une stratégie d'analyse de la représentativité des cas de MAAD inclus dans la BNA par rapport à la population française est incontournable pour rendre cette base de données opérationnelle pour le suivi épidémiologique des MAAD.

Une fois la représentativité des cas de MAAD estimée, le nombre de cas prévalent et incident de MAAD diagnostiqués à la première consultation dans la BNA ou le nombre de cas de MAAD diagnostiqué une année pourront être rapportés à la population française de la période considérée pour calculer des taux de prévalence corrigés. Ces indicateurs pourront alors être utilisés pour suivre l'évolution temporelle de la prévalence et de l'incidence en France et pourront être déclinés par âge, sexe et zones géographiques.

6.5.2 Chaînage des données de la BNA avec les bases de données médico-administratives

Le chaînage de la BNA avec les bases médico-administratives (Sniiram et causes médicales de décès avec le CépiDc) offrirait des opportunités d'analyses complémentaires et nécessaires pour la surveillance des MAAD. Les bases de données du Sniiram regroupent les informations relatives aux consommations de soins (médicaments, prestations, type de consultations (généraliste, spécialiste),...), à l'hospitalisation en court séjour (médecine, chirurgie et obstétrique), en service de soins de suite et de réadaptation, en services de psychiatrie et à domicile, et aux affections longue durée (ALD) de tous les affiliés à un régime d'assurance maladie en France, soit environ 98 % de la population française. Dans les bases du CépiDc sont répertoriés annuellement toutes les causes de décès de tous les décès ayant eu lieu sur le territoire français. Le chaînage de la BNA avec ces deux sources de données ouvrirait des perspectives inédites pour le suivi du pronostic de cas de MAAD bien validés ainsi que pour l'étude du parcours et de la consommation de soin des malades.

7 PERTINENCE DE LA BNA EN REGARD DU SYSTÈME DE SURVEILLANCE EXISTANT

7.1 Sources de données de surveillance actuelles

À ce jour, les estimations de fréquence de la MAAD proviennent essentiellement de l'étude de cohorte longitudinale Paquid [2]. Cette étude a débuté en 1988 et a inclus 3 777 personnes âgées de 65 ans et plus, résidant à domicile dans les départements de Gironde ou de Dordogne. Le suivi, encore en cours, de l'ensemble des sujets de la cohorte avec une recherche active des cas de MAAD à des intervalles de temps réguliers (2 à 3 ans) permet de produire des estimations du nombre de déments en France. De plus, cette étude de cohorte, a permis de produire d'importantes données concernant les facteurs de risque de MAAD, l'évolution des performances cognitives et fonctionnelles des sujets et leur pronostic de survie [15,17]. D'autres cohortes françaises participent également à la production de données épidémiologiques sur les MAAD. En particulier, la cohorte des Trois-Cités a inclus près de 10 000 sujets âgés de plus de 65 ans résidant dans les communes de Bordeaux, Dijon, et Montpellier et a pour objectif d'étudier les déterminants vasculaires des démences. Les résultats obtenus à partir de cette cohorte contribuent largement à l'amélioration des connaissances concernant le suivi des patients [18], les facteurs de risque vasculaires des MAAD [19,21], mais également d'autres facteurs de risque tels que la nutrition, la dépression, la fragilité ou la consommation de médicaments [22,24]. Cependant, la faible fréquence de la MAAD dans cette cohorte et la non représentativité de la cohorte liées au recrutement de volontaires, ne permettent pas de suivre l'évolution de la fréquence de la MAAD en France. De plus, le faible nombre d'individus qui y est inclus ne permet pas d'identifier suffisamment de sous-types de MAAD pour pouvoir en décrire la prévalence ou l'incidence, ainsi que pour étudier leurs facteurs de risque ou leurs pronostics. Enfin, ces cohortes vieillissent et ne pourront bientôt plus faire l'objet de suivi « pertinent » pour la production d'indicateurs de fréquence.

Récemment, la cohorte Constances a été mise en place et représente une opportunité pour la production de données de surveillance dans le champ du vieillissement et des démences en particulier. Cette étude prévoit d'inclure environ 200 000 individus de 18 à 69 ans suivis par questionnaires tous les ans, avec un examen de santé tous les 5 ans [25], et les outils nécessaires à l'analyse de l'évolution des performances cognitives de tous les sujets âgés de plus de 45 ans. Un atout majeur de cette nouvelle étude réside dans son chaînage avec les données du Sniiram. En effet, cette étude disposera d'un large échantillon d'individus pour lesquels, non seulement l'ensemble des informations liées à leur prise en charge médicale seront disponibles grâce au Sniiram, mais également une large série de données socio-environnementales et cliniques recueillies à des intervalles de temps réguliers. Concernant l'identification des cas de démences dans cette cohorte, il n'est pas prévu la mise en place d'un diagnostic pour tous les sujets de la cohorte. Néanmoins, une validation des cas potentiels de MAAD est prévue. Ces cas potentiels seront repérés par des informations en provenance de leurs médecins généraliste ou spécialiste ainsi que par des informations contenues dans le Sniiram (consommation de médicaments antidémenciels, présence d'une ALD pour démence ou diagnostic d'hospitalisation pour démence).

En parallèle des cohortes, l'utilisation des bases de données du Sniiram pour la surveillance épidémiologique de certaines maladies chroniques est en expansion, compte-tenu du caractère représentatif de la population concernée (>98 % de la population française), de la remontée systématique et dynamique de ces données et de son faible coût. Concernant

l'utilisation du Sniiram pour la surveillance des démences [26], on se heurte néanmoins au problème de la non-spécificité des critères utilisés pour repérer les cas de démences. Les algorithmes développés se basent sur la consommation de médicaments antidémence, la présence d'une ALD pour démence et l'hospitalisation avec un diagnostic de démence. Or, les médicaments antidémence ne sont pas prescrits systématiquement aux malades, l'ALD pour démence, en l'absence de la prise de médicaments spécifiques, n'est pas toujours attribuée et un malade n'est pas hospitalisé de façon systématique (pour effectuer un bilan, par exemple). De plus, il n'est pas exclu dans ces algorithmes d'inclure à tort de faux déments (ex : un code diagnostic d'hospitalisation erroné ou la prise d'un médicament à un stade prédémence). Ainsi, les cas de MAAD identifiés dans le Sniiram représentent seulement une partie des malades diagnostiqués et pris en charge pour leur maladie. Une récente étude [27], montre que, chez les individus de plus de 70 ans, les estimations de prévalence obtenues à partir de ces données sont inférieures d'environ 60% aux prévalences attendues à partir de la cohorte Paquid. En revanche, chez les moins de 70 ans, les estimations sont en adéquation avec ce qui était attendu et semble donc de bonne qualité. Une étude de validation directe de la qualité des algorithmes utilisés dans le Sniiram reste cependant incontournable avant de pouvoir les utiliser et les interpréter raisonnablement. Un intérêt du Sniiram pour la surveillance des MAAD concerne la possibilité de mettre en place des études de cas incidents dont le suivi au cours du temps permettra d'étudier la survenue d'événements indésirables, en particulier rares, tels que les hospitalisations, la survenue d'autres pathologies et le décès. Un écueil commun avec les données de cohortes concerne l'impossibilité d'étudier des sous-types de démences rares. En effet, même si des codes CIM-10 sont disponibles pour différencier les sous-types de MAAD identifiés via leurs ALD ou les hospitalisations, la fiabilité du codage reste à démontrer. De plus, plus d'un tiers des malades sont identifiés via leur consommation de médicaments qui ne permet pas de distinguer les sous-types de MAAD.

7.2 Apport de la BNA

Le principal atout de la BNA par rapport aux autres bases de données actuellement disponibles concerne la qualité et la précision des diagnostics et sous diagnostics de MAAD effectués. Dès aujourd'hui, la mise en place d'études spécifiques à partir des cas de MAAD identifiés dans les CM/CMRR est d'un grand intérêt pour la recherche. En ce qui concerne l'épidémiologie descriptive et la surveillance de la démence en France, l'intérêt de la BNA pourrait largement s'accroître si elle était chaînée aux bases de données médico-administratives (Sniiram, CépiDc...). En effet, cela permettrait d'envisager la mise en place d'un certain nombre de projets tels que :

- étudier le pronostic (hospitalisation, pathologies, mortalité) de cas bien validés de MAAD et en particulier de sous-types de MAAD rares ;
- compléter l'algorithme de définition des cas de MAAD utilisé dans le Sniiram pour améliorer le repérage des cas dans le Sniiram et à terme produire des données de surveillance (prévalence, incidence, pronostic, survie) ;
- caractériser les malades qui se rendent en CM/CMRR en vue d'améliorer la prise en charge des malades en population générale.

Une réflexion sur le chaînage de la BNA avec le Sniiram est actuellement menée dans le cadre du plan maladies neurodégénératives 2014-2019. Des premiers éléments de

réflexions sont en faveur de sa faisabilité. En effet, la mise en place du nouveau système national des données de santé permettra de disposer plus facilement des données chaînées du Sniiram et des causes médicales de décès, entre autre. De plus, la mise en application de la nouvelle loi de santé autorisant l'utilisation du NIR des individus avec plus de souplesse dans les études épidémiologiques est également un élément majeur d'avancée pour ce type de projet.

8 CONCLUSION DU RAPPORT

À partir de la base de données fournies par le CHU de Nice regroupant un corpus d'information minimal Alzheimer (CIMA) de près de 560 000 patients ayant consulté un ou plusieurs un CM ou CMRR entre le 1^{er} janvier 2010 et le 31 décembre 2014, les principales variables de la BNA ont été analysées afin dans un premier temps de s'assurer de la qualité de la base de données et de sa capacité à identifier tout patient atteint d'une MAAD ainsi que de décrire leur capacité à contribuer à la surveillance épidémiologique des démences en France. Au total, 1 511 664 lignes d'observations de 33 variables (enregistrées sous la forme de 118 modalités de variables soit 178 376 352 données) pour près de 560 000 patients identifiés différents répartis sur l'ensemble du territoire français. Ces données sont collectées de manière systématique dans tous les CM et CMRR de France ainsi que par certains neurologues libéraux. Les diagnostics de démence sont posés selon une démarche qui tend à être homogène et standardisée en reprenant les nomenclatures internationales existantes pour le codage des MAAD. Cette base de données contient également des données d'activité des centres et des informations sur liées à la MAAD, tels que certains facteurs de risque des MAAD comme le niveau d'études, la profession exercée, ainsi que des informations socio-économiques (APA, mise en tutelle ou curatelle, domicile).

L'appropriation de la base à des fins d'analyse n'a pas été facile, du fait d'un manque de cohérence entre les documents relatifs à la BNA eux-mêmes et entre les documents et la base elle-même. Par ailleurs, il a été observé des doublons provenant soit d'enregistrement multiple d'actes d'un même centre, soit d'un risque de collision (différents identifiants pour un même patient) lors de l'anonymisation du fichier ou encore des incohérences qui pourraient être éliminées par la mise en place de procédure de validation aux différentes étapes de constitution de la base.

La BNA est constituée de l'enregistrement du CIMA de tous les patients consultant un CM ou CMRR sur l'ensemble du territoire métropolitain. Elle pourrait donc être représentative de l'ensemble des patients consultants. Néanmoins, il n'a pas été possible de vérifier l'exhaustivité des cas enregistrés dans la BNA du fait de l'absence aujourd'hui d'une liste complète des CM et CMRR et de l'impossibilité de vérifier que tous les patients consultants ont été enregistrés. Pour améliorer la qualité de cette base, il est donc nécessaire dans un premier temps de s'assurer que la BNA remonte bien les données CIMA de l'ensemble des CM et CMRR et de tous les patients reçus en consultation, avec des informations sur les patients ayant refusé de participer (au moins leur nombre).

Même si la BNA est quasi-exhaustive de l'ensemble des patients ayant recours à un CM ou CMRR en France, elle reste non représentative des MAAD en population générale et ce biais n'est aujourd'hui pas caractérisé. Les caractéristiques des patients qui consultent en CM/CMRR devraient pouvoir être comparés avec les caractéristiques des MAAD en population générale, afin d'estimer les biais potentiels de ces données. Cette comparaison est néanmoins complexe pour plusieurs raisons. Il n'existe pas aujourd'hui de sources décrivant les principales caractéristiques des MAAD sur toutes les tranches d'âge: les cohortes (Paquid, étude des 3C) sont constituées de personnes âgées de plus de 65 ans et reposent sur des diagnostics de démence non homogène ; les bases de données médico-administratives, comme le Sniiram, représentent seulement une partie des malades, ceux diagnostiqués et pris en charge pour leur maladie (repérés dans ces bases par au moins un remboursement thérapeutique) et peu de données sont disponibles pour caractériser les malades dans ces bases. D'autre part, l'absence de suivi longitudinal pour tous les patients inclus dans la BNA (a minima le décès) empêche le calcul de taux de prévalence ou d'incidence de la MAAD. En effet, le dénominateur pour le calcul de ces taux est incertain :

on peut faire l'hypothèse que la probabilité pour qu'un patient diagnostiqué MAAD à la première consultation revienne à une nouvelle consultation est plus élevée que pour un patient diagnostiqué non-MAAD. Ces défauts de caractérisation des MAAD inclus et suivis dans la BNA par rapport à l'ensemble des MAAD sont des éléments majeurs empêchant l'estimation d'indicateurs de surveillance épidémiologique des démences en France.

Et c'est fort dommage car la BNA est la seule source qui propose un diagnostic homogène de la démence à partir d'une procédure d'identification simple, rapide et standardisée et qui tient compte de l'évolution des processus de diagnostic et de prise en charge. L'utilisation combinée d'un examen clinique guidée par une méthodologie standardisée et la réalisation d'examen cliniques complémentaires permet ainsi de confirmer les diagnostics et de préciser l'étiologie de la démence. Le partage, par tous professionnels contribuant à la BNA, d'une définition claire et partagée permet le suivi épidémiologique des démences enregistrés dans la BNA et notamment, du fait du nombre important de patients enregistrés, le suivi des cas de maladies rares.

Un centre peut donc mesurer et suivre au cours du temps le nombre de cas de MAAD, prévalent ou incident, et la part de ces cas identifiés lors d'une première consultation parmi l'ensemble des patients vus pour la première fois au sein du centre. Et du fait d'une définition diagnostic homogène, chaque CM/CMRR peut comparer, toutes choses égales par ailleurs, ces indicateurs et leur évolution par rapport aux données issues des autres CM/CMRR. Si l'intérêt de ces indicateurs est indéniable pour le suivi et la planification de l'activité des CM/CMRR, ils ne peuvent être utilisés pour le suivi épidémiologique des MAAD au niveau national, en l'absence d'information sur la représentativité de la population incluse dans la BNA par rapport à la population française.

Dans sa dimension actuelle, la BNA a un intérêt au niveau de la recherche pour comprendre ou avoir une analyse macroscopique du fonctionnement du système de soin national. Mais pour pouvoir intégrer la BNA au système de surveillance, différentes actions doivent donc être menées, que ce soit la mise en place de procédures de vérification au moment de la constitution de la base, le chaînage avec les bases de données médico-administratives administratives pour permettre le suivi longitudinal des patients, l'amélioration de l'enregistrement des nouveaux cas de démence ou encore la caractérisation des biais de sélection de cette population. En plus de ces améliorations nécessaires pour assurer une meilleure qualité des données de la BNA, la pérennisation du dispositif devra se faire en conformité avec la loi informatique et libertés. Ces points fondamentaux à l'utilisation de la BNA, notamment comme outil pour la surveillance épidémiologique des démences en France, sont actuellement discutés au sein de groupes de travail animés par l'équipe projet du plan dans le cadre du suivi de la mesure 16 « Pérenniser et améliorer la banque nationale de données Alzheimer et maladie apparentée » du plan des maladies neurodégénératives 2014-2019, groupes auxquels participe Santé publique France.

Références bibliographiques

- [1] Langmuir AD. *The surveillance of communicable diseases of national importance*. N Engl J Med. 1963, 288 : 182-192.
- [2] J.-F. Dartigues, M. Gagnon, P. Michel, *et al.* *Le programme de recherche Paquid sur l'épidémiologie de la démence. Méthodes et résultats initiaux*. Rev Neurol 1991 ; 147 : 225-230.
- [3] 3C study Group. *Vascular factors and risk of dementia: Design of the Three-City Study and baseline characteristics of the study population*. Neuroepidemiology 2003 ; 22 : 316-325.
- [4] Fondation Plan Alzheimer. *Cohorte nationale de patients : MEMENTO* [Internet] www.fondation-alzheimer.org (consulté le 20/02/2017).
- [5] Pérès K, Matharan F, Allard M *et al.* *Health and aging in elderly farmers : the AMI cohort*. BMC Public Health. 2012 ; 12 : 558.
- [6] Jacqmin-Gadda H, Alperovitch A, Montlahuc C, Commenges D, Leffondre K, Dufouil C, *et al.* *20-Year prevalence projections for dementia and impact of preventive policy about risk factors*. Eur J Epidemiol. 2013 ; 28 (6) : 493-502.
- [7] Ankri J *et* Van Broeckhoven C. *Rapport d'évaluation du plan Alzheimer 2008-2012*. Ministère des Affaires sociales, de la Santé et des Droits des femmes, 2013. 101 pages.
- [8] Ministère des Affaires sociales, de la Santé et des Droits des femmes. *Plan maladies neuro-dégénératives 2014-2019*. 2014, 124 pages. Disponible sur : http://social-sante.gouv.fr/IMG/pdf/Plan_maladies_neuro_degeneratives_def.pdf (consulté le 20/02/2017)
- [9] Le Duff F, Develay AE, Quétel J, *et al.* *The 2008-2012 French Alzheimer plan: description of the national Alzheimer information system*. Journal of Alzheimer's disease JAD. 2012; 29 (4) : 891-902.
- [10] Anthony S, Pradier C, Chevrier R, Festrats J, Tifratene K, Robert P. *The French National Alzheimer Database: A Fast Growing Database for Researchers and Clinicians*. Dementia and geriatric cognitive disorders. 2014 ; 38 (5-6) : 271-280.
- [11] Pradier C, Sakarovitch C, Le Duff F, *et al.* *The Mini Mental State Examination at the Time of Alzheimer's Disease and Related Disorders Diagnosis, According to Age, Education, Gender and Place of Residence: A Cross-Sectional Study among the French National Alzheimer Database*. PloS one. 2014 ; 9 (8) : e103630.
- [12] Tifratene K, Duff FL, Pradier C, Quétel J, Lafay P, Schück S, Benzenine E, Quantin C, Robert P. *Use of drug treatments for Alzheimer's disease in France: a study on a*

national level based on the National Alzheimer's Data Bank (Banque nationale Alzheimer). *Pharmacoepidemiol Drug Saf.* 2012 Sep ; 21 (9) : 1005-12.

- [13] Tifratene K, Robert P, Metelkina A, Pradier C, Dartigues JF. *Progression of mild cognitive impairment to dementia due to AD in clinical settings*. *Neurology*. 2015 Jul 28 ; 85 (4) : 331-8.
- [14] *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM IV)*. Washington: American Psychiatric Association, 1994.
- [15] Carcaillon L, Pérès K, Péré JJ, Helmer C, Orgogozo JM, Dartigues JF. *Fast cognitive decline at the time of dementia diagnosis: a major prognostic factor for survival in the community*. *Dement Geriatr Cogn Disord*. 2007 ; 23 (6) : 439-45. Epub 2007 Apr 26.
- [16] Edjolo A, Proust-Lima C, Delva F, Dartigues JF, Pérès K. *Natural History of Dependency in the Elderly: A 24-Year Population-Based Study Using a Longitudinal Item Response Theory Model*. *Am J Epidemiol*. 2016 Feb 15;183(4):277-85. doi: 10.1093/aje/kwv223. Epub 2016 Jan 28.
- [17] Amieva H, Mokri H, Le Goff M, Jacqmin-Gadda H, Foubert-Samier A, Orgogozo JM, Stern Y, Dartigues JF. *Compensatory mechanisms in higher-educated subjects with Alzheimer's disease: a study of 20 years of cognitive decline*. *Brain* 2014 ; 137 : 1167-75.
- [18] Helmer C, Peres K, Pariente A, Pasquier F, Auriacombe S, Poncet M, et al. *Primary and secondary care consultations in ederly demented individuals in France. Results from de Three-City Study*. *Demt Geriatr Cogn Disor*. 2008 ; 26 (5) : 407-15.
- [19] Carcaillon L, Plichart M, Zureik M, Rouaud O, Majed B, Ritchie K, Tzourio C, Dartigues JF, Empana JP. *Carotid plaque as a predictor of dementia in older adults: the Three-City Study*. *Alzheimers Dement*. 2015 Mar ; 11 (3) : 239-48. doi: 10.1016/j.jalz.2014.07.160. Epub 2014 Dec 12.
- [20] Alépérovitch A, Blachier M, Soumaré A, Ritchie K, Dartigues JF, Richard-Harston S, Tzourio C. *Blood pressure variability and risk of dementia in an elderly cohort, the Three-City Study*. *Alzheimers Dement*. 2014 Oct ; 10 (5 Suppl) : S330-7. Doi : 10.1016/j.jalz.2013.05.1777. Epub 2013 Aug 15.
- [21] Raffaitin C, Gin H, Empana JP, Helmer C, Berr C, Tzourio C, Portet F, Dartigues JF, Alépérovitch A, Barberger-Gateau P. *Metabolic syndrome and risk for incident Alzheimer's disease or vascular dementia: the Three-CityStudy*. *Diabetes Care*. 2009 Jan ; 32 (1) : 169-74. doi : 10.2337/dc08-0272. Epub 2008 Oct 22.
- [22] Féart C, Samieri C, Rondeau V, Amieva H, Portet F, Dartigues JF, Scarmeas N, Barberger-Gateau P. *Adherence to a Mediterranean diet, cognitive decline, and risk of dementia*. *JAMA*. 2009 Aug 12 ; 302 (6) : 638-48. doi: 10.1001/jama.2009.1146. Erratum in : *JAMA*. 2009 Dec 9 ;302 (22) : 2436.

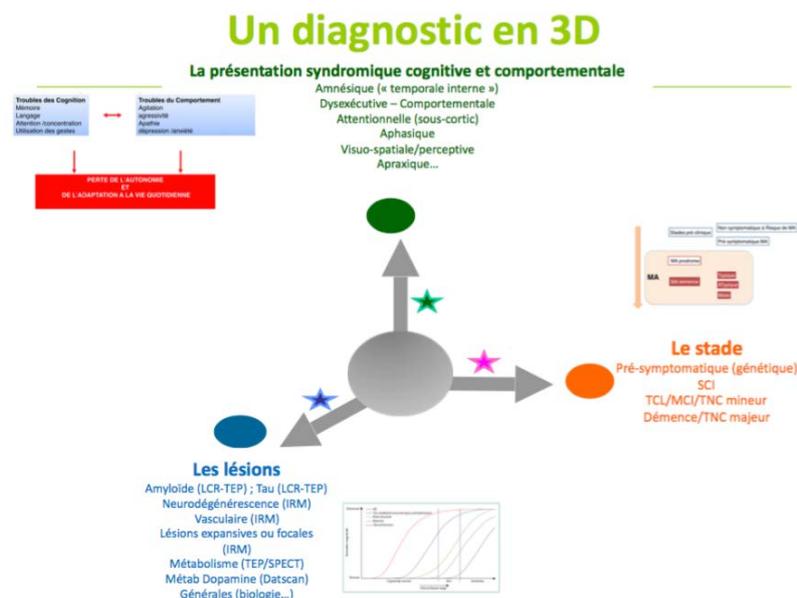
- [23] Godin O, Dufouil C, Ritchie K, Dartigues JF, Tzourio C, Pérès K, Artero S, Alperovitch A. *Depressive symptoms, major depressive episode and cognition in the elderly: the three-city study*. Neuroepidemiology. 2007 ; 28 (2) : 101-8. Epub 2007 Apr 4.
- [24] Mura T, Proust-Lima C, Akbaraly T, Amieva H, Tzourio C, Chevassus H, Picot MC, Jacquemin-Gadda H, Berr C. *Chronic use of benzodiazepines and latent cognitive decline in the elderly: results from the Three-city study*. Eur Neuropsychopharmacol. 2013 Mar;23(3):212-23. doi: 10.1016/j.euroneuro.2012.05.004. Epub 2012 Jun 15.
- [25] Zins M, Goldberg M, Carton M, Guéguen A, Henny J, Le Got S, *et al.*; équipe Constances et Centres d'exams de santé de la Sécurité sociale. *La cohorte Constances : une infrastructure pour la recherche et la santé publique*. Bull Epidemiol Hebd. 2016 ; (35-36) : 612-6. http://invs.santepubliquefrance.fr/beh/2016/35-36/2016_35-36_1.html
- [26] Tuppin P, Cuerq A, Weill A, Ricordeau P, Allemand H. *Maladies d'Alzheimer et autres démences : identification, prise en charge et consommation de neuroleptiques chez les bénéficiaires du régime général (2007-2009)*. Rev Neurol. 2012 ; 168 (2) : 152-60.
- [27] Carcaillon-Bentata L, Quintin C, Moutengou E, Boussac-Zarebska M, Moisan F, Ha C, *et al.* *Peut-on estimer la prévalence de la maladie d'Alzheimer et autres démences à partir des bases de données médico-administratives ? Comparaison aux données de cohortes populationnelles*. Bull Epidemiol Hebd. 2016;(28-29): 459-67. Disponible sur : invs.santepubliquefrance.fr/beh/2016/28-29/2016_28-29_1.html (consulté le 20/02/2017).

Annexes

Annexe 1. Diagnostic d'une MAAD dans la BNA

I FIGURE 4 I

Diagnostic en 3 dimensions d'une MAAD selon la révision 5 de la BNA qui propose 3 niveaux de diagnostic



1- Dimension 1 : Stade

Une seule réponse est autorisée :

- Absence de trouble mental (DSM V)
- Plainte cognitive isolée
- Trouble cognitif mineur (DSM V)
- Trouble cognitif majeur (DSM V)
- Non applicable

2- Dimension 2 : Diagnostic syndromique

Une seule réponse est autorisée

- 2.1 Présentation clinique à dominante cognitive :
 - Présentation à dominante amnésique
 - Présentation à dominante non amnésique
 - Troubles du langage
 - Trouble visuel
 - Syndrome dysexécutif cognitif
 - Apraxie progressive des membres
 - Syndrome cortico-basal
 - Autre

- Présentation cognitive « multidomaines »
- 2.2 Présentation clinique à dominante comportementale ou psychologique :
 - Syndrome dysexécutif comportemental
 - Troubles émotionnels ou affectifs
 - Troubles productifs (délire, hallucinations, agitation etc...)
 - Autre
- 2.3 Présentation clinique sans dominante cliniquement significative (association de déficits cognitifs et de troubles comportementaux et/ou psychologiques)
- 2.4 Non applicable

3 Dimension 3 : Diagnostic étiologique

Nombre maximal de 3 choix

- Maladie d'Alzheimer
 - o Cliniquement probable (sans biomarqueurs)
 - o Probable (avec biomarqueurs*) ¹³
 - o Certaine (génétique)
- Maladie d'Alzheimer dans le cadre d'un syndrome de Down
- Maladie à corps de Lewy
- Maladie de Parkinson
- Dégénérescence lobaire fronto-temporale
- Paralyse supranucléaire progressive
- Dégénérescence cortico-basale
- Maladie de Huntington
- Atrophie multisystématisée
- SLA et autres maladies du MN
- Sclérose en Plaques
- Troubles cognitifs vasculaires
- Angiopathie amyloïde
- Pathologies vasculaires non liées à athérome (CADASIL, vascularites, mitochondriopathies...)
- Séquelles d'accident vasculaire cérébral
- Maladie à prion
- Hydrocéphalie chronique
- Tumeur intracrânienne
- Épilepsie
- Encéphalopathie chronique post-traumatique
- Séquelles encéphaliques de traumatisme crânien
- Commotion cérébrale
- Maladie de Wilson
- Encéphalopathie d'origine métabolique (dont hypothyroïdie, troubles du métabolisme calcique...)
- Encéphalopathie d'origine infectieuse (dont VIH, neurosyph.)
- Encéphalopathie éthylique
- Encéphalopathie anoxique
- Encéphalopathie carencielle non éthylique
- Encéphalopathie toxique (dont iatrogénie)
- Encéphalopathie auto-immunes (dont E. limbiques, E. de Hashimoto, E. dans le cadre de MAI....)
- Syndrome des apnées du sommeil

¹³ Dans ce cas renseigner obligatoirement la rubrique examens complémentaires (Exam) dans le Corpus d'Information minimum.

- Autre trouble d'origine organique
- Troubles dépressifs (DSM 5)
- Trouble anxieux (DSM 5)
- Troubles psychotiques (DSM 5)
- Trouble spécifique des apprentissages (DSM 5)
- Trouble du spectre autistique (DSM5)
- Trouble du déficit de l'attention (DSM5)
- Autre trouble neurodéveloppemental (DSM5)
- Troubles obsessionnels compulsifs et connexes (DSM5)
- Troubles du spectre psychotique (DSM 5)
- Autre maladie mentale
- DIAGNOSTIC EN ATTENTE
- Non applicable

Annexe 2. Variables dans les bases de données transmises

Base principale

I TABLEAU 6 I

Liste des variables (par ordre d'ordonnement) et paramètres (type de variable et longueur du champ) dans la base transmise

N°	Nom de la Variable	Type de variable	Long.	N°	Nom de la Variable	Type de variable	Long.
1	ID	Num.	8	40	CIActiviteReseau	Texte	1
2	Application	Texte	10	41	CICollabCmCmrr	Texte	1
3	CentreAdCp	Texte	7	42	CICollabNeuroPsy	Texte	1
4	CentreAdRue1	Texte	21	43	CINbFormAn	Texte	1
5	CentreAdRue2	Texte	1	44	CIPratiqueETP	Texte	1
6	CentreAdVille	Texte	9	45	CITypeSpecialiste	Texte	1
7	CentreDepartement	Texte	4	46	CmrrActiviteAnimation	Texte	1
8	CentreHdj	Num.	8	47	CmrrActiviteEthique	Texte	1
9	CentreInclusion	Num.	8	48	CmrrActiviteFormation	Texte	1
10	CentreLabel	Num.	8	49	CmrrActiviteRecherche	Texte	1
11	CentreLabelAnnee	Texte	1	50	CmrrCmActiviteSoutien	Num.	8
12	CentreMail	Texte	24	51	CmrrCmActiviteStimul	Num.	8
13	CentreMultisite	Num.	8	52	CmrrFormationsUniv	Texte	1
14	CentreNom	Texte	50	53	CmrrPublications	Texte	1
15	CentreNomReseau	Texte	40	54	DateFinPeriod	Date MMDDYY10	8
16	CentreNum	Texte	14			.	
17	CentreNumAdeli	Texte	1	55	DclALD	Num.	8
18	CentreNumFiness	Texte	11	56	DclAPA	Num.	8
19	CentreNumRmes	Texte	1	57	DclAnneePremierDiag	Num.	8
20	CentreNumRpps	Texte	1	58	DclCmrrRecours	Texte	1
21	CentreOuvConsultation	Num.	8	59	DclCmrrRecoursDiagDif	Texte	1
22	CentrePersoARC	Texte	1				
23	CentrePersoASocial	Texte	1	60	DclCmrrRecoursPTech	Texte	1
24	CentrePersoAutre	Texte	1	61	DclCmrrRecoursSpExpert	Texte	1
25	CentrePersoDescriptionAutre	Texte	1				
26	CentrePersoGeriatre	Num.	8	62	DclCmrrRecoursSuivi	Texte	1
27	CentrePersoIDE	Texte	1	63	DclDateActe	Date MMDDYY10	8
28	CentrePersoKine	Texte	1			.	
29	CentrePersoNeurologue	Texte	1	64	DclDateDeces	Date. MMDDYY10	8
30	CentrePersoOrthophoniste	Num.	8			.	
31	CentrePersoPsychiatre	Texte	1	65	DclDateEntreeInstitution	Texte	1
32	CentrePersoPsychologue	Num.	8				
33	CentrePersoSecetaire	Num.	8	66	DclDateFinSuivi	Date. MMDDYY10	8
34	CentrePersoTEC	Texte	1			.	
35	CentreRegion	Num.	8	67	DclDiagnostic	Num.	8
36	CentreReseau	Num.	8	68	DclEffetIndesirable	Num.	8
37	CentreResponsable	Texte	16	69	DclEtatDemence	Num.	8
38	CentreType	Num.	8	70	DclFinSuivi	Num.	8
39	CIActiviteConsultMemoire	Texte	1	71	DclIADLConsequence	Texte	1
				72	DclIADLDetection	Num.	8

N°	Nom de la Variable	Type de variable	Long.
73	DclIPSRAccueilJour	Num.	8
74	DclIPSRaucun	Num.	8
75	DclIPSRAutre	Num.	8
76	DclIPSRclic	Num.	8
77	DclIPSRergotherapeute	Num.	8
78	DclIPSRgroupe	Num.	8
79	DclIPSRkinesitherapeute	Num.	8
80	DclIPSRmaia	Num.	8
81	DclIPSRorthophoniste	Num.	8
82	DclIPSRpsychologue	Num.	8
83	DclIPSRrepts	Num.	8
84	DclIPSRsiad	Num.	8
85	DclIMMSE	Num.	8
86	DclIMesureProtection	Num.	8
87	DclIModeDeVie	Num.	8
88	DclIPCAucun	Num.	8
89	DclIPCBioMarqueurs	Num.	8
90	DclIPCIRM	Num.	8
91	DclIPCET	Num.	8
92	DclIPCSPECT	Num.	8
93	DclIPScanner	Num.	8
94	DclIPR	Num.	8
95	DclIPRType	Texte	1
96	DclIPRVisite	Texte	1
97	DclIPAntagonisteNMD	Num.	8

N°	Nom de la Variable	Type de variable	Long.
	A		
98	DclTPAntidepressseurs	Num.	8
99	DclTPAnxiolitiques	Num.	8
100	DclTPAucun	Num.	8
101	DclTPAucunDansListe	Num.	8
102	DclTPAutres	Num.	8
103	DclTPHypnotiques	Num.	8
104	DclTPInhibAcetyl	Num.	8
105	DclTPNeuroleptiques	Num.	8
106	DclTPNootropes	Num.	8
107	DclTypeActe	Num.	8
108	NumFic	Texte	8
109	PatientAccompagnant	Num.	8
110	PatientAnneeNaissance	Num.	8
111	PatientEnvoyePar	Num.	8
112	PatientId	Texte	66
113	PatientNiveauEtudes	Num.	8
114	PatientProfession	Num.	8
115	PatientSexe	Texte	3
116	PatientSituationGeo	Num.	8
117	RetourMail	Texte	21
118	TsCreation	Date. DATETIME. ANYDTDTM 40.	8

Description des 30 fichiers accompagnants la base principale

Table PBNAACCOMPAGNANTS, fichier accompagnant la base principale

I TABLEAU 7 I

Structure de la table PBNAACCOMPAGNANTS

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 8 I

Modalités de la table PBNAACCOMPAGNANTS

ID	Nom	NomCourt	NumOrdre
1	Conjoint	CONJOINT	10
2	Enfant	ENFANT	20
3	Bel enfant	BELENFANT	30
4	Petit enfant	PTENFANT	40
5	Neveu ou nièce	NEVEU	50
6	Frère ou sœur	FRERE	60
7	Curateur	CURATEUR	70
8	Aucun	AUCUN	80
9	Autres	AUTRES	90
10	Sans Informations	SI	100

Table PBNAACTSCMCL, fichier accompagnant la base principale

I TABLEAU 9 I

Structure de la table PBNAACTSCMCL

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 10 I

Modalités de la table PBNAACTSCMCL

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20

Table PBNAACTSRESCL, fichier accompagnant la base principale

I TABLEAU 11 I

Structure de la table PBNAACTSRESCL

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 12 I

Modalités de la table PBNAACTSRESCL

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20

Table PBNAALDS, fichier accompagnant la base principale

I TABLEAU 13 I

Structure de la table PBNAALDS

Nom	Type	Longueur	Type	Libellé
ID	Numérique	8		ID
Nom	Texte	50		Nom
NomCourt	Texte	50		NomCourt
NumOrdre	Numérique	8		NumOrdre

I TABLEAU 14 I

Modalités de la table PBNAALDS

ID	Nom	NomCourt	NumOrdre
1	Non	N	10

2	Oui	O	20
3	Sans Information	SI	30

Table PBNAAPAS, fichier accompagnant la base principale

I TABLEAU 15 I

Structure de la table PBNAAPAS

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 16 I

Modalités de la table PBNAAPAS

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20
3	Sans Information	SI	30

Table PBNACOLLSCMCL, fichier accompagnant la base principale

I TABLEAU 17 I

Structure de la table PBNAACTSCMCL

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 18 I

Modalités de la table PBNAACTSCMCL

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20

Table PBNADÉPARTEMENTS, fichier accompagnant la base principale

I TABLEAU 19 I

Structure de la table PBNADÉPARTEMENTS

Nom	Type	Longueur	Type	Libellé
ID	Numérique	8		ID
Nom	Texte	50		Nom
NumOrdre	Numérique	8		NumOrdre
Numero	Texte	50		Numero
Region	Numérique	8		Region

I TABLEAU 20 I

Modalités de la table PBNADÉPARTEMENTS

ID	Nom	NumOrdre	Num.	Region	ID	Nom	NumOrdre	Num.	Region
1	Bas-Rhin	0	67	1	50	Gard	0	30	13
2	Haut-Rhin	0	68	1	51	Hérault	0	34	13
3	Dordogne	0	24	2	52	Lozère	0	48	13
4	Gironde	0	33	2	53	Pyrénées-Orientales	0	66	13
5	Landes	0	40	2	54	Corrèze	0	19	14
6	Lot-et-Garonne	0	47	2	55	Creuse	0	23	14
7	Pyrénées-Atlantiques	0	64	2	56	Haute-Vienne	0	87	14
8	Allier	0	03	3	57	Meurthe-et-Moselle	0	54	15
9	Cantal	0	15	3	58	Meuse	0	55	15
10	Haute-Loire	0	43	3	59	Moselle	0	57	15
11	Puy-de-Dôme	0	63	3	60	Vosges	0	88	15
12	Calvados	0	14	4	61	Ariège	0	09	16
13	Manche	0	50	4	62	Aveyron	0	12	16
14	Orne	0	61	4	63	Haute-Garonne	0	31	16
15	Côte-d'Or	0	21	5	64	Gers	0	32	16
16	Nièvre	0	58	5	65	Lot	0	46	16
17	Saône-et-Loire	0	71	5	66	Hautes-Pyrénées	0	65	16
18	Yonne	0	89	5	67	Tarn	0	81	16
19	Côtes-d'Armor	0	22	6	68	Tarn-et-Garonne	0	82	16
20	Finistère	0	29	6	69	Nord	0	59	17
21	Ille-et-Vilaine	0	35	6	70	Pas-de-Calais	0	62	17
22	Morbihan	0	56	6	71	Loire-Atlantique	0	44	18
23	Cher	0	18	7	72	Maine-et-Loire	0	49	18
24	Eure-et-Loir	0	28	7	73	Mayenne	0	53	18
25	Indre	0	36	7	74	Sarthe	0	72	18
26	Indre-et-Loire	0	37	7	75	Vendée	0	85	18
27	Loir-et-Cher	0	41	7	76	Aisne	0	02	19
28	Loiret	0	45	7	77	Oise	0	60	19
29	Ardennes	0	08	8	78	Somme	0	80	19
30	Aube	0	10	8	79	Charente	0	16	20
31	Marne	0	51	8	80	Charente-Maritime	0	17	20
32	Haute-Marne	0	52	8	81	Deux-Sèvres	0	79	20
33	Corse-du-Sud	0	2A	9	82	Vienne	0	86	20
34	Haute-Corse	0	2B	9	83	Alpes-de-Haute-Provence	0	04	21
35	Doubs	0	25	10	84	Hautes-Alpes	0	05	21
36	Jura	0	39	10	85	Alpes-Maritimes	0	06	21
37	Haute-Saône	0	70	10	86	Bouches-du-Rhône	0	13	21
38	Territoire de Belfort	0	90	10	87	Var	0	83	21
39	Eure	0	27	11	88	Vaucluse	0	84	21
40	Seine-Maritime	0	76	11	89	Ain	0	01	22
41	Paris	0	75	12	90	Ardèche	0	07	22
42	Seine-et-Marne	0	77	12	91	Drôme	0	26	22
43	Yvelines	0	78	12	92	Isère	0	38	22
44	Essonne	0	91	12	93	Loire	0	42	22
45	Hauts-de-Seine	0	92	12	94	Rhône	0	69	22
46	Seine-Saint-Denis	0	93	12	95	Savoie	0	73	22
47	Val-de-Marne	0	94	12	96	Haute-Savoie	0	74	22
48	Val-d'Oise	0	95	12					
49	Aude	0	11	13					

ID	Nom	NumOrdre	Num.	Region	ID	Nom	NumOrdre	Num.	Region
97	Guadeloupe	0	971	23	102	Mayotte	0	976	23
98	Martinique	0	972	23	103	Wallis-et-Futuna	0	986	23
99	Guyane	0	973	23		Polynésie			
100	La Réunion	0	974	23	104	Française	0	987	23
101	Saint-Pierre-et-Miquelon	0	975	23		Nouvelle-Calédonie	0	988	23

Num., Numéro

Table PBNADIAGS, fichier accompagnant la base principale

I TABLEAU 21 I

Structure de la table PBNADIAGS

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	200	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 22 I

Modalités de la table PBNADIAGS

ID	Nom	NomCourt	NumOrdre
1	Plainte mnésique	R413	10
2	MCI amnésique	F06.7A	20
3	MCI autres	F06.7B	30
4	Maladie d'Alzheimer	F00.1	40
5	Démence vasculaire	F01.9	60
6	Autre démence vasculaire	F01.8	70
7	Démence mixte	F00.2	80
8	Démence de la maladie de Parkinson	F02.3	100
9	Démence à corps de Lewy	F02.34	110
10	Démence fronto-temporale (DFT-vf)	F02.0	120
11	Démence sémantique (et autres DFT-vt)	G31.B	130
12	Aphasie progressive primaire	G31.A	140
13	Atrophie corticale postérieure	G31.C	150
14	Paralysie supra-nucléaire progressive	G23.1	160
15	Dégénérescence cortico-basale	G23.8	170
16	Maladie de Huntington	F02.2	180
17	Maladie de Creutzfeld Jakob	F02.1	190
18	Hydrocéphalie à pression normale	G91.2	200
19	Encéphalite limbique	G04.9	210
20	Démence du VIH	F02.4	220
21	Démences non classées ailleurs	F0.3	230
22	Troubles anxieux	F41.9	240
23	Trouble dépressif isolé	F32.9	250
24	Trouble dépressif récurrent	F33.9	260
25	Trouble anxieux / dépressif	F41.2	270
26	Troubles psychotiques	F29	280
27	État de stress post-traumatique	F43.1	290
28	Autres troubles psychiatriques	F44.9	300
29	Autres troubles neurologiques	F09	310
30	Lésion vasculaire (AVC)	F01.81	320
31	Traumatisme crânien	S06.9	330

ID	Nom	NomCourt	NumOrdre
32	Épilepsie	G40	340
33	SEP	G35	350
34	Maladie de Parkinson	G20	360
35	Trouble organique cérébral directement lié à une pathologie	F02.8	370
36	Trouble organique cérébral directement lié à l'utilisation d'alcool ou autres toxiques	F10.6	380
37	Intoxication	F13	390
38	Diagnostic en attente	R418	400

Table PBNAEFFETSIND, fichier accompagnant la base principale

I TABLEAU 23 I

Structure de la table PBNAEFFETSIND

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 24 I

Modalités de la table PBNAEFFETSIND

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20
3	Sans Rapport	SR	30

Table PBNAFINSSUIV, fichier accompagnant la base principale

I TABLEAU 25 I

Structure de la table PBNAFINSSUIV

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 26 I

Modalités de la table PBNAFINSSUIV

ID	Nom	NomCourt	NumOrdre
1	Suivi en cours	ENCOURS	10
2	Déménagement	DEM	20
3	Entrée en institution	INSTIT	30
4	Suivi par médecin traitant	MED	40
5	Suivi par un autre centre	CENTRE	50
6	Décès	DECES	60
7	Choix du patient	CHOIX	70

8	Sans Information	SI	80
---	------------------	----	----

Table PBNAHDJSCENTRE, fichier accompagnant la base principale

I TABLEAU 27 I

Structure de la table PBNAHDJSCENTRE

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 28 I

Modalités de la table PBNAHDJSCENTRE

ID	Nom	NomCourt	NumOrdre
1	Pas d'HDJ	1	10
2	HDJ dans l'établissement	2	20
3	HDJ dans un autre établissement	3	30

Table PBNAINCLUSCENTRE, fichier accompagnant la base principale

I TABLEAU 29 I

Structure de la table PBNAINCLUSCENTRE

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 30 I

Modalités de la table PBNAINCLUSCENTRE

ID	Nom	NomCourt	NumOrdre
1	Centre Inclus dans un CHU	CHU	10
2	Centre Inclus dans un CH	CH	20
3	Centre Inclus dans un CHS	CHS	30
4	Centre Inclus dans un hôpital Local	HL	40
5	Centre Inclus dans un PSPH	PSPH	50
6	Centre Inclus dans une structure privée	PRIV	60

Table PBNALABELSCENTRE, fichier accompagnant la base principale

I TABLEAU 31 I

Structure de la table PBNALABELSCENTRE

Nom	Type	Longueur	Type	Libellé
ID	Numérique	8		ID
Nom	Texte	50		Nom
NomCourt	Texte	50		NomCourt
NumOrdre	Numérique	8		NumOrdre

I TABLEAU 32 I

Modalités de la table PBNALABELSCENTRE

ID	Nom	NomCourt	NumOrdre
1	Centre non labellisé	N	10
2	Centre labellisé	O	20

Table PBNAMESPROT, fichier accompagnant la base principale

I TABLEAU 33 I

Structure de la table PBNAMESPROT

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 34 I

Modalités de la table PBNAMESPROT

ID	Nom	NomCourt	NumOrdre
1	Aucune	AUCUNE	10
2	Tutelle	TUTELLE	20
3	Curatelle	CURATELLE	30
4	Sans Information	SI	40

Table PBNAMODESVIE, fichier accompagnant la base principale

I TABLEAU 35 I

Structure de la table PBNAMODESVIE

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 36 I

Modalités de la table PBNAMODESVIE

ID	Nom	NomCourt	NumOrdre
1	À domicile, seul, sans famille à coté	DOMSEULSFA	10
2	À domicile, seul, avec famille à coté	DOMSEULAF	20
3	À domicile, avec conjoint	DOMCONJ	30
4	À domicile, dans famille	DOMFAM	40
5	À domicile, sans autres indications	DOMSAI	50
6	En famille d'accueil	FAMACC	60
7	En foyer logement	FOYER	70
8	En hospitalisation (court séjour, SSR, USLD)	HOSPI	80
9	En maison de retraite médicalisée	MRET	90
10	En EHPAD	EHPAD	100

11	En établissement psychiatrique	HPSY	110
12	Autre	AUTRE	120

Table PBNANEUROSCL, fichier accompagnant la base principale

I TABLEAU 37 I

Structure de la table PBNANEUROSCL

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 38 I

Modalités de la table PBNANEUROSCL

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20

Table PBNANIVEAUXET, fichier accompagnant la base principale

I TABLEAU 39 I

Structure de la table PBNANIVEAUXET

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 40 I

Modalités de la table PBNANIVEAUXET

ID	Nom	NomCourt	NumOrdre
1	Aucune Scolarisation	AUCUNE	10
2	Études Primaires	PRI	20
3	Études Secondaires 1 ^{er} Cycle	SEC1	30
4	Études secondaires 2 ^e Cycle	SEC2	40
5	Études supérieures	SUP	60
6	Non renseigné	NR	70

Table PBNAPROFESSIONS, fichier accompagnant la base principale

I TABLEAU 41 I

Structure de la table PBNAPROFESSIONS

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom

NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 42 I

Modalités de la table PBNAPROFESSIONS

ID	Nom	NomCourt	NumOrdre
1	Agriculteur	AGRI	10
2	Artisan	ARTISAN	20
3	Commerçant ou chef d'entreprise	COMMERCANT	30
4	Cadres et prof. intellectuelles supérieures	CADRE	40
5	Professions intermédiaires	INTER	50
6	Employé	EMPLOYE	60
7	Ouvrier	OUVRIER	70
8	Femme au foyer	FOYER	80
9	Autres	AUTRES	90
10	Sans Informations	SI	100

Table PBNAPROTOS, fichier accompagnant la base principale

I TABLEAU 43 I

Structure de la table PBNAPROTOS

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 44 I

Modalités de la table PBNAPROTOS

ID	Nom	NomCourt	NumOrdre
1	Étude clinique non thérapeutique	ECNT	10
2	Essai médicamenteux	EM	20
3	Essai non médicamenteux	ENM	30
4	Essai Global	EG	40

Table PBNAPROVSPAT, fichier accompagnant la base principale

I TABLEAU 45 I

Structure de la table PBNAPROVSPAT

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 46 I

Modalités de la table PBNAPROVSPAT

ID	Nom	NomCourt	NumOrdre
1	Médecin généraliste	MEDGEN	10
2	Neurologue	NEURO	20
3	Psychiatre	PSY	30
4	Gériatre	GER	40
5	Autre spécialiste	SPE	50
6	Direct	DIRECT	60
7	Ehpad	MRET	70
8	Centre mémoire	CM	80
9	Centre mémoire de ressources et recherche	CMRR	90
10	Service d'hospitalisation	HOSPI	100
11	Urgences médicales	URG	110
12	Services Sociaux	SS	120
13	Clic	CLIC	130
14	Maia	MAIA	140

Table PBNAREGIONS, fichier accompagnant la base principale

I TABLEAU 47 I

Structure de la table PBNAREGIONS

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	80	Nom
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 48 I

Modalités de la table PBNAREGIONS

ID	Nom	NumOrdre
1	Alsace	10
2	Aquitaine	20
3	Auvergne	30
4	Basse-Normandie	40
5	Bourgogne	50
6	Bretagne	60
7	Centre	70
8	Champagne-Ardenne	80
9	Corse	90
10	Franche-Comté	100
11	Haute-Normandie	110
12	Île-de-France	120
13	Languedoc-Roussillon	130
14	Limousin	140
15	Lorraine	150
16	Midi-Pyrénées	150
17	Nord-Pas-de-Calais	170
18	Pays-de-Loire	180
19	Picardie	190
20	Poitou-Charentes	200
21	Provence-Alpes-Côte-d'Azur	210
22	Rhône-Alpes	220
23	Outre-Mer	230

Table PBNARESEAUXCENTRE, fichier accompagnant la base principale

I TABLEAU 49 I

Structure de la table PBNARESEAUXCENTRE

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 50 I

Modalités de la table PBNARESEAUXCENTRE

ID	Nom	NomCourt	NumOrdre
1	Centre n'appartenant pas à un Réseau	N	10
2	Centre appartenant à un réseau	O	20

Table PBNASITESCENTRE, fichier accompagnant la base principale

I TABLEAU 51 I

Structure de la table PBNASITESCENTRE

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

Modalités de la table PBNASITESCENTRE

ID	Nom	NomCourt	NumOrdre
1	Centre mono-site	N	10
2	Centre multi-sites	O	20

Table PBNASITSGEO, fichier accompagnant la base principale

I TABLEAU 53 I

Structure de la table PBNASITSGEO

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 54 I

Modalités de la table PBNASITSGEO

ID	Nom	NomCourt	NumOrdre
1	Dans la ville du CMRR ou CM	VILLE	10
2	À plus de 50 km du CMRR ou CM	+50 KM	20
3	Dans une autre région	EXTREG	30
4	Dans un autre pays	EXTPAYS	40

Table PBNASOUTIENSCM, fichier accompagnant la base principale

I TABLEAU 55 I

Structure de la table PBNASOUTIENSCM

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 56 I

Modalités de la table PBNASOUTIENSCM

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20

Table PBNASPECIALSCL, fichier accompagnant la base principale

I TABLEAU 57 I

Structure de la table PBNASPECIALSCL

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 58 I

Modalités de la table PBNASPECIALSCL

ID	Nom	NomCourt	NumOrdre
1	Cabinet Isolé	ISOLE	10
2	Cabinet dans un groupe	GROUPE	20
3	Cabinet dans un établissement	ETAB	30

Table PBNASTIMULSCM, fichier accompagnant la base principale

I TABLEAU 59 I

Structure de la table PBNASTIMULSCM

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom

NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 60 I

Modalités de la table PBNASTIMULSCM

ID	Nom	NomCourt	NumOrdre
1	Non	N	10
2	Oui	O	20

Table PBNATYPESACTE, fichier accompagnant la base principale

I TABLEAU 61 I

Structure de la table PBNATYPESACTE

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 62 I

Modalités de la table PBNATYPESACTE

ID	Nom	NomCourt	NumOrdre
1	Consultation unique	CONS	10
2	Consultation multidisciplinaire	CONSMULTI	20
3	Bilan neuropsychologique	BILAN	30
4	Hôpital de jour	HDJ	40
5	Groupe de patients (GP)	GP	50
6	Groupe de familles (GF)	GF	60
7	Hôpital de jour - SSR	HDJSSR	45
8	Intervention infirmière	IINF	70
9	Intervention assistance sociale	IASS	80
10	Intervention psychologue	IPSY	90
11	Intervention orthophoniste	IORTHO	100
12	Consultation avec annonce Alzheimer ou apparentée	CONSANN	110
13	Consultation avec bilan neuropsychologique	CONSBIL	120
14	Bilan neuropsych. demandé par spé. libéral de ville	BILSLVIL	14

Table PBNATYPESCENTRE, fichier accompagnant la base principale

I TABLEAU 63 I

Structure de la table PBNATYPESCENTRE

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 64 I

Modalités de la table PBNATYPESCENTRE

ID	Nom	NomCourt	NumOrdre
1	Centre mémoire	CM	10
2	Centre mémoire de ressources et recherche	CMRR	20
3	Spécialiste libéral	CL	30

Table PBNATYPESVIS, fichier accompagnant la base principale

I TABLEAU 65 I

Structure de la table PBNATYPESVIS

Nom	Type	Longueur	Libellé
ID	Numérique	8	ID
Nom	Texte	50	Nom
NomCourt	Texte	50	NomCourt
NumOrdre	Numérique	8	NumOrdre

I TABLEAU 66 I

Modalités de la table PBNATYPESVIS

ID	Nom	NomCourt	NumOrdre
1	Screening	S	10
2	Inclusion	I	20
3	Visite de Suivi	VS	30

Annexe 3. Comparaison de la structure de la base de données avec les documents

I TABLEAU 67 I

Description des 24 variables présélectionnées et étudiées et description des difficultés identifiées pour leur prise en main.

Nom de la Variable <i>Libellé supposé de la variable</i>	Type <i>Longueur</i>	Informations complémentaires <i>(conditions/modalités)</i>	Difficultés retrouvées pour la prise en main de la base et des différents documents fournis
Variables géographiques			
CentreRegion <i>Région du centre</i>	Num. ¹ 8	<i>observé : 1 modalité pour chacune des 22 régions et 1 pour l'ensemble des DOM</i>	Variable non décrite dans les documents fournis ; Caractère obligatoire ou facultatif, non disponible ; Date de création non connue Rattaché <i>a priori</i> à la table PBNAREGIONS
CentreDepartement <i>Département du centre</i>	Texte 4	<i>Observé : 1 modalité pour chacune des 105 départements dont les 7 DOM-TOM</i>	Variable non décrite dans les documents fournis ; Caractère obligatoire ou facultatif, non disponible ; Date de création non connue Rattaché <i>a priori</i> à la table PBNADepARTEMENTS
Variables « patient »			
PatientId <i>Identifiant unique du patient</i>	Texte 64 ou 300	obligatoire	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Obtenu avec l'algorithme d'anonymisation (SHA-1 ou SHA-2 selon les documents). Variable de 64 caractères selon le SIC et de 300 dans la base
PatientSexe <i>Sexe du patient</i>	Texte 1 ou 3	obligatoire ; 2 modalités (M/F ; Femme, Homme)	Modalités Femme/Homme selon le SIC et F/H selon le CIMA. Format numérique selon le SIC et alphanumérique dans la base
PatientAnneeNaissance <i>Année de naissance du patient</i>	Num. ¹ 4 ou 8	obligatoire	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues.
DclAnneePremierDiag <i>Année de première consultation</i>	Num. ¹ 4 ou 8	obligatoire	Date de création non connue. Variable devenue obligatoire avec la version 5.5 ; le libellé est modifié avec la version 5.6. Les modifications avant la version 5.1 ne sont pas connues.
DclALD <i>ALD</i>	Num. ¹ 2 ou 8	obligatoire ; 2 ou 3 modalités (Oui, Non ± sans information)	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNAALDS
DclAPA <i>APA</i>	Num. ¹ 2 ou 8	obligatoire ; 2 ou 3 modalités (Oui, Non ± sans information)	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNAAPAS
PatientAccompagnant <i>Caractéristiques de l'accompagnant</i>	Num. ¹ 8 ou 9	obligatoire ; 10 modalités	Date de création supposée le 01/01/2011 ; Création avec la version 5.5. Variable modifiée avec la 5.7 (ajout de modalités*). Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table

Nom de la Variable <i>Libellé supposé de la variable</i>	Type <i>Longue</i>	Informations complémentaires (conditions/modalités)	Difficultés retrouvées pour la prise en main de la base et des différents documents fournis
PatientEnvoyePar <i>Patient envoyé par</i>	Num. ¹ <i>6 ou 8</i>	obligatoire ; 14 modalités	PBNAACCOMPAGNANTS Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNAPROVSPAT
PatientNiveauEtudes <i>Niveau d'étude</i>	Num. ¹ <i>6 ou 8</i>	obligatoire ; 6 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNANIVEAUXET
PatientProfession <i>Profession – dernier métier exercé</i>	Num. ¹ <i>8 ou 10</i>	obligatoire ; 10 ou 12 modalités	Date de création non connue. Création avec la version 5.5. Variable modifiée avec la 5.7 et 5.7 (ajout d'une modalité*). Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNAPROFESSIONS. Le CIMA indique 12 modalités contre 10 dans la base ; celles relatives au lieu de résidence ne sont pas retrouvées dans la base.
¹ Num., numérique * modalité précisée dans le SIC			
Variables « acte »			
DclDateActe <i>Date de l'acte</i>	Date <i>8</i>	obligatoire	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues.
DclTypeActe <i>Type d'acte</i>	Num. ¹ <i>8</i>	obligatoire ; 14 modalités	Date de création non connue. Variable modifiée avec les versions 5.2 (ajout d'une modalité*), 5.5 (ajout de modalités, sans précision) et 5.8 (ajout d'une modalité* qui existe dans la base depuis 2010) ; Apparition de la modalité « intervention Assistance sociale » dès 2011. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNATYPESACTE
DclEtatDemence <i>Diagnostic N1</i>	Num. ¹ <i>8</i>	obligatoire ; 3 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues.
DclDiagnostic <i>Diagnostic N2</i>	Num. ¹ <i>6 ou 8</i>	obligatoire ; 38 modalités	Date de création non connue. Variable modifiée avec les versions 5.1 (suppression du diagnostic G30.1, lésion vasculaire ou AVC sont codés F01.81) et 5.4 (suppression du diagnostic G30.8). Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNADIAGS
DclIADLDetection <i>IADL score détection</i>	Num. ¹ <i>2 ou 8</i>	facultatif, score sur une échelle de 4	Date de création non connue. Cette variable a été révisée avec la version 2 selon le glossaire du CIMA. Les modifications avant la version 5.1 ne sont pas connues dans le SIC.
DclIMMSE <i>Score au Mini Mental test</i>	Num. ¹ <i>2 ou 8</i>	facultatif, score sur une échelle de 30	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues.
DclMesureProtection <i>Mesure de protection</i>	Num. ¹ <i>8 ou 9</i>	obligatoire ; 4 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table

Nom de la Variable <i>Libellé supposé de la variable</i>	Type <i>Longueur</i>	Informations complémentaires <i>(conditions/modalités)</i>	Difficultés retrouvées pour la prise en main de la base et des différents documents fournis
DclModeDeVie <i>Mode de vie actuel</i>	Num. ¹ <i>8 ou 10</i>	obligatoire ; 12 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNAMODESVIE
DclIPCAucun <i>Examens dont les résultats ont été consultés le jour de l'acte</i>	Num. ¹ <i>1 ou 8</i>	obligatoire ; 2 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Modalités en « Oui/Non » dans le SIC et en « 0/1 » dans la base
DclTPAucun <i>Traitements pharmacologiques</i>	Num. ¹ <i>1 ou 8</i>	obligatoire ; 2 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Modalités en « Oui/Non » dans le SIC et en « 0/1 » dans la base
DclFinSuivi <i>Type de fin de suivi</i>	Num. ¹ <i>7 ou 8</i>	obligatoire ; 8 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Rattaché <i>a priori</i> à la table PBNAFINSSUIV. Variable décrite dans le SIC et la base ; deux variables peuvent y être rattachées dans le CIMA et dans le glossaire du CIMA (variables « Type de fin de suivi » avec 8 modalités dont « En cours » et « État du suivi »)
DclIPSRaucun <i>Traitements non pharmacologiques</i>	Num. ¹ <i>1 ou 8</i>	obligatoire ; 2 modalités	Date de création non connue. Les modifications avant la version 5.1 ne sont pas connues. Modalités en « Oui/Non » dans le SIC et en « 0/1 » dans la base

¹ Num., numérique

* modalité précisée dans le SIC

Annexe 4. Items retenus et cotations de ces items pour le score IADL

I TABLEAU 68 I

Score IADL

IADL - Activités courantes

Entourez le chiffre correspondant à l'état du patient et reportez le score à l'échelle

Score maximum 4

4 ITEMS CONSERVES

Aptitude à utiliser le téléphone

1	Se sert normalement du téléphone	1
2	Compose quelques numéros très connus	1
3	Répond au téléphone mais ne l'utilise pas spontanément	1
4	N'utilise pas du tout le téléphone spontanément	0
5	Incapable d'utiliser le téléphone	0

Moyens de transport

1	1. Utilise les moyens de transports de façon indépendante ou conduit sa propre voiture	1
2	2. Organise ses déplacements en taxi ou n'utilise aucun moyen de transport public	1
3	3. Utilise les transports publics avec l'aide de quelqu'un	1
4	4. Déplacement limité en taxi ou en voiture avec l'aide de quelqu'un	0

Responsabilité à l'égard de son traitement

1	Est responsable de la prise de ses médicaments (doses et rythmes corrects)	1
2	Est responsable de la prise de ses médicaments si les doses ont été préparées à l'avance	0
3	Est incapable de prendre seul ses médicaments même si ceux-ci ont été à l'avance	0

Score | __ |

Aptitude à manipuler l'argent

0	Non applicable, n'a jamais manipulé l'argent	
1	Gère ses finances de façon autonome	1
2	Se débrouille pour les achats quotidiens, mais a besoin d'aide pour les opérations à la banque et les achats importants	1
3	Incapable de manipuler l'argent	0

Annexe 5. Description des modalités et de leur répartition, 2010-2014

Variables qualitatives

I TABLEAU 69 I

Description des variables qualitatives, 2010-2014 (sous l'hypothèse que PatientID[#] est unique par patient)

Variable (<i>obligatoire réactualisable</i>) modalités	Effectif	%	manquant	%
Nombre de PatientID#	556 069	100%	0	0 %
Sexe (au premier acte)			0	0 %
Femme	348 375	62,7		
Homme	207 694	37,4		
État dément (au premier acte)			0	0 %
Dément	224 210	40,3		
Troubles cognitifs sans démence	201 292	36,2		
Autre cas	130 563	23,5		
Modalité non définie	4	0		
Diagnostic (au premier acte)			0	0 %
Dément ¹⁴	202 508	46,2		
Alzheimer	107 615	25,1		
Diagnostic en attente	188 406	33,9		
Score IADL Detection (1^{er} score renseigné)	296 992	53%	259 077	58,9 %
0	79 162	14,2		
1	43 969	7,9		
2	39 886	7,2		
3	41 130	7,4		
4	92 052	16,6		
score > ou = 5	793	0,1%		
Patient envoyé par (au premier acte)			0	0 %
Médecin généraliste	324 456	58,4		
Neurologue	50 220	9,0		
Psychiatre	11 822	2,1		

¹⁴ Si la variable **DcIDiagnostic** correspond aux modalités 4, 5, 6, 7, 8, 9, 10, 16, 17, 20, 21, 30 et 35 (cf. Modalités de la table PBNADIAGS)

Pour cette analyse, il est fait l'hypothèse que le numéro d'anonymisation PatientID est unique. Or, nous avons mis en évidence que cette hypothèse n'est pas vérifiée (cf. 4.2.2.3.1), les effectifs indiqués dans ce tableau sont donc erronés.

Variable (obligatoire réactualisable) modalités	Effectif	%	manquant	%
Gériatre	32 261	5,8		
Autre spécialiste	26 541	4,8		
Direct	27 877	5,0		
Ehpad	10 313	1,9		
Centre mémoire	2 693	0,5		
Centre mémoire de ressources et recherche	2 446	0,4		
Service d'hospitalisation	60 948	10,9		
Urgences médicales	3 684	0,7		
Services sociaux	1 652	0,3		
Clic	873	0,2		
Maia	283	0,05		
Accompagnant du patient (au premier acte)¹⁵	555 613	100%	456	0 %
Conjoint	55 481	10,0		
Enfant	55 156	9,9		
Bel enfant	3 076	0,6		
Petit enfant	1 485	0,3		
Neveu ou nièce	2 203	0,4		
Frère ou sœur	2 891	0,5		
Curateur	336	0,06		
Aucun ¹⁵	63 365	11,4		
Autres	10 582	1,9		
Sans information ¹⁵	361 038	0,6		
Niveau d'études (au premier acte)	483 111	87%	72 958	13,12
Aucune scolarisation	36 656	6,6		
Études primaires	220 137	39,6		
Études secondaires 1 ^{er} Cycle	99 126	17,8		
Études secondaires 2 ^e Cycle	63 647	11,5		
Études supérieures	63 545	11,4		
Profession du patient (au premier acte)	555 613	100%	456	0 %
Agriculteur	11 683	2,1		
Artisan	6 287	1,1		
Commerçant ou chef d'entreprise	11 291	2,0		
Cadres et prof. intellectuelles supérieures	20 780	3,7		
Professions intermédiaires	15 077	2,7		
Employé	56 061	10,1		
Ouvrier	25 759	4,6		
Femme au foyer	19 735	3,6		
Autres	13 172	2,4		
Sans Informations	375 768	67,6		

¹⁵ Toutes les modalités, notamment « aucun » et « sans information », sont présentes entre 2010 et 2014

Variable (obligatoire réactualisable) modalités	Effectif	%	manquant	%
Patient sous ALD (1^{re} fois renseigné)			0	0 %
Non	197 247	35,5		
Oui	193 449	34,8		
Sans Information	165 373	29,7		
Non	289 318	52,0		
Oui	86 611	15,6		
Sans Information	180 140	32,4		
Aucun traitement pharmaceutique (au 1^{er} acte)				
oui ? Non ?	328 759	59,1		
non? Oui ?	227 310	40,9		
Aucun Résultats Examens (au 1^{er} acte)				
oui ? Non ?	135 148	24,3		
non? Oui ?	420 921	75,7		
Fin de Suivi (au premier acte)	542 557	98%	13 512	2 %
Suivi en cours	531 675	95,6		
Déménagement	115	0,02		
Entrée en institution	516	0,1		
Suivi par médecin traitant	7 690	1,4		
Suivi par un autre centre	466	0,1		
Décès	1 651	0,3		
Choix du patient	444	0,1		
Mesure de protection (au premier acte)			0	0 %
aucune	444 500	79,9		
tutelle	7 312	1,3		
curatelle	6 474	1,2		
Sans Information	97 783	17,6		
Mode de vie (au premier acte)			0	0 %
À domicile, seul, sans famille à coté	55 138	9,9		
À domicile, seul, avec famille à coté	103 743	18,1		
À domicile, avec conjoint	250 647	45,1		
À domicile, dans famille	37 600	6,8		
À domicile, sans autres indications	22 957	4,1		
En famille d'accueil	1 138	0,2		
En foyer logement	9 139	1,6		
En hospitalisation (court séjour, SSR, USLD)	13 825	2,5		
En maison de retraite médicalisée	7 539	1,4		
En Ehpad	39 474	7,1		

Variable (obligatoire réactualisable) modalités	Effectif	%	manquant	%
En établissement psychiatrique	2 206	0,4		
Autre	12 663	2,3		
Patient n'a aucune prise en charge (au premier acte)				
oui ? Non ?	120 240	21,6		
non? Oui ?	435 829	78,4		
Type d'acte				
			0	0 %
Consultation unique	284 644	51,2		
Consultation multidisciplinaire	65 264	11,7		
Bilan neuropsychologique	79 886	14,4		
Hôpital de jour	57 129	10,3		
Groupe de patients (GP)	545	0,1		
Groupe de familles (GF)	290	0,05		
Hôpital de jour - SSR	1 874	0,3		
Intervention infirmière	1 159	0,2		
Intervention assistance sociale	152	0,03		
Intervention psychologue	785	0,1		
Intervention orthophoniste	2 932	0,5		
Consultation avec annonce Alzheimer ou apparentée	3 575	0,6		
Consultation avec bilan neuropsychologique	50 284	9,0		
Bilan neuropsych. demandé par spé. libéral de ville	7 550	1,4		
Type de centre				
			0	0 %
CM	387 970	69,8		
CMRR	151 873	27,3		
Médecin libéral	16 226	2,9		
Région (au dernier acte)				
			0	0 %
Alsace	19 970	3,6		
Aquitaine	29 541	5,3		
Auvergne	10 816	2,0		
Basse-Normandie	17 140	3,1		
Bourgogne	17 731	3,2		
Bretagne	34 572	6,2		
Centre	17 303	3,1		
Champagne-Ardenne	9 150	1,7		
Corse	1 912	0,3		
Franche-Comté	9 431	1,7		
Haute-Normandie	19 416	3,5		
Île-de-France	52 427	9,4		
Languedoc-Roussillon	26 549	4,8		
Limousin	10 854	2,0		
Lorraine	17 350	3,1		
Midi-Pyrénées	32 490	5,8		

Variable (obligatoire réactualisable) modalités	Effectif	%	manquant	%
Nord-Pas-de-Calais	39 911	7,2		
Pays-de-Loire	35 755	6,4		
Picardie	20 562	3,7		
Poitou-Charentes	15 592	2,8		
Provence-Alpes-Côte-d'Azur	43 773	7,9		
Rhône-Alpes	71 339	12,8		
Outre-mer	2 485	0,5		

Variables quantitatives

I TABLEAU 70 I

Description des variables quantitatives, 2010-2014 (sous l'hypothèse que PatientID# est unique par patient)

Variable (obligatoire réactualisable) modalités	Effectif	%	manquant	%	médiane	min	max
Nombre de PatientID#	556 069	100	0	0			
Âge au 1^{er} acte			0	0			
À partir de la variable PatientAnneeNaissance et de l'année de la date de l'acte, DclDateActe							
moins de 30 ans	4 468	0,8	dont moins de 16 ans		947	0,2	
30-39 ans	5 873	1,1					
40-49 ans	14 359	2,6					
50-59 ans	33 051	5,9					
60-69 ans	66 792	12,0					
70-79 ans	157 358	28,3					
80-89 ans	231 470	41,6					
90 ans et +	41 751	7,5					
Score MMSE (DclMMSE)	437 918	79	118151	21	23	0	30
de 0 à 10	28 499	5,1					
de 11 à 20	140 640	25,3					
de 21 à 30	268 779	48,3					
Âge au diagnostic (au premier diagnostic)	441 028	79 %	115041	21 %	78	110	121
À partir de la variable DclAnneePremierDiag et de l'année de la date de l'acte, DclDateActe							
Âge < 0	924	0,2 %					
0-10 ans	609	0,1 %					
11-19 ans	925	0,2 %					
20-29 ans	3 043	0,5 %					
30-39 ans	5 040	0,9 %					
40-49 ans	12 354	2,2 %					
50-59 ans	28 461	5,1 %					
60-69 ans	57 203	10,3 %					
70-79 ans	131 327	23,6 %					
80-89 ans	171 934	30,9 %					
90-99 ans	28 902	5,2 %					
100 ans et +	306	0,1 %					

Pour cette analyse, il est fait l'hypothèse que le numéro d'anonymisation PatientID est unique. Or, nous avons mis en évidence que cette hypothèse n'est pas vérifiée (cf. 4.2.2.3.1), les effectifs indiqués dans ce tableau sont donc erronés.

Citation suggérée : La Banque nationale Alzheimer, un outil pour la surveillance épidémiologique des démences en France ?
Saint-Maurice : Santé publique France, 2017. 86 p. Disponible à partir de l'URL : www.santepubliquefrance.fr

ISSN : EN COURS — ISBN-NET : 979-10-289-0328-2 — RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION, SANTÉ
PUBLIQUE FRANCE — DÉPÔT LÉGAL : MAI 2017