



**ARAMIS  
LAB**  
BRAIN DATA SCIENCE



**Inserm**



# Retour d'expérience sur l'utilisation des données du SNDS

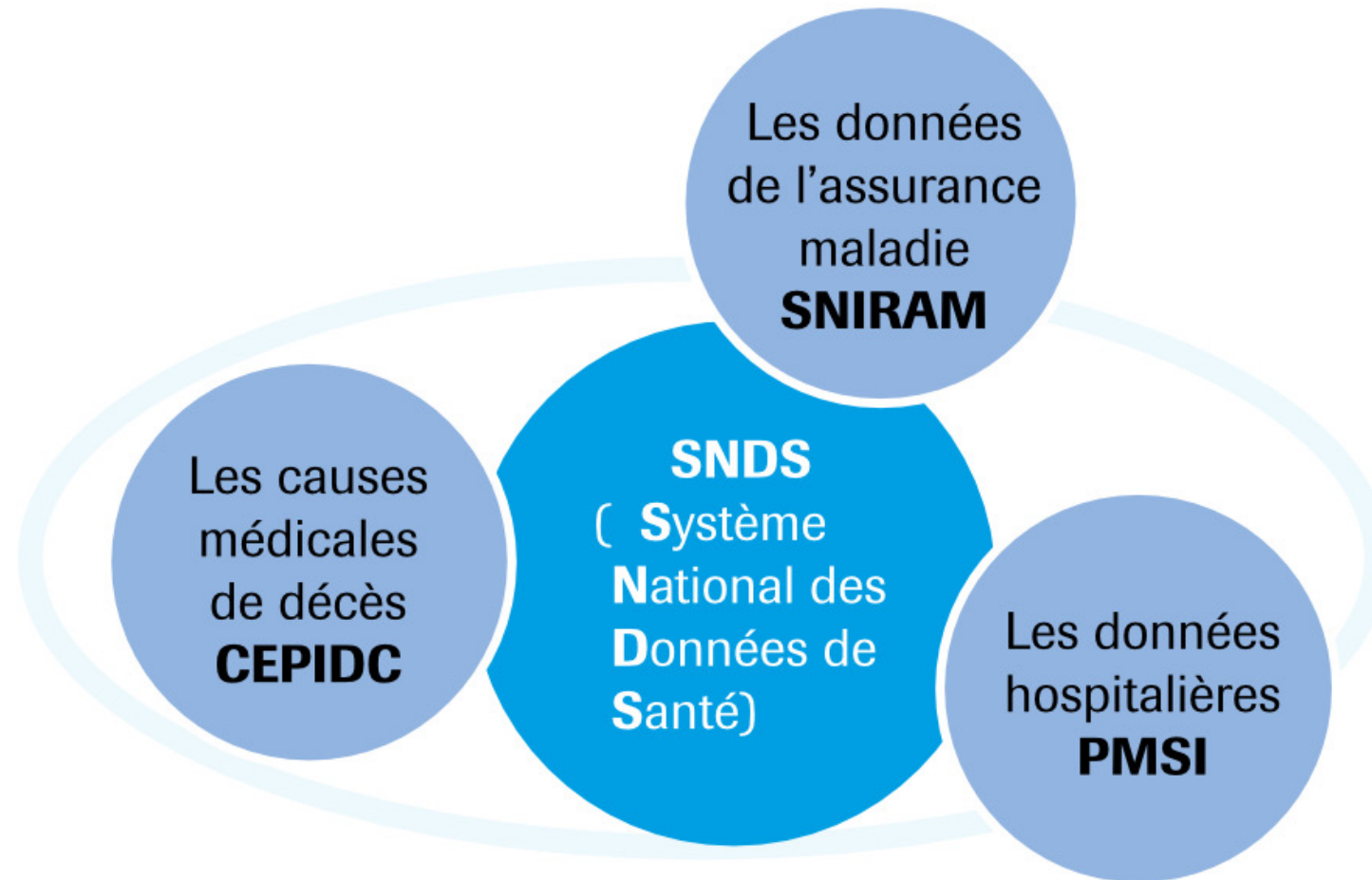
26 septembre 2023

Recontre Devtech santé numérique

Octave Guinebretiere

Doctorant au laboratoire Aramis (Institut du cerveau, Paris, France) sous la supervision de Thomas Nedelec et Stanley Durrleman

- I) **Présentation des données du SNDS**
- II) **Motivation/Méthodes**
- III) **Circuit d'accès**
- IV) **Résultats**
- V) **Futurs travaux**
- VI) **Difficultés d'utilisation**



## Ce que l'on a:

### Information socio-démographique:

- âge, sexe, département de résidence, proxy du niveau socio-économique (indice fdep), bénéficiaire de la complémentaire santé solidaire

### Remboursements en ville:

- Médicaments dispensés en ville (codé selon la classification ATC), dispositifs médicaux, spécialité de la consultation, actes médicaux (codés selon CCAM), actes de biologie

## Ce que l'on n'a pas

### Information socio-démographique:

- Niveau d'éducation, statut marital, catégorie socio-professionnelle, IMC, facteurs de risque comme tabac, alcool, activité physique etc.

### Remboursements en ville:

- Durée du traitement, dosage prescrit et indication, diagnostics pendant la consultation, médicaments disponibles en vente libre, médicaments non consommés, résultats d'examens biologiques

## Ce que l'on a:

### Activité à l'hôpital (séjours):

- Diagnostics séjours dans les champs MCO, SSR, HAD, Psy du PMSI (codé selon la classification ICD-10), date d'hospitalisation, actes médicaux, médicaments en SUS, dispositifs médicaux implantables

### Activité externe:

- Consultation médicale, actes médicaux, dispositifs médicaux implantables, spécialité du service médical

## Ce que l'on n'a pas

### Activité à l'hôpital (séjours):

- Médicaments dispensés, examens biologiques réalisés

### Activité externe:

- Spécialité du médecin, diagnostics

## Avantages

- Etudes à l'échelle populationnelle
  - Minimisation du biais de sélection
  - Représentativité
- Information collectée de façon prospective sans question de recherche prédéterminée
- Données individuelles médicalisées, structurées et codées de manière standardisée

## Inconvénients

- Non collectées pour un objectif de recherche
- Variables confondantes non mesurées
- La variable d'intérêt peut être absente
- Variable d'intérêt mal renseignée
  - Actes peu onéreux moins renseignés
  - Codes CIM-10 des hospitalisations imprécis
  - Sous-déclaration des ALD
- Retards diagnostics
  - Risques de biais de causation inverse

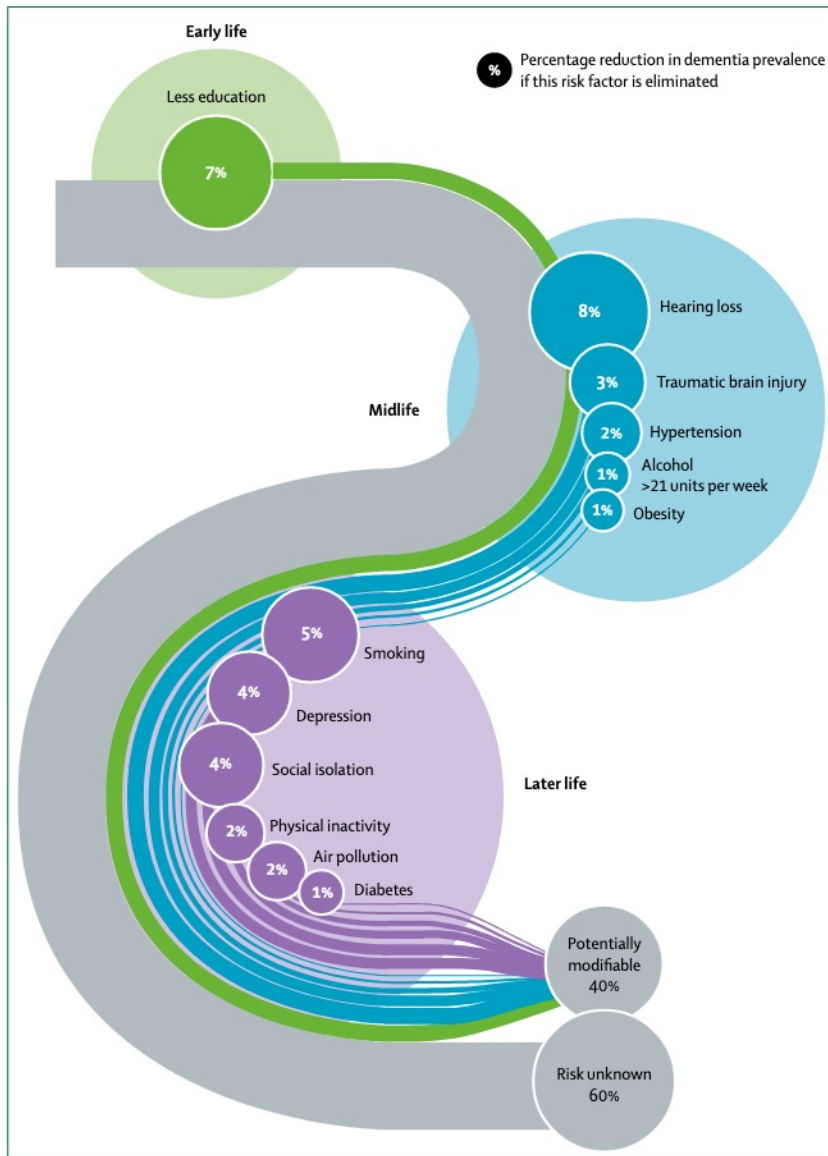


Figure 7: Population attributable fraction of potentially modifiable risk factors for dementia

*Dementia prevention, intervention, and care: 2020 report of the Lancet Commission. Livingston et al.*

- **Besoin d'améliorer la prévention des maladies neurodégénératives**
  - Certains maladies dégénératives prennent de **10 à 20 ans pour s'établir**.
  - Jusqu'à présent, les mesures de biomarqueurs connus de la maladie, tels que la réalisation d'une tomographie par émission de positons (PET scan) ou une ponction lombaire chez chaque individu de plus de 65 ans, ne sont **ni pratiques, ni éthiques, ni rentables**.
  - Mise en place de mesures de prévention en utilisant des données de santé collectées en routine :
    - Prévention primaire : **agir sur les facteurs de risque modifiables**.
    - Prévention secondaire : administrer des traitements **au stade le plus précoce de la maladie** pour ralentir sa progression.
    - Prévention tertiaire : **réduire les complications** de la maladie

*Fournir de nouvelles perspectives pour améliorer la prévention primaire, secondaire et tertiaire*

- Deux projets en cours basés sur l'utilisation des données du SNDS

**1) Comparaison des maladies neurodégénératives dans trois systèmes de santé :**

Prescription de médicaments précédant la maladie d'Alzheimer, la maladie de Parkinson et la sclérose latérale amyotrophique (SLA) - Une étude inter pays menée en Australie, en France et en Suède.

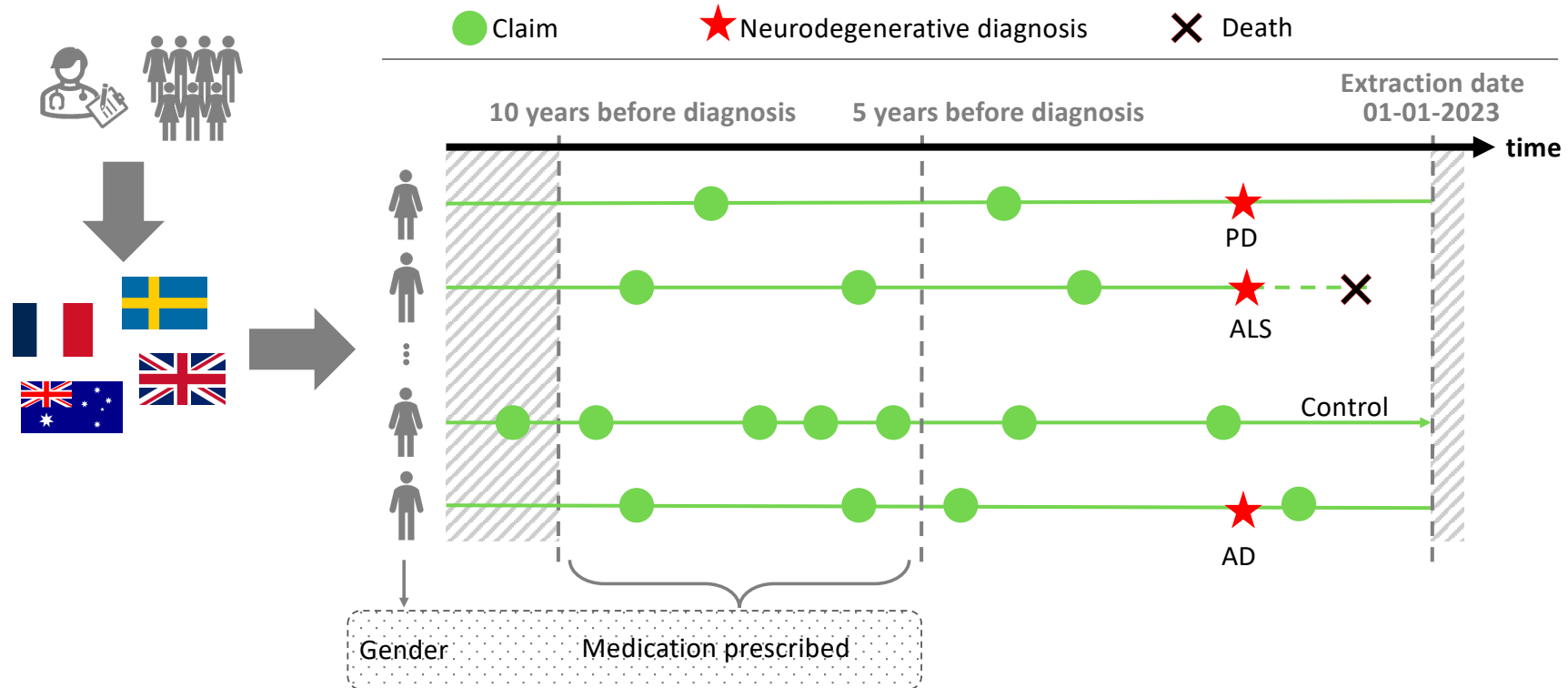
- Accès permanent qui permet d'**analyser de nombreuses maladies** si les informations du SNDS permettent d'identifier les patients
- **Base nationale**

**2) Validité des bases de données medico-administratives pour étudier les maladies**

**neurodégénératives** : Identification et détection du retard de diagnostic dans l'atrophie multi systématisée (AMS) dans le SNDS: une étude de validation sur un registre de l'AMS et une cohorte hospitalière.

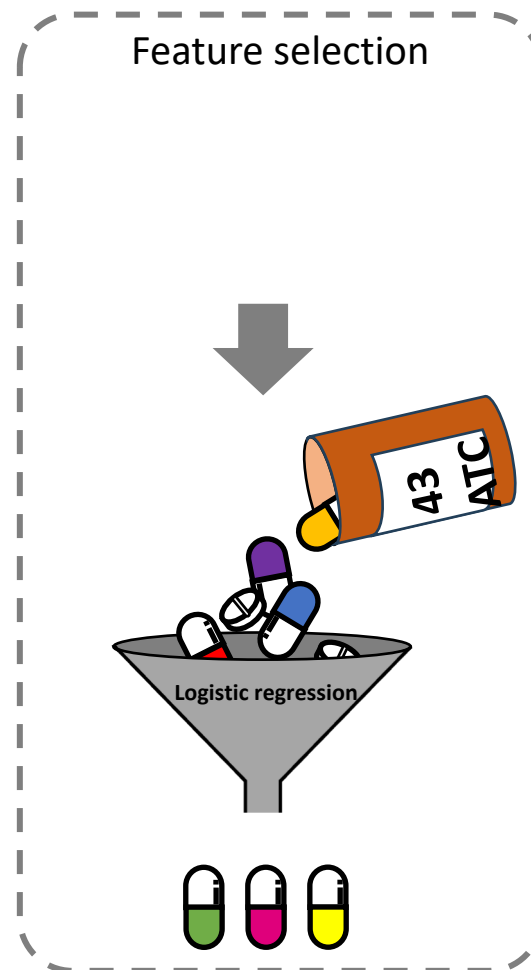
- Base de données nationale permet d'**étudier des maladies très rares**

- Comparaison des maladies neurodégénératives dans trois systèmes de santé



Abbreviations : PD : Parkinson's disease, ALS : amyotrophic lateral sclerosis, AD : Alzheimer's disease

- Comparaison des maladies neurodégénératives dans trois systèmes de santé



- Etude cas-témoin retrospectives sur les données longitudinales du SNDS.
- Test agnostique de 43 codes ATC.

- **Validité des bases de données medico-administratives pour étudier les maladies neurodégénératives : cas de l'AMS**
  - préoccupations concernant la classification erronée des maladies et son impact sur la validité des résultats des études
  - la précision du diagnostic de l'AMS au sein du SNDS n'a pas été vérifiée par rapport aux données cliniques.
  - objectif : évaluer la validité de l'algorithme d'identification de cas d'AMS basé sur les données de décharges hospitalières (PMSI) en utilisant deux cohortes cliniques : un registre de l'AMS et une cohorte hospitalière.

## *B. UNE UTILISATION SECONDAIRE DES DONNÉES ENCORE TROP COMPLEXE*

### **1. Un parcours d'accès qui présente encore des marges de fluidification**

La loi de 2019 a peu modifié la procédure d'accès aux données. **Le principal problème porté à la connaissance de la mission réside dans le délai d'obtention effective des données.**

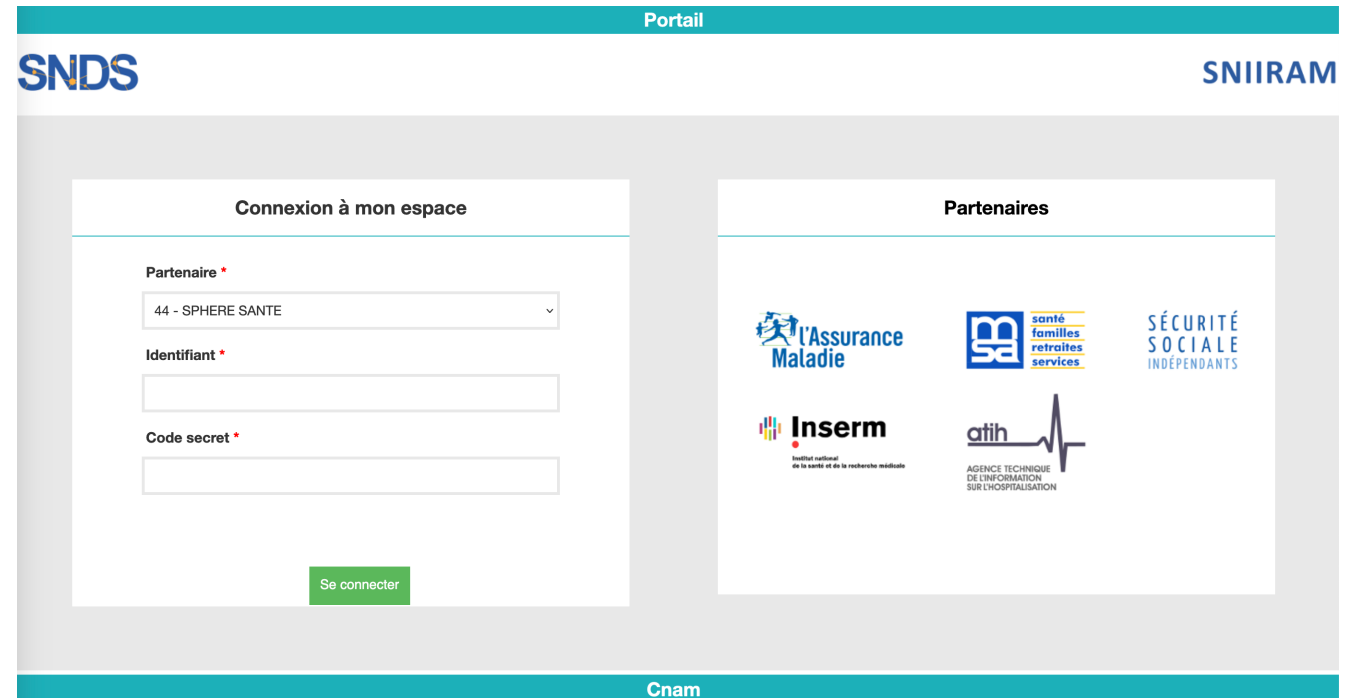
Si le Cesrees et la Cnil parviennent globalement à respecter les délais qui leur sont impartis, la mise à disposition, par la Cnam, de ses données peut prendre entre 10 et 12 mois, **ce qui porte le délai total de la procédure à 18 mois en moyenne.** Un tel délai est extrêmement préjudiciable à la bonne marche des projets de recherche, à l'attractivité de la France dans la compétition mondiale des chercheurs et à son rang dans la course à l'innovation.



**Délai moyen d'accès  
aux données  
du SNDS**

Extrait du rapport du sénat sur les données de santé (Deroche 2023)

- Inria a un accès permanent avec l'historique complet et données individuelles (décret R 1461-12 du 29 juin 2021)
- 6 mois de délai pour accéder aux données du SNDS
- Accès via la plateforme de la CNAM



The screenshot shows the 'Portail' (Portal) interface for SNDS (Système National des Données de Santé) and SNIIRAM (Système National d'Information sur l'Inscription, les Actes Médicaux et les Remboursements). The page is divided into two main sections: 'Connexion à mon espace' (Login to my space) and 'Partenaires' (Partners).

**Connexion à mon espace**

Partenaire \*  
44 - SPHERE SANTE

Identifiant \*  
[Input field]

Code secret \*  
[Input field]

Se connecter

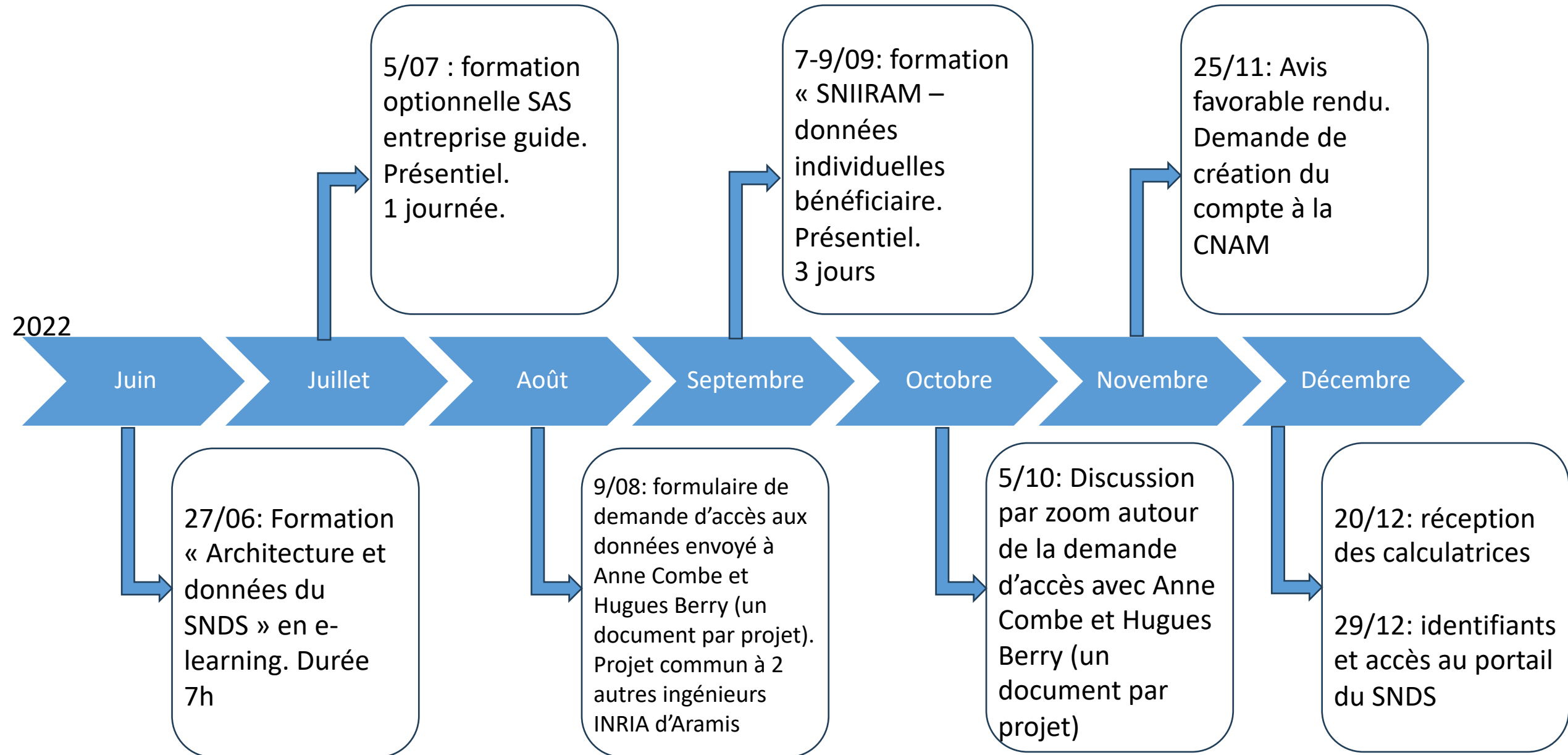
**Partenaires**

Logos of partner organizations are displayed, including:

- Assurance Maladie
- santé familles retraites services
- SÉCURITÉ SOCIALE INDÉPENDANTS
- Inserm (Institut national de la santé et de la recherche médicale)
- atih (Agence technique de l'information sur l'hospitalisation)

The footer of the page indicates 'Cnam'.

# Circuit d'accès (3/4)



Accueil

Statistiques

Dictionnaire

Forum

Tableaux et requêtes

## Outil Applicatif

Citrix

### Applications

Principal

#### Apps

All (6) Categories ▼



BRUS



Microsoft Excel 2010  
Microsoft



Microsoft PowerPoint 2010  
Microsoft



Microsoft Word 2010  
Microsoft



RStudio Workbench



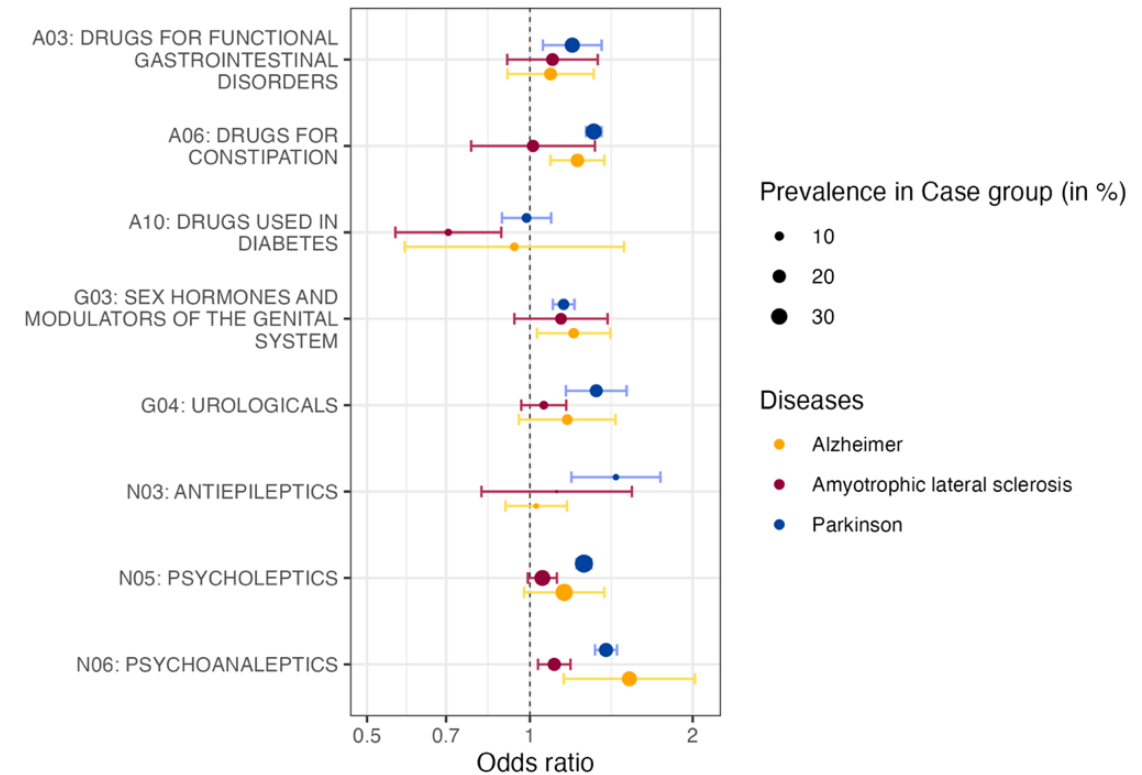
SASGuide 8

## • Comparaison des maladies neurodégénératives dans trois systèmes de santé

Characteristics	Alzheimer's disease			Parkinson's disease			Amyotrophic lateral sclerosis		
	France	The UK	Sweden	Australia	France	Sweden	Australia	France	Sweden
Number of cases	19,458	20,214	109,881		<b>195,575</b>	46,121	6621	<b>14,065</b>	6847
Age at diagnosis, mean (SD)/years	81 (76-85)	82 (76-87)	81.2 (8.0)		73.2 (9.7)	75.9 (9.7)	64.4 (11.9)	67.5 (12.5)	69.7 (12.1)
Sex, number (%)									
Men	7016 (37)	6108 (31)	40,686 (37.0)		111,223 (56.9)	32,630 (57.1)	3956 (59.7)	7803 (54.5)	3825 (55.9)
Women	12,442 (63)	14,106 (69)	69,195 (63.0)		84,352 (43.1)	24,491 (42.9)	2665 (40.3)	6262 (45.5)	3022 (44.1)
Year of birth, median (IQR)	1930 (1925-1935)	1926 (1921-1933)			1944 (1931-1953)	1925 (1916-1936)			
Year of diagnosis, number (%)									
<1990	0	11 (0.05)	14,683 (13.4)		0	5677 (9.9)	0	0	469 (6.8)
1990-1999	17 (0.09)	2045 (10.1)	31,893 (29.0)		0	17,652 (30.9)	0	0	1605 (23.4)
2000-2009	9247 (47.5)	10,602 (52.4)	35,579 (32.4)		0	21,291 (37.3)	2432 (36.7)	0	2537 (37.1)
≥2010	10,194 (53.4)	7556 (37.4)	27,726 (25.2)		195,575 (100%)	12,501 (21.9)	4189 (63.3)	14,065 (100)	2236 (32.7)
Years of data before diagnosis, mean (IQR/SD)	13 (8-18)	7 (4-10)	37.9 (9.3)		8.2 (2.3)	37.9 (8.5)	10.3 (4.5)	8.2 (2.2)	40.4 (8.5)

Characteristics of individuals with Alzheimer's disease, Parkinson's disease, or amyotrophic lateral sclerosis in Australia, France, the UK, and Sweden.

### Associations with neurodegenerative diseases (5-10 years to diagnosis)



Families of treatments associated with a future diagnosis of at least one neurodegenerative disease in the 5–10 years before diagnosis.

- Validité des bases de données medico-administratives pour étudier les maladies neurodégénératives : cas de l'AMS

		Gold standard 1 (MSA register)		Gold standard 2 (hospital-based cohort)		Validity index
		Positive	Negative	Positive	Negative	
MSA case finding algorithm	Positive	76		96	24	<b>PPV =</b> $96/(24+96)$ <b>= 80%</b>
	Negative	31	.*		.*	
	Validity index	<b>SE =</b> $76/(76+31)$ <b>= 71%</b>				

**Indices de validation de l'identification des cas d'AMS.** Abréviations: SE : sensibilité, PPV : Valeur predictive positive. .\* indique que le nombre n'a pas pu être calculé avec le gold-standard utilisé.

- Un futur projet qui a été validé par le comité scientifique de l'observatoire français de la sclérose en plaques (OFSEP)
  - ProdroSEP: Etude de la valeur pronostique de facteurs pre-diagnostics dans la sclérose en plaques à partir des données chaînées de l'OFSEP et du SNDS
- Le chaînage permet un enrichissement du SNDS (variables confondantes, mesures de score de handicap, etc.) et l'accès aux historiques médicaux des patients de l'OFSEP

- Trois projets d'utilisation des données du SNDS à l'équipe Aramis:
  - Un étudiant de première année de thèse (Marc Dibling) sous la supervision de Sophie Tezenas du Montcel: ***étude des parcours de soins des patients atteints de la SLA à partir du SNDS***. Début pour Janv. 2024
  - Un ingénieur de recherche sous la supervision de Thomas Nedelec et Jean-Christophe Corvol : ***comparaison de l'efficacité des traitements de premières lignes de la maladie de Parkinson (outcome: traitements de seconde ligne)***. Début pour Janv. 2024
  - Candidature à l'AAP Messidor (Sophie Tezenas, Thomas Nedelec, Jean-Christophe Corvol, Anne-Sophie Jannot) : ***Prévention personnalisée des maladies neurodégénératives à symptômes moteurs, par des approches médicamenteuses et non médicamenteuses***

- Base de données en étoile, format non standardisé à l'échelle internationale (projet en cours du HDH de passer au format OMOP)
- Peu de ressources de code en libre accès
- Gestion des identifiants dans le SNDS
  - Changements fréquents d'identifiants à cause de changements de régime ou autre -> dans certains cas perte irréversible de l'information
  - Etude qui nécessitent l'étude de **patients sains** : cas des personnes qui ne sont plus en contact avec la base : deux cas de figures : i) personnes saines qui n'ont pas d'interaction avec la base du SNDS, ii) personnes ayant déménagé à l'étranger par exemple et donc pas forcément saines.

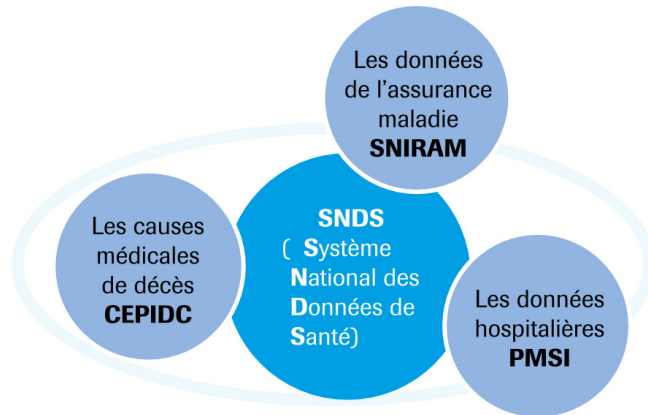
- Base de données de remboursement et non de recherche => collaborations avec des experts dans différents domaines indispensables pour savoir si la question scientifique posée peut être adressée à partir des données du SNDS.
  - Collaboration avec Sophie Tezenas du Montcel, médecin DIM (département information médicale) et PI au laboratoire Aramis
  - Collaboration avec une neurologue et un interne en neurologie pour le travail de validation de l'algorithme d'identification des cas d'AMS
- Si de l'information est manquante, cela remet en cause la validité et la fiabilité des résultats présentés

- Health Data Hub: répertoire gitlab de codes gérés par le HDD (codes des agences de santé, HAS etc.). Graphe interactif de la base du SNDS. Tutoriaux (fiches techniques)
- Formation données individuelles bénéficiaires (DIB) de la CNAM
- Document « SNDS ce qu'il faut savoir » de Santé Publique France
- Article « Réalisation d'études pharmacoépidémiologiques à partir du système national des données de santé (SNDS) : application des recommandations dans la pratique », de Germay et al. 2023

- Cours et ateliers sur l'utilisation de base de données medico-administratives (France, Italie et Danemark). Cours 4EU+. Responsable France: Julien Kirchgesner. Sorbonne Université.
- Congrès EMOIS: information médicale, parcours de soins, épidémiologie

- Notes de Jean Feydy sur le SNDS (chercheur INRIA équipe HeKa, Paris, France)
- Rapport du sénat 2023 sur les données de santé (Deroche)
- Présentation Aya Ajrouche sur la validation d'algorithme dans le SNDS

# MERCI POUR VOTRE ATTENTION



## Associations with neurodegenerative diseases (5-10 years to diagnosis)

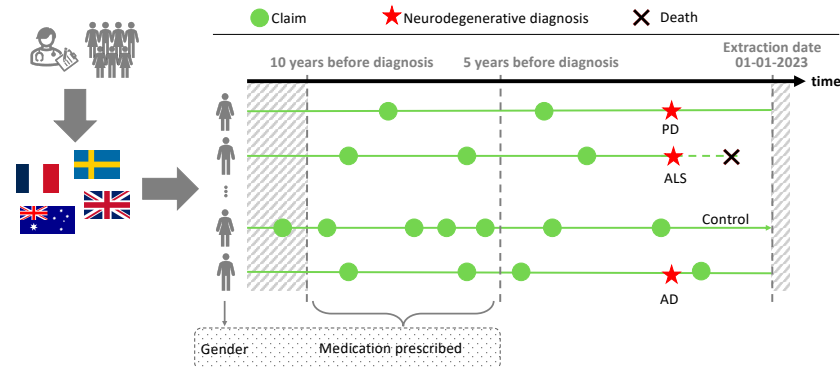
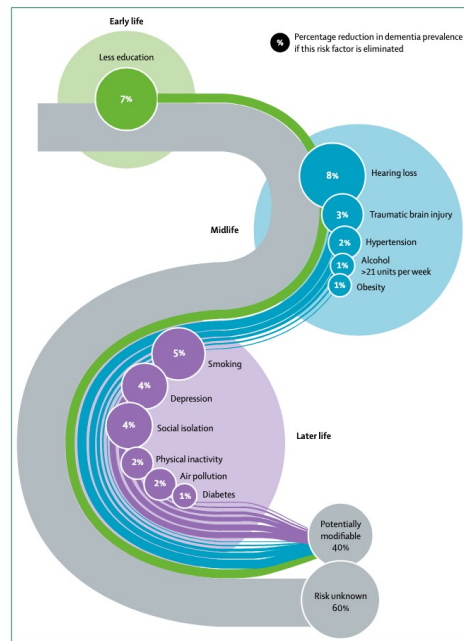
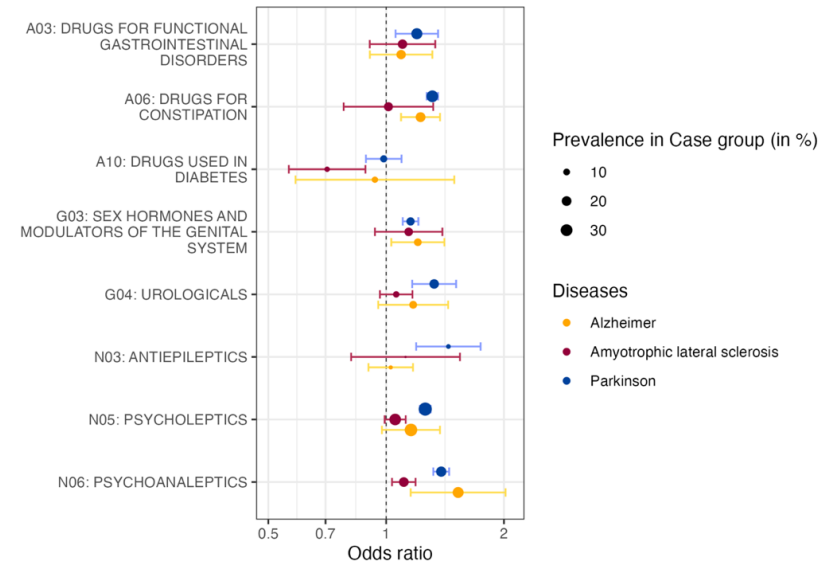


Figure 7: Population attributable fraction of potentially modifiable risk factors for dementia